Social network analysis in the context of information security risk management

R Serfontein

orcid.org / 0000-0002-0428-6494

Thesis accepted for the degree *Doctor of Philosophy in Computer Science* at the North-West University

Promoter: Prof HA Kruger

Graduation May 2020

21165750

OPSOMMING

Een van die primêre faktore wat die doeltreffendheid van inligtingsekuriteit bepaal is die aanspreek van die risiko's wat met menslike akteurs verband hou. Dit word gewoonlik bewerkstellig deur die gebruik van sekuriteitsbeleide wat daarop gemik is om gebruikers se gedrag te bestuur, asook veiligheidsbewusmakingsprogramme wat ontwerp is om die kennis wat gebruikers het oor bedreigings, asook hul gedrag, te verbeter. Ongelukkig, alhoewel hierdie metodes dikwels inligtingsekuriteit risiko's verminder, het hulle sekere tekortkominge wat 'n invloed kan hê op hoe doeltreffend hulle is om hierdie risiko's aan te spreek. Bewusmakingsprogramme, byvoorbeeld, spreek nie noodwendig nuwe risiko's aan nie, terwyl beleide wat te streng is tot sekuriteitsmoegheid kan lei. 'n Addisionele benadering is om Sosiale Netwerk Analise (SNA) te implementeer ten einde inligtingsekuriteitsrisiko's te identifiseer en te bestuur deur strukturele risiko's in die sosiale netwerke van organisasies aan te spreek. Hierdie sosiale netwerke beskryf die interaksies tussen mense, take en hulpbronne, en deur dit te ondersoek, kan verborge inligtingsekuriteit risiko's moontlik geïdentifiseer word. In hierdie studie word 'n raamwerk voorgestel wat daarop gemik is om SNA te gebruik om die inligtingsekuriteit risiko's wat in sosiale netwerke voorkom, te identifiseer. Die voorgestelde raamwerk bied ook 'n gestruktureerde benadering tot die ontwikkeling van risikobeperkende strategieë wat gebruik kan word om hierdie risiko's te verminder, asook om hierdie strategieë te implementeer. Ten einde 'n volledige raamwerk te ontwikkel, bied die studie ook 'n aantal metodes aan wat aangepas is vir gebruik met SNA. Hierdie nuwe toepassings sluit onder meer die implementering van self-organiserende kaarte in wat gebruik kan word om inligtingsekuriteit risiko's in 'n sosiale netwerk grafies te evalueer, en 'n aangepaste netwerkoptimiseringstegniek. 'n Regte wêreld netwerk, gebou met behulp van data uit 'n korporatiewe risikoverslag, word saam met verskeie kleiner netwerke gebruik om die geldigheid en nut van die raamwerk te demonstreer.

SLEUTELWOORDE:

Inligtingsekuriteit, risiko bestuur, sosiale netwerk analise, self-organiserende kaarte, netwerk optimisering, risiko vermindering strategieë, sekuriteit bewusmakingsprogramme

ABSTRACT

One of the primary factors that determines the efficacy of information security is addressing the risks associated with the human actors involved. This is usually accomplished through the use of security policies that aim to manage user behaviour, and security awareness programmes that aim to improve both the knowledge users have of information security threats, and their behaviour. Unfortunately, while these methods do often reduce information security risk, they have certain shortcomings that may have an impact on how effectively they can help mitigate these risks. Awareness programmes, for example, may not necessarily address new risks, whereas overreaching policies could lead to information security fatigue. An additional approach is to implement Social Network Analysis (SNA) in order to identify and manage information security risks by addressing structural risks in the social networks of organisations. These social networks describe the interactions between people, tasks, and resources, and by investigating them hidden information security risks can potentially be identified. In this study a framework is proposed that aims to use SNA in order to identify the information security risks present in social networks. The proposed framework also presents a structured approach to developing risk mitigation strategies that can be used to reduce these risks, as well as the implementation of these strategies. In order to develop a complete framework, the study also presents a number of methods that were adapted for use with SNA. These novel applications include, among others, an implementation of Self-Organising Maps that can be used to evaluate information security risks in a social network graphically, and an adapted network optimisation technique. A real-world network, built using data from a Corporate Risk Report, is used in conjunction with multiple smaller networks to demonstrate the validity and utility of the framework.

KEYWORDS:

Information security, risk management, social network analysis, self-organising maps, network optimisation, risk mitigation strategies, security awareness programmes

ACKNOWLEDGEMENTS

I would like to thank my study promoter, Professor Hennie Kruger, for all of the invaluable help he has given me over the past 4 years. Without his willingness to read draft after draft, to do so even over weekends, and his patience with some of my grammatical quirks, this study would likely not have been a successful endeavour.

I would also like to express my gratitude to my family: my father Wynand, my mother Elsje, and my sister Riana, for their support during this time. An undertaking of this magnitude is difficult to complete without a proper support structure, and I am deeply grateful to mine.

And last, but assuredly not least by any possible measure, I would like to thank the Lord God Almighty, the Holy Trinity, for the gifts, opportunities, and inspiration that allowed me to undertake this study.

TABLE OF CONTENTS

- PART I -

1. Introduction 4

1.1. Informal Problem Statement 4

1.2. Goals and Objectives 7

1.3. Statement of the Scope of the Study 8

1.4. Structure and Organisation of Study 8

1.5. Chapter summary 10

- PART II -

2. Information Security and Risk 14

2.1. Introduction to Information Security 14

2.2. CIA triad 14

2.2.1. Confidentiality 15

2.2.2. Integrity 17

2.2.3. Availability 18

2.3. Information Security Risk Management 18

2.3.1. Risk Identification 21

2.3.2. Risk Analysis 22

2.3.3. Risk Control 30

2.4. Human Aspects of Information Security 33

2.4.1. Information Security Culture 33

2.4.2. Information Security Knowledge 35

2.4.3. Information Security Behaviour and Attitude Theories 35

2.5. Chapter Summary 40

3. Social Network Analysis 42

3.1. Introduction 42

3.2. Graph Theory and Associated Principles 45

3.3. Visualisation of Networks 52

3.3.1. Introduction to Visualisation Techniques Used in the Literature.....53
3.3.2. Self-Organising Maps (SOMs) 57
3.4. Social Network Analysis: Metrics and measures.....58
3.4.1. Centrality.....58
3.4.2. Boundary Spanner 62
3.4.3. Shared Situation Awareness 63
3.4.4. Structural Holes Constraint.....63
3.5. Community Detection 64
3.5.1. Hierarchical Clustering.....65
3.5.2. Edge Removal.....66
3.5.3. Cooperative Game Method 66
3.5.4. Evolutionary Node Centrality Algorithm 67
3.6. Network Optimisation and Monitoring.....67
3.6.1. Optimisation of the Critical Diameter and average path of social networks 67
3.6.2. Computer Network Optimisation 68
3.6.3. Monitoring a Social Network 69
3.7. Chapter Summary.....71

4. Social Network Analysis in the Context of Information Security 74
4.1. Review of literature sources using SNA in the Context of Information Security 74
4.1.1. Organizational risk using network analysis.....74
4.1.2. Applying network analysis to investigate interpersonal influence of information security behaviours in the workplace.....75
4.1.3. Applying social network analysis to security 76
4.1.4. Understanding of Impact and Propagation of Risk based on Social Network Analysis 76
4.1.5. Applying network analysis to assess coastal risk planning 77
4.2. SNA Metrics and their Relationship to the CIA Triad 78
4.2.1. CIA Rationale for SNA metrics.....78
4.2.2. Illustrative Example Using a Simulated Network.....81
4.3. Tabulated Summary of Literature Sources 83
4.3.1. Information Security Studies 83

4.3.2. Social Network Analysis Studies 85
4.3.3. Studies Featuring Both SNA and Information Security 86
4.4. Summary of Part II 87
4.5. Chapter Summary 88

- PART III -

5. Research Method 92
5.1. Research Onion Model 92
5.1.1. Philosophies 93
5.1.2. Approaches to Theory Development 96
5.1.3. Methodological Choice 97
5.1.4. Research Strategies 98
5.1.5. Time Horizon 100
5.1.6. Techniques and Procedures 100
5.2. Research Approach Followed in this Study 102
5.3. Chapter Summary 104

- PART IV -

6. Methods and Adaptations 108
6.1. Optimisation of a Social Network Using Risk Metrics 108
6.2. Evaluation of the Risk in a Social Network using Self-Organising Maps 117
6.3. Improving Information Security Awareness using SNA 123
6.4. Chapter Summary 125

7. A Novel Framework for Addressing Information Security Risk using SNA 128
7.1. Method Framework 128
7.1.1. Phase 1: Relationship Graphing 129
7.1.2. Phase 2: Develop an Information Security Risk Profile 133
7.1.3. Phase 3: Structural Optimisation Using Risk Profile 136
7.1.4. Phase 4: Develop Risk Mitigation Strategies 139
7.1.5. Phase 5: Implementation and Monitoring 140
7.2. Illustrative Example 142

7.2.1. Phase 1 143
7.2.2. Phase 2 146
7.2.3. Phase 3 149
7.2.4. Phase 4 153
7.2.5. Phase 5 154
7.2.6. Summary of Illustrative Example 161
7.3. Large Dataset Example 162
7.4. Chapter Summary 166

8. Risk Management Network: Analysis and Optimisation 168
8.1. Overview of data and method of collection (Phase 1) 168
8.2. Graphical Evaluation (Phase 2) 172
8.2.1. Analysis of CRR Network Graph 172
8.2.2. SOM Analysis 174
8.2.3. Node-Level Risk Profile 179
8.3. Optimisation (Phase 3) 188
8.4. Final Reflection 197
8.5. Chapter Summary 201

- PART V -

9. Evaluation of the Framework 206
9.1. Expert Feedback 206
9.2. Critical Evaluation 208
9.3. Concluding Remarks 215
9.4. Chapter Summary 216

10. Summary and Conclusion 218
10.1. Synopsis of the Study 218
10.2. Contributions 222
10.3. Limitations and Future Work 223
10.4. Chapter Summary 224

Appendix A: Published Articles.....226

Appendix B: Chapter 7 Appendix - Phase 2 Risk Profile 252

Appendix C: Chapter 8 Appendix - Node-Level Risk Profile 254

REFERENCES.....266

LIST OF FIGURES

Figure 1.1: High level structure of study.....9
Figure 2.1 CIA Triad representing Confidentiality, Integrity and Availibility 15
Figure 2.2 Relationship between the various actors in information security risk (ISO/IEC 15408-1:2009 (2009)) 19
Figure 2.3 Risk management process (Yue *et al.*, 2007).....21
Figure 2.4 Graphical representation the CORAS method (Braber *et al.*, 2003) 23
Figure 2.5 The CIRA Procedure (Wangen, 2015) 24
Figure 2.6 Stakeholder prioritisation scheme (Wangen, 2015; Mitchell *et al.*, 1997) 25
Figure 2.7 ISRAM Flow diagram (Karabacak & Sogukpinar, 2005) 26
Figure 2.8 Example of an ISRAM risk table (Karabacak & Sogukpinar, 2005) 27
Figure 2.9 Levels of IS culture (Van Niekerk & von Solms, 2010; Schein, 2009; Schlienger & Teufel, 2003) 34
Figure 2.10 Structure of TRA [Brodowsky *et al.* (2018) and Khan *et al.* (2011)] 37
Figure 2.11 Structure of TPB [Brodowsky *et al.* (2018) and Khan *et al.* (2011)].....37
Figure 2.12 Basic principle of GDT [Moody *et al.* (2018), Willison *et al.* (2018), and Straub Jr (1990)].....38
Figure 2.13 Basic structure of PMT [Tesson *et al.* (2016)].....39
Figure 3.1: Fundamental network types. (a) is a social network of strangers; (b) is a social network of colleagues; (c) is a military information network 47
Figure 3.2 a) Reachability in symmetrical networks; b) Reachability in asymmetrical networks 47
Figure 3.3: Simple graph showing a clique with four nodes.....48
Figure 3.4: Graph showing multiple distances between similar nodes.....49
Figure 3.5 Distances in undirected graphs.....50
Figure 3.6: The different types of walks; note that only those nodes and edges that exist within a particular walk's collection are shown. (a) is the graph that contains all of the walks; (b) is a walk containing repeating paths and repeating endpoints (any of the nodes can

potentially be the start/end node); (c) is a trail with repeating endpoints (any of the nodes can potentially be the start/end node); (d) is a path with repeating edges (while most of the nodes can be either a start or end node, node B must be either a start or an end node) 51
Figure 3.7 Simple visualisation technique, with (a) using dots and (b) using squares as nodes 53
Figure 3.8 Large network visualised using simple method (Ying & Xiao, 2011) 54
Figure 3.9 Simple method colouring variations: (a) using colour to uniquely identify nodes and arcs; (b) using colour to draw attention to a specific node 54
Figure 3.10 Modified simple method using coloured nodes and arcs (Tsui & Liebowitz, 2005) 55
Figure 3.11 Simple method variation using colour to uniquely identify node types, and node size variations to identify important nodes (Dang-Pham *et al.*, 2017b) 56
Figure 3.12 Simple method variation using colour to identify nodes, and size variations on the node labels to identify important nodes (Dang-Pham *et al.*, 2017c) 56
Figure 3.13 Methods used to bundle arcs (edges) (Bach *et al.*, 2017) 57
Figure 3.14: Simulated network with three communities 64
Figure 3.15: An example of a small hierarchical clustering tree. The circles at the bottom represent nodes and the tree shows the order in which they join to form communities (Girvan & Newman, 2002). four communities of differing sizes have been encircled.....65
Figure 3.16: Interpreting control charts (Render *et al.*, 2012) 70
Figure 4.1: Network representing simulated SNC employees 81
Figure 5.1: The research onion [adapted from Saunders *et al.* (2019)] 93
Figure 5.2: Overall structure of the study.....103
Figure 6.1: SNC network before optimisation 113
Figure 6.2: SNC network after first optimisation step 115
Figure 6.3: SNC network after second optimisation step 116
Figure 6.4: Allocation of data points to topographical regions using SOM technique (López *et al.*, 2019) 118
Figure 6.5: Guide to reading a SOM.....120
Figure 6.6: SOM illustration network.....120
Figure 6.7: Process for developing a targeted security awareness programme using SNA ..124
Figure 7.1: Overall structure of framework 129
Figure 7.2: Overview of Phase 1 132
Figure 7.3: Overview of Phase 2 136
Figure 7.4: Overview of Phase 3 138
Figure 7.5: Overview of Phase 4 140
Figure 7.6: Overview of Phase 5 142
Figure 7.7: Overview of Phase 1 143
Figure 7.8: UD network without any layout changes. Students are coloured red and academic staff blue. The nodes in purple are post graduate students in academic staff positions.....145

Figure 7.9: UD network with a circular layout. The nodes with the highest betweenness are placed in the centre 145
Figure 7.10: Overview of Phase 2 146
Figure 7.11: SOM Betweenness centrality (BC) 147
Figure 7.12: SOM Closeness Centrality (CC) 147
Figure 7.13: SOM Eigenvector Centrality (EiC) 147
Figure 7.14: SOM Closeness Centrality (CC) 147
Figure 7.15: SOM Eccentricity Centrality (ecC) 148
Figure 7.16: SOM Structural Holes Constraint (SHC) 148
Figure 7.17: SOM Boundary Spanner (BS) 148
Figure 7.18: Overview of Phase 3 149
Figure 7.19: UD network after structural optimisation. The nodes that gained relationships are shown with a large rectangular form; the relationships that were added are highlighted. This graph does not show weighted links. 151
Figure 7.20: Micro-network containing nodes identified in phase 3. The relationships between these nodes are included. Students are coloured red, academic staff blue, and the nodes in purple are post graduate students in academic staff positions. 156
Figure 7.21: Trident Juncture Twitter network (Frankenstein *et al.*, 2016). The circular red nodes are individuals, the green damonds are tweets, the light green pentagons are topics, and the orange haxagons are locations 163
Figure 7.22: Zoomed out 3D version of the Trident Juncture Twitter network, showing the possible existence of at least 7 communities 164
Figure 7.23 (a) to (G): SOM for the Trident Juncture Twitter network, coloured using each of the 7 selected metrics 165
Figure 8.1: Relationship between the various types of nodes in the CRR network 169
Figure 8.2: crr network showing relationships between risks (red), risk controls (green), risk coordinators (yellow), risk owners (orange), control owners (pink) and the governance bodies (blue). 170
Figure 8.3: CRR network with a circular layout. Nodes with the highest betweenness are placed in the centre. 171
Figure 8.4: CRR network with a circular layout, with a zoomed-in portion of the network shown 171
Figure 8.5: SOM clusters 174
Figure 8.6: CRR network prior to optimisation 194
Figure 8.7: Links that were added to CRR network during optimisation 194
Figure 8.8: Links that were removed from CRR network during optimisation 194
Figure 8.9: CRR network before (a) and after (b) optimisation, with circular layout applied. Node with highest betweenness centrality is placed in the centre. 195
Figure 8.10: Overall progression of first three phases 198
Figure 8.11: Changes to the metrics for RISK COORDINATOR 8, and how these changes impact CIA risks. Blue indicates that a metric has no impact on the risk value of the node,

orange indicates an elevated risk, yellow indicates a low risk, and red indicates that the value for the metric has increased as a result of the optimisation. 199
Figure 8.12: Changes to the metrics for Control Owner 5, and how these changes impact CIA risks. Blue indicates that a metric has no impact on the risk value of the node, green indicates that the value of the metric has been reduced to zero, orange indicates an elevated risk, yellow indicates a low risk, and red indicates that the value for the metric has increased as a result of the optimisation. 200
Figure 9.1: Overall structure of the novel framework 210
Figure 9.2 Risk management process (Yue *et al.*, 2007) 210

LIST OF TABLES

Table 2.1 Comparison between CORAS, CIRA, ISRAM, and ISra (Agrawal, 2017) 29
Table 4.1: Summary of SNA metrics in the context of the CIA triad 80
Table 4.2: SNA metrics for SNC network 82
Table 4.3: Summary of information security studies 84
Table 4.4: Summary of SNA studies 85
Table 4.5: Summary of SNA studies that feature information security 86
Table 6.1: Normalised SNA metrics for SNC network. Total risk is calculated as the sum of the risk metrics 112
Table 6.2: Normalised SNA metrics for modified SNC network following first iteration 115
Table 6.3: Risk metrics following second iteration. Total reduction in risk of 14.04% 116
Table 6.4: SNA metrics for nodes in SOM demonstration network 121
Table 6.5: Nodes contained in each of the four clusters 121
Table 6.6: SOM of data in Table 6.4, coloured using each of the six metrics used 121
Table 7.1: Comparison of organisational characteristics that are appropriate for the selection of formal and informal networks 131
Table 7.2: Portion of Phase 2 Risk Profile 148
Table 7.3: New relationships created in UD network. The new relationship edges are indicated using their source node (high risk) and destination node (low risk) 152
Table 7.4: Monitoring data for the UD micro-network during the course of the 5th phase .159
Table 8.1: Number of each type of node in each cluster. To enhance readability, the cells with a value of 0 have been filled in using black. 174
Table 8.2: Weight scale for metrics. 181
Table 8.3: Summary of Self-Organising Map (SOM) weighting criteria 182
Table 8.4: Weight scale for nodes 182
Table 8.5: Risk Profile Weights for node types 184
Table 8.6: Risk Profile Weights for metrics 185
Table 8.7: Preference Scale for pairwise comparisons, adapted from (Taylor, 2013) 186

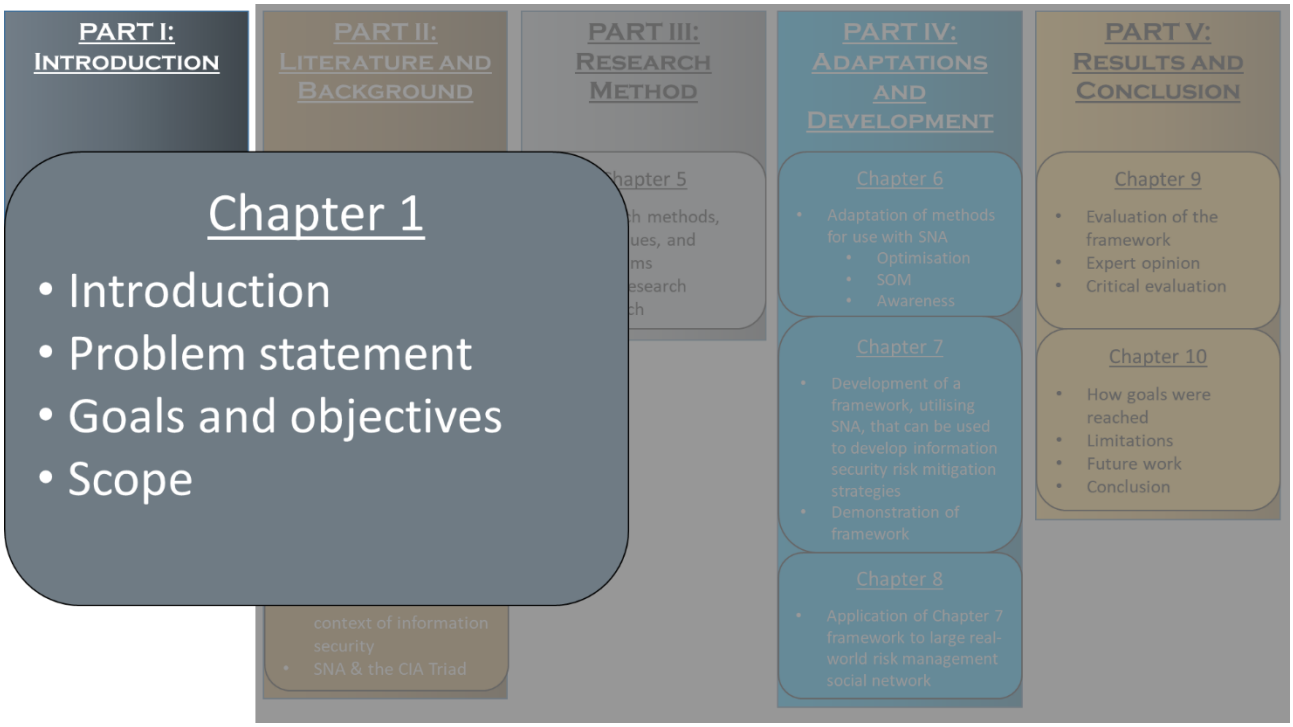
Table 8.8: Extract of node-level risk profile. The table shows the normalised values for betweenness centrality (BC), closeness centrality (CC), eccentricity centrality (ECC), eigenvector centrality (EiC), structural holes constraint (SHC), and boundary spanner (BS). the weight and risk values for these nodes are also shown, as well as the metric weights and the total risk value. 187
Table 8.9: Extract of node-level risk profile, showing all the nodes with a risk z-score greater than 2. The table shows the normalised values for betweenness centrality (BC), closeness centrality (CC), eccentricity centrality (ECC), eigenvector centrality (EiC), structural holes constraint (SHC), and boundary spanner (BS). the weight and risk values for these nodes are also shown. 189
Table 8.10: Data used to identify new relationships during optimisation process. Cells highlighted in red indicate the highest remaining z-score at the start of the step. Cells highlighted in yellow are the metrics that contribute the most to a node's risk value. “REMOVE” indicates that links have to be removed in order to reduce the metric, whereas “ADD” indicates that a metric value will be reduced if links are added.....191
Table 8.11: Summary of links that were added and removed 196
Table B.1: Phase 2 risk profile.....252
Table C.1: Risk Profile Weights for metrics.....254
Table C.2: Node-level risk profile.....254

PART I

Introduction

"The beginning is the most important part of the work."

- Plato



CHAPTER 1: INTRODUCTION

CHAPTER HIGHLIGHTS:

- What is the focus of this study?
- What are the problems that it aims to address?
- What are the goals of this study?
- What is the scope?
- How is the study structured?

1 INTRODUCTION

In this chapter the subject of the study, as well as the research question, will be introduced. In order to accomplish this effectively, the various topics are introduced in four primary points of discussion. The first is the problem statement, wherein the research question is also introduced. The second point is a statement of the goals and objectives of the study. Thirdly, the scope of the study is clarified. Finally, the overall structure of the study is discussed.

This study follows a positivist paradigm, with an experimental component. It should be clarified, however, that this chapter does not contain an in-depth discussion of the research approach, or how it was selected, as this is discussed in significant detail in Chapter 5, and is presented as a single contiguous unit. This chapter therefore only focusses on the core of the study, namely the problem statement, its goals, and overall structure.

1.1. INFORMAL PROBLEM STATEMENT

In the field of information security, one of the primary success factors is the human aspect (Shillair *et al.*, 2015). In fact, the human aspect is so important that past research has shown that a balanced approach wherein both technological and social aspects are addressed is crucial to maintaining information security (Soomro *et al.*, 2016; Parsons *et al.*, 2014). Despite repeated campaigns to educate users on information security however, a significant number of users still engage in risky behaviour (Widjaja *et al.*, 2019; Byrne *et al.*, 2016), and they are still considered the weakest link in information security (Chung, 2019; Arachchilage & Love, 2014). One of the best known traditional methods that are used to address this risk are security awareness programs (Kemper, 2019; Aloul, 2012; Thomson & von Solms, 1998). There are, however, a number of drawbacks to these awareness programs, e.g. the awareness programs might not be comprehensive enough (Siponen, 2000), they might not address new threats quickly enough when the risks change continuously (Kruger & Kearney, 2006), and the programs rely upon the users to consciously decide to comply with information security principles (Ng *et al.*, 2009). While a large amount of research is focused on attempting to address these shortcomings (Tsohou *et al.*, 2015), a possible way to improve on the situation might be to attempt to address the risks themselves in a more subtle manner, rather than only relying on awareness programs and human cooperation.

Another factor that may impact the information security in an organisation is the structure of the organisation itself. It is important to consider the structure of an organisation's social-

and working relationships, as these relationships may have an impact on, or may even cause, information security risks (Armstrong & McCulloh, 2010). This is especially evident when one considers the existence of so-called "shadow security" systems, which are informally developed by the users of a system in order to circumvent security measures that are deemed to be annoying or counterproductive (Dang-Pham *et al.*, 2017d; Dang-Pham *et al.*, 2017b). Addressing these structural risks can be complicated, as the culture of an organisation ultimately determines its members' willingness to cooperate with security measures and systems, and that culture can be influenced by people situated at critical points in the network.

Therefore, given the importance of the human aspect in information security and the potential problems with the current method of primarily using mass security awareness programs, a different approach is proposed. In this study the use of Social Network Analysis (SNA) as a technique to evaluate information security risk within an organisation, and then develop suitable strategies, will be discussed. SNA is a method, utilising graphs and graph theory, that can be used to represent a social organisation, such as a community or business, in such a way that the social interactions can be studied quantitatively (Scott & Carrington, 2011). The technique is suitable for use in environments where certain risks, including those risks associated with information security, are present, and has been used in the past to, among others:

- Identify core members and organizations within terrorist groups (Fu *et al.*, 2015);
- Assess organisational risk (Armstrong & McCulloh, 2010);
- Detect insider trading (Gupta & Hossain, 2011);
- Identify hierarchies in criminal DarkWeb forums (Philips *et al.*, 2015);
- Identify individuals that pose a high risk as cause of virological infection within social groups (Christley *et al.*, 2005); and
- Control Avian flu in poultry (Martin *et al.*, 2011).

The expected advantage of using SNA is that certain high risk groups and individuals can be identified. By evaluating the positions and powers (both formal and informal) of the individuals within a social network using SNA metrics, their overall risk to the organisation can be determined. Among these individuals are those that act as crucial intermediaries within the network, and whose possible removal could cause damage to the integrity of the information security systems. Another group are those that act as informal leaders and could therefore influence the information security culture of the organisation in positive or negative ways. It is necessary, in dealing with the human aspect of information security, to also take note of the security culture of an organisation, as this will ultimately have an impact upon any measures taken to improve the information security systems of the organisation (Da Veiga & Eloff, 2010; Thomson *et al.*, 2006).

SNA does however have a significant drawback when it comes to large networks, as these networks may have so many nodes and arcs that they are incomprehensible when

visualised. A number of studies have attempted to address this drawback, and generally employ methods that alter the appearance of nodes and arcs based on their attributes. Some of these methods include differentiating the colour of nodes and edges (Tsui & Liebowitz, 2005), sizing nodes according to certain metrics (Dang-Pham *et al.*, 2017b), and using labels of differing size depending on its importance (Dang-Pham *et al.*, 2017c). A somewhat more novel technique makes use of Self-Organising Maps (SOMs) to directly visualise network data (Boulet *et al.*, 2008). A SOM is an useful technique, as it can be used to not only visualise high-dimensional data, but to cluster it as well (Kohonen, 1998). This means that the SOM technique can be used to identify similar nodes within a social network, even in the presence of seemingly contradicting attributes, and present this data in a graphical way. The literature mentions several ways in which these maps can be applied to information security, from improving intrusion detection methods (De la Hoz *et al.*, 2015), to analysing information security behavioural data (López *et al.*, 2019; Hunt & Hill, 2015). However, while SOMs have previously been used to visualise social networks, there is little mention of an application whereby SOMs can be used to investigate possible information security risks that can be identified using SNA. As such an application could potentially allow for SNA identified risks to be investigated using visual analyses, the development and testing of such an approach should be considered.

Whilst identifying individuals that may pose a risk (hereafter referred to as at-risk individuals) using SNA is not new (Armstrong *et al.*, 2010), the literature is sparse when it comes to the development of strategies to mitigate the risks after they have been identified using SNA. It is in this regard that this study will aim to make its primary contribution: to develop strategies that can be used to mitigate information security risks identified using SNA. In order to develop these strategies, network optimisation techniques are proposed. While often used to determine improvements to physical networks such as computer networks (Rezazad, 2011), the literature makes no mention of the application of these techniques to social networks. This study will aim to demonstrate how network optimisation techniques, once adapted, can be used to develop risk mitigation strategies within a social network.

Based on this informal problem statement, the formal problem statement can be expounded on in the following manner: a number of techniques exist in the literature that can be used to address the information security risks posed by human users. The risks caused by social network structures, however, are not addressed by any of these techniques, as they mostly aim to directly correct user behaviour or culture, rather than focussing on less obvious causes, such as relationships. SNA can potentially be used to identify these risks, as well as provide guidance in developing risk mitigation strategies.

In summary, the core of the study lies in the following research question:

Can social network analysis be used to develop risk management strategies in an information security context?

In order to answer this question, a number of different aspects will have to be considered. The first deals with the application of SNA to identifying information security risks. Secondly, it should be determined whether or not network techniques, similar to those that were used to identify the risks, could also be used to identify areas of improvement. Finally, the viability and usability of the techniques should be investigated and established. These goals are discussed in greater detail in the next section.

1.2. GOALS AND OBJECTIVES

The primary goal of this study is to **demonstrate how SNA can be used in the context of information security risk management**. In order to achieve this goal, the following objectives are pursued:

- **Conduct** a review of the literature, and use the information and insight gained from it to compile informative sections on the following topics:
    - Information security, specifically focussing on its core principles, various approaches to addressing the human aspects, and risk management strategies;
    - Network theory, including network metrics and visualisation techniques, with special attention given to appropriate sections of graph theory and its relevance to SNA; and
    - Past instances of SNA being used to address information security risk.
- **Adapt** relevant existing methods so that they can be used in conjunction with SNA in the context of this study;
- **Develop** a novel method framework that incorporates SNA that can be used to evaluate information security risk in an organisation and propose improvements that will address those risks;
- **Demonstrate** how the novel method can be applied to real-world data;
- **Investigate** the viability of applying Self-Organising Maps (SOMs) as a visualisation- and data processing technique to the novel method;
- **Demonstrate** how the method, utilising SOMs, can be applied to real-world data;
- **Critically evaluate** the method and its applications, and provide an overview of its advantages and potential;
- **Identify** possible shortcomings in the method and propose how future research may improve the method;
- **Suggest** sensible and related future work that may follow from this study.

These nine goals summarise the objectives of this study, and should sufficiently support the primary goal.

1.3. STATEMENT OF THE SCOPE OF THE STUDY

The ultimate goal of this study is to determine if and how SNA can be used to improve information security risk management. As such, there are a number of aspects regarding the scope that need to be clarified beforehand:

- While a number of the techniques that can be used to collect social network data, such as entities, relationships, and the strength of those relationships, are discussed in due course, the focus of the study is not on these techniques per se. Subsequently, no special data collection methods will be developed, nor will any data collection techniques be compared to determine their individual validity.
- Every organisation is unique, and will therefore likely have different ways in which they can develop effective risk mitigation strategies. The intent is to investigate how SNA can be used to inform and improve the development of these strategies, rather than develop the strategies themselves. This does not mean that example strategies will be excluded, but rather that the development and implementation of strategies for specific organisations lie outside the scope of this study.
- The study investigates the application of a technique that utilises SNA in order to evaluate information security risk. The purpose is therefore to introduce new methods, establish their viability and usability, and then use them to evaluate selected real-world networks. The scope is limited in this aspect to the evaluation of the selected networks and an appraisal of the results of the evaluation.

The study will aim to adhere to this scope, so that the research can be properly narrowed to the topics of investigation. This should also aid in making the techniques and methods proposed in this study more applicable to real-world situations, as the scope is also intended to minimise the amount of subjective, or application-specific, knowledge that is required.

1.4. STRUCTURE AND ORGANISATION OF STUDY

This study is presented in ten chapters, which are organised into five main parts: **Introduction**, **Literature and Background**, **Research Method**, **Adaptations and Development**, and **Evaluation and Conclusion**. The structure of the study is presented visually in Figure 1.1, and the contents of each of the five main parts will be discussed briefly.

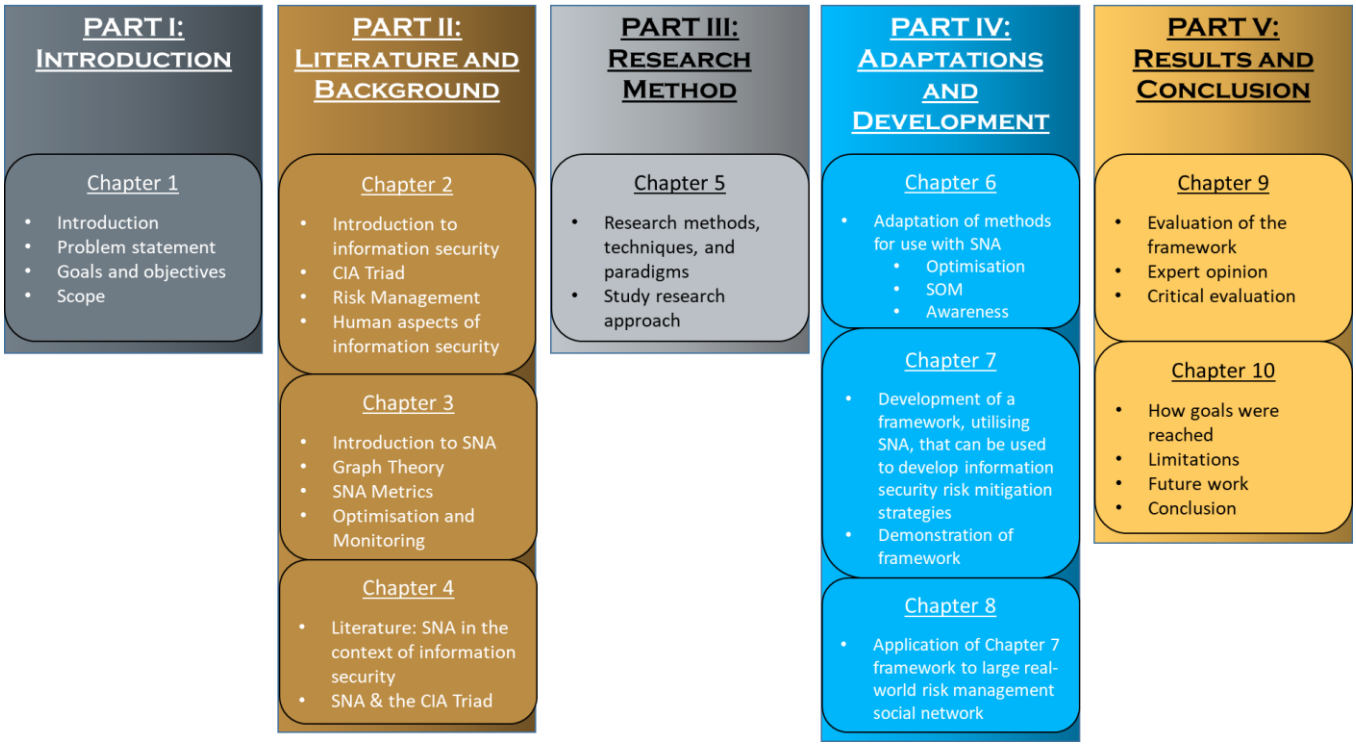


FIGURE 1.1: HIGH LEVEL STRUCTURE OF STUDY

Part I – Introduction

The first part contains the introduction to the study. This encompasses the problem statement, goal and objectives, and discussion of the scope. Part I contains only one chapter, namely Chapter 1.

Part II – Literature and Background

Part two of the study deals with past research, and an investigation of the implications and conclusions of previous work. This part contains three chapters, namely Chapters 2, 3, and 4, and covers topics ranging from information security principles (Chapter 2), to SNA (Chapter 3), to the relationship between SNA and information security (Chapter 4).

Part III – Research Method

Like Part I, this section of the study contains only one chapter, namely Chapter 5, and deals with the various topics relevant to selecting a research approach. Some of the topics covered in this section are research paradigms, methods, and techniques. The overall approach to the study is also discussed in greater detail.

Part IV – Adaptations and development

The penultimate part contains most of the novel contributions of this study. In this section the ways in which existing techniques are adapted for use with SNA is discussed in Chapter 6. The novel use of techniques such as SOMs are also explored. This part also introduces a novel framework in Chapter 7 that implements SNA in order to develop risk mitigation strategies, and demonstrates the framework using data from multiple real-world sources. To validate that the framework can be used to assess large real-world networks, it is also applied in depth to a large risk management network in Chapter 8.

Part V – Evaluation and Conclusion

In the final part the work presented in Part IV is evaluated critically, and the final comments and conclusions surrounding the work are provided. This forms the bulk of Chapter 9. Part V then concludes with Chapter 10, wherein a summary of the ways in which the various goals presented in this chapter were achieved. A brief consideration for possible future work is also provided.

In summary, the study consist of 10 chapters organised into five parts based on the contents of the chapters. Part I deals with introductory concepts, whereas Part II is focussed on literature and background. The third part, Part III, deals with the background and selection of an appropriate research approach. Part IV, which contains the bulk of the contributions, deals with the adaptation of methods, the development of new techniques, and the testing of said techniques. Finally, Part V deals with the evaluation of the contributions and concludes the study as a whole.

1.5. CHAPTER SUMMARY

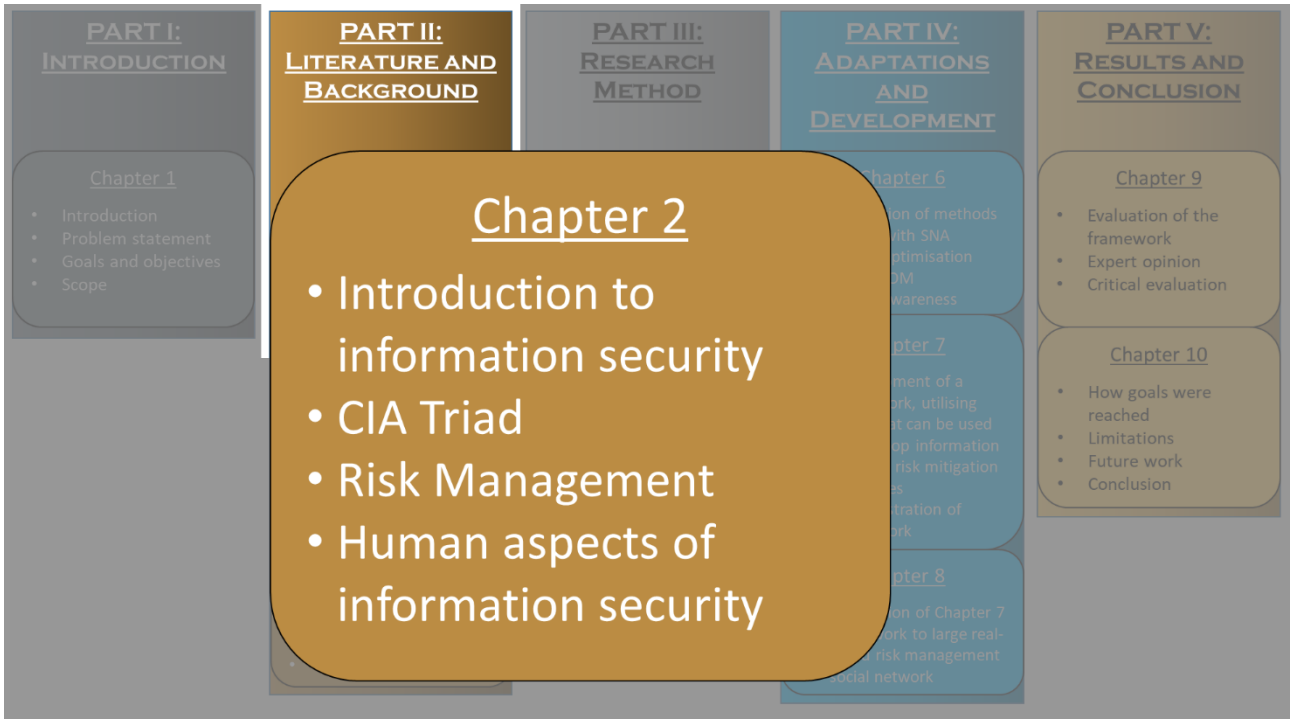
In this chapter the core of the study is introduced. The problem statement, as well as the goals and objectives, are discussed. Following this, the scope of the study is clarified. The chapter then concludes by providing an overview of the structure of the study by highlighting the overall themes for each of the parts, and briefly mentioning which topics are found in each of the chapters.

PART II

Literature & Background

"Literature adds to reality, it does not simply describe it. It enriches the necessary competencies that daily life requires and provides; and in this respect, it irrigates the deserts that our lives have already become."

- C. S. Lewis



CHAPTER 2: INFORMATION SECURITY AND RISK

CHAPTER HIGHLIGHTS:

- What are the basic principles of information security?
- What is the CIA Triad?
- What are the basic principles of risk management?
- What are the human aspects of information security?

2 INFORMATION SECURITY AND RISK

The purpose of this study is to demonstrate how Social Network Analysis (SNA) can be applied to risk management within an information security context. It is therefore important to provide an introductory background on the subject of information security. In this chapter the various principles relating to information security and information security risk will be discussed. The overview given in this chapter will aim to place these principles within the context of this particular study. Information security as a field of research encompasses quite a significant number of subjects and a complete overview of the field lies outside the scope of this study.

The chapter will begin with a brief discussion of the basic principles of information security based on the Confidentiality, Integrity, and Availability (CIA) triad. The focus will then shift to risk management strategies, and a discussion of the human aspects of information security will conclude the chapter.

2.1. INTRODUCTION TO INFORMATION SECURITY

The term information security, as the name implies, concerns the protection of information assets. Stated more formally, information security involves the preservation of the confidentiality, integrity, and availability of information (ISO, 2016). The concept of information, however, is a relatively vague one as it can refer to almost anything – bank statements, corporate reports, academic papers etc.; it can even refer to some physical assets such as laptops (Da Veiga & Eloff, 2010). The vagueness of this concept is further compounded by the fact that new types of information assets are added to the list as new research is conducted, and the human race grows. Studying information security therefore requires a number of specific structures and concepts. In this section a number of these concepts will be discussed.

2.2. CIA TRIAD

One of the older and most well-known principles of information security is known as the CIA Triangle (Whitman & Mattord, 2011), or CIA Triad (Au *et al.*, 2016). The CIA in this instance is an abbreviation for the three corners of the triangle: Confidentiality, Integrity, and

Availability, which are considered to be the key characteristics of information (Posthumus & von Solms, 2004). Each of these three concepts will now be discussed in turn. An illustration of the triangle is shown in Figure 2.1.

2.2.1. CONFIDENTIALITY

According to ISO/IEC 27000:2016 (2016), Confidentiality is defined as the property that information is only made available to authorised individuals, entities and processes.

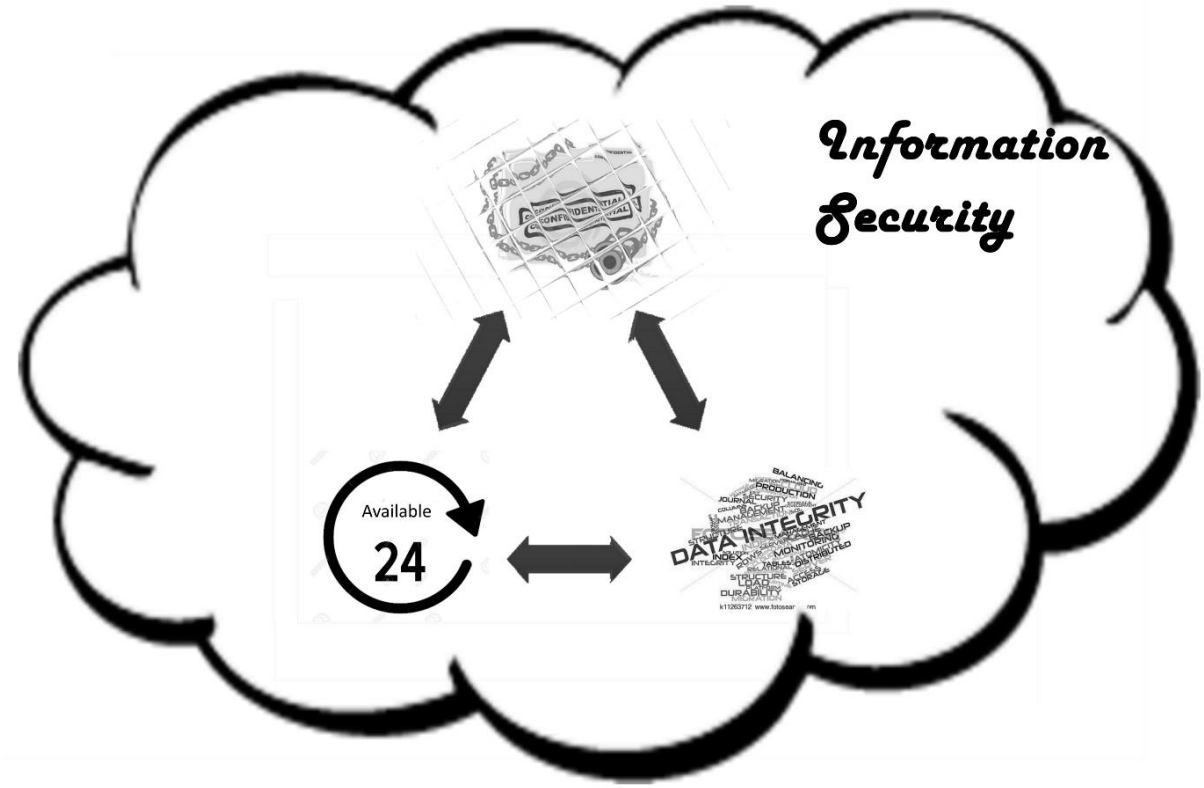


FIGURE 2.1 CIA TRIAD REPRESENTING CONFIDENTIALITY, INTEGRITY AND AVAILIBILITY

From an organisational standpoint, confidentiality can therefore be described as organisational privacy. Whitman and Mattord (2011) mention that the confidentiality of information can be protected using any number of measures, such as information classification, secure document storage, application of general security policies, and the education of information custodians and end users. Confidentiality is therefore dependent upon the flow of information, as both of the concept of confidentiality, and the measures used to protect it, show that inaccessible information has a high level of confidentiality. It

furthermore stands to reason that encrypted information, even if leaked to an unauthorized entity, has not necessarily lost its confidentiality. Because of this, secret information, even if obtained by an unauthorized party, will remain confidential so long as it is inaccessible (Posthumus & von Solms, 2004). It should also be kept in mind that data collected as a result of user interaction might also have to be kept confidential (Wu *et al.*, 2017).

According to a study conducted by Tankard (2017), confidentiality is crucial in handling big data sets. In this study two core approaches are proposed in order to ensure data confidentiality: firstly, to encrypt the data both in transit and at rest and, secondly, to design the databases used with confidentiality in mind. This implies that confidentiality can be addressed in an active manner, such as by using encryption, or in a more passive way, such as designing systems with confidentiality in mind.

One of the ways in which data confidentiality can be protected is described by Eloff and Eloff (2005). This method, which involves the implementation of an Information Security Architecture (ISA), is ultimately a management process that aims to provide a systematic approach to adapt to new risks on an ongoing basis. This serves to substantiate the notion that management and managerial processes have a substantial impact on information security.

In the work done by Olawumi *et al.* (2017) the consequences of data confidentiality being compromised is clearly demonstrated. Olawumi *et al.* (2017) demonstrated how an eavesdropping attack was used to obtain confidential information, which could then be used to launch more in depth attacks. Some of the types of data that were identified as being vulnerable to this type of attack were e-mail messages, Internet surfing details, and phone conversations. As any of these data types might contain crucial information, they should all be considered sensitive and therefore be protected.

Shedden *et al.* (2016) point out that there are typically three instances where the confidentiality of data might be compromised: when the data is at rest, when it is in transit, and when it is in use. This is further expanded upon by illustrating how certain backup practices may increase the number of possible threats to the confidentiality of the backed up data. In the case mentioned by Shedden *et al.* (2016), an individual working at the company used personal devices to make data backups that were stored at his office, private residence and in his car. Whilst true that the data has a greater availability when backed up in this manner, the greater number of access points means that there are a greater number of attack vectors that can be used to compromise the data's confidentiality. This ultimately means that a balance between confidentiality, integrity, and availability is crucial in maintaining information security.

In conclusion, confidentiality deals with how secret the information being protected is. It is a crucial aspect that must be addressed, as breaches in confidentiality can have disastrous implications. As far as this study is concerned, confidentiality deals broadly with the flow of

information: if usable information is transferred to any entity that is not supposed to have access to that information, then there is a problem with regard to confidentiality.

2.2.2. INTEGRITY

Integrity, as defined by ISO/IEC 27000:2016 (2016), is the property of information that describes how accurate and complete the information is. Whitman and Mattord (2011) expand upon this concept by explaining that information has integrity when it is whole, complete, and uncorrupted. This is further supported by Posthumus and von Solms (2004) who state that by maintaining the correctness and comprehensiveness of an information resource, its integrity is preserved. Pfleeger and Pfleeger (2006), focusing more on attacks, state that assets have integrity when they can only be modified by authorised parties, or in authorised ways. Integrity is therefore focused on maintaining information: information should be kept valid by only modifying it in valid ways, expanding on it using valid data, and using new and valid data to keep it up to date. It should be pointed out that, even if the original information was incorrect, data integrity requires that the incorrect information be considered valid.

In Tchernykh *et al.* (2016), integrity is defined as the assurance that information is trustworthy and accurate. This ultimately means that ensuring data integrity involves maintaining data consistency, accuracy, and trustworthiness.

Ensuring the integrity of data can sometimes be complicated by any number of factors. Whilst cloud computing offers a number of advantages, Yu *et al.* (2016) mention that one of the reasons why users tend to avoid cloud services is that they are concerned about the integrity of their outsourced files. These files may be arbitrarily deleted by the cloud provider, or the backup policies in place might not be sufficient to ensure data integrity. This may be further compounded by the fact that cloud providers are not immune to hardware and software failures. While these providers have designed a number of protection methods, these methods are not necessarily adequate (He *et al.*, 2017).

Integrity, in closing, is a very important aspect of information, as it determines how accurate and complete the information is and, ultimately, how useful it is. Inaccurate and incomplete information, while not necessarily useless, can have a negative impact on the information's owners or organisation when it is used. As such, in this study, integrity will also be understood to refer to the trustworthiness of the information, in addition to its completeness and accuracy.

2.2.3. AVAILABILITY

The definition of Availability, as given by ISO/IEC 27000:2016 (2016), is the property of information that describes how accessible and usable it is upon demand by an authorized entity. This, by extension, means that any information that is available can be accessed without interference or obstruction in the required format (Whitman & Mattord, 2011), and in a timely manner (Posthumus & von Solms, 2004). Maintaining availability is one of the great challenges in information security (Pfleeger & Pfleeger, 2006), as it goes hand in hand with confidentiality. Indeed, it seems as though controlling availability is one of the primary methods of maintaining confidentiality. Unfortunately, this means that certain strict methods of confidentiality control may have a negative impact on availability, as confidentiality controls such as encryption may make timely access infeasible. This once again demonstrates that maintaining information security requires a balance between availability and confidentiality, and ultimately integrity as well (Karlsson *et al.*, 2017; Olivier, 2002).

Cloud computing demonstrates how focussing on availability may compromise the security of the data. As pointed out by Halabi and Bellaiche (2017), companies that have adopted cloud computing have experienced an increase in the availability of their data, but referred to statistics that indicate that information security concerns are one of the main impediments to adoption of cloud computing.

Availability, in conclusion, is one of the most difficult aspects of information to protect. By applying controls to protect integrity and confidentiality, availability will almost always be impacted negatively. Because of this, this study will only focus on availability in a broad sense: if an entity has access to information, it will be assumed that the information has at least some measure of availability.

2.3. INFORMATION SECURITY RISK MANAGEMENT

ISO/IEC 27000:2016 (2016) goes to some length in attempting to define risk. According to this standard, risk is the effect of uncertainty on objectives. This definition, being somewhat vague, is further expanded upon by clarifying that information security risk is the potential that threats will exploit vulnerabilities of the information assets and thereby cause harm to an organisation. From this definition, it is clear that any information security risk is an inherent exploitable vulnerability that, if exploited, will cause harm to an organisation. It is important to make this clarification as, logically speaking, not all vulnerabilities are necessarily exploitable or harmful. This conclusion is made based on the fact that an exploit involves information security assets being compromised. As not all information security

assets are always directly valuable, and not all organisations need to concern themselves with information security risk management, the loss of information security assets might not automatically equate to harm to the organisation (Broderick, 2001). ISO/IEC 15408-1:2009 (2009) mirrors this sentiment in its description of the relationships between owners, assets, risks, threats, threat agents, and countermeasures. These relationships, and the relevant ontology, is shown graphically in Figure 2.2.

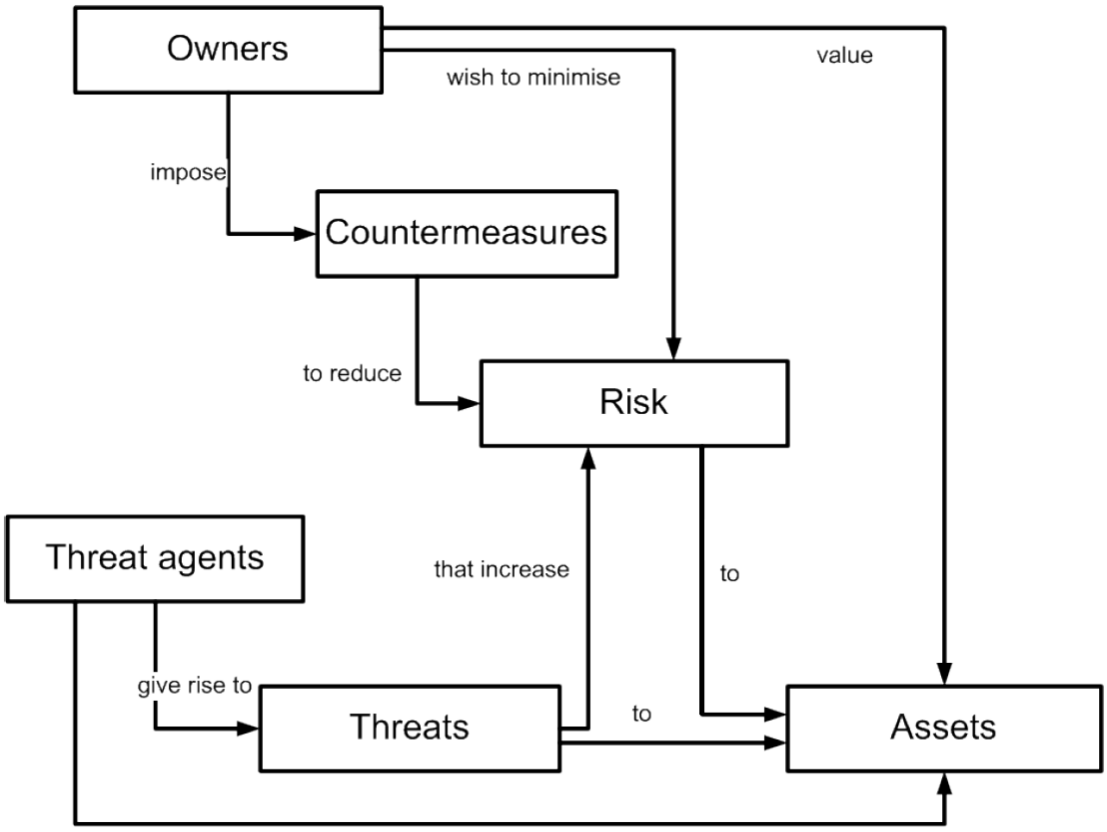


FIGURE 2.2 RELATIONSHIP BETWEEN THE VARIOUS ACTORS IN INFORMATION SECURITY RISK (ISO/IEC 15408-1:2009 (2009))

In order to illustrate the use of an ontology, such as the one shown in Figure 2.2, consider the following example: a software development company, which own and *value* a number of confidential, proprietary software optimisation algorithms, i.e. *assets*, have become aware of new vulnerabilities in their security systems. While not necessarily a problem, the vulnerabilities of the system they are known to have, have been published extensively and, as such, a number of hackers are known to have knowledge of the vulnerabilities. These *threat agents give rise to* the *threat* that the system might be penetrated and that in turn *increases* the *risk to the assets*. As the owners *wish to minimise the risk to the assets,* they appoint a task force to address the vulnerability in the system. Following the recommendations of this task force, the owners *impose* new *countermeasures that aim to reduce the risk to the assets*.

In addressing information security risk, both ISO/IEC 27000:2016 (2016) and Whitman and Mattord (2011) describe a number of risk treatment strategies. The strategies are:

- Defend against the risk by implementing controls that prevent exploitation of the vulnerability that is associated with the rest;
- Take the risk, or even increase the risk, in order to pursue an opportunity;
- Completely remove the risk;
- Change the likelihood of the risk having an impact, i.e. reduce the risk of the vulnerability being exploited;
- Change the consequences of the vulnerability being exploited. This strategy is typically subdivided into one of four categories:
    - Risk mitigation;
    - Risk elimination;
    - Risk prevention; and
    - Risk reduction.
- Share the risk with, or transfer the risk to, a third party; and
- Accept the risk as being part of the operating environment.

The abovementioned strategies are implemented using countermeasures, as shown in Figure 2.2. These countermeasures, also called controls, are typically security mechanisms, policies, procedures, devices or practices that modifies the risk to information assets (ISO, 2016; Whitman & Mattord, 2011; Pfleeger & Pfleeger, 2006). As mentioned in Section 2.2, the key aspects to information security are those represented by the CIA triad. Depending on the controls used, the risk to one or two of these aspects may be increased in order to change the risks associated with a third aspect. An example of this would be standard database access control: by enforcing access controls to the database, such as limited access and table locks, the risk to both the confidentiality and the integrity of the information in the database is reduced. As these controls may both restrict access and have an impact on the amount of time required to access the data, the availability of the data is negatively affected. The statement, as alluded to in ISO/IEC 27000:2016 (2016), that controls might not have the desired effect, is therefore reasonable.

The selection and implementation of controls are typically done during the third phase of information security risk management planning, also referred to as the *risk control* phase. The first phase, *risk identification*, typically involves identifying the information security assets, as well as the threats to them. The second phase, *risk analysis*, focusses on analysing the threats identified during phase one, as well as analysing the vulnerabilities to the information security assets. This process is shown in Figure 2.3 (Yue *et al.*, 2007). From its overall structure, it can be deduced that this process aims to investigate the environment in order to identify risks, determine what threat those risks pose, and then implement countermeasures to control the identified risks.

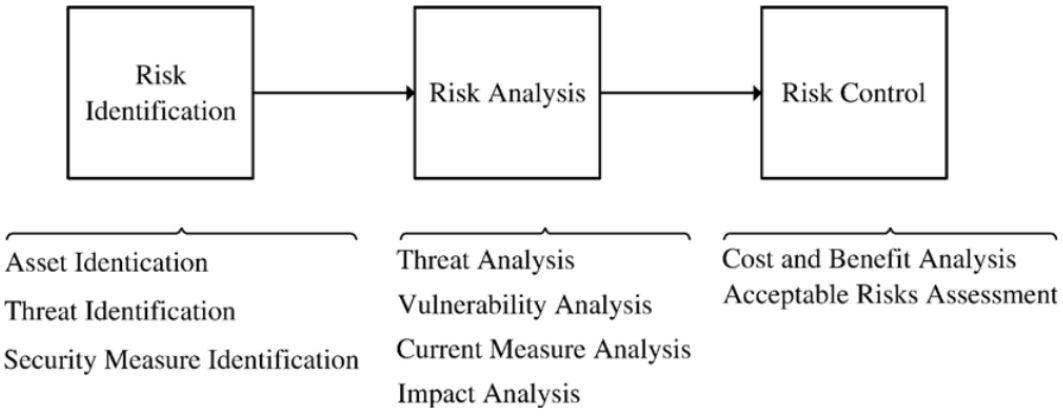


FIGURE 2.3 RISK MANAGEMENT PROCESS (YUE *ET AL.*, 2007)

In order to provide a better overview of what the risk management process entails, each of these three phases will now be briefly discussed in turn.

2.3.1. RISK IDENTIFICATION

As mentioned, the risk identification phase in the risk management process is concerned with identifying valuable information security assets and their associated risks. The first step in this phase, namely the **identification of information security assets**, is exceptionally important as properly identifying the crucial assets can make a significant impact when it comes to identifying threats later on (Kong *et al.*, 2017; Beckers *et al.*, 2011). Identifying these assets is not necessarily a straightforward process however, as certain organisations have a too narrow definition of the concept of information security assets (Shedden *et al.*, 2016). This is compounded by the fact that an asset's value is typically not only determined by the three CIA attributes of confidentiality, integrity and availability (Huang *et al.*, 2019), but also by how its loss or disruption will impact the organisation (Fernandez & Garcia, 2016; Suh & Han, 2003). Calculating an asset's value may be further complicated by the fact that not all of an asset's properties are necessarily equally important (Tondel *et al.*, 2008). Because of this, various methods related to identifying information security assets are actively being researched. Indeed, at time of access on 2 June 2017, a Google Scholar search using the keywords "information security "asset identification", and filtered to only return results dated since 2017, returned 59 results. As this study is not focussed on identifying assets, an in-depth discussion of the techniques used to identify assets are outside the scope of the study. For further information regarding information asset identification, the work done by Fernandez and Garcia (2016), Shedden *et al.* (2016), Beckers *et al.* (2011), and Palmer and Potter (1989) can be consulted.

Threat identification generally follows the risk identification step. Unlike risk identification, however, threat identification is a continuous process that re-evaluates threats on a regular basis. As a result, threat identification is usually conducted using threats and vulnerabilities that have been exploited in the past. The information concerning these past threats can come from a number of sources, such as security software logs, access logs, reports on halted security breaches, etc. The frequency of these threats should also be taken into consideration, as the frequency will help to determine the probability of a risk being realised during the Risk Analysis phase, as well as to select the appropriate strategies and controls during the Risk Control phase (Palmer & Potter, 1989).

The importance of conducting threat identification during the first risk management phase is demonstrated by the significant number of threats that exist. Most of these threats, such as viruses, spyware, hackers, Trojan horses, phishing, pharming, and social engineering, each have different attack approaches and philosophies, and therefore require different defence approaches (Ahmad, 2012; Whitman & Mattord, 2011). The specific defence approaches impact strongly on the controls used to protect the identified assets. Section 2.3.3 deals with the risk control phase, wherein these controls are discussed.

2.3.2. Risk Analysis

The risk analysis phase, as the name implies, is the phase wherein the risks that have been identified are analysed. Some of the analyses that can be conducted are (Yue *et al.*, 2007):

- **Threat analysis:** the actual threat of a risk being exploited;
- **Vulnerability analysis:** the practicable vulnerability associated with the risks;
- **Current measure analysis:** the controls already in place to manage the risks; and
- **Impact analysis:** the impact it will have should a risk be exploited.

A number of techniques exist that can be used to analyse these risks. Within the specific context of information security, four of these techniques will be discussed briefly. The four techniques that will be discussed are CORAS, CIRA, ISRAM, and ISRA. A comparison of the four methods is shown in Table 2.1 in Section 2.3.2.5 (Agrawal, 2017).

2.3.2.1. CORAS

The CORAS method, named after the Centre for Operational Research and Applied Statistics at Salford University in the United Kingdom which developed the method, is a qualitative model-based risk analysis method that is based on Unified Modelling Language, wherein

diagrams are used to illustrate relationships and dependencies (Shukla & Kumar, 2012; Lund *et al.*, 2011; Den Braber *et al.*, 2007; Aagedal *et al.*, 2002). CORAS functions in two main phases. In the first phase, a broad investigation is done to gain a general insight into the risks being evaluated. This first phase therefore corresponds to a low-level investigation into the risk "target area". The purpose of this phase is to determine the scope and purpose of the overall analysis. The second phase of the method evaluates each of the identified risks with an in-depth evaluation according to the scope as determined during the first phase. The basic structure of the CORAS method is shown in Figure 2.4.

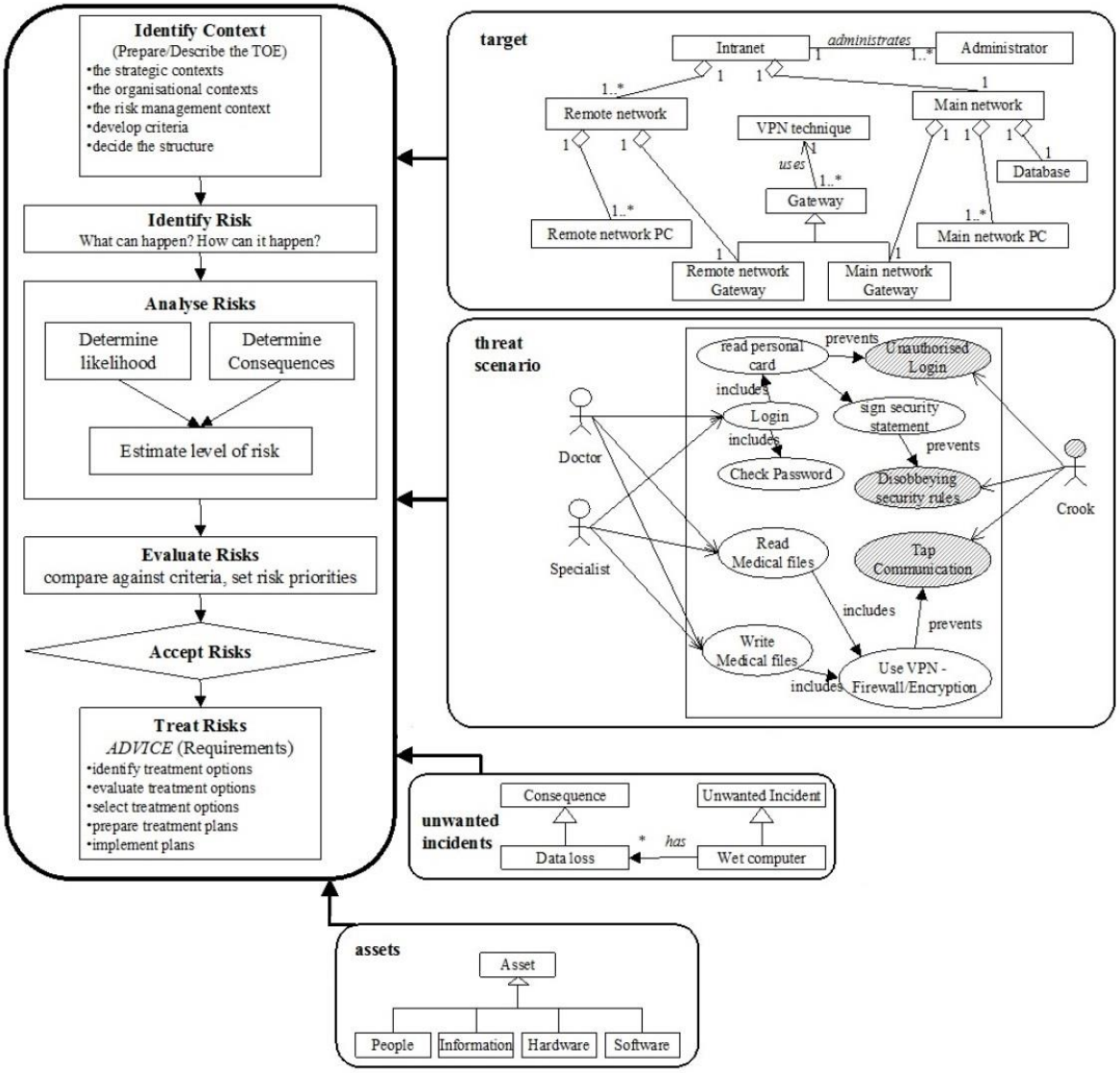


FIGURE 2.4 GRAPHICAL REPRESENTATION THE CORAS METHOD (BRABER *ET AL.*, 2003)

Figure 2.4 also broadly references the Risk Identification and Risk Control phases, and how they integrate with the CORAS risk evaluation methods. The diagram, taken directly from

Braber *et al.* (2003), uses a threat scenario wherein a medical doctor and specialist have access to the same system in order to access a patient's medical files. The threat involves the existence of a "crook" that seeks to gain access to the same information without authorisation. The CORAS method, which is shown on the left-hand side in Figure 2.4, is shown to make use of "target", "threat scenario", "unwanted incidents" and "assets" evaluations, all shown on the right-hand side. These evaluations all form part of the *Risk Identification* phase as shown in Figure 2.3.

CORAS, in conclusion, is a qualitative model-based technique, based on UML, which can incorporate threat scenarios in order to evaluate and identify risks. While the technique is scalable and can be used in both large and small organisations, the technique is time-consuming and requires specialist support, which may limit its use in small organisations.

2.3.2.2. CIRA

The Conflicting Incentives Risk Analysis (CIRA) method, developed by Rajbhandari and Snekkenes (2012), is a qualitative method that evaluates human-related risks by identifying actions, stakeholders and expected consequences associated with risks (Rajbhandari & Snekkenes, 2013). The stakeholders themselves are subdivided into those individuals who own the risks, and those who own the strategies developed to control them. The core premise of CIRA is that, by evaluating the incentives each of the stakeholders have with regard to controlling the risks, it is possible to identify individuals who pose a greater risk to the information assets being protected. The CIRA procedure is shown in Figure 2.5.

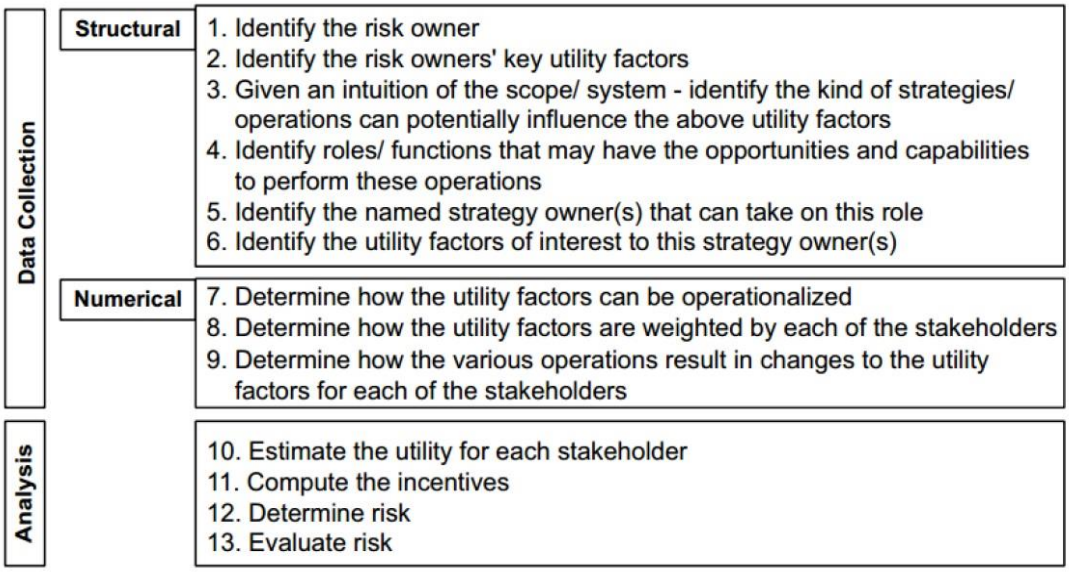


FIGURE 2.5 THE CIRA PROCEDURE (WANGEN, 2015)

Wangen (2015) expands upon this by incorporating the work done by Mitchell *et al.* (1997), wherein a prioritisation scheme is introduced that clarifies how managers prioritise the various stakeholders they interact with. This prioritisation scheme is shown graphically in Figure 2.6. In this scheme, each stakeholder has three largely subjective properties that determine its saliency: **power**, which is the ability of a stakeholder to force his will upon other members within the organisation; **legitimacy**, which describes the relationships the stakeholder has with other members of the organisation, and the **urgency** of the stakeholder's claims.

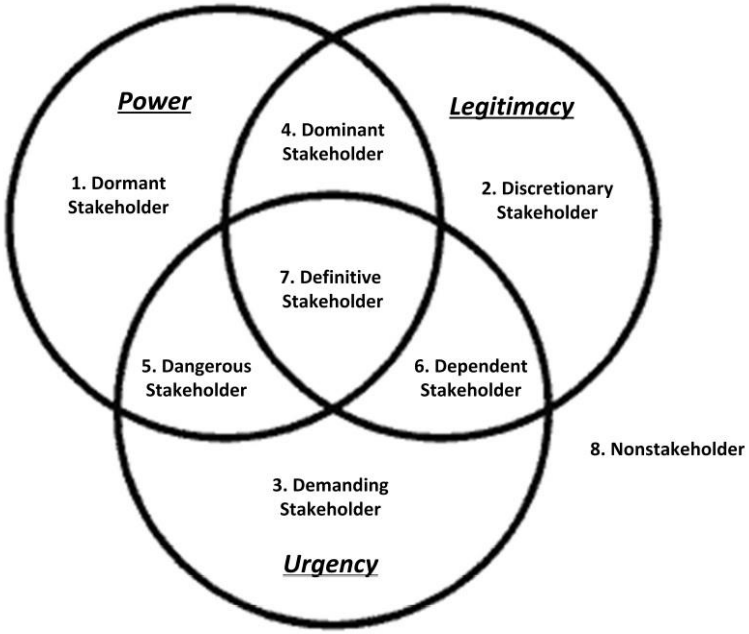


FIGURE 2.6 STAKEHOLDER PRIORITISATION SCHEME (WANGEN, 2015; MITCHELL *ET AL.*, 1997)

CIRA is therefore very useful in evaluating how risks are addressed in practice, as the impact of the various stakeholders on the risks are taken into account. CIRA, like CORAS, is a qualitative technique, but differs from CORAS in that it takes a much less technical approach. This technique does, however, not scale well and, because of the stakeholder evaluations, requires expert input in order to be used effectively.

2.3.2.3. ISRAM

The Information Security Risk Analysis Method (ISRAM), first proposed by Karabacak and Sogukpinar (2005), is a quantitative, survey-based approach that involves most, if not all, members of staff in evaluating security risk. The method utilises two independent types of

surveys, one evaluating the consequences of a risk being realised and the other evaluating the probability of the risk occurring, to assign a numerical risk level to each of the risks. The method is therefore basically used to prepare and conduct surveys. The ISRAM method can be used to evaluate complex information systems and, because the method uses independent surveys, the method naturally limits evaluation bias. The basic functioning of the ISRAM method is shown in Figure 2.7. Step 1 ties into the Risk Identification phase shown in Figure 2.3, as threat identification is crucial in becoming aware of information security problems. Steps 2 to 4 deal with the preparation of the surveys; on the left-hand side of Figure 2.7 the steps involved in preparing the probability surveys are shown, whereas the right-hand side shows the preparation of the consequences survey. The risk table mentioned in Step 4 provides for a way to assign a single digit numerical value to a survey result; an example of such a risk table is shown in Figure 2.8.

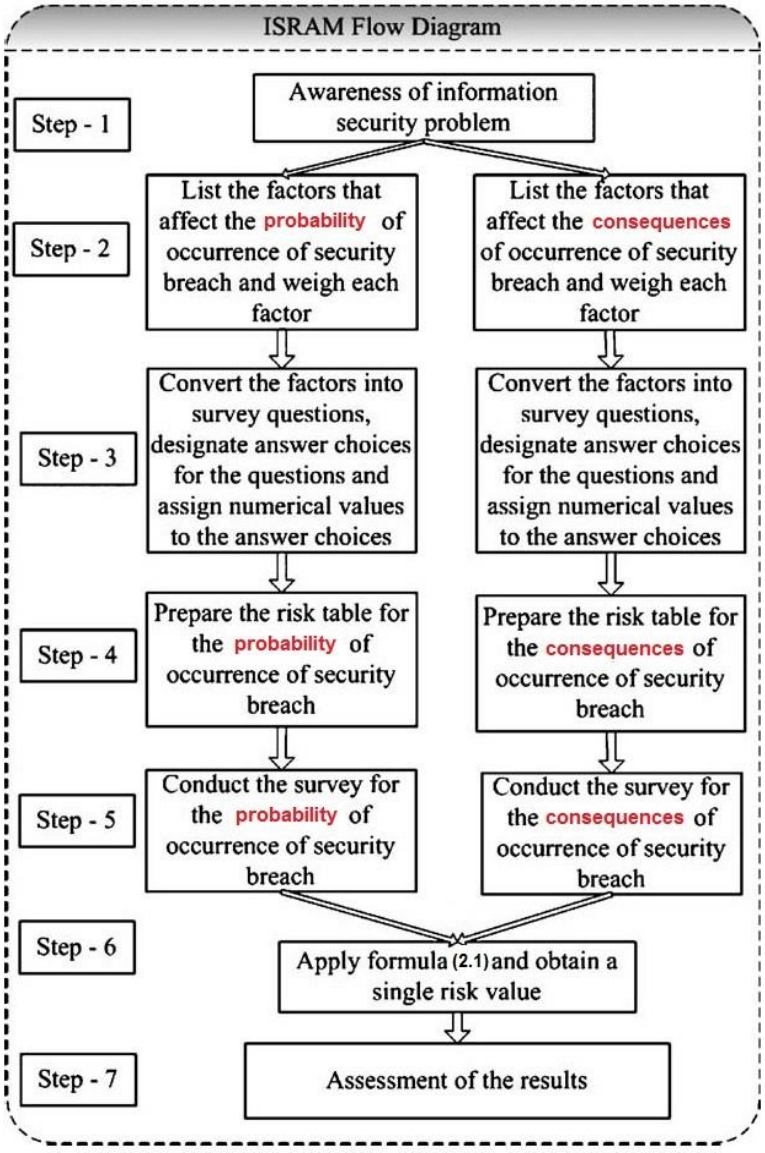


FIGURE 2.7 ISRAM FLOW DIAGRAM (KARABACAK & SOGUKPINAR, 2005)

Risk table for the survey of probability of infection parameter

Qualitative scale	Quantitative scale
Very low probability	1
Low probability	2
Medium probability	3
High probability	4
Very high probability	5

FIGURE 2.8 EXAMPLE OF AN ISRAM RISK TABLE (KARABACAK & SOGUKPINAR, 2005)

Once these steps have been completed, the results of the surveys need to be evaluated mathematically. In step 6, the following formula is used to calculate a numerical value for the overall risk:

$$Risk = \left(\frac{\sum_m [T_1(\sum_i w_i p_i)]}{m}\right)\left(\frac{\sum_n [T_2(\sum_j w_j p_j)]}{n}\right) \quad (2.1)$$

where

- i is the number of questions for the survey of probability of occurrence, determined at Step 2;
- j is the number of questions for the survey of consequences of occurrence, determined at Step 2;
- m is the number of participants who participated in the survey of probability of occurrence, a number that becomes definite at Step 5;
- n is the number of participants who participated in the survey of consequences of occurrence; this number becomes definite at Step 5;
- w_i and w_j are the weights of the questions i and j, determined at Step 2;
- p_i and p_j are the numerical values of the selected answer choice for question i and question j, determined at Step 3;
- T_1 is the risk table for the survey of probability of occurrence, constructed at Step 4; and
- T_2 is the risk table for the survey of consequences of occurrence, constructed at Step 4.

ISRAM, in conclusion, is a quantitative method that utilises questionnaires in order to evaluate both the probability and the consequences of risks being realised. As this method

relies on a broad evaluation, rather than a management-only evaluation, it is likely that the use of ISRAM will produce a much clearer image of the risks an organisation face. Also, because of the relative simplicity of the method compared to CORAS and CIRA, it is possible to utilise CIRA without specialist training.

2.3.2.4. ISRA

The Information System Risk Analysis method (ISRA), first described by Suh and Han (2003), is a quantitative method that was developed in order to include the impact of loss due to suspended operations in the risk analysis process. Originally based on a business model, the method utilises the Analytic Hierarchy Process (AHP) to determine the relative importance of a number of operational functions. The AHP is a multi-criteria technique that assigns weights to the various criteria used during evaluation and produces a single weighted list of values that can be used in further analyses (Tadisina *et al.*, 1991). The results of the AHP are used in a more classic analysis, whereby assets and risks are identified and associated. During the final stage of the analysis, the impact of each of the business functions being disrupted is associated with the identified information security assets to obtain a quantitative impact value.

ISRA is therefore a quantitative method that incorporates the possible losses an organisation might face when a risk is realised. The technique does not scale well, unfortunately, and is only really useful for risk evaluation in large organisations. It also requires specialist level effort and, as with most techniques that incorporate the AHP, is very time consuming (Ishizaka & Lusti, 2004).

2.3.2.5. COMPARISON BETWEEN CORAS, CIRA, ISRAM, AND ISRA

The four methods discussed in Section 2.3.2, namely CORAS, CIRA, ISRAM, and ISRA, each have a different approach when dealing with risk analysis. CORAS, for instance, has a very structured approach that follows well-known project management principles (Bentley & Whitten, 2007), whereas CIRA focusses on the relationships between the various stakeholders in an organisation. ISRAM, by comparison, uses surveys to determine the probability and impact of various risks being realised, whilst ISRA calculates the effect of a disruption on the organisation itself. Each of these techniques are therefore unique in their focus areas, and selecting one of the techniques is largely dependent upon the situation within the organisation, the type of organisation, the expertise available, and management preferences. Table 2.1 provides a tabulated comparison between these four methods, with specific reference to the purpose of the model, the input required, the effort needed, the

outcome obtained, the scalability of the method to smaller and larger organisations and finally, the pros and cons of each method.

TABLE 2.1 COMPARISON BETWEEN CORAS, CIRA, ISRAM, AND ISRA (AGRAWAL, 2017)

Criteria	CORAS	CIRA	ISRAM	ISRA
Methodology	Qualitative	Qualitative	Quantitative	Quantitative
Purpose	Model-based method to conduct risk analysis through strong business and asset orientation	Non-Technical risk analysis and decisions	Analyse the risks at complex information systems by allowing the participation of managers and staff	Calculate Annual loss expectancy
Input	Direct asset, vulnerability, threat scenario, risk, risk evaluation matrix, likelihood scale	Stakeholders, strategies, utility factors, weight, initial values	Questions for the survey, weight of the question, risk table	Business function, assets, importance of asset
Effort	Specialist level and time-consuming	Specialist level and time-consuming	Standard level, less time-consuming	Specialist level and time-consuming
Outcome	identify potential risks and assess potential treatments for unacceptable risks	Strength of Stakeholders incentive or changes in utility	Single numeric value for representing the risk, Annual loss expectancy	Calculation of loss from disruption of operations in determining the value of IS assets
Scalability	Yes	No	Yes	No
Pros	Free tool support, Facilitates iterative communication and collaboration between various stakeholders, suitable for security-critical systems and large organizations	Tool support, useful for both small and large organizations, Detects human risks in Information security	It is self-directed i.e. can be carried out by small teams of the organization's own employees, Ease of use, does not require dedicated tools	Based on a business model where importance of business functions and assets are evaluated, especially useful for large enterprise organizations
Cons	Requires expert knowledge from various backgrounds, extra efforts required, Time-consuming	Expert knowledge required, Full assessments of the method can be time-consuming and overly complex	Preparation phase and initial data collection phase can be time-consuming, absence of any expert can make the whole task complex	Expert knowledge required, Full assessments can be an extremely lengthy process and time-consuming

As the focus of this study is to develop a method that exploits Social Network Analysis within the context of information security risk management, the CIRA method mentioned earlier needs special mention. The CIRA method, as discussed, makes use of the broad relationships between the various stakeholders within an organisation to determine the probability of various risks being realised. The SNA method presented in this study evaluates the relationships between the members of an organisation quantitatively in order to identify at-risk entities. The two methods therefore have similar approaches, in that both methods evaluate the relationships within the organisation. The existence of the CIRA method, which investigates relationships to evaluate risk, therefore supports the supposition that the use of SNA is valid in this context.

2.3.3. Risk Control

Once the risk identification and analysis phases have been concluded, the risks that have been identified have to be managed in some way. As mentioned earlier in this section, there are a number of strategies that can be employed to manage the identified risks. As some of the actions suggested, such as taking the risk, avoiding the risk, and mitigating the risk are highly dependent on the risk being evaluated and the information security policies in place, this section deals primarily with information security controls and the factors that influence their selection and implementation.

All InfoSec controls can be placed into one of two categories: **technical**, which contains physical and logical controls, and **human**, which contains a number of controls such as education, awareness, and operational controls (Whitman & Mattord, 2011; Theoharidou & Gritazalis, 2007). Each of these categories will now be briefly defined in turn. As there are likely to be thousands of controls and types of controls in each category, the focus is on clarifying each of the categories based on their individual properties. The controls discussed here can be categorised as either technical or human (AlHogail, 2015).

2.3.3.1. Technical controls

Technical controls, in a broad sense, are controls that implement certain tools and techniques to address information security risk (Ma *et al.*, 2009). There are primarily two different types of technical controls, namely Physical and Logical.

Physical controls deal with the physical aspects of security, hence the name. Physical controls, such as security doors, cameras, keypads, etc., are used to control and monitor the physical access to information security assets (Whitman & Mattord, 2011). Physical security is also concerned with protecting physical assets from damage. Such damage can be caused by acts of terrorism, bombing, fire, floods, natural disaster, etc. It is important to protect

physical assets against these disasters, as the loss of a physical asset may well equate to the loss of an information security asset (Palmer & Potter, 1989). Disaster recovery plans, that describe how disasters such as fires and floods should be handled, can also be considered physical controls (Smith *et al.*, 2018). Business continuity plans, which detail the steps that need to be taken to recover after a disaster, may also incorporate physical controls. However, because of the broad scope modern business continuity plans need to have, it is highly likely that such a plan will reference other control types as well (Supriadi & Pheng, 2018). As physical controls tend to address threats such as theft and physical damage, these controls are mostly preventative in nature and therefore aim to either eliminate or reduce the risks involved.

Logical controls are controls that deal with the software aspects of protecting security assets (Whitman & Mattord, 2011). These controls usually involve the use of access control methods such as passwords, log-in privileges, database integrity systems, intrusion detection systems, and anti-malware systems. Controls that are used to protect against hardware based attacks such as *Spectre* and *Meltdown* (Watson *et al.*, 2018) are also technical in nature, as these controls deal with data rather than physical access. Logical controls address a wide range of threats such as malware and hackers, and therefore tend to incorporate a wide variety of risk management approaches, from reducing the impact of a threat being realised, to eliminating the threat entirely.

2.3.3.2. HUMAN CONTROLS

Human controls are differentiated from technical controls only in that they deal specifically with addressing risks caused by human nature or negligence. These controls typically involve the management procedures, as well as continuous education and awareness, and are often largely overlooked during security planning (Wiant, 2005).

Continuous Education, as the name implies, involves a process whereby critical users of a system are continuously educated and trained in information security (Whitman & Mattord, 2011). This control provides a human based security aspect for a system.

Security awareness, in contrast, usually does not involve formal training or education, and instead mostly utilises less formal techniques such as newsletters, posters, etc. to inform system users about security topics (Whitman & Mattord, 2011).

Security Policies are used to describe appropriate behaviour for users of an information security system (Moody *et al.*, 2018). As these types of policies are determined by management, and may influence processes, they typically overlap with operational controls.

Security Standards are used to describe security systems in general. These descriptions typically include standard definitions, generic processes, and specifications for certain types of methods. It is important that these standards are correct and implemented correctly, as

this ensures that improving and maintaining information security systems is a simple and reliable process (Siponen & Willison, 2009; von Solms, 1999).

Operational controls is a relatively broad category that includes all controls that in some way affect, or are affected by, human behaviour (Whitman & Mattord, 2011). These controls may be security policies and procedures, the administration of controls, allocation of access restricting roles such as user and administrator, etc. The implication of this is that operational controls are deeply connected to management; all technical and physical controls have to be managed in order to be effective (Khajouei *et al.*, 2017), and these management processes can all be categorised as operational controls. It is therefore reasonable to state that a significant part of information security risk management lies with operational controls and, by association, with managing human behaviour. As operational controls tend to deal with threats such as social engineering, the approaches used tend to be as broad as with logical controls.

The implementation of the various types of controls mentioned above is strongly influenced by the two Risk Control methods shown in Figure 2.3. The first of these methods, **cost-benefit analysis**, is a rational economic technique that has received wide-scale acceptance in the form of the Gordon-Loeb model (Gordon *et al.*, 2018; Weishäupl *et al.*, 2018). The basic principle of the Gordon-Loeb model is that the assets being protected should be protected according to their rational value, whilst acknowledging that it is likely impossible to completely protect the assets. As such, the costs involved in implementing the necessary controls should be rational and based on a "good enough" principle. The application of a cost-benefit analysis will therefore have a dramatic effect on the controls that are finally implemented.

The second technique, **acceptable risks assessment**, is generally used in conjunction with cost-benefit analysis. The acceptable risks assessment is used to distinguish between those risks that should be mitigated, removed, or accepted, by comparing the outcome of the application of a certain control with the outcomes required by management or an information security policy (Shukla & Kumar, 2012).

Section 2.3 focussed on some of the various aspects of the risk management process. Whilst this process is usually adapted to suit specific situations, the essence of information security risk management can be summarised into five points (Rosenquist, 2016):

- *Know* the value of your data;
- *Know* who has access to the data;
- *Know* where the data is;
- *Know* who is protecting the data; and
- *Know* how well it is protected.

By addressing these five points, which is referred to as the **Five Knows of Information Security**, it is possible to address most, if not all, of the crucial risk management aspects discussed in this section.

In summary, Section 2.3 focussed on the information risk management process. The discussion of this topic mainly centred upon the three phases shown in Figure 2.3, namely the Risk Identification, Risk Analysis, and Risk Control phases. It is however important to note that the efficacy of these three phases are strongly influenced by the people involved, hence the strong focus on the human aspect of information security in the literature. The next section focusses on discussing these human aspects.

2.4. HUMAN ASPECTS OF INFORMATION SECURITY

In the field of information security, one of the primary success factors is properly addressing the human aspect of information security. Past research has shown that a balanced approach wherein both technological and human aspects are addressed is crucial to maintaining information security (Soomro *et al.*, 2016; Parsons *et al.*, 2014). Despite repeated campaigns to educate users on information security however, a significant number of users still engage in risky behaviour (Byrne *et al.*, 2016) and are still considered the weakest link in information security (Arachchilage & Love, 2014). In this section, some of the aspects that influence this human aspect, such as security culture, knowledge, behaviour, and attitude are discussed in greater detail.

2.4.1. INFORMATION SECURITY CULTURE

Information security culture broadly determines how the users of information security assets interact with the controls meant to protect those assets (Da Veiga & Eloff, 2010). By not having a good information security culture, users make mistakes in applying the controls, which may negatively impact the information security assets (Sohrabi Safa *et al.*, 2016; Metalidou *et al.*, 2014). This information security culture is typically influenced by the visible organizational structures and processes, also called artefacts, the values of the people that influence and are influenced by the culture, the shared assumptions of the community, and finally the underlying information security knowledge (Van Niekerk & von Solms, 2010). These various aspects each influences the information security culture on different levels. The levels of information security culture are represented graphically in Figure 2.9. Information security culture is also strongly influenced by the traits of the individual members of the community (Gratian *et al.*, 2018), thus demonstrating that

information security culture is determined by the members of the community: the personal traits of the members of the community influences the security culture, whilst the security culture influences the behaviour within the community.

A study done by Albrechtsen (2007) aimed to determine what the views were that various people and groups had with respect to information security. As these views influence both the knowledge basis for information security through willingness to learn, as well as the espoused values of the members of the community, it is important to mention the various points that the study identified with regard to information security culture. These points are only listed along with a reasonable declaration of the level of culture they influence. The points, in no particular order, are:

- Information security awareness (knowledge);
- Users are motivated for individual information security work, but do not have clear instructions on what to do (artefacts);
- There is a latent conflict of interest between information security and functionality (tacit assumptions); and
- A lack of exposure to established information security policies and guidelines for expected behaviour (knowledge, artefacts).

Based on the literature mentioned above, it becomes evident that cultivating an effective information security culture is both crucial and difficult. This is further supported by Martins and Eloff (2002), who support the standpoint that information security culture can take years to form. Furthermore, like all cultures, information security culture can be difficult to change (Schneider *et al.*, 1996).

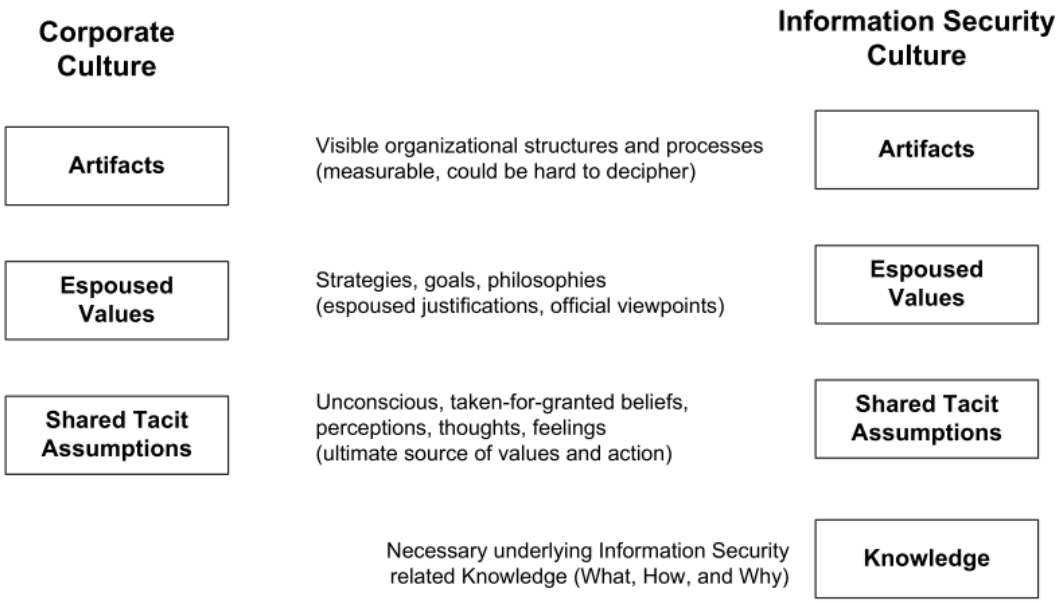


FIGURE 2.9 LEVELS OF IS CULTURE (VAN NIEKERK & VON SOLMS, 2010; SCHEIN, 2009; SCHLIENGER & TEUFEL, 2003)

2.4.2. Information Security Knowledge

As shown above, information security knowledge is deeply linked to information security culture. However, because of the importance of this aspect, it is discussed separately from information security culture.

Information security knowledge, as it is understood in this study, is knowledge of various information security threats and countermeasures. Information security knowledge would therefore include such knowledge as hacking, malware, anti-malware, privacy protection, etc. One of the best known methods used to improve information security knowledge is the use of awareness programs (Aloul, 2012; Chen *et al.*, 2008; Thomson & von Solms, 1998). There are however a number of significant drawbacks to these awareness programs, e.g. the awareness programs might not be comprehensive enough (Siponen, 2000), they might not address new threats quickly enough as the risks change continuously (Kruger & Kearney, 2006), and the programs rely upon the users to consciously decide to comply with information security principles (Ng *et al.*, 2009). Based on what was discussed in section 2.4.1 above, it stands to reason that these shortcomings are tied directly into the effect information security culture has on the overall habits of the community. While a significant amount of research is focused on attempting to address the shortcomings in awareness programs (Tsohou *et al.*, 2015), it can be argued that a more holistic approach that aims to address both information security knowledge and culture simultaneously may deliver better results. This argument is supported by a recent study by Mamonov and Benbunan-Fich (2018) that found that computer users exposed to news articles about corporate security breaches adjusted their behaviour, i.e. changed their personal information security culture mind-set, to be more conscientious of privacy and security concerns. This also, to a lesser extent, demonstrates how the medium used to transfer the knowledge can influence how the knowledge is received and perceived; a security awareness program might make use of hypothetical situations, whereas a news article relays actual events in a manner that is designed to be enjoyable.

2.4.3. Information Security Behaviour and Attitude Theories

While the security culture model discussed earlier provides a good insight into how role-players and information security assets and -processes interact with one another, one of the aspects of human nature that is not clearly emphasized is that of behaviour. Behaviour, as the common understanding implies, deals with the day-to-day actions taken by individuals. Behaviour can therefore have a significant impact on information security. There are a number of studies that have shown how behaviour, as it pertains to information security, seemingly contradicts information security knowledge and even the espoused values of a

corporate information security culture (Kokolakis, 2017; Crossler *et al.*, 2013; Cox, 2012). The result of this is that, of the significant number of theories that have been developed over the years in an attempt to describe behaviour, several have been used to specifically describe information security behaviour. Lebek *et al.* (2013) identify more than 54 different studies that have aimed to develop behaviour theories. While each of these studies have a different approach, the four theories that have enjoyed the most attention are (Kearney & Kruger, 2016):

- The Theory of Reasoned Action (TRA);
- The Theory of Planned Behaviour (TPB);
- The General Deterrence Theory (GDT); and
- The Protection Motivation Theory (PMT).

The first two theories in this list, namely TRA and TPB, both incorporate attitude as a determining factor in behaviour; because of this behaviour and attitude will be examined simultaneously within the context of these theories.

2.4.3.1. The Theory of Reasoned Action and The Theory of Planned Behaviour

The Theory of Reasoned Action (TRA) and the Theory of Planned Behaviour (TPB) are two closely related theories that both incorporate attitude and subjective norms (Gundu & Flowerday, 2013; Lebek *et al.*, 2013). Attitude impacts on an individual's decision to engage in an act, i.e. it determines how the personal "should I, should I not" questions are answered. The subjective norms, by contrast, impact how supportive the individual believes other relevant individuals, such as co-workers, will be of the decision to act. In TRA, these two aspects are considered to be the determining factors in choosing to take action, which in turn determines behaviour (Brodowsky *et al.*, 2018; Khan *et al.*, 2011). TPB, which can be seen as an extension of TRA, also includes behavioural control, i.e. the freedom an individual has to choose to behave in a certain manner, as a determining factor in behaviour (Kearney & Kruger, 2016; Ifinedo, 2012). The basic structure of these theories is shown in Figure 2.11 and Figure 2.11.

2.4.3.1. The General Deterrence Theory

The General Deterrence Theory (GDT) postulates that unwanted behaviour can be minimised by implementing deterrents. These deterrents generally take the form of punishments and have proven somewhat effective in curbing information system abuse (Willison *et al.*, 2018; Kearney & Kruger, 2016; Straub Jr, 1990). However, as certain studies

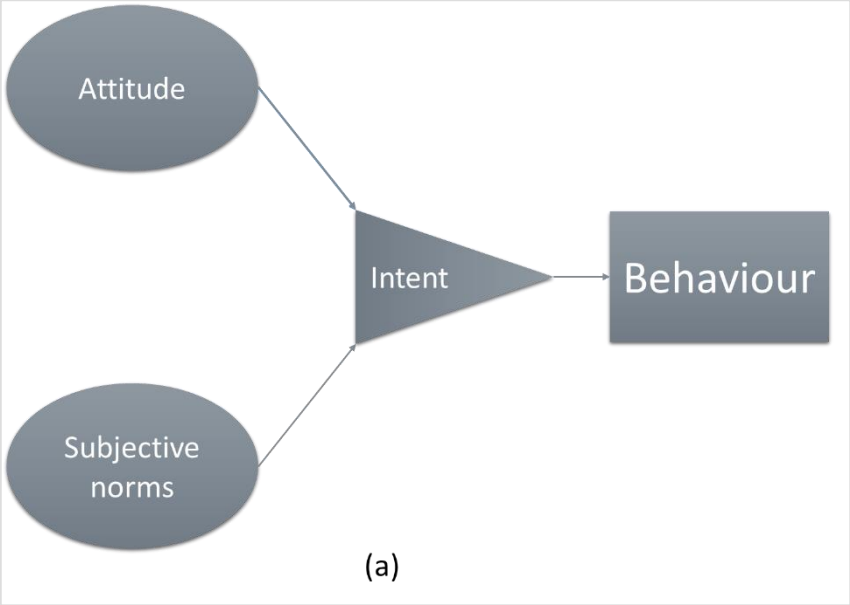


FIGURE 2.10 STRUCTURE OF TRA [BRODOWSKY *ET AL.* (2018) AND KHAN *ET AL.* (2011)]

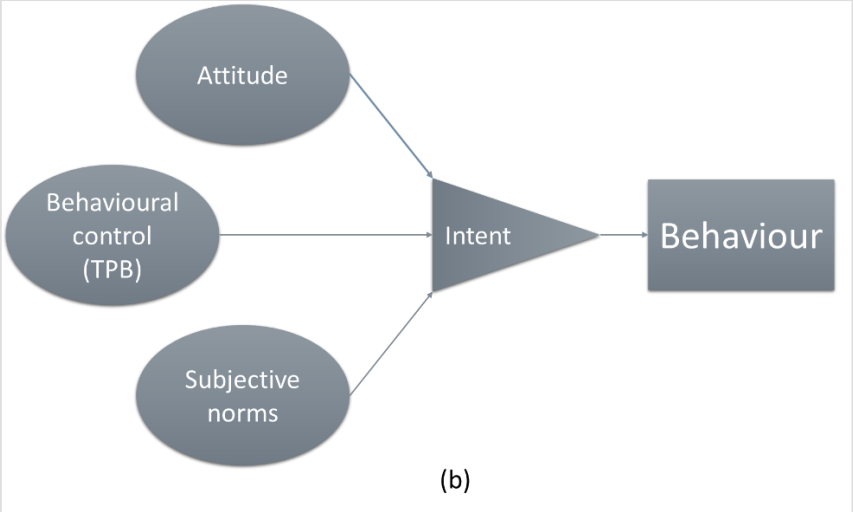


FIGURE 2.11 STRUCTURE OF TPB [BRODOWSKY *ET AL.* (2018) AND KHAN *ET AL.* (2011)]

have shown, human behaviour is such that deterrence only works if the deterrent has a significant enough impact on the individual. In a study done by Curtis *et al.* (2018), the efficacy of banning patrons from certain establishments as a deterrent is evaluated. While it is not an information security study, it serves to demonstrate one of the caveats of applying GDT in principle: a deterrent may not modify behaviour at all if it is not effective enough. A recent study by Moody *et al.* (2018) support this by pointing out that, due to the limited number of studies and the mixed results obtained, it is difficult to set effective deterrents in place. The basic principle of the GDT is shown in Figure 2.12, which also illustrates the impact of an ineffective deterrent: such a deterrent may still influence behaviour, even if

unacceptable behaviour is not eliminated entirely. The application of GDT, in conclusion, is heavily dependent upon selecting and implementing the correct deterrents. It should also be mentioned at this point that GDT deals with whether or not individuals decide to act on their intentions; this stands in contrast to TRA and TPB that deal with the circumstances that influence the intention itself. It is therefore possible to implement all of these theories simultaneously when addressing information security behaviour problems.

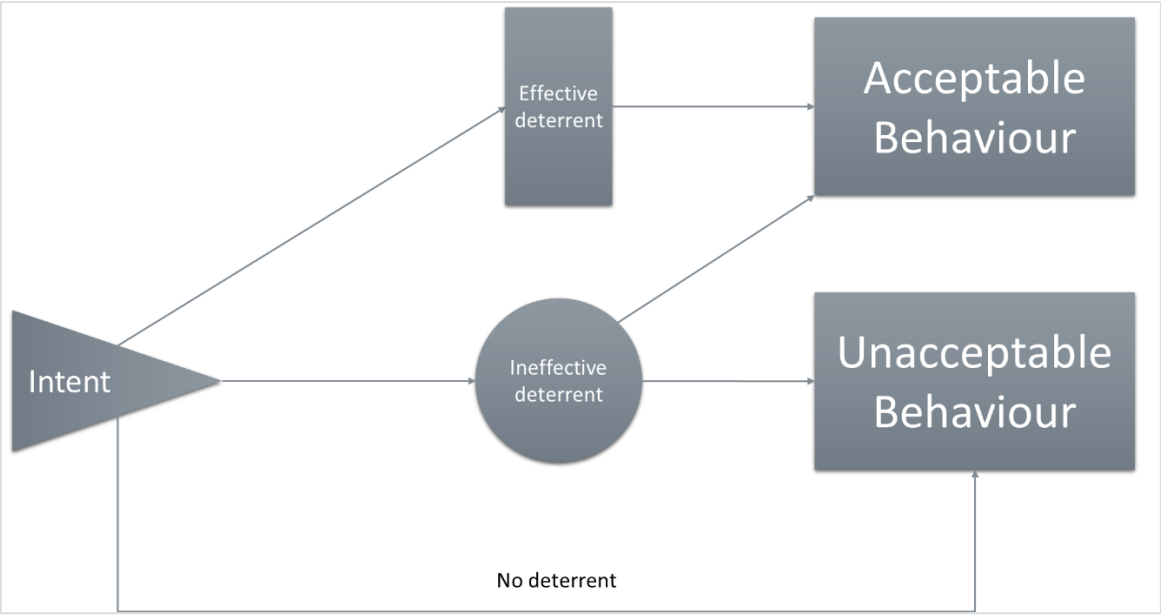


FIGURE 2.12 BASIC PRINCIPLE OF GDT [MOODY *ET AL.* (2018), WILLISON *ET AL.* (2018), AND STRAUB JR (1990)]

2.4.3.2. THE PROTECTION MOTIVATION THEORY

The Protection Motivation Theory (PMT) was originally a general theory of communication based on persuasion, specifically using fear as a driver (Kearney & Kruger, 2016; Boer & Seydel, 1996). The theory is often considered one of the best predictors of behavioural change, as it incorporates two aspects: threat appraisal, which deals with an evaluation of the threat level, and coping appraisal, which is the measure of how well an individual can cope with or avoid the consequences of an appraised threat (Gundu & Flowerday, 2013; Prentice-Dunn *et al.*, 2009). Depending on how this theory is implemented, it can also be used in addition to GDT, as the deterrent may be perceived as a threat by an individual with the intent to engage in unacceptable behaviour. PMT can therefore be used to determine why a GDT deterrent is ineffective, or even select an appropriate deterrent. Figure 2.13 provides a graphical representation of the PMT process. The threat appraisal is done based on a collective evaluation of both the individual's vulnerability to the threat, as well as the perceived severity of the threat. Coping appraisal, on the other hand, makes use of a form

of introspective analysis: the individual determines their personal ability to cope with the threat by determining their response- and self-efficacies. This coping ability is then compared to the costs involved in dealing with the threat, should it be realised. These costs can involve any form of reasonable cost, such as loss of money, time, data, reputation, reliability, etc. The final coping appraisal is therefore a representation of both the individual's personal coping ability and the willingness of the individual to carry the costs of a threat being realised. The final personal analysis, whereby the threat- and coping appraisals are compared, determines the ultimate intent to act.

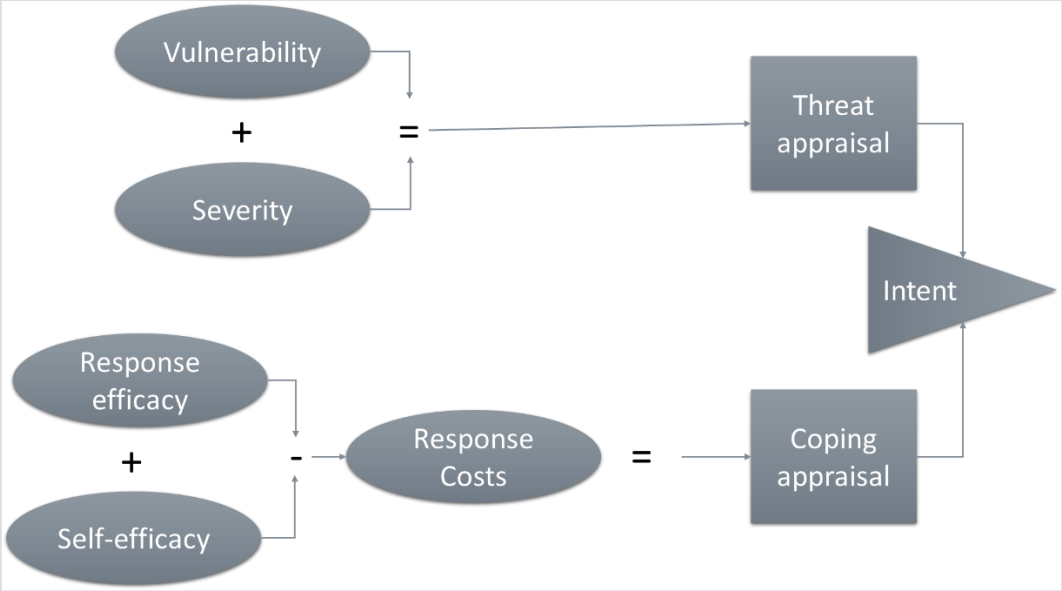
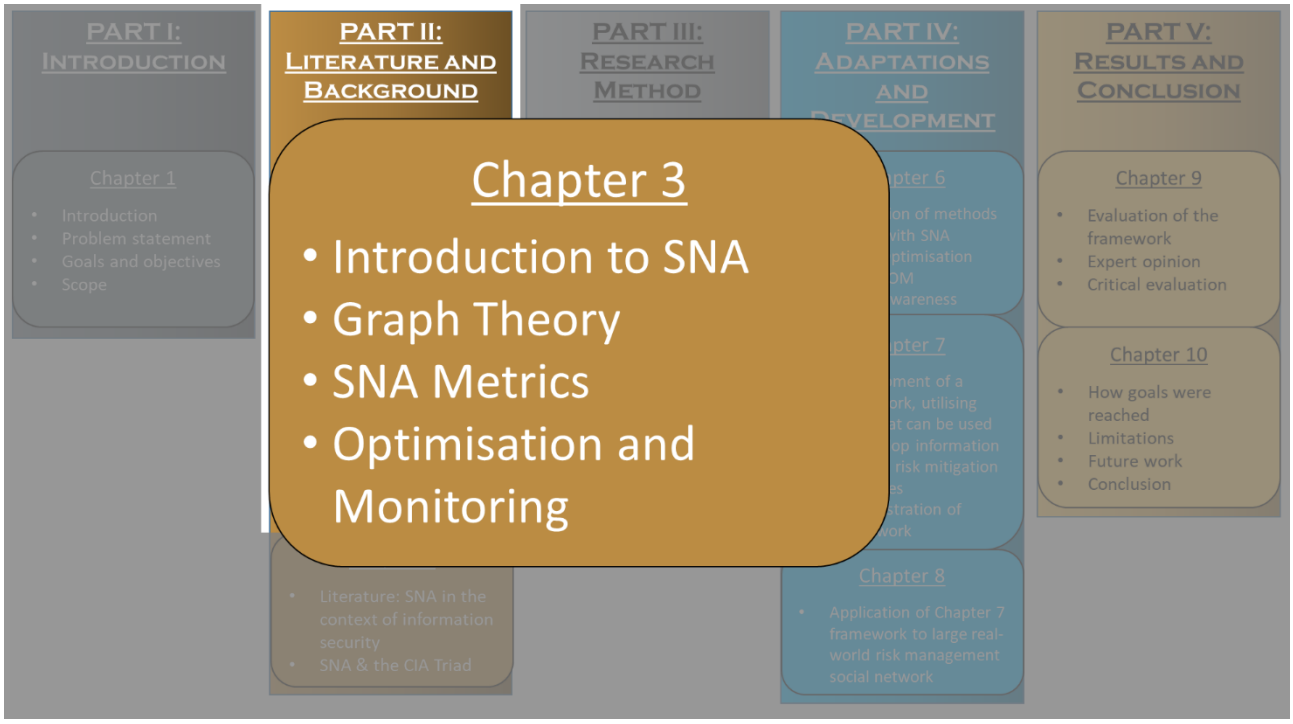


FIGURE 2.13 BASIC STRUCTURE OF PMT [TESSON *ET AL.* (2016)]

Based on the relatively brief discussion in this section, it becomes clear that the study of the human aspects of information security is both a very important and a very complex field of study. Among some of the topics that were not discussed are risk homeostasis, availability heuristics, optimism bias, cumulative risk, omission bias, the influence of familiarity, the influence of framing, and personality and cognitive styles. The purpose of this section was to provide a significant enough introduction to the human aspect of information security to enable a sufficient understanding of the remainder of the study. A discussion of these topics, while both fascinating and relevant, is outside the scope of this study. There are a number of literature sources, such as the work done by Parsons *et al.* (2010), that can be consulted for more information on these topics.

2.5. CHAPTER SUMMARY

In this chapter, some of the various topics of information security, such as the CIA triad, information security risk management, and approaches to dealing with the human factor were discussed. The topic of this study is however not only information security, but the application of social network analysis to information security. Several examples appear in the literature of how SNA can be used in this manner, such as Dang-Pham *et al.* (2017a) and Armstrong *et al.* (2010). In the interest of discussing these literature sources in greater detail, the principles of social network analysis are presented first in the next chapter. The following chapter, Chapter 4, concludes the literature review with a discussion of the existing work that combines social network analysis and information security. The approaches and contributions made in these studies are used to inform the development of the novel framework presented in Chapter 7.



CHAPTER 3: SOCIAL NETWORK ANALYSIS

CHAPTER HIGHLIGHTS:

- What is Social Network Analysis (SNA)?
- What are the basic principles of SNA?
- What are some of the most important concepts of graph theory?
- How can graphs be visualised?
- What are Self-Organising Maps?
- Which SNA metrics are most relevant to this study?
- How can communities be detected in social networks?
- How can networks be optimised using graph theory?
- How can changes be detected in social networks?

3 SOCIAL NETWORK ANALYSIS

The primary focus of this study is to demonstrate how Social Network Analysis (SNA) can be used to manage risk within an information security context. To this end, it is important to provide some background on the subject.

In this chapter the basic principles of SNA are therefore discussed. The chapter focusses on a number of things, starting with an introduction on the primary principles of SNA. Secondly, the visualisation of social networks is discussed. In this, the second section of the chapter, various visualisation elements, such as visualisation methods and software, are described. In the third part of the chapter, a significant number of metrics and measures that are found in SNA are mentioned and, where appropriate, described in detail. The chapter then concludes with a brief discussion of optimisation- and monitoring techniques that can be used with social networks.

3.1. INTRODUCTION

In SNA, social life is assumed to be composed primarily of relationships (Wasserman & Faust, 1994). Because of this, any social group can be quantitatively evaluated using graph theory. Graphs in this case are developed by using entities, such as people and resources, as nodes and the relationships between them as edges. As this approach has a different perspective than individualist and attribute-based evaluations, a significant amount of subjective data is excluded. SNA can therefore be said to be a quantitative technique that can be used to evaluate qualitative data (Scott & Carrington, 2011; Loosemore, 1998).

The first challenge is typically to select which nodes to include in the network, also known as **bordering**. In some cases, because of the possible influences outside sources may have on the entities in the network, selecting which nodes to ignore is not as straight forward as it seems. There are primarily three methods that can be used to select the nodes that should be included (Laumann *et al.*, 1989):

- Only include those entities that are formally part of the network. Examples of these types of entities would be formal employees of an organisation and office resources, such as printers.
- Include those entities that have a significant impact with regard to events that are believed to have defined the organisation. An example in this regard would include

people that have shared experiences, such as two IT researchers having visited the same security conference.

- Start with a so-called seed group that serves as the primary evaluation point of the network. From here, only include those entities that have specific types of relationships with the members of the seed group. This network can then be expanded by following the same process for the added entities. An example of a network developed in this way would be one where the IT researchers who published in the same journal are taken as the seed group, with all of the co-authors as new nodes. From here the network can be expanded to include the researchers who were co-authors on different articles with the aforementioned co-authors as authors.

In addition to these three methods, SNA also has a number of guiding principles. The first principle is that SNA is based in sociology, and not psychology (Loosemore, 1998). The attributes of the individual are therefore considered irrelevant with regard to causation, and causation is therefore assumed to be based in the social network and not the individual. **The first principle of SNA is subsequently that relationships, not attributes, are ultimately evaluated.**

SNA is focussed on dealing with the interactions specific entities have with one another. As it is next to impossible to assign all entities to isolated, mutually exclusive groups, SNA avoids dealing with pre-defined groups by working with bounded networks. In this manner, **SNA evaluates networks that can be bounded using certain suppositions, rather than artificially isolated groups.**

Relationships are not the only things that can be studied with SNA; it can also be used to study the patterns of similar relationships. **SNA therefore studies relationships within a relational context.** It can furthermore be used to describe how certain networks, or an actor's position within a network, can influence particular outcomes. According to Borgatti *et al.* (2009), there are four main categories of interaction than can occur within a network, namely transmission, adaption, binding and exclusion.

With **transmission** interactions, each relationship can be viewed as a pipeline whereby such things as knowledge, values, rumours, etc. can be transmitted to other actors within the network. The best targets for novel or unique ideas would therefore be those actors with minimal network exposure. Similarly, those actors with the greatest number of links to other actors may prove the most effective target when the goal is the introduction of new knowledge. The basic premise of **adaption**, on the other hand, is that two actors with similar network positions and with similar situational constraints will take similar actions because they have similar conditions. This means that it is possible to predict the actions of an actor if the actions taken by another actor in the same circumstances is known.

Binding occurs when multiple actors operate as a single entity due to the social interactions that exist between them. In such cases a bound network can be expected to operate almost as a single actor. **Exclusion** is encountered when an actor's preference for one relationship precludes the existence of another relationship. In these cases the precluded relationship is likely to influence the network as whole.

Due to its ability to allow researchers to study social groups in a quantitative way, SNA has seen a broad range of applications in the literature. One of these applications is found in the study of virology, which deals with the way viral and bacterial infection spread. In a study done by Christley *et al.* (2005), SNA was used to identify individuals that pose a high risk when it comes to the spread of infectious diseases. Other studies, such as those conducted by Martin *et al.* (2011), confirm that the spread of diseases within population groups can be described and controlled using SNA.

SNA has also seen use in the field of fraud prevention. In a study done by Gupta and Hossain (2011) SNA is used to identify individuals that may be guilty of insider trading. SNA has furthermore proven useful in identifying important individuals in criminal networks. Philips *et al.* (2015) demonstrated how SNA can be used to identify the hierarchies in criminal Dark-Web forums, and Fu *et al.* (2015) propose a method that can be used to evaluate terrorist organisations and predict their behaviour.

There are, therefore, a large number of applicable uses for SNA. As pointed out by Dang-Pham *et al.* (2017b) however, one of the lesser researched applications of SNA is in the field of information security. This notion is supported by Google Scholar searches conducted on 29 August 2018: one search, requiring the terms "information security" and "social network analysis" in the title, returned only eight results, whereas another broader search requiring the terms "information security" and "network analysis" in the title returned only 22 results. As SNA is useful for evaluating the qualitative data of human relationships quantitatively, there is ample room to expand on the use of SNA in the field of information security.

The interactions, or relationships, between entities are studied in SNA in order to evaluate or predict their behaviour. This is done by applying graph theory to a social network wherein the individuals, as well as other important entities such as resources, are represented as nodes and the relationships between them are represented by edges. These entities and relationships are specially selected using a technique called bordering, which applies criteria to a social network and prevents the entire universe from potentially being added to the network. As SNA relies on the principles of graph theory to make inferences, the next section focusses on introducing some of the various topics, principles, and concepts of graph theory.

3.2. GRAPH THEORY AND ASSOCIATED PRINCIPLES

In this section, some of the various principles and definitions from graph theory that are used in SNA are introduced. Graph theory is the mathematical study of graphs, which are models of pairwise relationships between objects or entities (Gross *et al.*, 2013). While graph theory is highly mathematical in nature, the purpose of this discussion is not an in-depth discussion of its mathematical foundations, but rather some of the most relevant concepts and principles with regard to SNA. As such, this section mainly focusses on providing broad, explanatory descriptions. This section is primarily based on the work of a number of authors, specifically Clemente *et al.* (2016), Rubinov and Sporns (2010), Gross *et al.* (2013), Wasserman and Faust (1994) and Wasserman (1994). The work done by Wasserman and Faust is especially authoritative, and a Google Scholar search conducted on 23 May 2018 indicated that the work was cited at least 30648 times. The article by Rubinov and Sporns was cited 3998 times, with Gross *et al.* being cited 1730 times.

In graph theory, any entity can be represented as a **node**[1], whereas the relationships between them can be represented as **edges** connecting them. Formally, a graph is defined as follows: The graph *G* is the ordered pair $G = (V, E)$, where V is the collection of nodes, or vertices, in the graph, E is the collection of edges in the graph, $E \subseteq [V]^2$, alternatively $E \subseteq \{\{u, v\} : u, v \in V\}$, and $V \cap E = \emptyset$.

Building on this definition, a number of statements can be made about the nodes and edges within a graph:

- An edge will always connect two nodes to one another. Subsequently, each edge will always have two **endpoints** that are **adjacent**. When referring to the endpoints themselves, they may be said to be **neighbours** of one another.
- Because an entity's existence is not dependent upon its relationships to other entities, it is possible to have an unconnected node in a graph. The same is not true of edges, as edges specifically represent relationships between nodes.
- A single node can at most be uniquely connected once to all of the other nodes within the graph. Therefore, the maximum number of edges that can exist in a graph is $n(n-1)$, where n is the number of nodes in the graph. **Density** is a measure of how close the number of edges in the graph is to this absolute maximum: the greater the number of edges, the higher the density.

One of the main principles to be discussed deals with the fundamental relationships that exist between the entities a graph is based on. As mentioned earlier, a graph is a model of the pairwise relationships between entities and, as such, a graph must be able to represent the various relationships that may exist between entities. In order to properly illustrate the

[1] Some source refer to vertices rather than nodes

difference between the different types of graphs, consider three distinct situations, where each situation can be modelled as a social network. As a graph can represent any relationships, the term **network** is used in this context to denote a social network specifically. In the first situation, a group of six people who have no prior connections to one another is placed in the same room. In this instance, each of the people in the room has the same relationship to one another, as none of them have ever met before. The only connection they have in common is that they are now in the same room at the same time. Their presence in the same room at the same time therefore creates a weak, albeit universally shared, relationship to one another. The graph of this network is shown in Figure 3.1(a), where the reciprocal relationships are shown as simple edges connecting the nodes. In the second situation, a group of six colleagues are placed in the same room. In this instance, the relationships between the people in the room will vary according to their personal history and, therefore, the connections between them will differ. The graph of this network is shown in Figure 3.1(b); note that some of the edges are darker to indicate a stronger relationship. For the third situation, shown in Figure 3.1(c), consider a military network wherein each superior officer has the authority to choose how much information a subordinate has access to. In this network, the arrows indicate the direction of the relationships, i.e. in which direction information can flow from one node to the next.

The graph of the first network, consisting of a room of strangers, is known as a **simple** or **symmetrical** graph. This contrasts to the other two graphs in that the relationships between the various nodes in the graph are all similar, specifically that all of the relationships are fully reciprocal and equally strong or meaningful. The second graph, referred to as a **weighted** graph, provides more information about the strength of the relationship between the nodes. In general, the higher the weight of the edge connecting two nodes, the stronger the relationship. Note that, in Figure 3.1(b), the stronger relationships are graphically represented as edges that are “thicker”, whereas weaker relationships are presented as “thinner”. The third type of graph, known as a **directed** or **asymmetrical** graph, provides information about directed relationships, i.e. relationships that are not reciprocal. An example of such a relationship is one that exists between a manager and an employee: while the manager can give instructions to the employee and decide how much information the employee has access to, the same cannot be said in reverse. The employee, after all, cannot formally give instructions to the manager. These types of graphs have additional restrictions on the relationships between nodes, as not all nodes are necessarily **reachable** within the graph. **Reachability** is a measure of how accessible a node is within the graph. While all of the nodes in all of the graphs in Figure 3.1 are reachable, the same cannot be said of the nodes in Figure 3.2. In Figure 3.2(a), all of the nodes are reachable, whilst in Figure 3.2(b), only one of the nodes is reachable.

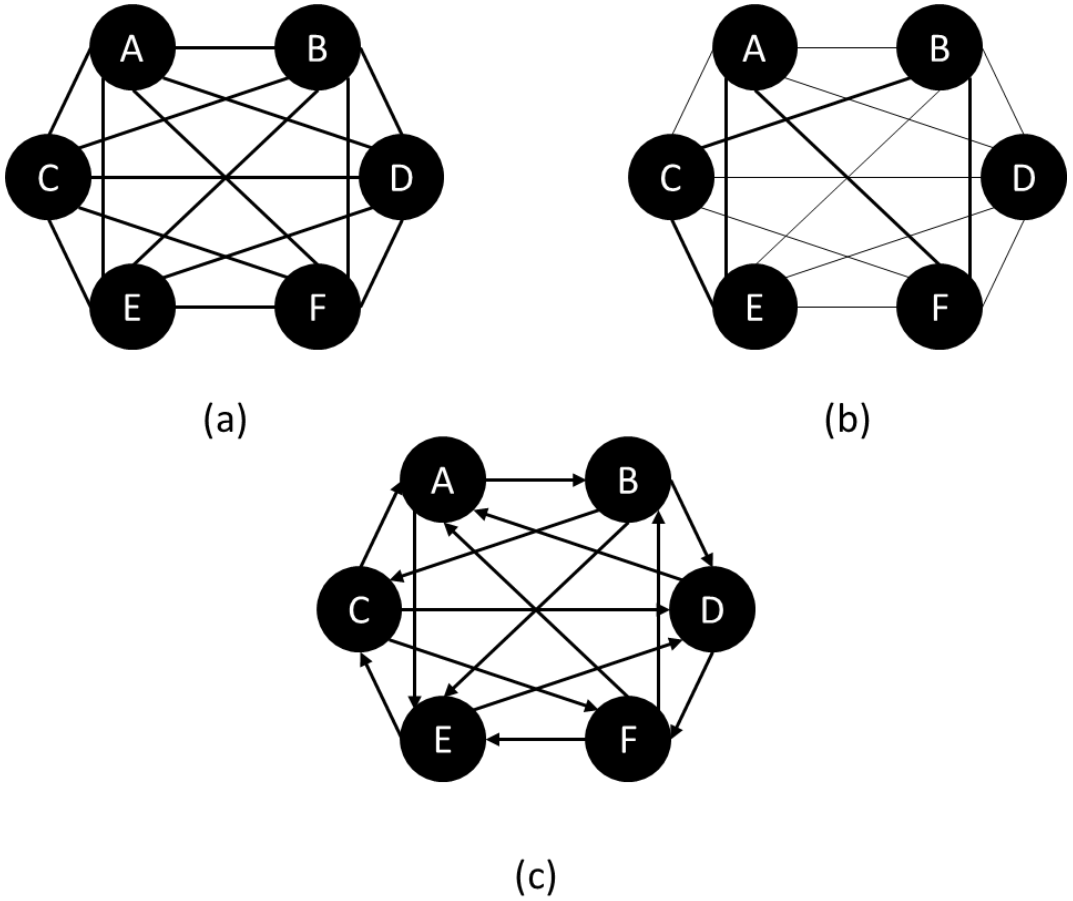


FIGURE 3.1: FUNDAMENTAL NETWORK TYPES. (A) IS A SOCIAL NETWORK OF STRANGERS; (B) IS A SOCIAL NETWORK OF COLLEAGUES; (C) IS A MILITARY INFORMATION NETWORK

a) A and B are both reachable

b) In this asymmetrical network B is reachable, but A is not

FIGURE 3.2 A) REACHABILITY IN SYMMETRICAL NETWORKS; B) REACHABILITY IN ASYMMETRICAL NETWORKS

In a directed graph, or **digraph**, a distinction exists between an edge and an **arc**. An edge, like in a weighted or simple graph, does not have a direction to indicate connectivity. An arc, in contrast, has a direction. A digraph specifically only has arcs.

The final fundamental type of graph is a combination of the directed and weighted graph types mentioned above. These graphs are known as **mixed** graphs. In these graphs, both edges and arcs may be found, with some arcs having both a weight and a direction.

Any graph has the potential to become massive in size, with millions of nodes and billions of edges. It is therefore useful to define specific types of sub-graphs in order to simplify explanation. There are two types of sub-graphs that are specifically defined: dyads and triads. These two types of sub-graphs are essentially the same, with only the number of nodes they contain being different. A **dyad** is defined as any graph or sub-graph that contains two nodes, along with edges that represent their relationships to one another. A dyad may contain two nodes that do not necessarily have a relationship to one another. It is therefore possible to have a dyad that contains only two nodes and no edges. Also, as the relationship might not be reciprocal, it is possible to have two unique edges connecting the nodes if the dyad is a digraph. A dyad therefore always contains two nodes, but may contain up to two edges, or none at all. A **triad** is essentially similar to a dyad, but with three nodes instead of two. A triad may similarly have up to six edges, or none at all.

In addition to dyads and triads, it is possible to describe another type of sub-graph, namely a **clique**. A clique is defined as any sub-graph that is complete. A graph is considered **complete** when all of the nodes in that graph are adjacent. An example graph containing a single clique with four nodes is shown in Figure 3.3.

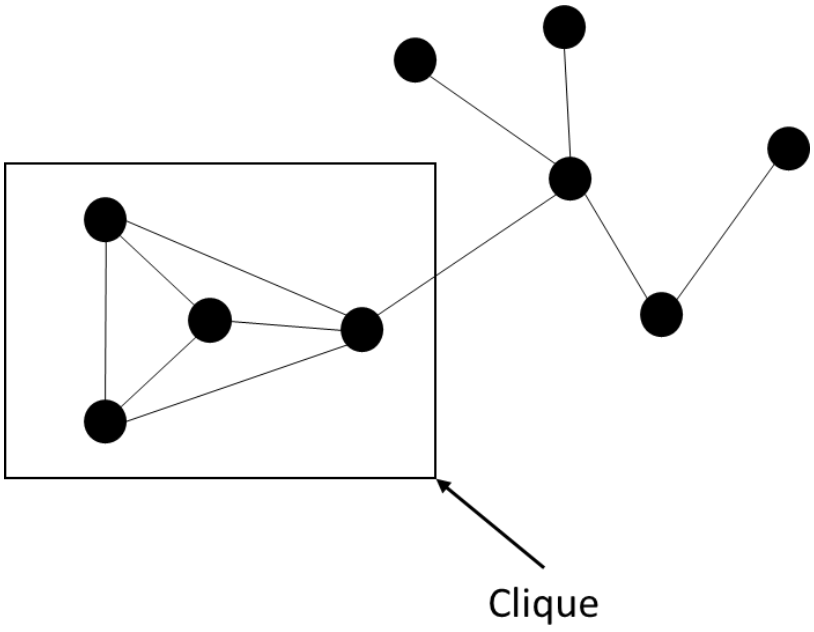


FIGURE 3.3: SIMPLE GRAPH SHOWING A CLIQUE WITH FOUR NODES

Another aspect that is important in graph theory is distance. In the same way that the weight of an edge can indicate the strength of a relationship between two nodes, the distance between them can also be used to evaluate their relationship. The measure **distance** is determined by the number of edges, or “jumps”, that exist between two nodes

in simple networks. In weighted networks, where the strength of the relationships differs. from edge to edge, the distance is additionally impacted by the weight, or **nearness**, of an edge. The stronger the relationship, the closer the nodes are to one another. This ultimately means that, while all edges have the same length in simple graphs, namely one "jump", in weighted graphs this length is additionally impacted by the strength of the connection between the two neighbours. This is illustrated in Figure 3.4, where a very strong reciprocal relationship exists between nodes A and B, a strong relationship exists between nodes B and C and a weak relationship exists between nodes A and C. Depending on how much stronger the relationship between A and B is compared to the relationship between B and C, it is possible for the effective distance between A and C to be shorter through B than it is directly. This is caused by the effect of the nearness that exists between A and B and B and C: because A and B have a very strong relationship, they are much closer to one another than A and C is, even though the number of "jumps" between them are the same. In weighted graphs, it is therefore more accurate to state that the distance between two nodes is equal to the sum of the weights of the edges connecting them. It is subsequently possible, especially in complex weighted graphs, for there to exist a number of distances between two nodes. Because of this, the measure **geodesic distance** is specially defined. The geodesic distance between two nodes is the shortest distance that exists between those two nodes. Figure 3.5 provides a graphical illustration of how simple distance, as well as geodesic distance, is determined. It is possible in directed networks for the geodesic distance of a path that leads from node A to B to differ from the geodesic distance of the path that leads from node B to node A, as some relationships are not reciprocal and alternate, possible longer, paths are needed to connect node B to node A.

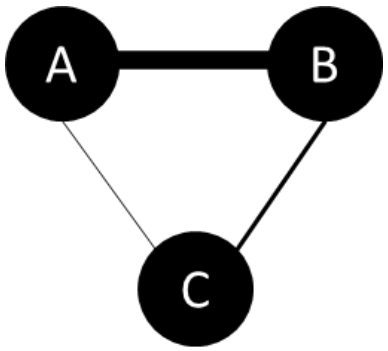
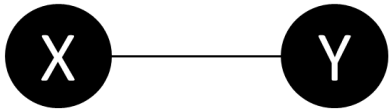


FIGURE 3.4: GRAPH SHOWING MULTIPLE DISTANCES BETWEEN SIMILAR NODES

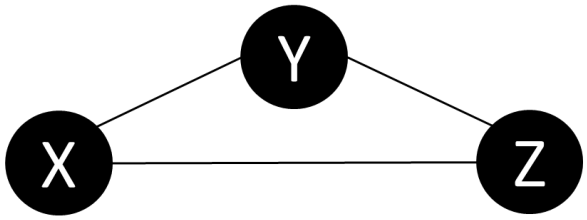
In addition to the geodesic distance between two nodes, a further geodesic measure can be used to evaluate a node. The **eccentricity** of a node is the greatest geodesic distance it has to any other node it is even remotely connected to, whereas the greatest eccentricity among all the nodes in the network represents the network's **diameter** (CASOS, 2019).

The measure **connectivity**, which is similar to distance, is a measure of the number of edges that can be removed from the path connecting two nodes before the nodes become disconnected. This measure is therefore an indication of how well connected two nodes are to one another. If, for example, node X is connected to node Z only by virtue of a shared relationship with node Y, then the connectivity between X and Z would be 1. In Figure 3.5, nodes X and Y in network (a) and (b) both have a connectivity of 1, whereas the nodes in network (c) have a connectivity of 2 between X and Z, as well as between Y and X, and Y and Z.



a) Distance of 1 between X and Y

b) Distance of 2 between X and Z



c) There are two measures of distance from X to Y: longest distance, via Z, which is 2, and geodesic distance which is 1

FIGURE 3.5 DISTANCES IN UNDIRECTED GRAPHS

In order to more easily describe the relationship between various nodes in a graph, a number of terms are defined to describe the connections that exist between them. The simplest of these descriptions is a walk. A **walk** is a sequence of nodes and edges, starting and ending with a node, such that each of the nodes in the sequence is adjacent to the previous node in the sequence by virtue of the edge connecting them. This means that all of the endpoints of the edges in the walk are also contained in the walk. A **trail** is a walk wherein no edges are repeated and a **path** is a walk where the walk itself has two different endpoints. If the two endpoints of a walk are the same, then the walk is said to be **closed**; otherwise it is **open**. A path is therefore always an open walk. The four different types of walks are shown in Figure 3.6, where (a) is the main graph, (b) is a normal walk, (c) is a trail, and (d) is a path.

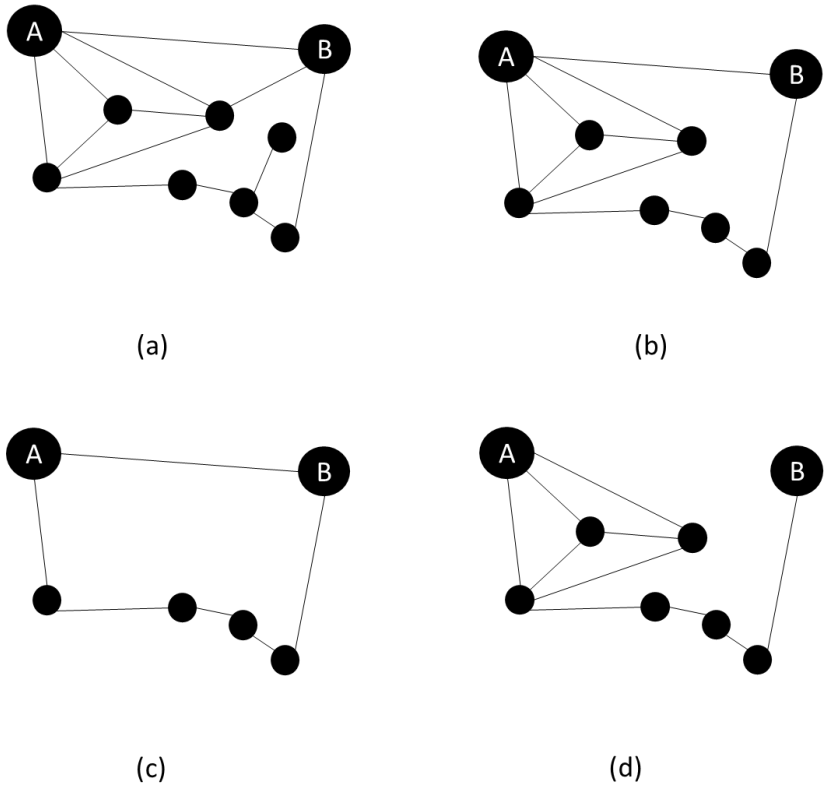


FIGURE 3.6: THE DIFFERENT TYPES OF WALKS; NOTE THAT ONLY THOSE NODES AND EDGES THAT EXIST WITHIN A PARTICULAR WALK'S COLLECTION ARE SHOWN. (A) IS THE GRAPH THAT CONTAINS ALL OF THE WALKS; (B) IS A WALK CONTAINING REPEATING PATHS AND REPEATING ENDPOINTS (ANY OF THE NODES CAN POTENTIALLY BE THE START/END NODE); (C) IS A TRAIL WITH REPEATING ENDPOINTS (ANY OF THE NODES CAN POTENTIALLY BE THE START/END NODE); (D) IS A PATH WITH REPEATING EDGES (WHILE MOST OF THE NODES CAN BE EITHER A START OR END NODE, NODE B MUST BE EITHER A START OR AN END NODE)

When collecting and processing network data, a **relational matrix,** also known as an **adjacency matrix,** is most often used. Such a matrix, which is always a $[n\ x\ n]$ matrix, with n being the number of nodes in the graph, plots all of the connections between the nodes in the graph. In an unweighted graph, the only values in the matrix will be 0, representing a non-existent connection, and 1, representing a connection. In a weighted graph, the value represents the strength, or weight, of the connection. The elements in an adjacency matrix are always real numbers. In a symmetric graph, the adjacency matrix will necessarily be symmetrical, whereas the adjacency matrix of a digraph is not necessarily symmetrical. The various types of adjacency matrices for the networks in Figure 3.1 are shown in Matrices 3.1, 3.2, and 3.3 below. Matrix 3.1 shows the adjacency matrix A for network (a), Matrix 3.2 shows the adjacency matrix A for network (b), and Matrix 3.3 shows the adjacency matrix A for network (c). In all of these matrices, the nodes in the rows are the "from" nodes, whereas the nodes indicated in the columns are the "to" nodes. These rows and columns are organised in the same manner, so that both the rows and the columns represent nodes A, B, C, D, E, and F in order. While not an important distinction for a simple graph, in a

digraph, this becomes crucial. In Matrix 3.3, for example, the connection (A, C) has a value of 1, indicating a connection, whereas (C, A) has a value of 0, which indicates that there is no connection. These connections are indicated in bold in Matrix 3.3.

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \varepsilon\ \mathbb{R}^{6\times 6} \tag{3.1}$$

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 & 2 & 2 \\ 1 & 0 & 2 & 1 & 1 & 2 \\ 1 & 2 & 0 & 1 & 2 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 2 & 1 & 2 & 1 & 0 & 1 \\ 2 & 2 & 1 & 1 & 1 & 0 \end{bmatrix} \varepsilon\ \mathbb{R}^{6\times 6} \tag{3.2}$$

$$A = \begin{bmatrix} 0 & 1 & \mathbf{0} & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ \mathbf{1} & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \end{bmatrix} \varepsilon\ \mathbb{R}^{6\times 6} \tag{3.3}$$

In this section, some of the principles of graph theory were discussed in an explanatory, non-mathematical way. In the next section some of the ways that networks can be visualised as graphs are discussed.

3.3. VISUALISATION OF NETWORKS

In this section, some of the various techniques used to visualise graphs, as well as techniques used to emphasise certain features in visualised graphs, are described briefly.

This section starts with a description of simple graph representation methods. This is followed by a brief evaluation of some of the methods that can be used to simplify complex networks, and the section concludes with a brief introduction to Self-Organising Maps (SOMs).

3.3.1. Introduction to Visualisation Techniques Used in the Literature

The most elementary visualisation technique used in the literature involves using simple shapes and straight lines, all in the same colour, to represent the network. For the purposes of this discussion, this technique is referred to as the **simple method**. The basic principle of this visualisation technique is shown graphically in Figure 3.7, where nodes are represented by the dots and the relationships between them are indicated with the arcs connecting them. However, as shown in Figure 3.8, there is a risk that using this technique with large datasets will result in networks that are extremely difficult, if not impossible, to evaluate graphically.

One of the techniques used to make large networks visualised using the simple method more legible, is to alter the colour of the node and arcs, as shown in Figure 3.10. While this does make large networks significantly easier to follow, it is potentially limited in that it can either be used to distinguish between nodes and their associated arcs, as shown in Figure 3.9 (a), or be used to draw attention to a specific node, or a number of important nodes, as shown in Figure 3.9 (b). In order to address this issue, specifically that colouring cannot be used to draw attention to specific nodes if all nodes have a unique colour, sizing is employed in order to draw attention to important nodes. Two different sizing techniques are used in the literature to draw attention to important nodes: the first technique, shown in Figure 3.11, resizes the nodes themselves according to their relative importance, and the second technique, shown in Figure 3.12, resizes the labels associated with the nodes.

(a) (b)

FIGURE 3.7 SIMPLE VISUALISATION TECHNIQUE, WITH (A) USING DOTS AND (B) USING SQUARES AS NODES

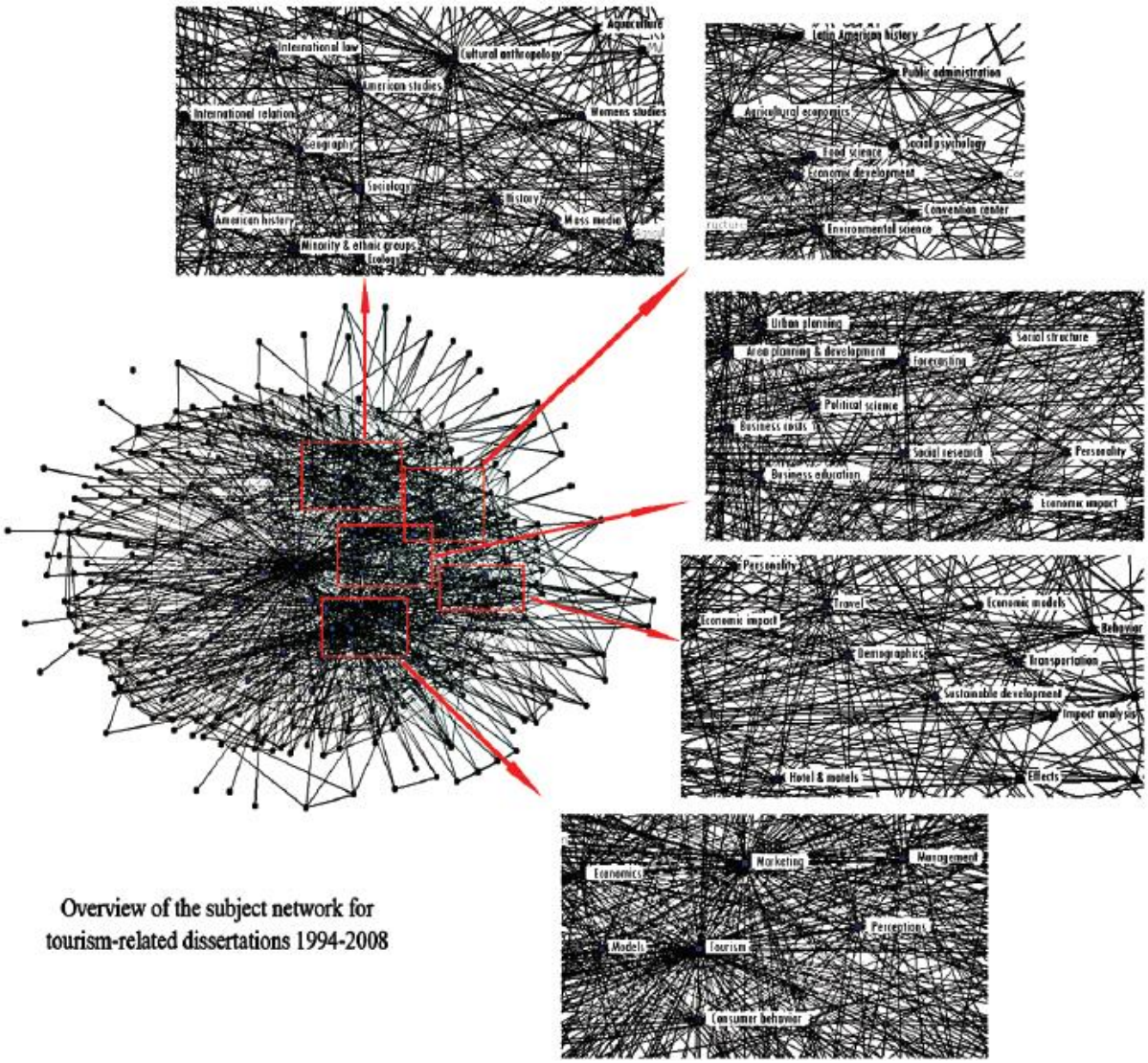
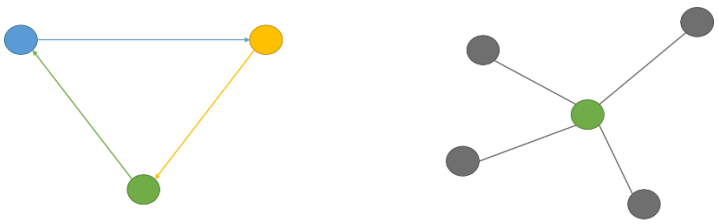


FIGURE 3.8 LARGE NETWORK VISUALISED USING SIMPLE METHOD (YING & XIAO, 2011)



(a) (b)

FIGURE 3.9 SIMPLE METHOD COLOURING VARIATIONS: (A) USING COLOUR TO UNIQUELY IDENTIFY NODES AND ARCS; (B) USING COLOUR TO DRAW ATTENTION TO A SPECIFIC NODE

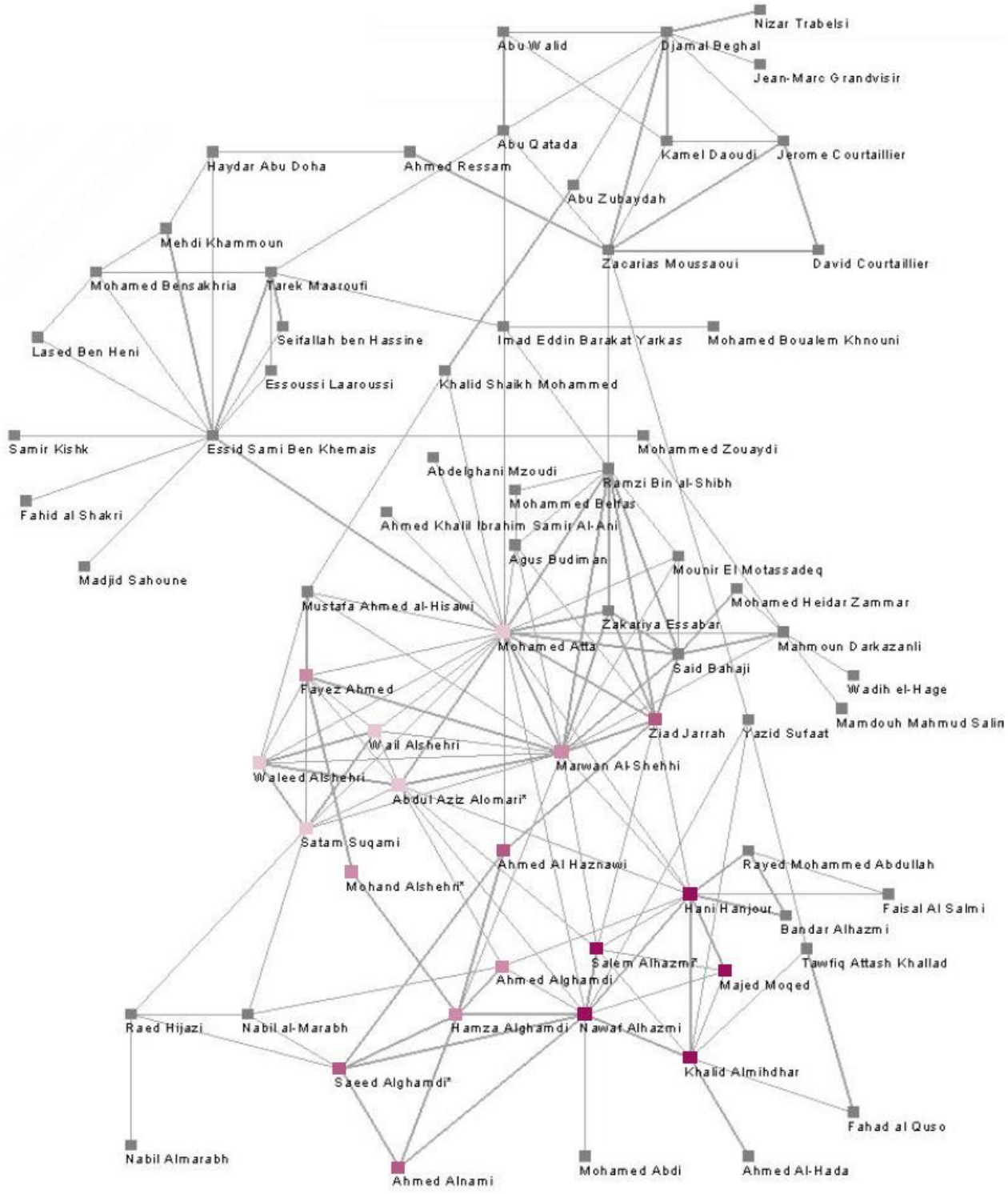


FIGURE 3.10 MODIFIED SIMPLE METHOD USING COLOURED NODES AND ARCS (TSUI & LIEBOWITZ, 2005)

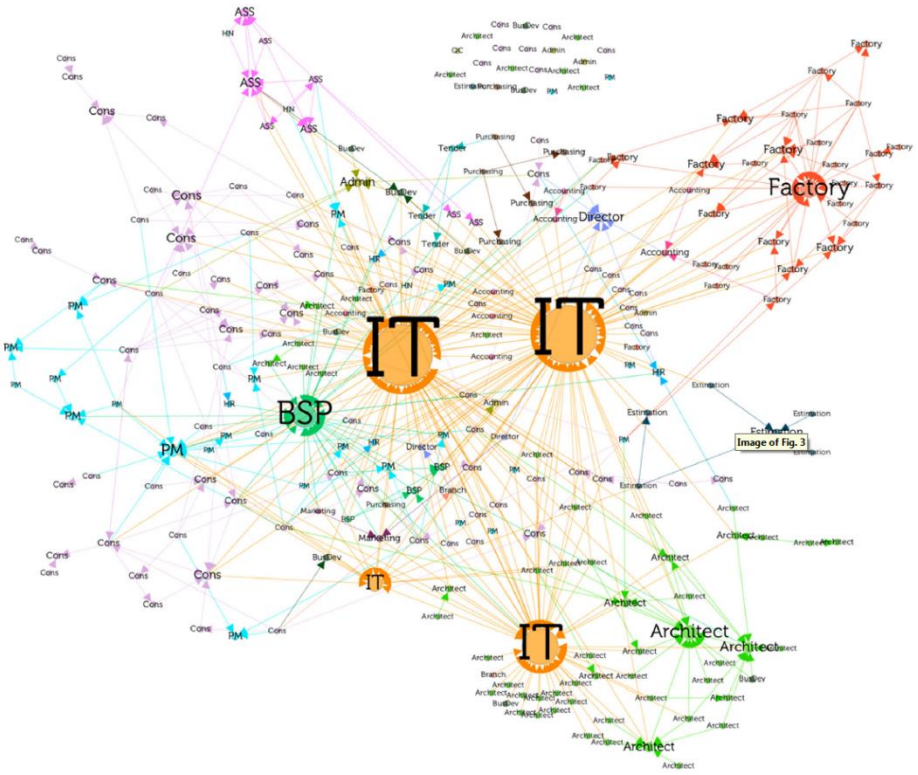


FIGURE 3.11 SIMPLE METHOD VARIATION USING COLOUR TO UNIQUELY IDENTIFY NODE TYPES, AND NODE SIZE VARIATIONS TO IDENTIFY IMPORTANT NODES (DANG-PHAM *ET AL.*, 2017B)

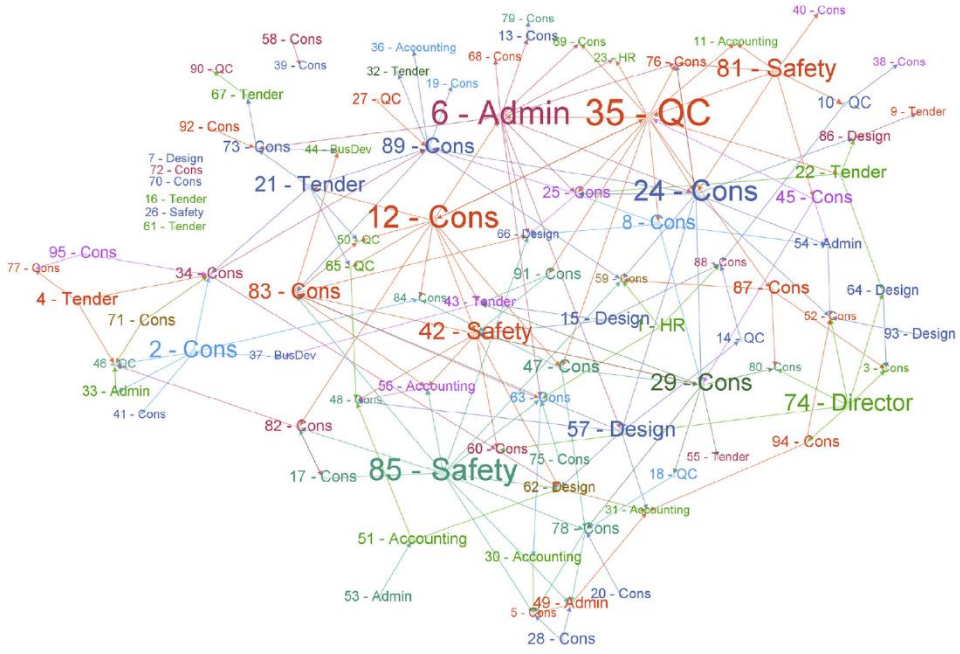


FIGURE 3.12 SIMPLE METHOD VARIATION USING COLOUR TO IDENTIFY NODES, AND SIZE VARIATIONS ON THE NODE LABELS TO IDENTIFY IMPORTANT NODES (DANG-PHAM *ET AL.*, 2017C)

While colouring and resizing techniques contribute significantly to making large networks easier to follow graphically, two general limitations to their use can be reasonably deduced:

- Using colouring methods may not be feasible if the networks are to be included in reports that are exclusively black and white; and
- Similarly, resizing the nodes or node labels to the point where the network is legible might result in a graph that is too big to be included in reports.

As is seen in Figure 3.8, one of the main reasons that large networks are difficult to follow graphically is that all of the arcs are shown. One of the techniques that can be used to reduce the number of arcs that are shown graphically is known as edge bundling. This technique involves combining adjacent arcs so that distinct arcs are only shown when necessary. The basics of this technique are shown in Figure 3.13.

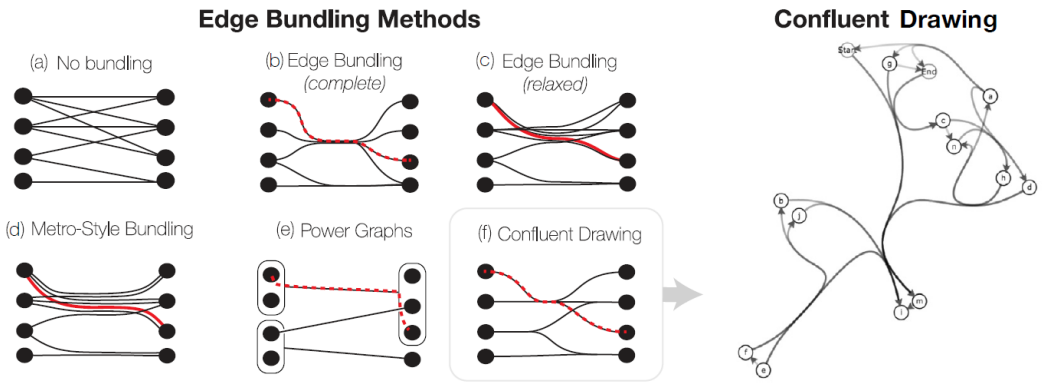


FIGURE 3.13 METHODS USED TO BUNDLE ARCS (EDGES) (BACH *ET AL.*, 2017)

The visualisation techniques presented in this section are representative of those most often used in the literature. The purpose of these techniques is to present a visual representation of the graphs, rather that the information that can be obtained from the graphs. In the next section, a technique that can be used to visualise this information is briefly introduced.

3.3.2. Self-Organising Maps (SOMs)

One of the visualisation techniques that are not commonly associated with SNA metrics in the literature are self-organising maps, or SOMs. A SOM is produced by applying an algorithm that projects a data set in a non-linear clustered way. The standard algorithm

makes use of the Euclidean structure of the data space of a data set and can therefore not be used to project a network directly (Boulet *et al.*, 2008). However, as the metrics of the network are used in this study to evaluate risk, rather than the network edges themselves, this does not pose a problem. The technique is described in more detail in Chapter 6. The software used to develop the SOMs in this study is the evaluation version of Viscovery.net's SOMine. The maps developed using this software use the metrics developed using ORA-Lite as input data.

3.4. SOCIAL NETWORK ANALYSIS: METRICS AND MEASURES

In this section some of the various metrics and measures that are used in SNA are discussed. There are a large number of metrics, so only those most relevant to this study, i.e. those that describe readily usable traits, will be discussed in detail. As the software suite ORA-Lite (CASOS, 2019) is the primary mode of calculation, the formulas provided in the ORA-Lite help files are also included where available.

3.4.1. CENTRALITY

In dealing with networks, one of the most common ways to evaluate the importance, or lack thereof, of a node is to determine how central a node is within the network. A node's ability to influence a particular network is governed by its position within the network, and this in turn is referred to in general as the centrality measure (Armstrong & McCulloh, 2010). This centrality can be calculated in any number of ways, either using specific geometric techniques, or evaluating the impact the node has on the network or its neighbours (Gross *et al.*, 2013). In this section, five of the many centrality measures are briefly discussed, specifically Degree-, Closeness-, Betweenness-, Eccentricity-, and Eigenvector Centrality.

3.4.1.1. DEGREE CENTRALITY

The degree centrality measure is concerned with an individual node and more importantly the particular node's position within the network (Armstrong & McCulloh, 2010; Hanneman & Riddle, 2005). The core principle of centrality is that a node that is located more centrally, i.e. has more connections than other nodes, will have a greater influence on the network as a whole. The quantitative measure used to describe the influence of such a node is referred to as its degree centrality, and is calculated by using several node properties, such as the number of connections leading into the node, the number of connections leading out of the

node, and the sum of the aforementioned connections (Armstrong & McCulloh, 2010). A node with a high degree centrality would likely be at risk as a target for attacks that propagate through a network, such as chain phishing and email-based ransomware. The calculation of degree centrality is mathematically expressed as follows (Clemente *et al.*, 2016):

Given node n_i in an unweighted graph *G*, the degree centrality of n_i, given as $C_D(n_i)$, is calculated as

$$C_D(n_i) = \sum_{j=1}^{n} a_{ij} = \sum_{j=1}^{n} a_{ji} \tag{3.1}$$

where a_{ij} and a_{ji} are elements of the adjacency matrix of *G*.

The degree centrality index of a node can alternatively be calculated using formula 3.2 (CASOS, 2019). This formula also accounts for weighted edges, and therefore the maximum weight of any edge in the graph is used. This value is referred to as the maximum link value.

Let *A* be the input network with *N* nodes and maximum link value *V* (for weighted networks). The degree centrality for a node *i* is calculated as follows:

$$Total\ Degree\ Centrality\ for\ node\ i\ = \frac{\sum_1^N A(i,:) + \sum_1^N A(:,i) - \sum_1^N A(i,i)}{2V(N-1)} \tag{3.2}$$

where $\sum A(i,:)$ is the total of all arcs beginning at node i $(out\ degree)$, $\sum A(:,i)$ is the total of all arcs leading into i $(in\ degree)$, and $A(i,i)$ is the total of all loop arcs, i.e. all arcs where node i is both the source and destination node of an arc.

3.4.1.2. CLOSENESS CENTRALITY

Closeness centrality is calculated by determining all the geodesic distances (i.e. the shortest distances) to all other nodes within the network (CASOS, 2019), and takes all indirect ties a node possesses, together with all direct ties, into account. Degree centrality, in contrast, places focus only on adjacent nodes (Hanneman & Riddle, 2005; Wasserman & Faust, 1994). A node that has a high closeness centrality value is considered to be a good source of information, whereas nodes with a high degree centrality aid in the diffusion of information throughout the entire network. This means that analysis of the nodes with the greatest closeness centrality values should provide the best information with regard to the information in the network. The mathematical formulation of closeness centrality is as follows (Clemente *et al.*, 2016):

Given two nodes n_i and n_j in a graph *G* with *n* nodes and a geodesic distance between them of $d(n_i, n_j)$, the closeness of node n_i, given as $C_{(C)}(n_i)$, is calculated as

$$C_{(C)}(n_i) = \left[\sum_{\substack{j=1 \\ i \neq j}}^{n} d(n_i, n_j)\right]^{-1} \quad (3.3)$$

The closeness centrality of a node can alternatively be calculated using the following formula (CASOS, 2018):

Let A be the unimodal input network with N nodes (i.e. a network where all nodes have a singular value), minimum link value v, and maximum link value V, and let D be the distance network defined as:

- $D(i,j)$ = shortest path length from i to j, if a path exists from i to j
- $D(i,j) = NV$, if no path exists from i to j
- $D(i,i) = 0$

The following computes the sum of the geodesic path lengths from node i to all other nodes (d):

$$d = \sum D(i,j), \quad for\ all\ nodes\ j \quad (3.4)$$

$$Closeness\ Centrality\ value\ for\ node\ i = \frac{V(N-1)}{d} \quad (3.5)$$

3.4.1.3. Betweenness Centrality

When examining interactions between two non-adjacent nodes, the nodes that lie between the paths connecting the two nodes have some control over the interaction between the two nodes (Wasserman & Faust, 1994). The betweenness centrality measure is a representation of the number of times that a particular node finds itself in the geodesic path of other nodes within the entire network(Armstrong & McCulloh, 2010). This measure is reflective of the number of indirect nodes that are connected to a particular node. A node with a high betweenness measure is therefore at risk of being overburdened, as such a node would spend a portion of its time facilitating interactions between other nodes.

A node that finds itself as an intermediary in an information exchange relationship between two nodes is also considered to be in a position of power, as any information exchanged between the two nodes will likely have to go through the intermediary. The intermediary has a unique position of power in this instance, as it can determine not only the fidelity of the information being exchanged, but also whether information is exchanged at all. Thus, as the number of nodes that rely on such an intermediary increases, so too does the relative power the intermediary node possesses (Hanneman & Riddle, 2005). Betweenness centrality is mathematically calculated in the following manner (Clemente *et al.*, 2016):

Given node n_k in a graph $G = (V, E)$, the betweenness centrality of n_k, given as $C_b(n_k)$, is calculated as

$$C_b(n_k) = \sum_{\substack{n_i, n_j \in V \\ i \neq n \\ j \neq k}} \frac{g_{ij}(n_k)}{g_{ij}} \tag{3.6}$$

where $n_i, n_j, n_k \in V$, $i, j, k = (1,2, \dots, n)$, $g_{ij}(n_k)$ is the number of geodesic paths between n_i and n_j that pass through n_k, and g_{ij} is the number of geodesic paths between n_i and n_j.

The betweenness centrality of a node can alternatively be calculated using the following formula (CASOS, 2019):

Let D be the distance network for the input network; *D(i,j)* = shortest path distance from i to *j*, and zero if no path exists, and let C be the network of the number of shortest paths for the input network (so that *C(i,j)* = number of shortest paths from i to *j*, and zero if no path exists.

The following computes the total fraction of shortest paths that node i lies on:

$$\sum \frac{C(u,i)C(u,v)}{C(u,v)} \quad for\ all\ (u,v)\ where\ D(u,v) = D(u,i) + D(i,v) \tag{3.7}$$

This value is then normalised by the maximum number of shortest paths possible to get the betweenness centrality.

3.4.1.4. Eccentricity Centrality

The eccentricity centrality of a node describes how far removed the node is from the other nodes in the network. A highly eccentric node, i.e. one that is far away from other nodes in the network, will still have a low eccentricity centrality value; the eccentricity of a node is calculated as the inverse of its eccentricity centrality. The eccentricity centrality index of a node is calculated in the following way (Clemente *et al.*, 2016):

Given two nodes n_i and n_j in a graph *G* with *n* nodes and a geodesic distance between them of $d(n_i, n_j)$, the eccentricity centrality of node n_i, given as $C_{(ecc)}(n_i)$, is calculated as

$$C_{ecc}(n_i) = \frac{1}{\max[d(n_i, n_j)]} \tag{3.8}$$

3.4.1.5. Eigenvector centrality (PageRank & Katz)

Eigenvector centrality measures the extent to which a particular node is connected to other nodes that are considered to be highly connected or are of some particular importance

(Armstrong & McCulloh, 2010). Nodes that have a high eigenvector index are important to note since they are considered to possess emergent leadership properties (Borgatti, 2005). Nodes with a high eigenvector centrality are therefore also considered risks, as they could potentially not only influence information security culture (which is what ultimately determines compliance with policies and controls), but also the information handled by those they have influence over. The calculation of eigenvector centrality follows the following procedure:

Let *A* be the unimodal input network (if the network is not symmetric, links are added to make it symmetric) with *N* nodes, let *K* be the number of components in the network *A*, let N_k be the nodes in the k^{th} component, and let V_k be the dominant eigenvector computed on the sub-network induced by the nodes N_k. The V_k values are then scaled by multiplying them by $\frac{N_k}{N}$. The Eigenvector Centrality is then computed as the combined vector of all scaled component V_k values.

Alternatively, the eigenvector centrality of a node can be calculated as follows (Clemente *et al.*, 2016): Given node n_i in a graph $G = (V, E)$, the eigenvector centrality index of n_i is defined as the i^{th} component of the eigenvector $\vec{x}$, corresponding to the greatest eigenvalue of the characteristic equation $A\vec{x} = \vec{x}\lambda$, where *A* is the adjacency matrix of *G*.

PageRank centrality, which is a special form of eigenvector centrality, determines the importance of a node by calculating how often a node would be visited if a random walk was taken through the network (Tu *et al.*, 2018), and is calculated as follows (Clemente *et al.*, 2016): Given node n_i in an unweighted digraph $G = (V, E)$, the PageRank centrality index of n_i, given as $PR(n_i)$, is calculated as

$$PR(n_i) = p \sum_{j \neq 1} \frac{a_{ji}}{k_j^{out}} PR(n_j) + q, \tag{3.9}$$

where *a* is the adjacency matrix of *G*, p is the heuristic probability of node n_i connecting to another node, and q is the basic popularity of each node.

3.4.2. Boundary Spanner

A boundary spanner node is quite simply a node that, when removed, will cause the creation of two disconnected sub-graphs. A boundary spanner is therefore a node that acts as the only connection between who groups within a network (Cormen *et al.*, 2001). In order to find such a node, a depth-first search is executed on the whole network. A boundary spanner, alternatively called an articulation point, is a node that has at least two children. If all of the walks that connect those children to one another include the parent node, then that parent is a boundary spanner. Alternatively, if there is no walk that connects

the children save for those that include the parent node, then the parent node is an articulation point.

The boundary spanner value for a node is binary, i.e. either 1 or 0. The reason for this is that a node cannot be a partial boundary spanner, as nodes cannot be partially removed from the network. Therefore, the value for a node's boundary spanner metric is either 1, which indicates that it is a boundary spanner, or 0, which indicates that it is not a boundary spanner.

3.4.3. Shared Situation Awareness

Shared situation awareness is a measure of how well an individual node knows what other nodes in the network are doing (Graham, 2005). A node that is high in shared situation awareness is therefore well informed of what other similar groups are doing.

In order to determine shared situation awareness, the following procedure is followed (CASOS, 2019):

Let:

- $\alpha, \beta, \delta, \gamma, \mu$ be decimal numbers;
- A be an Agent x Agent interaction/communication network;
- P be an Agent x Agent physical proximity network;
- S be an Agent x Agent social demographic similarity network;
- e be the eigenvector centrality measure computed on A; and
- G be the geodesics between agents computed on A.

Then the shared situation swareness SSA_{ij} between agents i and j is:

$$SSA_{ij} = \alpha e_i e_j + \beta P_{ij} + \frac{\delta S_{ij}}{\gamma G_{ij}} + \mu A_{ij} A_{ji} \tag{3.10}$$

3.4.4. Structural Holes Constraint

Structural holes constraint is used to determine if a node is prevented, i.e. constrained, from performing its duties as a result of crucial connections being missing (CASOS, 2019). The measure is calculated as follows (Burt, 1992):

Given three nodes i, j, and q in graph G, and a value p_{xy} indicating the strength of the connection between any two nodes x and y in graph G, the Structural Holes Constraint between nodes i and j SHC_{ij} is calculated as

$$SHC_{ij} = \left(p_{ij} + \sum_{q} p_{iq} p_{qj} \right)^2 , \qquad q \neq i, j \tag{3.11}$$

It should be emphasised once more that this is not an exhaustive list in any way; as of 2019, ORA-Lite allows for the calculation of over 170 metrics (CASOS, 2019). The eight metrics described in this section were selected either due to their relevance to this study, or as a result of their known association with structural risks in the literature (Armstrong & McCulloh, 2010).

3.5. COMMUNITY DETECTION

As with any society, a social network can be divided into a number of communities. A community, within this context, describes a portion of a network wherein the nodes are more densely connected to one another than to the rest of the network (Jonnalagadda & Kuppusamy, 2018). An example graph of a network with three communities is shown in Figure 3.14.

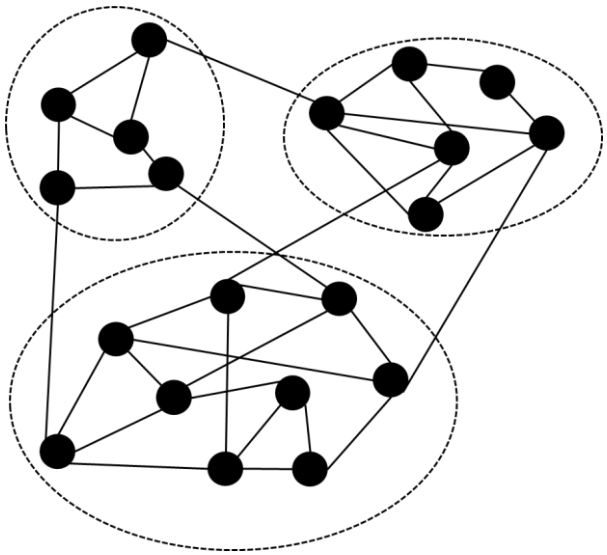


FIGURE 3.14: SIMULATED NETWORK WITH THREE COMMUNITIES

Identifying these communities has practical value, as the member nodes of a community may have similar characteristics. Studying the interactions between various social groups in a social network may also be difficult if these groups are not clearly identified (Zhou *et al.*,

2016). It is therefore important to take note of community detection techniques, as it may help in simplifying evaluation processes by dividing a social network into meaningful sub-networks. In this section, some of the methods that can be used to detect communities in a social network are discussed. The specific techniques introduced are the hierarchical clustering and edge removal methods (Girvan & Newman, 2002), as well as the cooperative game method (Zhou *et al.*, 2016) and the evolutionary node centrality algorithm (Žalik, 2019).

3.5.1. HIERARCHICAL CLUSTERING

The hierarchical clustering technique, as the name suggests, allows for the detection of communities by determining the relative strength of the relationships between nodes and then uses these strengths to produce a hierarchical structure, such as the one shown in Figure 3.15. This structure can then be used to identify the various communities that exist within the network. Because the method organises nodes into a hierarchy, the existing edges are only used to determine a weight for the relationships between each of the nodes. Once these weights have been determined, the edges are removed and new edges are added. The process for adding new edges involves identifying the relationships with the highest weights and then connecting their associated nodes to one another. The process is then repeated for each of the nodes until all of the nodes are connected to one another.

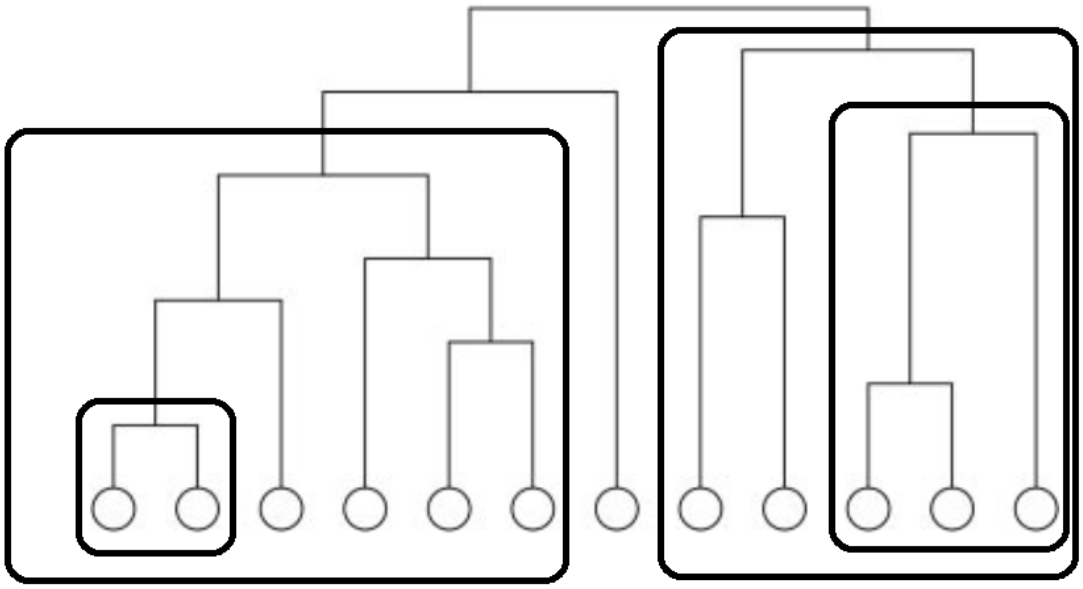


FIGURE 3.15: AN EXAMPLE OF A SMALL HIERARCHICAL CLUSTERING TREE. THE CIRCLES AT THE BOTTOM REPRESENT NODES AND THE TREE SHOWS THE ORDER IN WHICH THEY JOIN TO FORM COMMUNITIES (GIRVAN & NEWMAN, 2002). FOUR COMMUNITIES OF DIFFERING SIZES HAVE BEEN ENCIRCLED.

Once the relationship hierarchy for the network has been created, grouping the nodes into communities is done by traversing the hierarchy, or hierarchical clustering tree. Each node is placed into a group along with the nodes that are closest to it in the tree. Depending on how granular the communities have to be, any number of groupings can be identified. If possible, the ideal is to use a graphical method to approximate the number of extant communities, and then group the nodes into that number of communities. This is not always possible for very complex networks. One of the limitations of the hierarchical clustering method is that it does not account for the existence of nodes that may simultaneously be members of multiple communities.

3.5.2. EDGE REMOVAL

The second technique used to detect communities makes use of the betweenness centrality measure. As nodes in communities tend to be more closely connected to one another than to the rest of the network, it is reasonable to assume that the edges with the highest betweenness measures represent the relationships that connect communities together. The edge removal method exploits this by incrementally removing the edges with the highest betweenness centrality until an arbitrary threshold is reached. The nodes that remain connected to one another are members of the same community. This method is computationally more reliable than the hierarchical clustering technique, as it relies on calculating a known graph measure, rather than on weights that have to be defined beforehand. The method is arguably more adaptable as well, as other meaningful metrics such as boundary spanner measures can be used to divide the network into communities.

3.5.3. COOPERATIVE GAME METHOD

The cooperative game method aims to use concepts from game theory to identify communities. Specifically, the rules and behaviour inherent to cooperative games are used to predict the structure of communities. This is necessary as the edge removal and hierarchical clustering methods have difficulty in fully identifying communities in networks where nodes are members of multiple communities. In a cooperative game, players have to form coalitions in order to win, and each member has a certain value within a coalition. Each of these coalitions, once fully formed, represent a distinct community. By calculating the advantages a node will get by joining each of the communities, it is possible to determine the community the node will likely join. The assumption is therefore made that each node will join the community that is personally most advantageous. Additionally, if two communities are equally advantageous, it is possible for a node to be a member of both

communities. All of the communities that exist in the network are identified once all of the nodes have joined their respective communities.

3.5.4. Evolutionary Node Centrality Algorithm

The most recent of the community detection methods introduced in this chapter makes use of the Net-Degree algorithm, which approaches community detection as a multi-objective optimisation problem. In order to solve this problem, the Net-Degree algorithm adapts an evolutionary algorithm that uses the centrality measures of each node to assign a fitness value to each node. The goal is then to maximise the fitness of every node in the network by using genetic techniques, such as crossover and mutation; these techniques are outside the scope of this study, but the paper by Žalik (2019) provides more detail. Once the optimisation problem is solved, and the maximum fitness of each of the nodes has been determined, the nodes are grouped according to their dominance within the network. These final groupings represent the different communities within the network.

3.6. Network Optimisation and Monitoring

One of the aims of this study is to demonstrate how network optimisation, when applied to social networks, can be used to help address information security risks. In this section some of the various network optimisation techniques will be discussed.

3.6.1. Optimisation of the Critical Diameter and average path of social networks

One of the simpler ways to optimise a social network is to minimise the Average Shortest Path Length (ASPL) of a network (Du *et al.*, 2017). The ASPL, which represents the average geodesic distance between all nodes in a network, is a good measure of connectivity and network robustness. Most of the techniques that aim to improve the ASPL of a network focus on adding a number of so-called shortcut edges to the network, thereby reducing the geodesic distance between certain nodes; the selection of the shortcut edges is known as the shortcut selection problem (Meyerson & Tagiku, 2009). As Du *et al.* (2017) demonstrate, there exists a turning point with shortcut addition algorithms where the ASPL begins to decrease linearly with the addition of new edges. This is as a result of the increase in the

diameter of the network. The network diameter measure at this turning point is subsequently called the critical network diameter.

The method proposed by Du *et al.* (2017) is aimed at optimising ASPL by finding the critical diameter of a network. The proposed method, which utilises a genetic algorithm that is outside the scope of this study, randomly adds and selectively removes edges until the critical network diameter is obtained.

3.6.2. Computer Network Optimisation

While computer networks and social networks differ in the groups they represent, they both share similar properties due to the nature of graph theory, upon which network theory is based. As a consequence of this, computer networks can be described using familiar metrics, such as Average Shortest Path Length (ASPL), Network Diameter (ND), Degree Centrality and Betweenness Centrality (Rezazad, 2011). Because of these similarities, the principal methods of computer network organisation could be used to propose improvements to a social network, if meaningful metrics are selected as part of the optimisation.

One of the optimisation methods that utilise network metrics was proposed by Rezazad (2011). In this method, various edges are selected for Link Transfer, which combines Link Addition (LA) with Link Removal (LR). The basic premise is that, by transferring edges, the metrics of certain nodes can be altered to improve the state of the overall network. The two main adaptations that can be made, namely LA and LR, occur under the following circumstances:

- **LR:** Within this method, nodes with high overall centrality are considered a risk due to the negative implications their loss would have. Additionally, nodes with a significant number of connections may be too well connected, indicating a certain level of unnecessary redundancy. When selecting a node for LR, the algorithm first selects the two most central nodes with the highest degree centrality that are interconnected. The edge connecting these nodes is then removed. The algorithm iterates through nodes with the highest degree centrality, removing edges where necessary, until the overall degree centrality per node has been minimised.
- **LA:** If edges were removed from the network, it is likely that certain vital relationships were affected. As this algorithm focusses on optimising computer networks, the ideal is for all nodes to have roughly the same degree centrality, as this would balance redundancy with efficiency. Therefore, with edges having been removed to decrease the degree centrality of certain nodes, it is only reasonable that the same number of edges should be added to increase the degree centrality of

under-connected nodes. The process for selecting which edges to add broadly follows that of the process used to select edges for LR: the two least central nodes with the lowest degree centralities are selected and an edge is added to connect the two nodes.

While not perfectly applicable to SNA, this method provides insight into how a network can be optimised using network metrics as a guide. An adapted method based on this algorithm is presented in Chapter 6.

3.6.3. Monitoring a Social Network

In a doctoral thesis by McCulloh (2009), a method is introduced whereby changes in a social network can be identified quickly. The method makes use of statistical control charts (SCC), which is a quality control technique used to detect anomalous changes. SCCs measure statistical "normality" within an environment and can be used to detect any significant departures from typical behaviour. The study mentions three types of control charts, namely the cumulative sum control chart, the exponentially weighted moving average control chart, and the scan statistic chart. The use of these charts provides a way to monitor social networks in real time, and identify when and if significant changes have occurred. Furthermore, by evaluating a network over time, the stability of the network can be determined: if a network has a high frequency of abnormal statistical shifts, determined using the network's metrics over time, then there is a good chance that the network is unstable. If, for example, the overall average eigenvector centrality in the network fluctuates violently, it may mean that there is no coherent or consistent leadership structure in place. As an unstable network may be at risk simply due to its instability, this allows for an additional measure of risk analysis. This increases the usability of the techniques described in later chapters, as it allows for adaptive changes to be made only when necessary, and it may be possible to stabilise an unstable network using the proposed methods. The basic principle of statistical control charts will now be briefly introduced, based on the work by Render *et al.* (2012).

In order to demonstrate the basic functioning of a control chart, the mean-control chart will be discussed. This particular control chart, which utilises a sample mean $\bar{x}$ and a sample standard deviation $\sigma_{\bar{x}}$, makes use of the central limit theorem. The central limit theorem states that the distribution of sample means will follow a normal curve as the sample size grows. The theorem also states that the mean of the distribution of sample means, which is denoted as $\mu_{\bar{x}}$, will equal the mean μ of the overall population. $\sigma_{\bar{x}}$, on the other hand, is the population standard deviation σ_x divided by the root of the sample size n. This means that calculating the sample distribution standard deviation is done using the following formula:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \tag{3.12}$$

The mean control chart uses these values to define ranges wherein sample values should be if no abnormal activity is present. The values of these ranges are calculated using the following formulas:

$$Upper\ control\ limit = \bar{\bar{x}} + z\sigma_{\bar{x}} \tag{3.13}$$
$$Lower\ control\ limit = \bar{\bar{x}} - z\sigma_{\bar{x}} \tag{3.14}$$

where $\bar{\bar{x}}$ is the mean of the sample means and z is the number of normal standard deviations (2 for 95.5% confidence and 3 for 99.7% confidence).

Once these values have been calculated, the sample mean values can be plotted on a chart such as those shown in Figure 3.16.

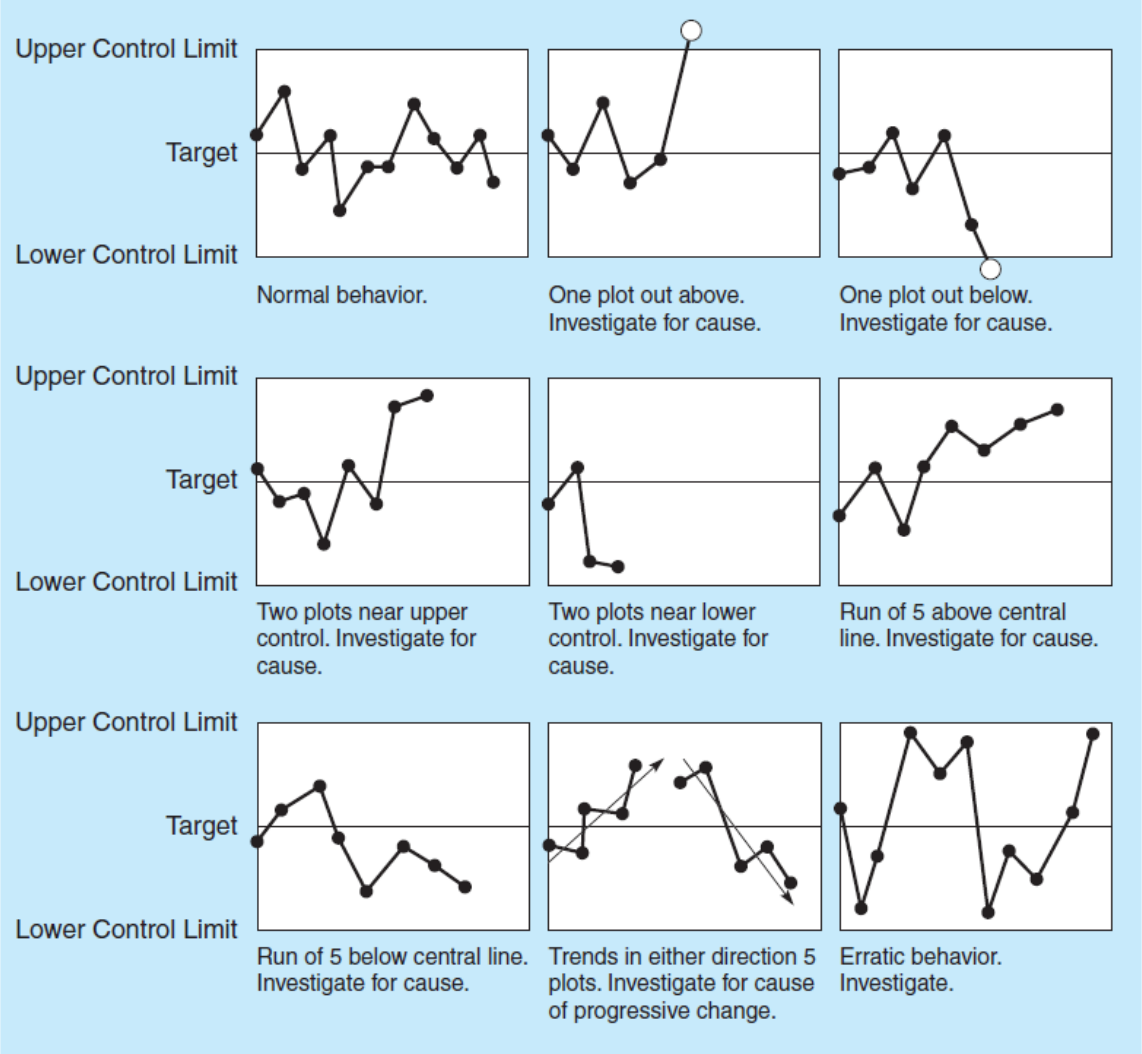


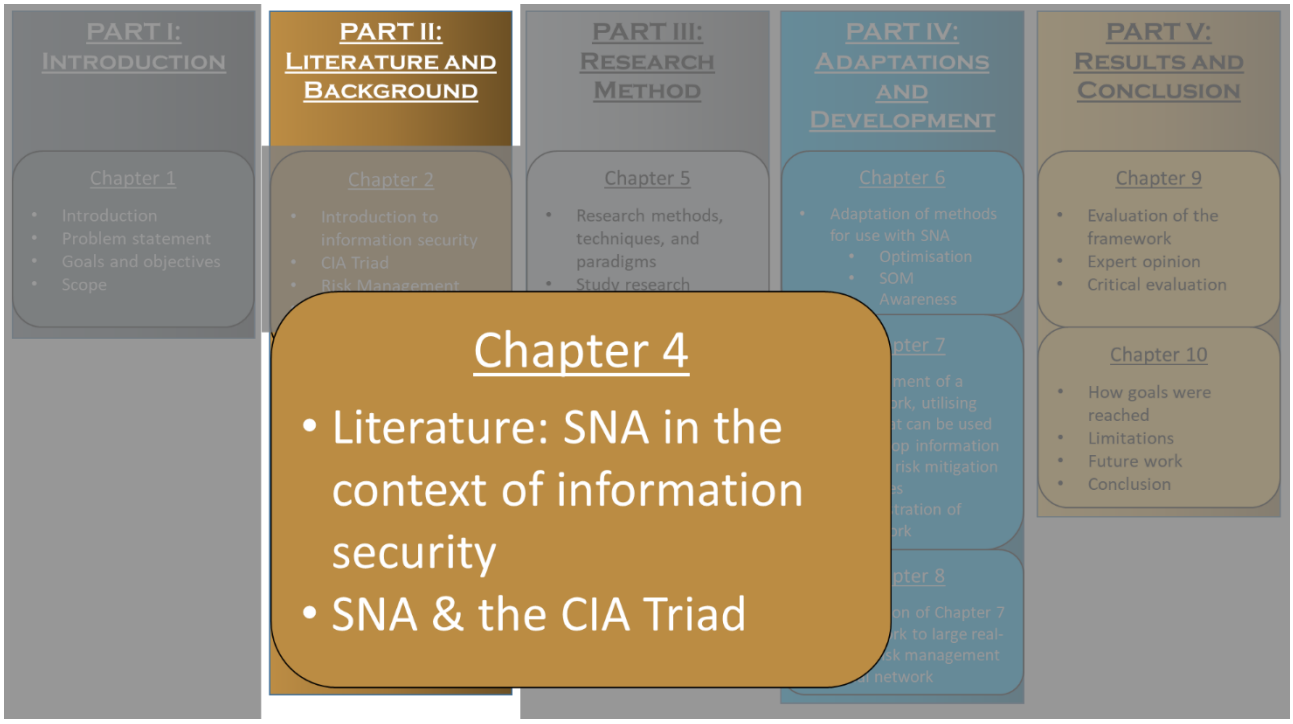
FIGURE 3.16: INTERPRETING CONTROL CHARTS (RENDER *ET AL.*, 2012)

When this technique is applied to social networks, a specific set of metrics are calculated for the network at certain times. By using the values of these metrics as sample values, the control limits for a SCC can be calculated using equations 3.13 and 3.14. By then plotting all of the calculated metric values on the control chart, the specific time at which a significant change occurred can be identified. In addition, due to the nature and use of SCCs, it may be possible to identify positive or negative trends in the changes to the network. All of this makes SCCs a valuable tool when monitoring social networks, as they can be used to quickly identify important changes.

3.7. CHAPTER SUMMARY

In this chapter, some of the various principles of SNA and graph theory were discussed. The chapter started with a discussion of the basics of SNA, followed by an overview of selected graph theory topics, including metrics. These metrics are used in Chapter 7 and 8 in order to calculate the risk level for nodes in a social network. Some of the techniques that can be used to visualise graphs were then described, followed by a selection of the methods that can be used to detect communities in networks. The use of these community detection techniques within the context of this study is revisited in Chapter 7. This chapter then concluded by introducing techniques that could be used to optimise and monitor social networks.

In the next chapter some of the ways in which SNA can be, and has been, applied to information security are discussed.



CHAPTER 4: SOCIAL NETWORK ANALYSIS IN THE CONTEXT OF INFORMATION SECURITY

CHAPTER HIGHLIGHTS:

- In which ways have information security studies implemented SNA in the past?
- What is the relationship between certain SNA metrics and the CIA Triad?

4 SOCIAL NETWORK ANALYSIS IN THE CONTEXT OF INFORMATION SECURITY

In the first two chapters of Part II, some of the various principles and techniques used in information security and Social Network Analysis (SNA) were discussed. In this study SNA is used to evaluate information security risk and, as such, a discussion of how SNA is used in information security in the literature is necessary. This chapter focusses on the discussion of this topic. The chapter starts with an overview of a selection of literature sources that use SNA to evaluate information security risk. After this, an evaluation of how SNA metrics can be used to determine CIA risk follows. The penultimate section of this chapter includes a list of various studies that are relevant to the broad topics of SNA and information security, but lie outside the scope of this study. The chapter then concludes Part II with a summary of the important facts contained in Chapters 2-4.

4.1. REVIEW OF LITERATURE SOURCES USING SNA IN THE CONTEXT OF INFORMATION SECURITY

A selection of literature sources that use SNA in the context of information security is discussed in this section. For each source, the discussion starts with a description of what was done in the study. This is followed by an explanation of the approach and method published in each source. Each section will then conclude with an appraisal of the contributions of the discussed source. While this selection of sources is not exhaustive, the number of studies that use SNA in the context of information security are limited, and the sources that were selected are presented as examples of research that has been done in the context of SNA and risk. Table 4.5 also contains a summary of various additional sources that use SNA in the context of information security and risk. For each of the following sources, the title of the publication is given in the heading, while the reference is provided in the first sentence of each section.

4.1.1. ORGANIZATIONAL RISK USING NETWORK ANALYSIS

In a paper by Armstrong and McCulloh (2010), entitled “Organizational risk using network analysis“, SNA is presented as a method that can be used to evaluate organisational risk. The paper starts by making a well-reasoned argument for the use of SNA to address organisational risk, before moving on to a description of some of the metrics found in SNA.

Most of the metrics mentioned in the paper, specifically degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality, and how they relate to information security, are further discussed in section 4.2.1. The metrics themselves are discussed in Chapter 3.

Following the discussion on the meaning of the various metrics, the paper moves on to discuss SNA in more detail, and provides arguments for how the metrics should be evaluated. In the conclusion, the authors state that the usability of the technique relies on the quality of the data. The authors also propose that managers should use software tools specifically tailored to their situations when using SNA to evaluate organisational risk.

Appraisal: The paper provides a solid basis for the process that can be used to determine which metric can be used to evaluate which type of organisational risk. No proposals are made that deal with how risks are managed or addressed after they have been identified, and the reader is left with the impression that it is up to a manager to figure out how to address any risks that have been identified.

4.1.2. Applying network analysis to investigate interpersonal influence of information security behaviours in the workplace

A publication by Dang-Pham *et al.* (2017b) starts by comparing workplace realities to managerial ideals, specifically by stating that workers tend to approach information security in a much more relaxed fashion than management would like. The discussion then proceeds to provide a substantiation for the study, resting on two points: the first is to study the interaction between employees with regard to information security, and the second is to expand the use of SNA within the field of behavioural information security research. An overview of relevant literature is provided, followed by a description of various hypotheses that would be tested during the course of the study. The method used in the research is discussed next: the study made use of a questionnaire consisting of eight questions that was sent to employees at an organisation. They obtained a response rate of 71%. The degree centrality was calculated for each of the responding employees and Exponential Random Graph Modelling (ERGM) was applied to test the various hypotheses stated earlier in the paper. The specific results obtained in the study are then discussed. The paper concludes by stating that one of the aims is to provide recommendations to security managers. A number of factors that can influence which employees are considered reliable sources for information security knowledge are highlighted. It is also emphasised that the network structure was influenced by the unique culture of Vietnam, which is where the evaluated company resides. The authors conclude by stating that an SNA approach deserves greater attention in the field of behavioural information security, as it could help develop so-called “shadow security” networks, which should in turn help increase workplace security.

Appraisal: This particular study provides a good substantiation for the use of SNA to improve information security in the workplace. There are limits to the study, however, in that it only uses degree centrality measures as inputs for the ERGM. This is not a problem here, as the intent was to evaluate the impact of influence on security culture exclusively. The study therefore does not address risks as such, and only aims to review existing interactions and social influences.

4.1.3. Applying social network analysis to security

The work done by Philips *et al.* (2015) aims to use SNA to determine what kind of information an attacker could obtain from a member of a social network, specifically the hierarchical importance of the member. The research makes use of data obtained from emails to build a social network and, ultimately, determine SNA metrics. The metrics were used independently to determine how well they can be used to classify each of the nodes into one of seven categories, with each category representing a distinct managerial level. The results obtained show that certain metrics are more suited for classification purposes depending on the particular properties of the node that the metric evaluates. It is also mentioned that certain traits, which can be used to identify insider threats, were identified using the SNA data. The study also demonstrates the use of SNA metrics in machine learning.

Appraisal: The paper demonstrates how SNA can be used to identify important individuals in a social network. Sadly, and despite the title, the study does not really attach SNA to information security, and merely makes a closing reference to insider threat identification. Nevertheless, the work done serves the vital role of confirming that email data can be used to build social networks, that it is important to select specific SNA metrics using their traits and intended use, and that SNA can be used to identify specific members in a social network who have particular traits.

4.1.4. Understanding of Impact and Propagation of Risk based on Social Network Analysis

In a paper by Ongkowijoyo and Doloi (2018), entitled “Understanding of Impact and Propagation of Risk based on Social Network Analysis”, a method is proposed whereby SNA is used to capture, draw, and simulate the patterns inherent to risk impact propagation. The paper starts by discussing urban infrastructure and the complexities involved in evaluating the associated risks. The proposed method is then discussed, which features the use of SNA

to model risk relationships. The goal is to develop a single network wherein cause-effect relationships form arcs, and the various identified risks form nodes. The third part of the study describes how the model was validated. Specially designed questionnaires were used that were sent to industry participants, and 126 individuals from eight stakeholder groups participated. The model implemented a number of metrics, such as degree centrality, betweenness centrality, and eigenvector centrality in order to determine the risk relationships. The authors conclude that the model is a viable method to evaluate urban infrastructure risk, and should prove invaluable in improving infrastructure management procedures.

Appraisal: While this study does not deal with information security risk, the approach taken in order to manage risk is broadly similar. The proposed model makes use of SNA and associated centrality metrics in order to evaluate risk, and demonstrates that the SNA risk model is valid compared to real-world expectations. The paper therefore supports the notion that SNA can be used to model and manage urban infrastructure risk situations, which is similar to information security risk scenarios.

4.1.5. Applying network analysis to assess coastal risk planning

Roca *et al.* (2018) propose SNA as a method to evaluate the risk management processes involved in coastal risk planning. The paper focusses on the situation as it relates to the management of the Catalan coast, and attempts to determine the impact the various risk management structures have on the coast. The various coastal risks that have to be managed are introduced, and point out that certain risks are approached centrally, while other risks are managed more locally. The collected data are then described, specifically the relationships that exist between risks, stakeholders and plans. The risks are evaluated using the various stakeholders and plans, and the management structures are identified. The article concludes by identifying various inherent structural risks, and supports the use of SNA to assess risk management processes.

Appraisal: This study, like the one done by Ongkowijoyo and Doloi (2018), does not focus specifically on information security risk, but does support the use of SNA to assess and manage risk strategies. Additionally, the paper demonstrates how risk data involving risks, stakeholders and plans, alternatively controls, can be evaluated using SNA. In conclusion, this paper has merits on a number of grounds, despite not dealing directly with information security risk management.

As demonstrated by the number of studies that use SNA to evaluate risk, SNA and risk management are very compatible. The limited number of SNA studies dealing specifically

with information security risk do however show that there is still a significant amount of work that can be done in this area.

4.2. SNA METRICS AND THEIR RELATIONSHIP TO THE CIA TRIAD

In this section, the SNA metrics, discussed in Chapter 3, are discussed in terms of the CIA triad, which is described in Chapter 2. The intention in section 4.2.1 is to demonstrate how SNA metrics can be used to evaluate information security risk using these metrics. A table summarising the metrics and their relationships to the CIA triad concludes the section. In Section 4.2.2 a simple illustrative example using simulated data is used to demonstrate how the information security risk of a social network can be evaluated using the metrics discussed.

4.2.1. CIA RATIONALE FOR SNA METRICS

The rationale behind the association of certain metrics to specific members of the CIA triad is discussed in this section. As described in Chapter 2, the CIA triad, which features Confidentiality, Integrity, and Availability, forms the core of information security theory. It should be noted that, while all members of the CIA triad can arguably be associated with each of the metrics discussed, the intention is to associate each of the metrics to the members of the CIA triad that are most relevant to the specific metric.

Degree Centrality: Both Armstrong and McCulloh (2010) and Dang-Pham *et al.* (2017b) state that nodes with a high degree centrality are highly influential in the network, and that these nodes are generally responsible for diffusion of information in the network. As a result of these aspects, nodes with a high degree centrality are considered a security risk: their loss has a detrimental effect on the network, and they may possibly have the power to influence information in the network. These nodes are therefore associated with the **Integrity** aspect of information security. The metric can also be associated with the **Reliability** aspect of the extended CIA model, which is a sub-member of **Availability**. There are three primary forms of degree centrality: **in-degree centrality**, which counts the number of edges leading into a node, **out-degree centrality**, which measures the number of edges leaving a node, and **total-degree centrality**, which takes both in- and out-degree centrality into consideration.

Closeness centrality: Nodes with a high closeness centrality are typically the best sources of information in the network. When closeness centrality is approached in the context of risk, these nodes can be identified as risks to **Confidentiality** due to their access to information.

It can be argued that, should a node with a high closeness centrality fall victim to a phishing- or social engineering attack, the impact may be much more profound than if a node less central had been attacked.

Betweenness centrality: When nodes have a high measure of betweenness centrality, and therefore act as brokers, they can be considered a risk to both the **Confidentiality** and **Integrity** of information in the network. The rational is that such a broker, which may possibly have trust in the network, is considered crucial to the flow of information. Because of this, a node with a high betweenness centrality may have a significant amount of power with regard to who has access to which information. As the broker can then potentially modify the information or choose who has access to it, the risks may be realised. A node with a high betweenness centrality may also potentially be a risk to **Availability**, as the loss, either temporary or permanent, of such a node may cripple information flow in the network.

Eccentricity centrality: The eccentricity centrality of a node is a measure of how distant it is in the network. While a node with a high measure of eccentricity may not pose an immediate threat to confidentiality or integrity, due to being removed from most of the information, it may still pose a threat to availability. The argument can be made that, especially if a corporate network is considered, each node has a reason for being in the network. If a node is both highly eccentric and crucial to business processes, this may cause delays in the transfer of information, which is a risk to **Availability**.

Eigenvector centrality: A node with a high eigenvector centrality is highly connected to other highly connected nodes. This typically indicates that the node is an emergent leader and is expected to be able to “get things done”. This metric, as a result, may indicate an overall risk to all three part of the CIA triad. The argument in this case is that an informal leader may potentially not only be able to influence information security culture (which is what ultimately determines compliance with policies and controls), but also the information handled by those they have influence over. A node with a high eigenvector centrality can therefore be considered an overall risk to **Confidentiality**, **Integrity** and **Availability**, but can also be considered a potential champion.

Boundary spanner: When a node has a high boundary spanner value, it acts as a sole connection between various sub-networks. The loss of such a node is therefore likely to completely isolate parts of the network and the node should therefore be considered a high risk to **Availability**. Furthermore, similar to betweenness centrality, such a node may act as a broker between the sub-networks. It is possible, however, that such a node may not have a high betweenness centrality. The associated risks to integrity and confidentiality may therefore not necessarily apply to a boundary spanner node.

Shared situation awareness: Shared situation awareness is a measure of how well a node is informed about circumstances in other similar parts of the network. As such, a node with a

high shared situation awareness may be considered an opportunity, whereas the absence of any nodes with high shared situation awareness poses a problem. It can be argued that the absence of any nodes with high shared situation awareness is an indicator of a lack of cooperation, or a high rate of isolation between similar groups. Based on this, it is possible to associate a low overall amount of shared situation awareness with a lack of **Availability.**

Structural holes constraint: If a node has a high measure of structural holes constraint it will find itself unable to execute its duties properly as a result of missing, crucial, connections. Consequently, this metric indicates an immediate organisational risk to productivity, but also poses a risk to nodes that depend on information from such a node. While being unable to complete necessary tasks may not necessarily impact on confidentiality, an individual forced to work in such circumstances may resort to data fabrication. A high measure of structural holes constraint can therefore indicate that a node is a risk to both **Availability** and **Integrity**. Table 4.1 provides a summary of the metrics discussed above. A short version of the rationale for each metric is also included.

TABLE 4.1: SUMMARY OF SNA METRICS IN THE CONTEXT OF THE CIA TRIAD

SNA Metric	CIA Triad			Rationale
	C	I	A	
Degree centrality		X	X	Nodes with a high degree centrality have influence in the network and are connected to a significant portion of the network; their loss may be detrimental
Closeness centrality	X			A node with a high closeness centrality has access to a significant amount of information in the network.
Betweenness centrality	X	X		A high betweenness centrality may indicate a node that may be prone to information brokering and tampering.
Eccentricity centrality			X	A high eccentricity centrality in a node may cause delays with the transfer of information
Eigenvector centrality	X	X	X	Nodes with a high eigenvector centrality are considered emergent leaders and, depending on their influence and attitude, may be a general risk (either as an opportunity or a threat, depending on the circumstances)
Boundary spanner			X	The loss of a boundary spanner node may cause isolation in the network, which would negatively impact the flow of information
Shared situation awareness			X	A low measure of shared situation awareness in a network in general may indicate a high measure of effective isolation between groups
Structural holes constraint		X	X	Nodes constrained by structural holes may resort to information fabrication; the inability of a structurally constrained node to perform its duties may impact on the speed at which information becomes accessible.

It is therefore clear that there is a rational way to associate each of these eight SNA metrics to particular aspects of the CIA triad. To illustrate how these metrics can be used to identify potential information security risks, an illustrative example using a simple, simulated network will now be presented.

4.2.2. Illustrative Example Using a Simulated Network

In order to demonstrate how the metrics discussed in the previous section can be applied to a network, a simple example using a simulated network is presented in this section. The network, shown graphically in Figure 4.1, represents the organogram of a fictitious Simulated Networks Company (SNC), which has three directors, two project managers, and six project team members. SNC has a company policy that determines this structure: each project team has three members, who report exclusively to a project manager. The two project managers report to the directors, but are also expected to cooperate with one another. As a result of this setup, the company effectively has three cliques, with each project manager being a member of both one project clique and the management clique. In order to evaluate this network using SNA, the measures shown in Table 4.2 were generated using ORA-Lite. In this table BC is the betweenness centrality, CC is the closeness centrality, EcC is the eccentricity centrality, EiC is the eigenvector centrality, SHC is the structural holes constraint, TDC is the total degree centrality, and BS is the boundary spanner.

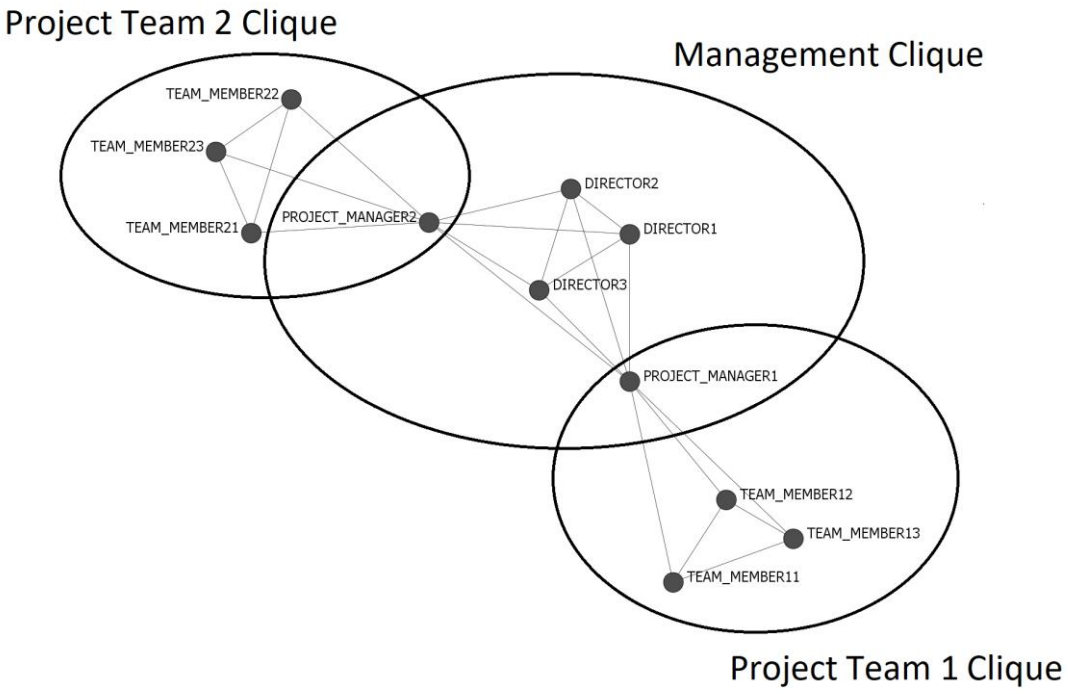


FIGURE 4.1: NETWORK REPRESENTING SIMULATED SNC EMPLOYEES

TABLE 4.2: SNA METRICS FOR SNC NETWORK

Node Title	BC	CC	EcC	EiC	SHC	TDC	BS
DIRECTOR 1	0	0.556	2	**0.567**	0.663	0.4	0
DIRECTOR 2	0	0.556	2	**0.567**	0.663	0.4	0
DIRECTOR 3	0	0.556	2	**0.567**	0.663	0.4	0
PROJECT MANAGER 1	**0.522**	**0.769**	2	0.526	0.446	**0.7**	**1**
PROJECT MANAGER 2	**0.522**	**0.769**	2	0.526	0.446	**0.7**	**1**
TEAM MEMBER 11	0	0.455	**3**	0.284	**0.819**	0.3	0
TEAM MEMBER 12	0	0.455	**3**	0.284	**0.819**	0.3	0
TEAM MEMBER 13	0	0.455	**3**	0.284	**0.819**	0.3	0
TEAM MEMBER 21	0	0.455	**3**	0.284	**0.819**	0.3	0
TEAM MEMBER 22	0	0.455	**3**	0.284	**0.819**	0.3	0
TEAM MEMBER 23	0	0.455	**3**	0.284	**0.819**	0.3	0

Based on the metrics shown in Table 4.2, a number of things can be inferred about SNC. Firstly, only the project managers have a **betweenness centrality** value greater than 0, and are also the only nodes that are **boundary spanners**. This means that the project managers effectively function as exclusive information brokers between management and the project teams. The project managers also have the highest **closeness centrality** measures, meaning they have access to the largest amount of information. Not only does this make it possible for these managers to manipulate information moving in either direction, but it may also give them access to information they may not necessarily be entitled to. The **total degree centrality** values of these two managers also support the notion that they can strongly influence the information available in the network. Because they are boundary spanners, they are also availability risks. In this network, the loss of a project manager would have a significant impact on the flow of information between a project team and management, which may necessitate appointing a new "liaison". Both project managers therefore pose risks to **confidentiality**, **integrity**, and **availability**.

With regard to leadership, inferred from the **eigenvector centrality** measure, it is clear that the directors are the ultimate leaders in this network, followed closely by the project managers. While this should not pose a negative information security risk, it does mean that the directorate is ultimately the greatest role-player with regard to information security management. There are indications of possible organisational risks, however, as it is clear that there are no institutional leaders amongst the project team members. This suggests that SNC has a rigid corporate structure, which may be a risk for a company with only 11 employees.

The **structural holes** constraint of the network is at its greatest amongst the project team members. As with the availability issues mentioned earlier, there is a risk that either of the teams may become disconnected from the network if a manager is suddenly unavailable.

The SHC also shows that each project team member is already at high risk of being constrained due to this bottleneck. While it may not always have an effect, the moment any form of collaboration between the two teams is desired, the structural holes will severely limit the amount of information flow that is available. This is supported by the **eccentricity centrality** of the team member nodes, which show that these members are the furthest removed from the network. The project team members are therefore at risk of suffering from lack of **integrity**, due to the unavailability of up-to-date, relevant information, and a lack of crucial **availability** of information from the rest of the network.

In summary, the nodes that pose the greatest risk in this simple network are the project managers, followed by the project team members. It should be noted that this is an over-simplified simulation network that was created specifically to demonstrate how SNA metrics can be used to evaluate the risk of individuals in a network. Most networks will likely not be as simple as this simulated network, as more advanced networks can also include resources such as office equipment and shared utilities. Depending on the bordering used to create the network, it could also include some, or all, of the tasks that the members have to perform.

4.3. TABULATED SUMMARY OF LITERATURE SOURCES

The purpose of this section is to tabulate the summaries of a number of studies that are relevant to SNA and information security. The tables are used in an attempt to provide the information in as concise a manner as possible. The relevance of each study can also be presented much easier in such a tabulated format. Three tables are provided, with each table containing summaries of a different topic. The first table deals with relevant studies involving information security. The second table, in contrast, deals with relevant SNA studies. The third table provides summaries of a selection of SNA studies that deal specifically with either information security or risk management.

4.3.1. INFORMATION SECURITY STUDIES

Table 4.3 contains summaries of relevant information security studies. The table provides the title of each study, a brief summary, and an indication of why the study is relevant. Seven selected studies are presented in a highly summarised form, with each of the studies being evaluated for their mode of relevance to this thesis.

TABLE 4.3: SUMMARY OF INFORMATION SECURITY STUDIES

Title of Associated Publication	Citation	Brief Summary of Contents	Mode of Relevance
Toward A Unified Model of Information Security Policy Compliance	(Moody *et al.*, 2018)	Development of a unified model for information security policy compliance, based on 11 well-known models. The new model implements a variety of factors, such as rewards, punishments, social factors, environment and habits.	Human aspect, specifically behavioural
Taxonomy of mobile users' security awareness	(Bitton *et al.*, 2018)	A taxonomy that can be used to develop awareness models for different modes of attack is introduced, specifically for mobile attacks, as a significant amount of research has already gone to developing awareness models for more traditional uses.	Human aspect; Security awareness
Confidentiality and Privacy for Smartphone Applications in Child and Adolescent Psychiatry: Unmet Needs and Practical Solutions	(Wu *et al.*, 2017)	The impact of confidentiality and privacy of smartphone apps in order to determine if they are appropriate for healthcare use is evaluated. Special focus is given to topics such as information sharing, trust, and risk-benefit analyses.	Possible impact of a lack of confidentiality to usability of certain systems
Practice-based discourse analysis of information security policies	(Karlsson *et al.*, 2017)	Eight quality criteria are proposed to improve the practical comprehensibility of information security policies.	Human aspect; demonstrates how a lack of information security culture can help undermine risk controls
Information Security Risk Assessment: A Method Comparison	(Wangen, 2017)	Three methods that can be used to assess information security risk are described and compared. The methods were evaluated by giving criteria to novices, who then implemented them and provided feedback.	This paper provides additional background for information security risk assessment
An analysis of multiple factors relating to teachers' problematic information security behaviour	(Chou & Chou, 2016)	An evaluation of how an overall information security culture impacts on the adoption of information security principles among teachers. This is important, as pupils often look to teacher as role models.	Human aspect; impact of culture on behaviour; Protection Motivation Theory
A comparative study on information security risk analysis practices	(Shukla & Kumar, 2012)	Evaluates and compares various risk evaluation methodologies, and provides a framework for how to compare them.	Information security risk assessment

This list is not presented as an exhaustive list of information security studies, but merely as an example of literature sources that are relevant. In addition to this list, two further lists of literature source are presented. The first list contains examples of SNA literature that is relevant to this study, whereas the second list focusses on extant literature where SNA was applied to improve information security.

4.3.2. Social Network Analysis Studies

Table 4.4 contains summaries of relevant SNA studies. The table provides the title of each study, a brief summary, and an indication of why the study is relevant. A total of seven selected studies are presented. There are a vast number of studies in the literature that deal with SNA, so these seven were selected as examples of relevant sources.

TABLE 4.4: SUMMARY OF SNA STUDIES

Title of Associated Publication	Citation	Brief Summary of Contents	Mode of Relevance
The transition towards a bio-based economy: A comparative study based on social network analysis	(Imbert *et al.*, 2019)	The development of the bioplastics niche in Germany and Italy is investigated. SNA is used to compare and evaluate the industries in both countries.	Real-world use of SNA; comparing of separate networks; impact of powerful actors
Defending one's friends, not one's enemies: A social network analysis of children's defending, friendship, and dislike relationships using XPNet	(Oldenburg *et al.*, 2018)	The behavioural patterns of elementary school pupils, specifically focussed on bullying, is investigated. The ability of SNA to be used in predicting behaviour is demonstrated, specifically as it relates to the effect personal relationships have on actions such as bullying and defending	Demonstrates SNA inference's real-world validity; predicting individual behaviour
Social network analysis: Characteristics of online social networks after a disaster	(Kim & Hastak, 2018)	Introduces a method that applies SNA to social media in order to propose improved disaster mitigation plans. The paper highlights the unique characteristics of social media interaction during disasters.	Risk management (mitigation); information flow evaluation using SNA
Using internal link and social network analysis to support searches in Wikipedia: A model and its evaluation	(Wu & Wu, 2011)	The use of SNA to improve Wikipedia searches by calculating the strength of connections between articles is investigated. Topics in strongly connected articles are presented together, aiding unstructured, unfamiliar decision-making tasks.	Application of SNA; calculation of link strengths
Visualization of the Chinese academic web based on social network analysis	(Bo *et al.*, 2010)	The use of SNA to investigate the level of collaboration between universities using their websites is presented. Universities with greater prestige were successfully identified using a described method.	Application of SNA to large networks; data verification methods
Social Network Analysis in Human Resource Development: A New Methodology	(Hatala, 2016)	The use of SNA to address human resource development in the literature is investigated. A number of techniques are introduced and illustrated, and the limitations of SNA are discussed	Methods, uses, and limitations
Social Network	(Scott,	The progress and use of SNA is investigated.	Methods, principles,

Analysis	2016)	A number of models, methods, and principles are described. Certain shortcomings are also discussed.	and shortcomings

As with the list of information security studies, this list is not presented as an exhaustive list, but as a list containing examples of relevant SNA literature. This list also clearly demonstrates the wide variety of fields wherein SNA is relevant.

4.3.3. Studies Featuring Both SNA and Information Security

In Table 4.5 a number of studies that apply SNA to information security are summarised. The table provides the title of each study, a brief summary, and an indication of why the study is relevant. The literature in this regard is very limited, which is why the different work of certain authors is presented more than once. The table also includes summaries of some of the studies that were discussed in greater detail in Section 4.1.

Table 4.5 contains the summaries of six relevant studies where SNA is used either in the context of risk management, or information security specifically.

TABLE 4.5: SUMMARY OF SNA STUDIES THAT FEATURE INFORMATION SECURITY

Title of Associated Publication	Citation	Brief Summary of Contents	Mode of Relevance
Leveraging Social Network Analysis and Cyber Forensics Approaches to Study Cyber Propaganda Campaigns	(Al-Khateeb *et al.*, 2019)	A method employing cyber-forensics and SNA is proposed that can be used to identify primary disseminators of malicious propaganda on social media.	SNA used to identify individuals; dissemination of information through a social network
Investigation into the formation of information security influence: Network analysis of an emerging organisation	(Dang-Pham *et al.*, 2017c)	Exponential random graph modelling (ERGM) is used to predict information security influence in an organisation. Various properties of individuals with influence are identified, and strategies that can be used to influence information security behaviour are proposed.	SNA applied to security; use of ERGM in prediction algorithm
Applications of social network analysis in behavioural information security research: concepts and empirical analysis	(Dang-Pham *et al.*, 2017a)	The application of SNA to information security awareness is discussed. The study also proposes strategies that can be used to conduct behavioural SNA research.	Real-world demonstration of SNA applied to information security; use of specific measures in inference
Applying network	(Dang-Pham *et al.*,	The impact of interpersonal relationships on information security is investigated. The	Use of SNA to identify crucial

analysis to investigate interpersonal influence of information security behaviours in the workplace	2017b)	study demonstrates the creation of hierarchical security influence networks in corporate environments.	individuals; SNA applied to security
Applying social network analysis to security	(Philips *et al.*, 2015)	The information that an attacker can obtain from a social network is investigated; specific attention is given to social hierarchies.	Inference using SNA metrics; prediction of information available in network
A Course Applying Network Analysis to Organizational Risk in Information Security	(Armstrong *et al.*, 2010)	The inclusion of SNA in a postgraduate course focusing on information security and risk is discussed. The effectiveness of the course is also discussed.	SNA applied to security; using SNA metrics to infer risks

Similar to Table 4.3 and Table 4.4, this table does not provide a complete list of studies where SNA is applied to information security. This list does however clearly demonstrate that it is feasible to use SNA in this context.

4.4. SUMMARY OF PART II

The purpose of Part II is to provide a working background of information security, risk analysis, network theory, SNA, and SNA in the context of information security. This was achieved in three chapters.

Chapter 2 introduced the CIA triad as a basis for describing information security. The chapter also provided information about information security risk management, and special attention was given to four different risk analysis techniques (CORAS, CIRA, ISRAM, and ISRA). The principles behind risk control were also introduced. The second part of the chapter focussed on the human aspects of information security, specifically information security culture, knowledge and behaviour.

Chapter 3 focussed on introducing the various topics that are important to networks in general, and SNA in particular. The chapter started by describing the principles of SNA, and clearly showed why graph theory is so crucial to SNA. The chapter's focus then shifted to graph theory and provided a clear, non-mathematical explanation of the most important topics as they are relevant to SNA. Following this, the methods that can be used to visualise social networks were shown and special mention was made of SOMs. The chapter then discussed some of the extant SNA metrics in more mathematical detail, and finally discussed

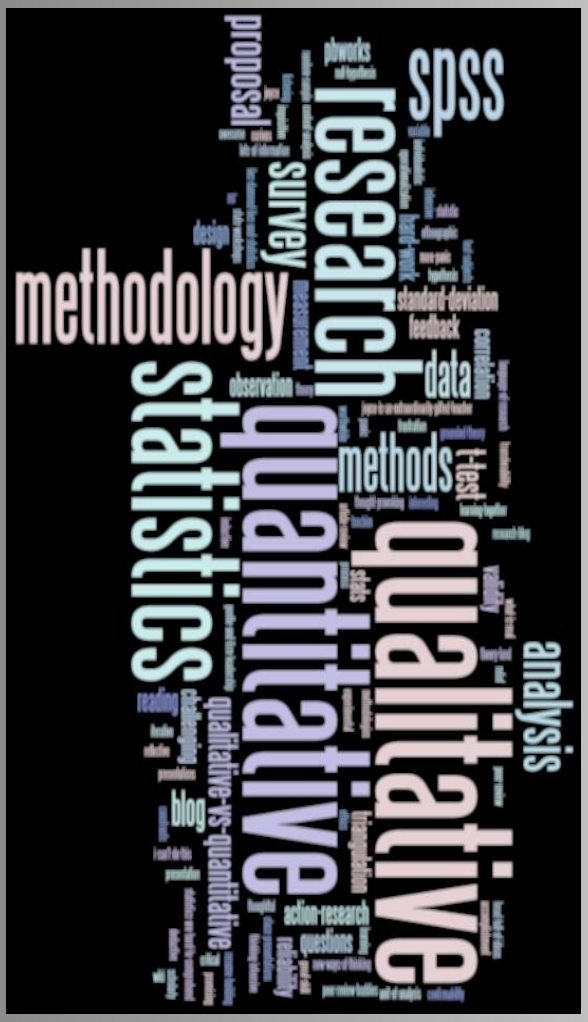
some of the advantages and shortcomings of SNA. The chapter then concluded with a discussion of various network optimisation techniques.

In **Chapter 4** the ways in which SNA can be used in information security were investigated. The chapter started with a brief review of select literature sources, followed by a discussion of how SNA metrics can be used to evaluate information security risk. A simple, illustrative example was then presented to demonstrate the discussed principles. The chapter concluded with three tables summarising various literature sources.

4.5. CHAPTER SUMMARY

In this chapter, some of the various studies dealing with SNA in the context of risk management were discussed. The methodology whereby SNA metrics are used to manage risk was also introduced. These studies showed that SNA is both valid and useable within the context of risk management. In addition, they presented a clear link between identifiable risks and SNA metrics, which in turn allowed for the identification of the relationships between these metrics and CIA risks. These relationships are used in later chapters to quantify the risks that may be present in social networks.

Chapter 4 concludes Part II of the thesis. Part III, which contains Chapter 5, deals with the research method followed during the course of the study.



PART III

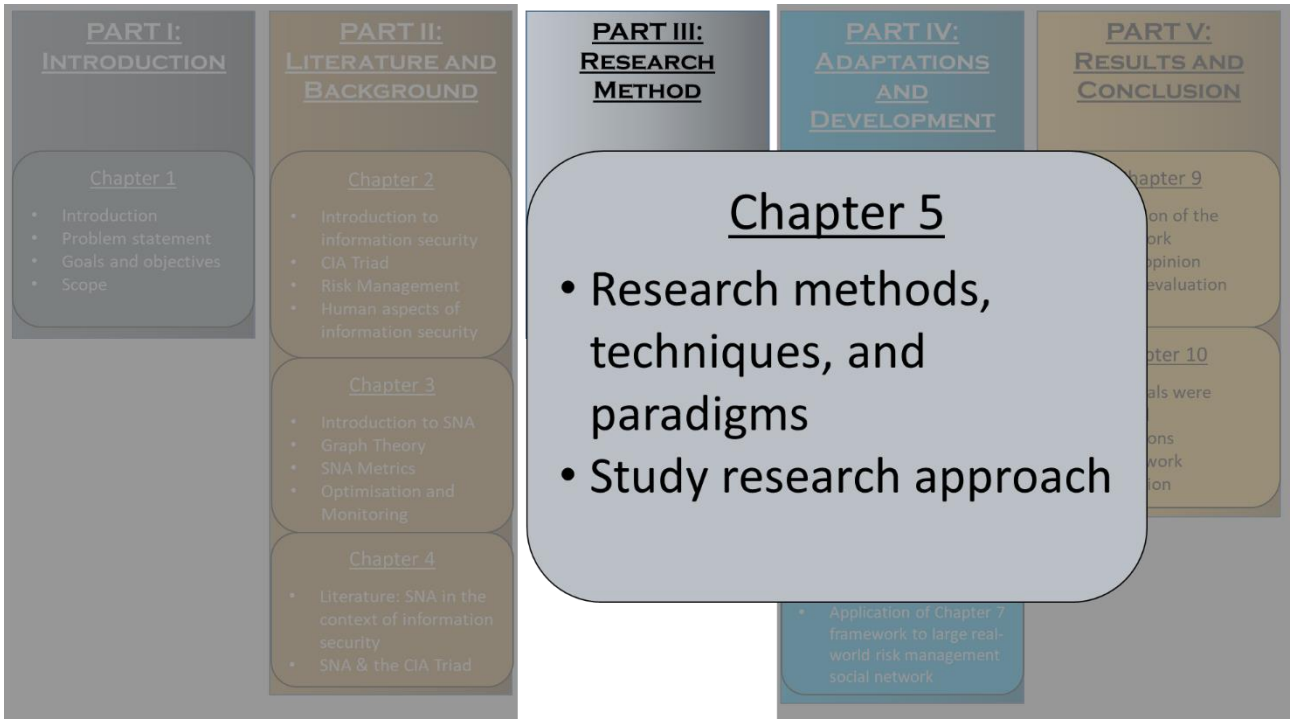
Research Method

"Where actual evidence had been a bit sparse he had, in the best traditions of the keen ethnic historian, inferred from revealed self-evident wisdom[†]

[†] *Made it up and extrapolated from other sources*[‡]

[‡] *had read a lot of stuff that other people had made up, too."*

-Terry Pratchett



CHAPTER 5: RESEARCH METHOD

CHAPTER HIGHLIGHTS:

- What is the Research Onion model?
- What are research philosophies, approaches, methodologies, strategies, time horizons, and techniques?
- What is the approach followed in this study?
- How is the study structured?

5 RESEARCH METHOD

Every research study should follow a structured approach based on a sound scientific research method, so as to ensure that the results obtained are usable and reliable. In this chapter the research approach followed in this study is discussed. The discussion starts by introducing a number of concepts, such as research philosophies, -strategies, and methodologies. In order to aid this discussion, and provide sufficient background on the relevant topics, the research onion model (Saunders *et al.*, 2019) is used. Following this introductory discussion, the focus of the chapter shifts to identifying a research approach that is appropriate to this study. Finally, the research approach for this study is discussed, and the chapter concludes with a description of the research process followed in this study.

5.1. RESEARCH ONION MODEL

The research onion model is a well-known and established model that describes research approaches in a logical and structured way. This model is used in this chapter to help describe the research approach for a number of reasons. The first is that it presents the process whereby a research approach is developed through progressive steps, starting with the broad conceptual, and then becoming more and more specific. This aids both in describing the various topics in a logical way, and helps to select the most appropriate research approach. The second reason is that it is a very complete model, in that it describes most, if not all, of the various aspects of a research method and the relationship between these aspects. The model describes the process whereby a research strategy is developed in six layers, where each layer contains the various ways in which each of the research method's aspects can be approached. A graphical representation of the model is presented in Figure 5.1.

The six layers contained in the research onion model are, from the outside in:

- Philosophies, which describe the broad approach to data and inference;
- Approach to theory development, wherein the way in which inferences are drawn from the data is discussed;
- Methodological choice, where the types of data collection techniques selected are broadly identified;
- Research strategies, which deal with the type of data collected;
- Time horizon, wherein the impact of time on the study is identified; and

- Techniques and procedures, where the specific data collection and –analysis techniques are identified and selected.

The approach built into the onion peeling method, whereby the selection of the research method for a study proceeds from broad conceptual to specific, greatly simplifies the task of selecting a research method.

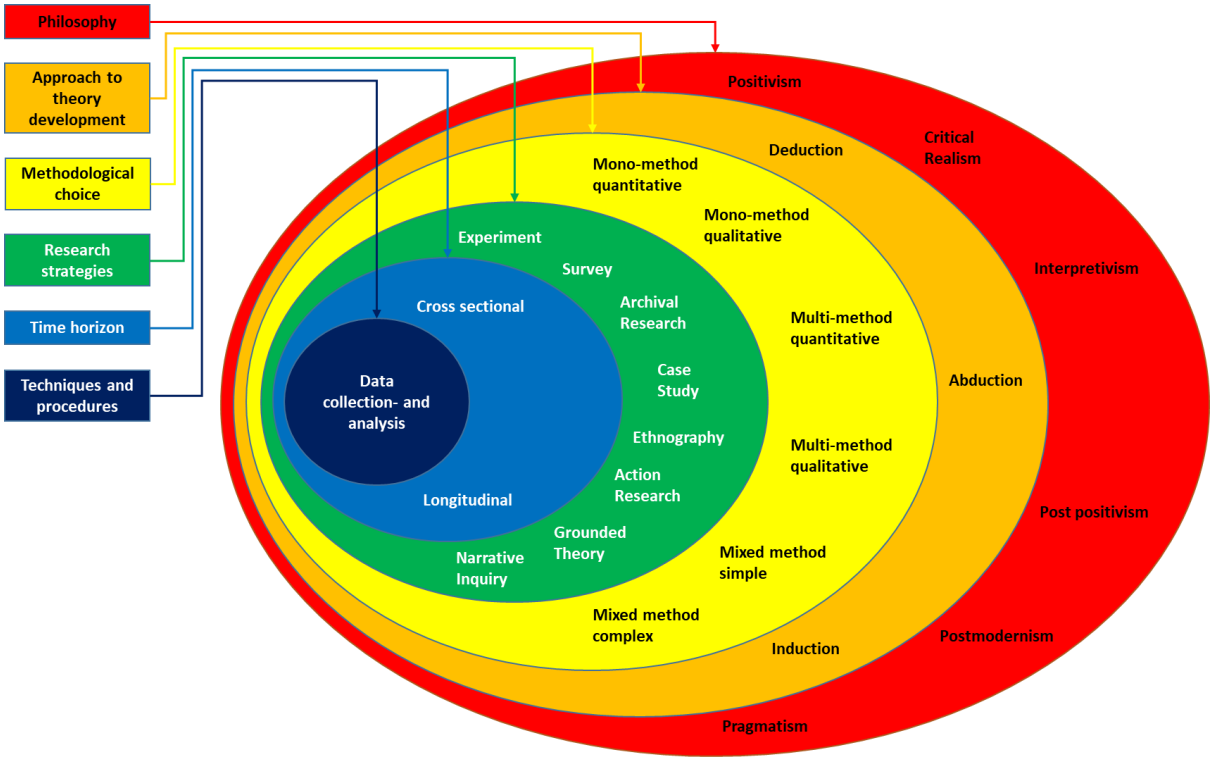


FIGURE 5.1: THE RESEARCH ONION [ADAPTED FROM SAUNDERS *ET AL.* (2019)]

The onion, when read from the outside inwards, presents increasing levels of detail that describe a research method. Each of these layers will now be discussed in order, from the outside inwards. The discussion is based primarily on the work done by Saunders *et al.* (2019), Creswell *et al.* (2012), and Oates (2005), with additional sources used for clarification.

5.1.1. PHILOSOPHIES

In the version of the research onion discussed here, six different research philosophies are included. These philosophies, named positivism, critical realism, interpretivism, post positivism, postmodernism, and pragmatism, represent most of the research philosophies

currently employed in active research. While other research philosophies exist, these six will be briefly discussed to provide a comprehensive overview of research philosophies in general.

Positivism is essentially a research philosophy that maintains that scientific study is based on observable and repeatable facts, and that the universe is not dependent upon the existence of an observer. Subsequently, the impact of the observer on the object of study should either be fully accounted for or eliminated completely. This philosophy is founded on an empirical observation of facts, i.e. that which can be observed with the senses (Angers, 2013), and generally works towards the development of natural laws. This trend towards the development of natural laws comes from the need for scientific facts to be both 'trans-situational' and 'trans-temporal', i.e. that they are valid everywhere in the universe, rather than just the "here and now" (Prendergast, 1979). Because of the structured nature of positivism, it is generally regarded as the paradigm most associated with empirical scientific study.

Critical realism is similar to positivism in that it maintains that natural laws are not dependent upon an observer. Positivism is however limited in that it does not allow for an observed fact to differ from reality. Critical realism, by contrast, allows for a perceived reality to differ from the underlying truth. To demonstrate how this is relevant to real-world situations, consider vision. When a human being looks at an object, they are in fact using the light reflected from the object to perceive it. The perceived object will likely be bathed in all kinds of light, including infrared and ultraviolet light. As the human eye cannot process ultraviolet or infrared light, it cannot be used to observe the full optical state of the object. The object, as it is perceived, therefore differs from the object as it is. Critical realism allows for this apparent inconsistency between observation and truth, and provides for the opportunity to investigate how a single truth can be perceived in multiple ways.

Interpretivism stands in stark contrast to positivism and critical realism, as interpretive research is used to identify and explore the interrelated factors within a research domain that cannot be described using empirical approaches. This type of research is therefore often associated with fields such as sociology, anthropology, and psychology, as these fields focus on exploring the subjective and often ambiguous facts surrounding human actions and understanding. It also differs from positivism in that interpretivistic research is more ideographic in nature, i.e. it is often more concerned with the aspects of a single entity than of the possibly non-existent aspects of the group within which the entity is found (Creswell *et al.*, 2012). This type of research also tends to focus on more subjective subjects, such as human perception. In this, interpretivism can be used to describe any number of subjective "realities" and the results of interpretive research are therefore often not universally applicable, or only applicable in some select cases. It also generally produces more than one valid explanation for the results of the study, and the outcome of a particular study therefore tends to itself be subjective. As this type of research also tends to be more

subjective in nature, and aims to establish meaning, it is primarily associated with qualitative data. Because of this, there are two main research strategies that are strongly associated with interpretivistic research, namely the evaluation of case studies and ethnography (Oates, 2006).

Post positivism, like interpretivism, is concerned with establishing meaning, rather than just pure fact (Henderson, 2011). However, unlike interpretivism, which is mainly qualitative, post positivism makes use of both qualitative and quantitative methods, and follows a predominantly positivist approach to scientific study. This, however, does not mean that it is merely the process of positivism with an interpretive goal, as it draws on the principles of other philosophies, such as pragmatism and modernism, as well. The implication of this is that a post positivistic philosophy does not restrict data collection or analysis in any way, apart from requiring that the methods that are used are correct and relevant for the data. Additionally, whereas positivism considers the observer to be completely irrelevant to proper study, post positivism recognises the value and impact the observer's motivations and experiences may have on the research being conducted. Finally, as the aim of this type of research is to find meaning alongside fact, it can be applied to determine how theory and practice relate to one another.

Postmodernism is a somewhat revolutionary philosophy, as it maintains a number of suppositions that are inconsistent with other philosophies. The first of these suppositions is that there is no absolute reality, but that realities are created through the experiences that individuals have. The second is that knowledge is itself fluid, changeable, and provisional. The third supposition is that the study of a plural society requires the use of interpretive and critical methods. The fourth, and final, supposition is that there is no set of values that is by definition better than any other set of values (Kroeze, 2012). The implication of this is that postmodernism lends itself more to the study of signs and simulations, while the study of a universal reality is left to positivist and post positivist styled research. One of the characteristics of the postmodernist philosophy is the wide range of research methods that can be used, and methods from positivist, interpretivist, critical reality, and even the humanities are used in postmodernist research. This does not mean that postmodernism is a philosophical approach wherein everything goes and all answers are correct: even if multiple viewpoints, or realities, are addressed in the same study, they must be evaluated using the same theory, method, and strategy, which in turns allows for scientific findings to be corroborated. The postmodernist approach, much like the post positivist one, therefore allows for almost any correct method to be used when conducting research.

Pragmatism differs substantially from all of the aforementioned philosophies in that its focus is not on how research is conducted, but on the value the research has in practice. A pragmatist is likely to disregard any rules that limit specific techniques to certain approaches, and will instead use any appropriate techniques that will eventually produce something that is of value. The core focus of pragmatism is found in this sense of value:

where positivists conduct research in order to describe the universe and discover its inner workings, and interpretivists study phenomena in order to ascertain the meaning behind those phenomena, pragmatists conduct research in order to solve real-world problems. The pragmatic approach, in that sense, is not concerned with the whys and wherefores of theory and meaning, but only on how it can be applied to bring about real-world change.

The selection of an appropriate research philosophy is a relatively important choice, as it informs the overall structure of the study. While no research philosophy can be claimed as being more correct than any other, certain types of research lend themselves more naturally to certain philosophies. It is therefore important to select the philosophy that is most suited to the research being conducted, even if that philosophy introduces certain limitations that makes conducting the study more tedious.

5.1.2. Approaches to Theory Development

The second layer of the onion describes how the underlying theory of a study is developed. The three approaches, namely deduction, abduction and induction, are all founded in formal logic and describe the relationship between premises and conclusions (Russell & Norvig, 2016).

When a **deductive** approach is followed in order to develop a theory, the conclusion is deconstructed into premises, of which all are developed in such a way that the conclusion will necessarily be true if all of the premises are true. The deductive approach is therefore well suited to experimental studies, as the truth of the conclusion can be determined by testing each of the premises. If the premises are rather complex, they too can be deconstructed into simpler premises to test. This allows for more general premises to be tested, so that a specific conclusion can be reached that is known to be true. The deductive approach, which is in common use in the natural sciences, also makes the process of falsification, whereby a conclusion can be proven to be false, relatively simple. As the conclusion requires all of the premises to be true, falsifying the conclusion may only require that one of the supporting premises is shown to be false.

The **abductive** approach, in contrast, does not seek to prove a conclusion, but draws a conclusion from a mostly incomplete set of premises that are known to be true. The premises are not dependent upon the conclusion and, as it is always possible that future evidence will serve as premises that disprove the conclusion, the conclusion itself cannot be proven to be absolutely true under any circumstances. Abduction aims to describe specific premises in general. The abductive approach aims to move towards an inductive result by repeatedly moving from conclusion to premise, and then back to conclusion. In this way, the

limited amount of data available can be increased by evaluating how the conclusion corresponds to reality.

The third approach, **induction**, is essentially similar to abduction, with the primary distinction being that inductive reasoning requires a mostly complete set of true premises. This differs from the abductive approach, which can be used if only an incomplete set of true premises are available, and can produce incomplete conclusions. Like abduction, induction develops a conclusion that aims to describe the general relationship between specific premises, and then makes the supposition that the general conclusion developed is universally valid, based on the "most likely" or "best available theory" arguments.

Distinguishing between induction and abduction can at times be difficult, as it requires a certain amount of information with regard to the size of the full set of premises. Deduction, by contrast, requires a full set of premises, and may therefore be difficult to apply to situations that do not involve controlled laboratory situations. The selection of a theory development approach will therefore depend on the information available, the environment wherein the research is conducted, and the expected nature of the outcome.

5.1.3. Methodological Choice

In the first layer of the research onion, an appropriate research philosophy is selected, followed in the second layer by the selection of a theory development approach. In the third layer, which is the final conceptual layer, the general types of data collection and analysis methods that will be used are selected. There are mainly three categories in this layer, namely quantitative, qualitative, and mixed method.

Quantitative methods, as a category, includes all methods that use quantitative data and statistical analysis. In essence, a quantitative method uses numbers and mathematically provable relationships exclusively. This differs fundamentally from **qualitative** methods, where data that is not necessarily quantitative or tangible is collected and used. This demonstrates the essential difference between qualitative and quantitative methods, in that quantitative methods produce data that is meaningful when evaluated mathematically, and qualitative methods produce data that is meaningful when evaluated humanistically. As a result, both of these types of methods are associated with certain philosophies: positivism makes us of quantitative methods almost exclusively, whereas interpretivism mainly uses qualitative data. The **mixed method** methodology is used when both qualitative and quantitative data is required. This may involve combinations such as the use of quantitative methods to evaluate qualitative data, or the collection of quantitative data to determine patterns, and qualitative data to ascertain the meaning behind the patterns.

As shown in Figure 5.1, each of these methodologies have two variations. For quantitative and qualitative, the methodology can involve either a **mono-method** approach, whereby only one data collection method is used, or **multi-method**, where more than one data collection method is used. A multi-method approach can employ both qualitative and quantitative methods, but the data from those methods are not combined and are processed and evaluated separately. This differs from the **mixed method** methodology, where qualitative and quantitative data are combined into one dataset before being processed (Flick, 2015; Feilzer, 2010). The mixed method approach can either be **simple**, where only one point of integration is used to combine the qualitative and quantitative data into the single data set, or **complex**, where multiple points of integration are used.

The selection of a methodology informs the research strategies that are used to collect and analyse data. In the same way that qualitative and quantitative methodologies are associated with different philosophies, so too are research strategies associated with particular methodologies. Choosing the correct methodology for a particular study is therefore just as important as choosing the correct philosophy, as choosing the wrong type of methodology may result in collecting the wrong type of data, which may in turn render months, or years, of research invalid.

5.1.4. Research Strategies

Once the first three layers of the onion have been peeled, i.e. the philosophy has been identified, the theory approach has been chosen, and the methodology has been chosen, the research strategy – or strategies, as is the case with multi-method and mixed method methodologies – has to be selected. In Figure 5.1 eight different strategies are named. Each of these strategies will now be described briefly.

With an **experimental** strategy, a research process is created whereby a hypothesis is developed, the factors needed for the hypothesis to be true are identified, and the factors are then tested individually in order to determine if the hypothesis is true or false. Experimental strategies tend to follow deductive approaches, in that a hypothesis is considered to be true if and only if all of the causal factors are correctly identified and investigated.

Surveys are usually associated with quantitative methodologies, and involve investigating a population by collecting a subset of data from that population. While surveys are often associated with questionnaires, other techniques such as random sampling and observations are also valid. As it involves the use of selected data in order to develop generalisations about a larger group, the strategy is necessarily either inductive or abductive.

Archival research, also known as document research, is the study of existing documents in order to obtain new knowledge or information. One of the most common forms of this type of research is literature study, which is typically undertaken as preparation for a larger research project. However, other forms of document-based study, such as historical research and empirical reviews, are also considered archival research.

A **case study** features the in-depth investigation of a single entity, or case, in order to identify its unique features and attributes, and then, depending on the type of study, use those identified features in order to make generalisations about similar cases (Gerring, 2007). This means that, depending on the type of study, it can be used with any of the three theory development approaches discussed in Section 5.1.2. Case studies are often used in research that aims to find meaning, such as interpretivistic or post positivistic research, but a case can itself be investigated using both qualitative and qualitative methods.

Ethnography involves the descriptive study of peoples or cultures. Ethnographic research focusses on collecting data about a culture, identifying unique aspects within that culture, comparing cultures to one another, reflecting on how the researcher impacted the culture being investigated, etc. As the goal of an ethnographic study is to investigate and describe a specific culture, it can be considered a special form of case study.

Action research necessitates a researcher actively taking part in a group being studied, and then bringing about change in the group. The impact of the changes is then investigated. This type of research can potentially be employed in any research that targets people, as the various requirements for the differing philosophies can all be accounted for if the research is conducted correctly.

Grounded theory is an inductive, qualitative strategy whereby a research subject is meticulously identified and studied. As the subject is studied, various attributes and patterns may be identified that can be used to develop theories about the research subject. As the study progresses, the theory is continuously tested against the empirical data, until a point is reached where the theory is complete and descriptive enough to “make sense” within the context of the research subject.

Narrative inquiry is a form of qualitative research where narratives, i.e. stories, are used as both a research method and the phenomena being investigated. This type of study seeks to investigate the experience of individuals through stories, and to identify the meaning behind them. This strategy can involve the use of qualitative techniques from other strategies, such as archival research.

With the selection of an appropriate research strategy, most of the descriptive work for the study has been done. All that remains, is to determine the time horizon for the study, and to select the specific techniques that will be used to collect and evaluate the data. As these

decisions are strongly informed by the philosophy, approaches and strategies selected, the last two layers are much simpler to work through within a conceptual planning context.

5.1.5. TIME HORIZON

The time horizon for a study deals with how the progression of time is accounted for in a study. It is important to correctly identify the time horizon for a study, as certain research projects require that data be collected over time in order to produce useable results. Two types of time horizons are shown in Figure 5.1, namely cross sectional and longitudinal.

Cross sectional studies are studies where the data is evaluated as it is observed at a single point in time. The name cross sectional comes from the idea that a cross section of time is taken, and the only observations used are those taken at that exact moment in time (Olsen & St George, 2004).

Longitudinal studies, by contrast, involve the subject of research being investigated over time. Part of the goal of longitudinal studies is to determine how time affects the subject, or how a particular change to the subject impacts how it changes over time.

The specific nature of a study will have an impact on how appropriate a certain time horizon is, as certain kinds of studies require a particular time horizon. To determine how happiness is affected by aging, to use an arbitrary example, a longitudinal study is required by default, as happiness by its nature is too fluid to evaluate cross sectionally. The time horizon may also impact on the selection of techniques and procedures considered in the final layer, as certain techniques may not be logistically or financially feasible for use in longitudinal studies. However, if a particular type of data is absolutely required by the study as determined by the strategies selected in the previous layer, and the method whereby such data is to be collected is not feasible for a particular time horizon, it may be necessary to structure the study in such a way that the relevant time horizon is be maintained without the research result being adversely affected.

5.1.6. TECHNIQUES AND PROCEDURES

The final layer of the research onion involves selecting the techniques and procedures that will be used to collect and evaluate data in the study. The selection of the techniques and procedures is informed by the philosophy, approach, strategies, methodologies, and time horizon selected when the first five layers were peeled. As there are a vast number of data

collection and analysis techniques that can be used, a few selected examples are presented to illustrate how data collection techniques and procedures are determined and used.

Questionnaires centre on the development, distribution, and collection of a list of relevant questions that is specifically compiled and structured to collect data that is both relevant and necessary to a study. As they always involve people, questionnaires have to conform to certain ethical standards, such as confidentiality and informed consent. While it is possible to collect qualitative data using this technique, it is most often associated with the collection of quantitative data.

Simulations are used to evaluate a process, system or technique using either real-world or hypothetical data. The use of real-world data in simulations allows for a way to test a process or technique without any risk to a real-world subject. Additionally, depending on the circumstances, it is more expedient to use real-world historical data in a simulation than to collect new data. They are often associated with experimental strategies, and are typically quantitative in nature.

The study and evaluation of **documents** in order to collect data is most often associated with archival research strategies. The data collected is usually qualitative in nature, and a variety of methods, such as hermeneutics and content analysis, can be used to group, sort, and evaluate the data contained in the documents.

When a researcher needs to collect data about a subject without the subject being influenced in any way, **observations** can be used to collect the data. The data that is collected can be either quantitative or qualitative, depending on the study and the subject being observed.

In studies where the point of view or opinion of an individual is used as data, **interviews** can be used to collect such data. An interview is a two-way conversation between the researcher and a research participant, which can be conducted in one of three ways. Open-ended and semi-structured interviews are used to collect qualitative data that either investigates a phenomena, or is used to corroborate other data. A structured interview, by contrast, can be quantitative in nature and is used to collect specific, consistent, and verifiable information.

Selecting appropriate data collection and analysis techniques, in conclusion, is influenced by the philosophy followed, the research approach, the research strategy, and the methodology used. Once all of these factors have been identified, the study can be structured in such a way that the correct data can be collected, and then be used effectively.

5.2. RESEARCH APPROACH FOLLOWED IN THIS STUDY

In the previous section, the various kinds of research philosophies, strategies, methodologies, etc. are discussed. In this section, the contents of Section 5.1 are applied to this study, in order to determine the research approach and methodology that should be followed. The research approach is discussed first, starting with an application of the research onion model to the overall research question and –objectives introduced in Chapter 1. Following this, the section introduces the overall structure of the research process and how each of the various chapters fit into that structure.

The first question to answer is which philosophy should be followed. The central theme is to investigate how Social Network Analysis (SNA) can be used in the context of information security risk management and, more specifically, if it is possible to use SNA to develop risk mitigation strategies. The study is therefore not concerned with meaning in the interpretivist sense, where the more abstract aspects of certain phenomena are investigated, but rather with investigating possible new uses for existing techniques. This suggests that a **positivist** philosophy is most suitable.

The proposed theory is that SNA can be used to develop risk mitigation strategies. This theory is tested using a novel framework that is then applied to real-world data. As it is impossible to determine if SNA can be applied to develop risk mitigation strategies for all social groups in existence, the theory focuses on using a specific demonstration to predict a general outcome. It is therefore **inductive** in its approach.

The positivistic research philosophy is almost exclusively associated with quantitative data. As SNA is wholly quantitative after a network has been graphed, and quantitative data can be used to graph the networks, the only question that remains is whether a mono-method or multi-method approach is warranted. Because quantitative network data can be obtained using a variety of methods, such as questionnaires, analysis of organograms, and even document- and correspondence analysis, the most suitable description of the methodological approach is **multi-method quantitative.**

Due to the fact that the study aims to determine if risk mitigation strategies can be developed using SNA, the most appropriate data collection strategy is **experimentation**. The argument for this is that, once the proposed technique, which can be used to produce risk mitigation strategies using SNA, has been developed, its usability and validity with regard to real-world data will need to be established. This can be accomplished using either real-world implementations of the technique, or critiqued simulations of the technique using real-world data. Both of these testing methods are experimental in nature.

In order to establish a time horizon, the stated goal needs to be considered once more. As the purpose of the study is to investigate the possibility of developing risk mitigation

strategies using network data taken at a single point in time, after which the strategies can be implemented, the time horizon for the study is **cross-sectional** in nature.

The final decision deals with the actual data collection and analysis techniques that will be used. With the selection of an experimental strategy with a cross-sectional time horizon, the best technique to use is **critiqued simulation**, whereby real-world data is used to simulate an application of the technique to real-world scenarios, and the outcome of the application is critically evaluated to determine its effectiveness and usefulness.

In conclusion, this is a positivist, inductive, multi-method quantitative, experimental, cross-sectional study that makes used of critiqued simulations. With the philosophical details and research approaches established, the overall structure of the study will now be discussed.

The study is essentially divided into four phases, namely **Investigation**, **Adaptation**, **Development**, and **Testing**. This structure is presented graphically in Figure 5.2.

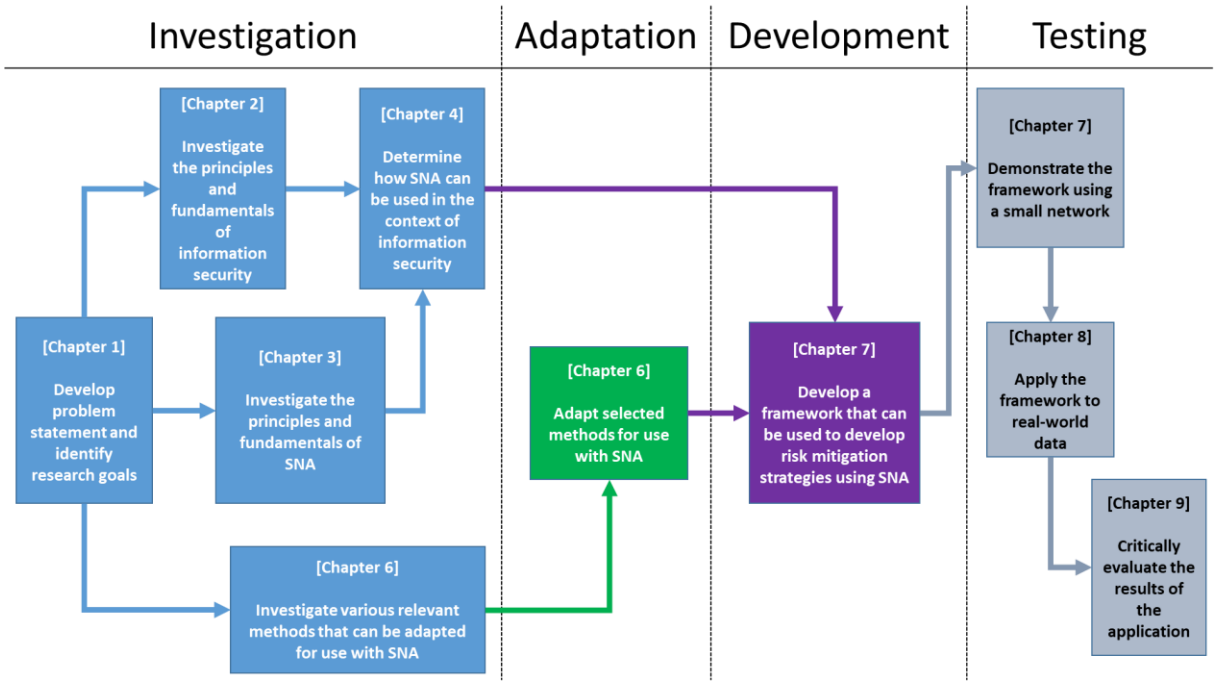


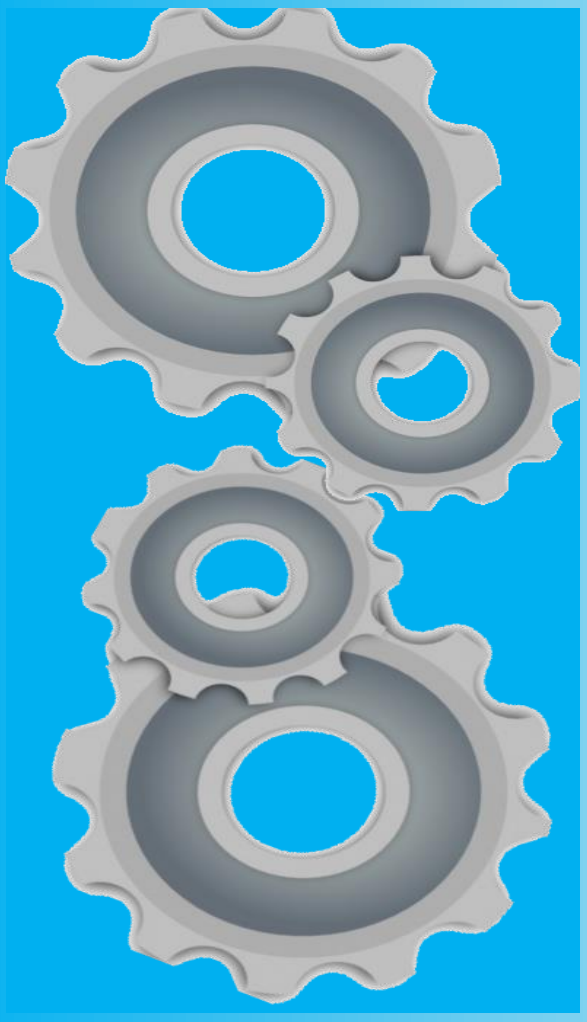
FIGURE 5.2: OVERALL STRUCTURE OF THE STUDY.

In the Investigation phase, a problem is identified and a problem statement is formulated. Following this, literature studies are conducted in order to determine the basic principles of SNA and information security. The relationship between SNA and information security in the literature is also investigated. In addition to these topics, methods that can be used to evaluate information security risk in a network, as well as network optimisation methods, are evaluated. Once these methods have been evaluated, they are adapted to fit exactly

with the requirements of this study. This adaptation is done during the second phase, i.e. the Adaptation phase. In the Development phase, a framework that can potentially be used to address the research question is developed. This framework makes use of the adapted methods obtained during the second phase. The final phase, testing, involves two separate steps: in the first step, the framework is demonstrated using a small and familiar real-world network. The application of the framework to this network's data is used to illustrate how the method works. During the second step, the framework is applied to a much larger real-world network, and the outcome of the step is evaluated critically in order to determine the framework's usefulness.

5.3. CHAPTER SUMMARY

In this chapter some of the various philosophies, theory development approaches, research methodologies, research strategies, time horizons, and data collection- and analysis techniques that can be used to conduct research were discussed. This was done so that a complete and appropriate research method could be identified for this study. The chapter concluded with a discussion of how these topics apply to this study, and a brief description of how the research approach of this study is structured. The approach presented in this chapter informs the progression and structure of the entire study, as well as the collection and evaluation of data.

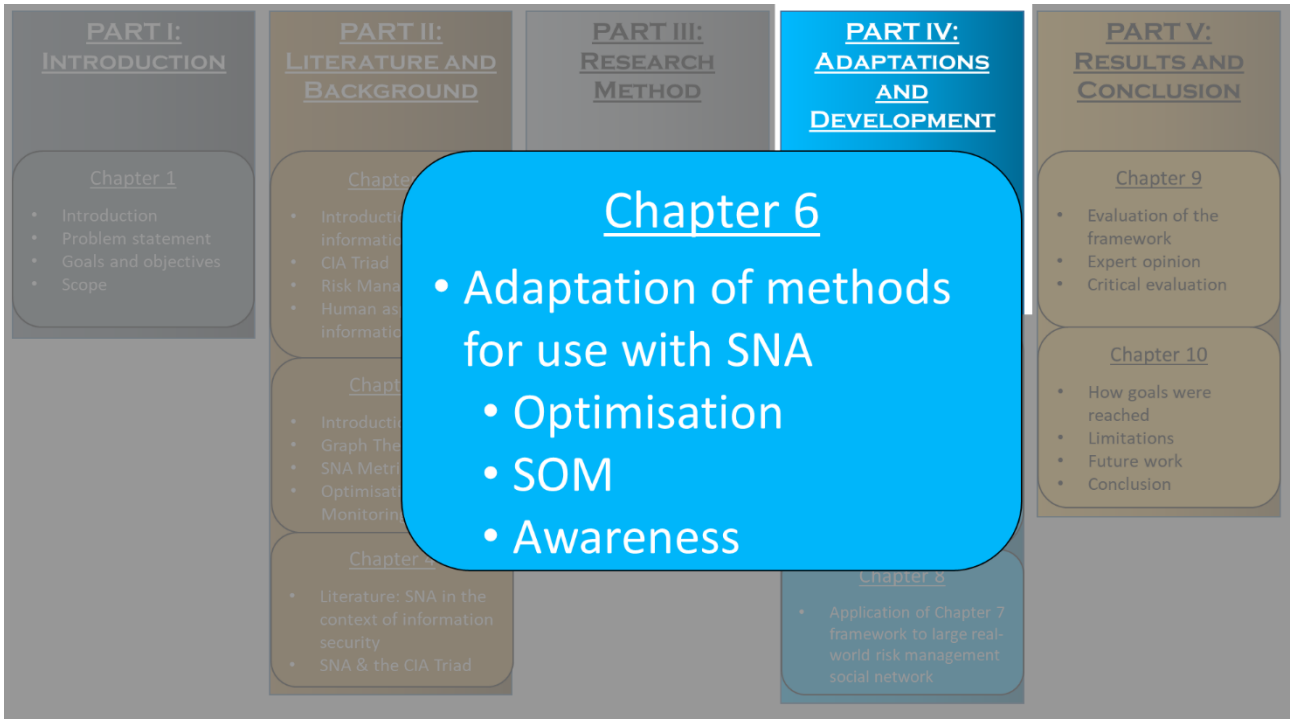


PART IV

Adaptations & Development

"The measure of intelligence is the ability to change"

- Albert Einstein



CHAPTER 6: METHODS AND ADAPTATIONS

CHAPTER HIGHLIGHTS:

- How can the optimisation techniques introduced in Chapter 3 be adapted for use with social networks?
- How can Self-Organising Maps (SOMs) be used to evaluate risk in a social network?
- How can information security awareness programmes be improved using SNA?

6 METHODS AND ADAPTATIONS

In this chapter, three SNA methods, each of which can potentially be used to improve the overall information security state of an organisation, are discussed. The first method is a network optimisation technique that specifically targets risk in social networks. The second method is an application of the Self-Organising Map (SOM) method that can be used to group various nodes based on their risk, and present the results graphically. The third, and final, method discussed uses SNA to address risk by improving information security awareness.

Each of these three methods can be implemented in different and distinct ways, and targets different aspects of information security within an organisation. The first method, which aims to optimise a social network using risk metrics, is implemented in the framework which is discussed in Chapter 7. The purpose of this method is to identify areas where information security risk is caused by the characteristics of the organisation's social network. The second method, which implements SOM, is intended to help identify similar nodes within a social network so that the risks they pose as a group can be addressed simultaneously. The method can also be implemented as a "quick-and-dirty" way of determining how much risk is present within a network.

The third method differs from the first two in that it follows a fundamentally different approach. Whereas the first two methods aim to identify risks that exist in an organisation, the third method aims to improve the information security culture of an organisation by improving its overall level of information security awareness. By combining all three of these methods it should be possible to address the overall information security risk in an organisation by means of a two-pronged approach: whereas the first two methods identify risks that can be addressed managerially, and therefore may produce top-down results, the third method may help improve the overall information security culture of the organisation, which in turn should produce bottom-up results.

6.1. OPTIMISATION OF A SOCIAL NETWORK USING RISK METRICS

In Chapter 3, two methods that can be used to optimise networks are discussed. The first method, which focusses on reducing the average path length in the network, makes use of a genetic algorithm to randomly add and selectively remove edges until the diameter of the network is optimised. The second method, which is aimed at computer networks, rearranges edges in order to minimise certain metrics; in doing so, the network is optimised

for its intended use. While these methods differ greatly in their approaches, they have two things in common. First, they both add and remove edges in order to improve the network and, secondly, they both use specific, meaningful metrics when selecting edges to add or remove.

Improving the risk in a real-world social network by optimising the network, however, is not quite as straight-forward as adding or removing relationships. Any algorithm that aims to improve the structure of a social network has to take a number of things into consideration. Firstly, not all relationships can be removed. Certain relationships, such as those that exist between the members of a board of directors, for example, are integral to the functioning of the organisation. Of those relationships that remain, not all are easy to remove: friendships, rivalries, and other similar personal relationships can survive for years depending on the situations. Secondly, destabilising a network does not improve it – any optimisation algorithm should therefore aim to, at worst, maintain the network's stability. Ideally, the overall stability of the network should be improved. Thirdly, it takes time to change the state of a relationship. Even more formal relationships, such as those that exist within a management structure, take time to change. Finally, it is important to note that some personal relationships are never formed strongly, or even at all.

The presence of any of these occurrences is highly dependent upon the type of relationships used to construct the network. In a network with formal relationships, such as those described solely by a managerial structure, certain relationships may be very easy to add and remove through restructuring or reassignment. In a more personal network, however, it may only be possible to add a relationship, while removing relationships may be impossible.

Another aspect that needs to be considered is which metric is being targeted when a relationship is being added or removed. Certain metrics, such as betweenness, are lowered by adding relationships, while other metrics, such as degree centrality, are lowered by removing relationships. It is also possible to either create or eliminate border spanners by adding or removing certain relationships. As border spanners are possible indicators, or even causes, of an unstable network, creating them should be avoided as much as possible.

Based on these observations, an optimisation method that aims to optimise a social network should follow a number of rules, namely:

1. The type of relationship used to construct the network must be taken into account.
2. The choice of whether to add or remove a relationship must be based on both a targeted metric and the type of relationships used.
3. It cannot be assumed that relationships will definitely cease to exist, or come into existence, because of managerial decisions. The optimisation process should therefore always select relationships based on the original, unaltered network.
4. The network may never be destabilised. This means that no border spanners are formed and the network diameter is decreased rather than increased.

5. The “optimised” network does not have to be perfect, only better than the original.

An optimisation algorithm, based on the computer network optimisation algorithm described in Chapter 2, and modified to take the above five rules into account, will now be described briefly. The algorithm is also presented more formally as Algorithm 6.1.

Step 1: Select the node with the highest determined risk value.

Step 2: Evaluate the node to determine if its relationships to other nodes can be added or removed. If the node’s state is fixed (i.e. no relationships can be added or removed), select the node with the next highest determined risk value. Repeat Step 2 until a node with an unfixed state is found.

Step 3: Identify the metric most responsible for the risk value of the node selected in the previous step. If there are multiple options, and one is a centrality measure, select the centrality measure; otherwise select the measure that contributes the most to the overall risk in the network. The centrality measure is selected first, as modifying this measure may have a greater overall impact on the network than the alternatives. If there are multiple centrality measures, select the one that contributes the most to the overall risk in the network. If multiple options still exist, select one arbitrarily.

Step 4: Determine if the metric can be improved by adding or removing relationships. If adding improves the metric, select the node with the lowest value for the identified metric. If multiple nodes exist that have the same low value, select the node with the lowest overall risk. If multiple nodes exist with the same metric and risk value, select the one that is closest in walk distance from the high-risk node. If there are still multiple options available, select one arbitrarily.

If removing a relationship will improve the metric, select the neighbour that has the highest value for the same identified metric. If there are multiple options, select the node with the highest risk value. If there are still multiple options, select a node arbitrarily. If removing a relationship will create a border spanner or an isolated sub-network, select the next relationship found using the same procedure. This is the only instance where previous changes are taken into consideration: if a node, were all the changes to take effect, becomes a border spanner, then another relationship should be selected.

Step 5: Select the node with the next highest risk value. Repeat steps 2 to 5 until the minimum risk threshold has been reached, or all of the nodes have been evaluated. If relationships can be assumed to come into existence, or cease to exist, almost immediately, as is the case with certain formal networks, then the metrics are recalculated for the entire network and the new risk values are used to select a node. Otherwise, if there is a chance that a relationship may not come into existence, or cease to exist, or if such a relationship may take time to develop or disappear, then the risk metrics of the original, unaltered network are always used to select the nodes.

ALGORITHM 6.1: Optimisation of Social Network using Risk Metrics

Input: A social network *N* with *n* nodes and a list *R* containing the risk values for *N*; a risk threshold value *r*
Output: A list *L* with relationships to add or remove; new overall risk value *T*

Variables:

r_i – Total risk value of node *i*
h - Currently selected node with highest risk
c - Currently selected metric
b_{ij} – Neighbour *j* of node *I*
v_{mi} – weighted value of metric *m* for node *i*
l - Currently selected node with lowest risk
d_{ij} – geodesic distance between nodes *i* and *j*
$hmax$ – number of neighbours node *h* has

1. while $(T > r)$ or nodes to be considered remain
2. select $\max(r_1, r_2, \dots, r_n)$
3. set $h = [2]$
4. select $\max(v_{1h}, \dots, v_{mh})$
5. set $c = m$ of nodes returned in [4]
6. if *c* of type Add to Reduce
7. select $\min(v_{c1}, v_{c2}, \dots, v_{cn})$
8. if [7] returns multiple options
9. select $\min(r_1, \dots, r_{[7]})$
10. if [9] returns multiple options
11. select $\min(d_{h1}, \dots, d_{h[9]})$
12. add "Add: $h \leftrightarrow [11]$" to *L*
13. else
14. add "Add: $h \leftrightarrow [9]$" to *L*
15. end
16. else
17. add "Add: $h \leftrightarrow [7]$" to *L*
18. end
19. else if c of type Remove to Reduce & network type allows for removal of relationships
20. select $\max(v_{cb_{h1}}, v_{cb_{h2}}, \dots, v_{cb_{h\,hmax}})$
21. if [20] returns multiple options
22. select $\max(r_1, \dots, r_{[15]})$
23. if removal of relationship $h \leftrightarrow [22]$ does not destabilise
24. add "Remove: $h \leftrightarrow [22]$" to *L*
25. else
26. remove [22] from set returned in [20]
27. go to [22]
28. end
29. else
30. if removal of relationship $h \leftrightarrow [20]$ does not destabilise
31. add "Remove: $h \leftrightarrow [20]$" to *L*
32. end
33. end
34. end
35. recalculate $T = \sum_1^n r_i$
36. remove *h* from future consideration
37. end
38. return *L*

In order to demonstrate how this method can be applied to a network, consider the Simulated Networks Company (SNC) network first introduced in Chapter 4. This network, shown in Figure 6.1, represents a simulated organisation with three directors, two project managers, and six project team members. Recall that, according to the process followed in Chapter 4, the two project managers were identified as the nodes representing the highest risk, as they had the highest values for the discussed risk metrics.

To adequately demonstrate how the optimisation method accounts for the presence of nodes that may not be removed, it is assumed that the SNC directorate and management structure cannot be changed in any fundamental way. This means that none of the relationships between any of the directors or the project managers may be removed. It is also assumed that the structure of the project teams may be modified. The normalised metrics values for the network, which are used to compare the metrics to one another, are given in Table 4.2. The exact normalisation method used is discussed in Chapter 7. The total risk value is calculated simply as the sum of all the normalised risk metrics. The metrics are not weighted in this example, as each of the metrics are considered to be equally important. The calculation of a weighted risk value using weighted risk metric values is also discussed in Chapter 7.

For the remainder of the chapter, the following abbreviations are used for each of the risk metrics: BC is the betweenness centrality, CC is the closeness centrality, EcC is the eccentricity centrality, EiC is the eigenvector centrality, SHC is the structural holes constraint, TDC is the total degree centrality, and a BS value greater than 0 indicates a boundary spanner.

TABLE 6.1: NORMALISED SNA METRICS FOR SNC NETWORK. TOTAL RISK IS CALCULATED AS THE SUM OF THE RISK METRICS

Node Title	BC	CC	EcC	EiC	SHC	TDC	BS	Total Risk
DIRECTOR 1	0	0.321656	0	1	0.581769	0.25	0	2.153425
DIRECTOR 2	0	0.321656	0	1	0.581769	0.25	0	2.153425
DIRECTOR 3	0	0.321656	0	1	0.581769	0.25	0	2.153425
PROJECT MANAGER 1	1	1	0	0.855124	0	1	1	4.855124
PROJECT MANAGER 2	1	1	0	0.855124	0	1	1	4.855124
TEAM MEMBER 11	0	0	1	0	1	0	0	2
TEAM MEMBER 12	0	0	1	0	1	0	0	2
TEAM MEMBER 13	0	0	1	0	1	0	0	2
TEAM MEMBER 21	0	0	1	0	1	0	0	2
TEAM MEMBER 22	0	0	1	0	1	0	0	2
TEAM MEMBER 23	0	0	1	0	1	0	0	2
OVERALL NETWORK RISK								**28.17052**

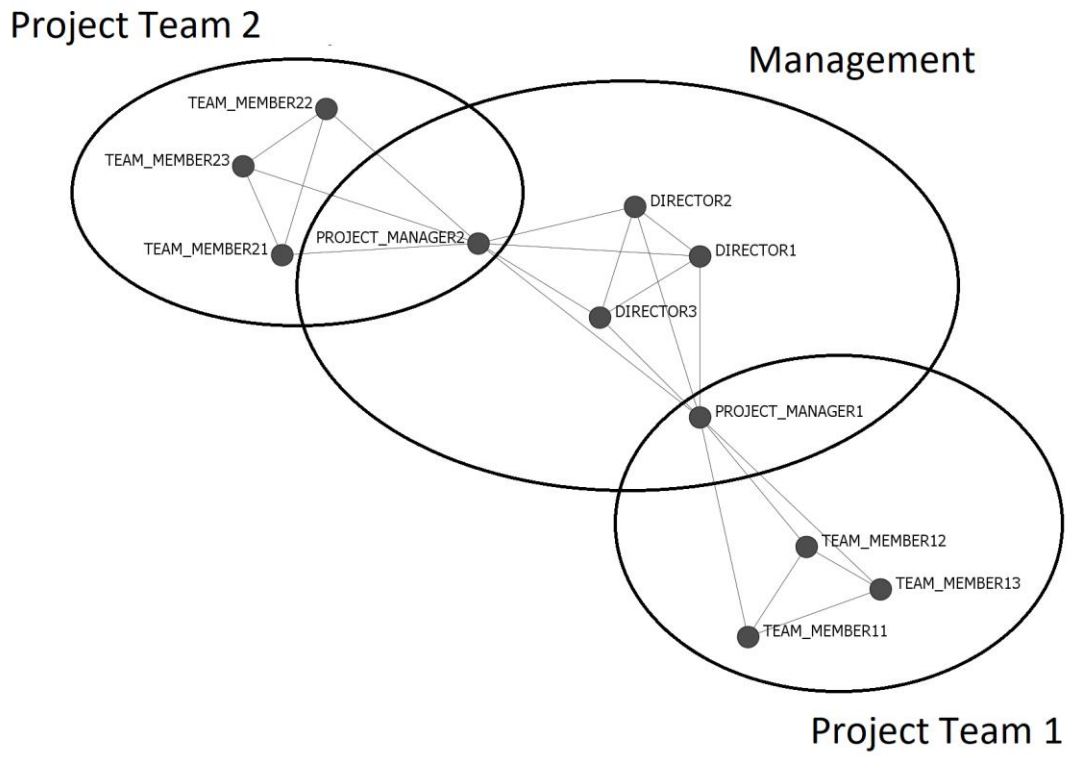


FIGURE 6.1: SNC NETWORK BEFORE OPTIMISATION

Step 1: Any of the two project managers can be selected, as both nodes have an equal amount of risk (4.855124). Project Manager 1 (PM1) is selected arbitrarily.

Step 2: As stated, the fundamental management structure cannot be changed, but the composition of the project teams may be changed. This means that, while no connections between PM1 and any other node may be removed, connections may be added between any two nodes.

Step 3: Because no connections may be removed, only metrics that can be reduced by adding connections directly to the node should be considered. The metrics that are valid in this instance are EcC and SHC. BS can unfortunately not be addressed by adding connections to the node, and merely indicates that no connections can be removed from PM1 without potentially destabilising the network. For PM1, because EcC and SHC are equal in value, the centrality measure EcC is selected.

Step 4: Due to the evaluation done in Steps 2 and 3, it is already known that a relationship needs to be added according to the EcC value. As a higher EcC indicates a higher risk, the aim should be to lower the node's EcC value. Because the normalised EcC value for PM1 equals 0, there are no nodes with a lower EcC value. This is not a problem, however, as the goal of the entire process is to reduce the overall risk in the entire network. Creating a

relationship with a node that has a higher EcC is therefore not a problem. Subsequently, a relationship between PM1 and the node with the lowest EcC from the remaining nodes should still be considered. As PM1 is already connected to all of the directors and the other project manager, the only connectable nodes are the project team members. From these, only the members of the other team are viable as, again, PM1 is already connected to all of the team members of Project Team 1. As all of the team members have the same EcC, anyone from Project Team 2 can be chosen arbitrarily; Team member 21 is selected and a new relationship is added. The normalised metric values for the network, with the new relationship added, are shown in Table 6.2. The modified network is shown in Figure 6.2.

To demonstrate the positive impact this relationship may have on the network, consider the risk value using the normalised risk metrics. Before the relationship was added, the overall risk in the network was calculated as 28.17052. After the addition of the new relationship, the value changes to 27.50091, which is a reduction in risk of 0.66961 or, more meaningfully, 2.4%. Furthermore, because the new relationship connects one of the project team members to someone other than that team's project manager, at least one of the boundary spanners is resolved as a result. This is also shown in Table 6.2, where Project Manager 2's BS value has been changed from 1 to 0. The greatest impact of this is that, if PM2 is removed, Project Team 2 will no longer be completely isolated. The overall stability of the network is improved as a consequence of the boundary spanner being resolved, along with a reduction in overall risk in the network. This also demonstrates the limitations in using a simple risk formula to calculate the overall risk in the network: depending on the relationship between the various individuals, someone with access to too much information may be a greater risk than a boundary spanner. The use of a weighted summation may therefore help to better reflect the risk in the network.

Following this first iteration, PM1's risk has increased from 4.855124 to 5. This increase, while not ideal, is acceptable as the overall risk has been decreased. It is also important to note that, as this is an iterative process, the risk some nodes pose may be increased after individual iterations. The goal remains to reduce the overall risk in the network; if individual iterations increase a node's risk, then more iterations should be considered.

Step 5: Now that a new connection has been added for PM1, the node with the next highest risk value is selected; note that the original risk metric values, i.e. those that represent the unaltered network, are always used to select nodes. This is done because there is no guarantee that the proposed relationships will come into existence. If the new relationships are used as a base for improvements, and they do not come into existence, then the risk in the network may be increased as potentially non-optimal relationships are added.

The next selected node is Project Manager 2. The process is then repeated for each of the nodes, until at most one connection is added per source node, or the risk threshold is reached. For this illustration, a risk reduction of at least 10% is considered sufficient. This requires one further step, wherein a relationship is added between Project Manager 2 and

Team Member 11. The new network that results from this second iteration is shown in Figure 6.3, and the new risk metrics are shown in Table 6.3.

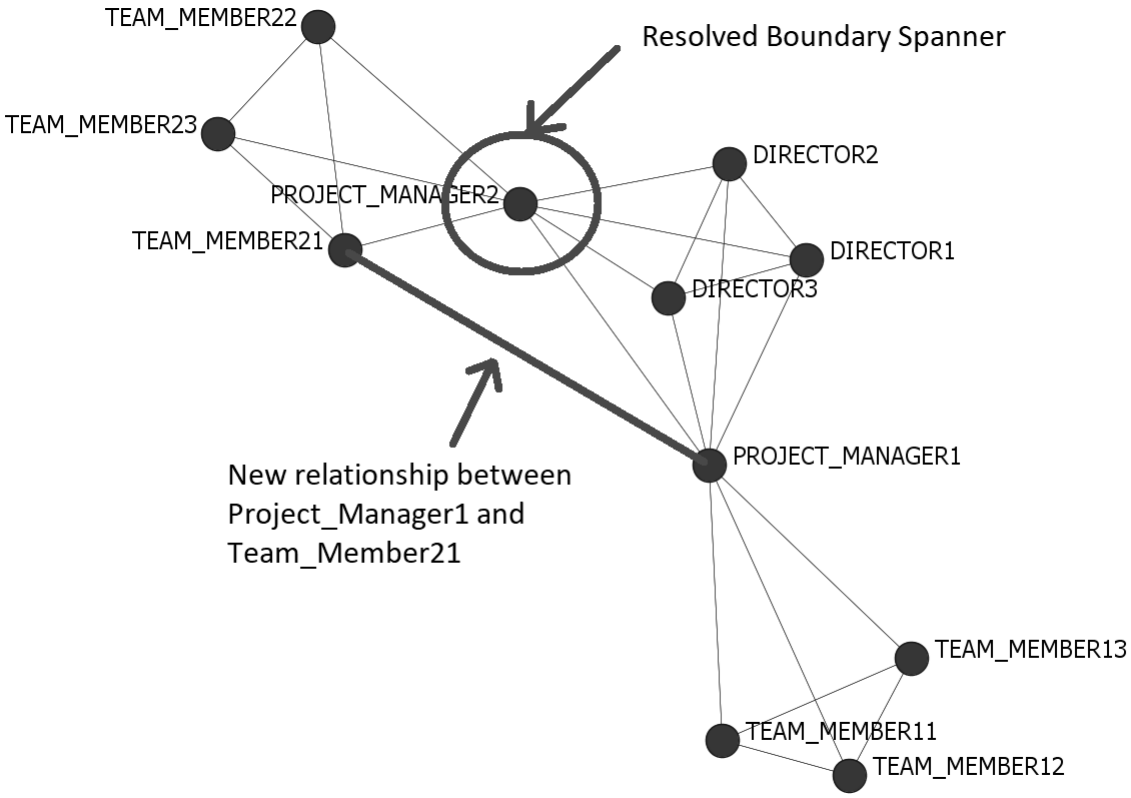


FIGURE 6.2: SNC NETWORK AFTER FIRST OPTIMISATION STEP

TABLE 6.2: NORMALISED SNA METRICS FOR MODIFIED SNC NETWORK FOLLOWING FIRST ITERATION

Node Title	BC	CC	EcC	EiC	SHC	TDC	BS	Total Risk
DIRECTOR 1	0	0.321656	0	0.879433	0.528926	0.221963	0	1.951977
DIRECTOR 2	0	0.321656	0	0.879433	0.528926	0.221963	0	1.951977
DIRECTOR 3	0	0.321656	0	0.879433	0.528926	0.221963	0	1.951977
PROJECT MANAGER 1	1	1	0	1	0	1	1	5
PROJECT MANAGER 2	0.855288	1	0	0.992908	0.126033	0.890187	0	3.864415
TEAM MEMBER 11	0	0	1	0	0.849174	0	0	1.849174
TEAM MEMBER 12	0	0	1	0	0.849174	0	0	1.849174
TEAM MEMBER 13	0	0	1	0	0.849174	0	0	1.849174
TEAM MEMBER 21	0	0.321656	0	0.861702	0.997934	0.221963	0	2.403255
TEAM MEMBER 22	0	0	1	0.414894	1	0	0	2.414894
TEAM MEMBER 23	0	0	1	0.414894	1	0	0	2.414894
OVERALL NETWORK RISK								**27.50091**

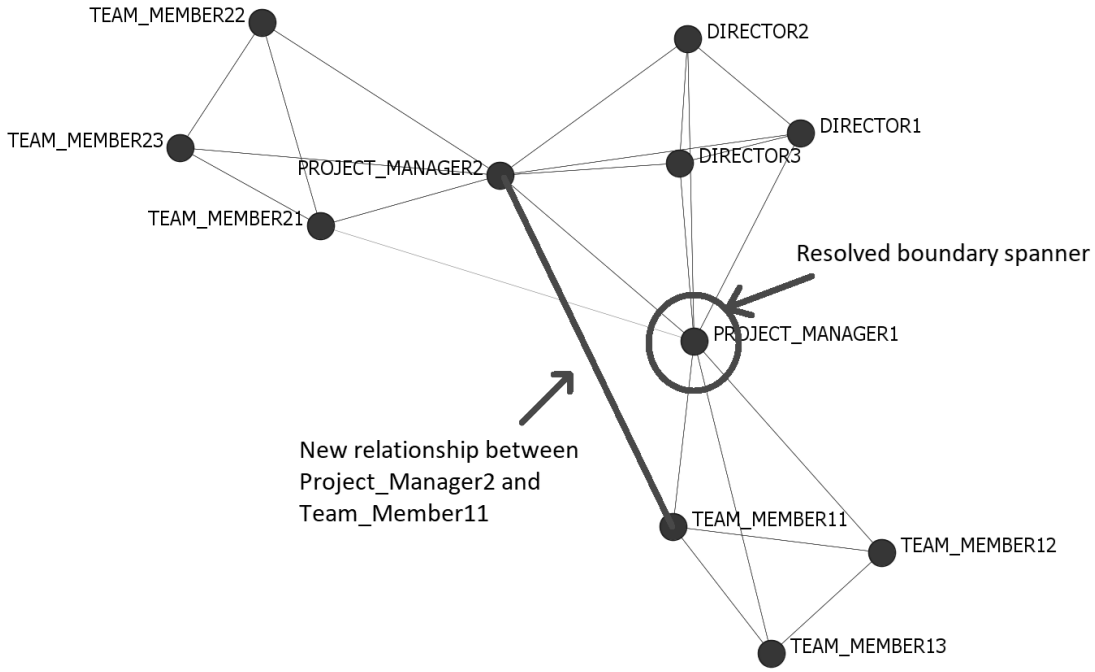


FIGURE 6.3: SNC NETWORK AFTER SECOND OPTIMISATION STEP

TABLE 6.3: RISK METRICS FOLLOWING SECOND ITERATION. TOTAL REDUCTION IN RISK OF 14.04%

Node Title	BC	CC	EcC	EiC	SHC	TDC	BS	Total Risk
DIRECTOR 1	0	0.267196	0	0.547009	0.521561	0.19958	0	1.535345
DIRECTOR 2	0	0.267196	0	0.547009	0.521561	0.19958	0	1.535345
DIRECTOR 3	0	0.267196	0	0.547009	0.521561	0.19958	0	1.535345
PROJECT MANAGER 1	1	0.830688	0	0.706553	0	0.89916	0	3.4364
PROJECT MANAGER 2	1	1	0	1	0.178645	1	0	4.178645
TEAM MEMBER 11	0	0	1	0.207977	0.581109	0.098739	0	1.887826
TEAM MEMBER 12	0	0	1	0	0.780287	0	0	1.780287
TEAM MEMBER 13	0	0	1	0	0.780287	0	0	1.780287
TEAM MEMBER 21	0	0.267196	0	0.54416	0.991786	0.19958	0	2.002722
TEAM MEMBER 22	0	0.055556	1	0.216524	1	0	0	2.27208
TEAM MEMBER 23	0	0.055556	1	0.216524	1	0	0	2.27208
OVERALL NETWORK RISK								**24.21636**

The final result of the optimisation process is a network that has no boundary spanners, and features a 14.04% reduction in risk. Both project managers are also a lower risk than they used to be, as their values went from 4.855124 prior to the process, to 3.4364 and 4.178645, respectively. In summary, in order for SNC to obtain a greater than 10% reduction in risk in its network, relationships need to be added between each of the project managers

and a member of the other manager's team. These proposed changes take into consideration that no relationships may be removed, and that some relationships may take time to form. As an added benefit, by adding these relationships, all of the boundary spanners in the network are resolved, thus improving the stability of the network. This is evidenced by the fact that there are no instances of any nodes within the new network where the removal of a node will create an isolated sub-network. Should a project manager be taken ill, for example, all of the team members will now still have contact with a project manager and, through him, the directorate.

It should be stressed, in conclusion, that this was a simple example using simulated data, and was created specifically to demonstrate the interaction between SNA and information security. Real-world networks, especially of the size found in most organisations, will likely be far more complex and require a significant number of iterations before the desired risk reduction threshold is reached, if it is reached at all. It is also important to note how the specific nature of a network influences the optimisation process. In the SNC example, relationships could be added, but not removed. This had a significant impact on both the metrics that could be used and the relationships that could be added. The optimisation method should therefore always be used within the context of the network it is applied to, otherwise the modifications it proposes may be unfeasible in the real world. This optimisation method is applied to a real-world network in Chapter 7.

6.2. Evaluation of the Risk in a Social Network using Self-Organising Maps

The use of Self-Organising Maps (SOMs) as a visualisation technique is briefly described in Chapter 3. In this section SOMs are discussed in greater detail, and a method, utilising SOMs, that can be used to evaluate risk in a social network is presented. This method, developed for this study, was presented at the WISE conference in Lisbon in 2019, and the paper was published as a chapter in Volume 557 of the IFIP Advances in Information and Communication Technology series. The pre-publication manuscript version of the article, cited as (Serfontein *et al.*, 2019), is presented in Appendix A.

The self-organising map (SOM), in essence, is a neural network technique that can be used to visualise and evaluate high-dimensional data (Bäck *et al.*, 2012). The SOM technique uses given data to produce a self-organising neural network wherein the data points are clustered into topographical regions (Pal *et al.*, 2018). A graphical representation of the functioning of this technique is shown in Figure 6.4. This visualisation technique has a wide range of known applications, from evaluating comparable biological adaptations (Nakayama *et al.*, 2018) and improving optimisation algorithms (Gu & Cheung, 2018; Kuo *et al.*, 2018),

to clustering data for problem-solving purposes (Lee, 2019; Pal *et al.*, 2018). One of the greatest advantages SOM has over other high-dimension visualisation techniques is that it produces a two-dimensional topographical map that can be evaluated and interpreted without any special knowledge or skills. In addition to clustering known data points, depending on the data, the technique can also be used in vector quantisation, and as a regression modelling technique (Bäck *et al.*, 2012). While all of these methods can arguably be used to obtain valuable information about data, the technique is applied here primarily in order to cluster nodes according to their risk metrics.

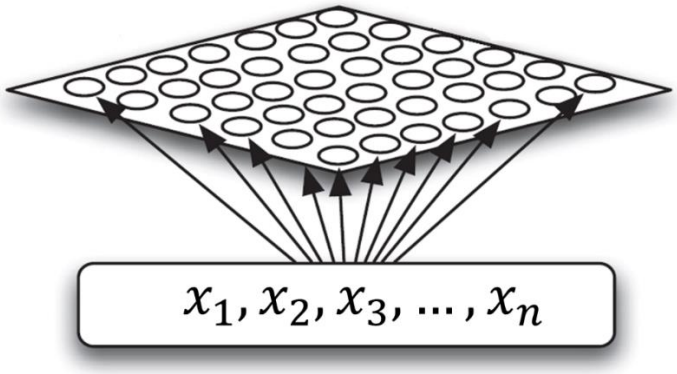


FIGURE 6.4: ALLOCATION OF DATA POINTS TO TOPOGRAPHICAL REGIONS USING SOM TECHNIQUE (LÓPEZ *ET AL.*, 2019)

As SOM makes use of complex mathematical neural network techniques, the derivation and development of which fall outside the scope of this study, only an overview of the algorithm will be provided. The version of the SOM algorithm presented here is based on the work done by Fausett (1994), and a more in-depth description of SOM algorithms can be found in this source. Algorithm 6.2 presents the Kohonen SOM, which clusters input vectors according to how well they match topographical area weight vectors. As the number of vectors in each area increases, the area itself is shrunk to increase the number of areas that vectors can be clustered into.

Because SOMs can be used to cluster high-dimensional data on a graphical two-dimensional map, it is especially valuable to those in managerial positions, as these individuals may not have the time to study large reports and data sets in detail. By making use of the various SNA risk metrics discussed in Chapter 4, the nodes in a social network can be clustered into groups that share similar risk patterns, and the result can be presented graphically. Additionally, due to the use of SOMs, a number of at-risk nodes may be identified that would not necessarily have been evident through the use of more traditional visualisation techniques such as bar-graphs. It is, for argument's sake, possible for a node to have all the traits of a high-risk node and not be obvious from the data itself. In these instances, a

clustering technique such as SOM can be used to identify nodes that have similar, possibly hidden, attributes.

ALGORITHM 6.2: Kohonen Self-Organising Map (Fausett, 1994)

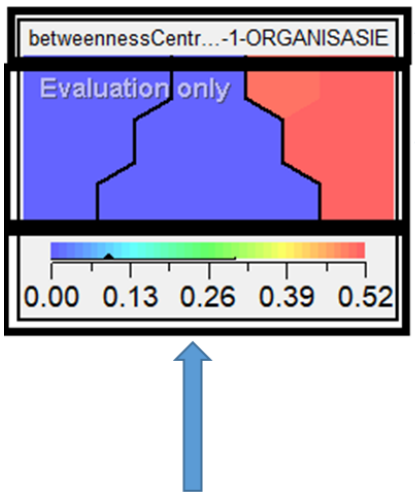
Input: Dataset *N*
Output: A topographical map *M* containing the data from *N*, sorted into topographical areas

Variables:
w_{ij} – Weight vector describing topographical area ij; either randomised or defined at start
x – An input vector contained in *N*
α – Learning rate that is a slowly decreasing function of time

```
while Stop condition is false
    For each x in N
        For each vector j
            Compute D(j) = ∑_i (w_ij − x_i)^2
        end
        Find index J such that D(J) is a minimum
        For all units j in a topographical area J, and for all i:
            Compute w_ij(new) = w_ij(old) + α[x_i − w_ij(old)]
        end
    end
    Update α
    Reduce radius of topographical area at specified times
    Test Stop condition
end
return M
```

In order for a SOM to be especially useful with regard to clustering, each of the nodes should be described using at least three attributes. In the case of the simulated network, the attributes for each node is the risk metrics for that particular node. The data is then processed using Viscovery's SOMine software suite in order to obtain the maps. One map is produced that contains the nodes clustered into topographical regions. This map, which can be referred to as the base map, can then be coloured to illustrate the relationship between a particular attribute and the nodes on the map. A coloured version can therefore be generated for each of the attributes used. Reading the maps is subsequently very straight-forward. Each of the regions represents a cluster, and the colour of each section of the map indicates the value the nodes in that section have with regard to each specific attribute. The value is matched to a colour scale, which is shown below each map. Figure 6.5 provides a graphical guide to evaluating the coloured SOMs produced by SOMine.

To demonstrate how these maps can be used to evaluate the risk in a social network, a simulated network is used. This second simulated network is shown in Figure 6.6. This network contains 12 nodes and 53 links. Seven metrics were calculated for each of the nodes, and the values of these metrics are shown in Table 6.4.



Name of the attribute used to colour the regions

SOM with three regions (clusters)

Each region contains a number of nodes. If an entire region is coloured red, then all of the nodes in that region have a betweenness centrality value of ~0.52. If a region contains one section where the colour Is different, then that section corresponds to a node that has other attributes that make that particular region the closest match.

Colour scale for attribute values. Colour on map corresponds to colours on scale, which correspond to the range of values the attribute has on this map. On this map, a red colour indicates a betweenness centrality value of ~0.52, whereas blue indicates a betweenness centrality value of ~0.00. as none of the nodes clustered have a betweenness centrality value between 0.13 and 0.39, turquoise, green, yellow, and orange are not used.

FIGURE 6.5: GUIDE TO READING A SOM

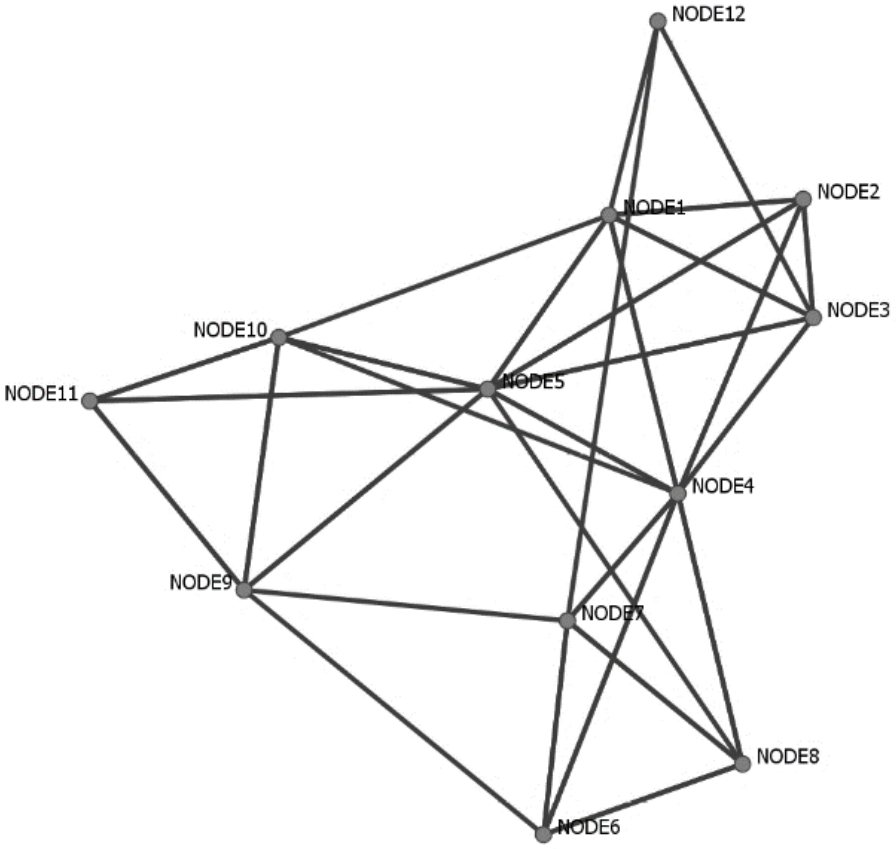


FIGURE 6.6: SOM ILLUSTRATION NETWORK

TABLE 6.4: SNA METRICS FOR NODES IN SOM DEMONSTRATION NETWORK

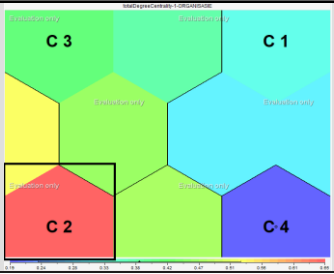
Node	Betweenness Centrality	Closeness Centrality	Eccentricity Centrality	Eigenvector Centrality	Structural Holes Constraint	Total Degree Centrality
NODE1	0.03	0.262	2	0.475	0.522	0.435
NODE2	0	0.262	2	0.403	0.618	0.348
NODE3	0.03	0.262	2	0.424	0.575	0.391
NODE4	0.367	0.297	2	0.482	0.383	0.652
NODE5	0.394	0.306	2	0.494	0.399	0.652
NODE6	0.003	0.239	3	0.348	0.649	0.304
NODE7	0.015	0.239	3	0.369	0.565	0.348
NODE8	0	0.239	3	0.351	0.643	0.304
NODE9	0.021	0.268	2	0.474	0.709	0.391
NODE10	0	0.268	2	0.448	0.633	0.348
NODE11	0	0.244	3	0.34	0.816	0.261
NODE12	0	0.440	3	0.184	0.485	0.130

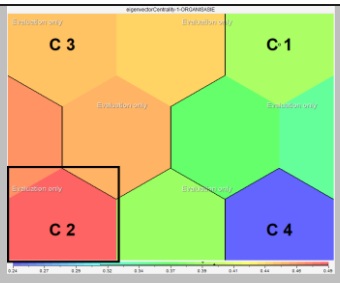
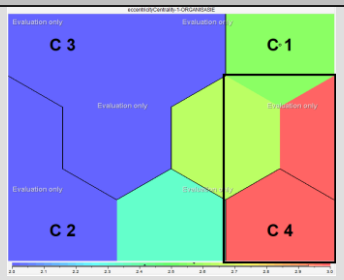
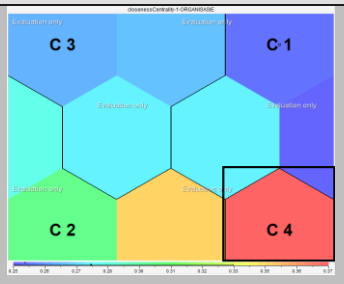
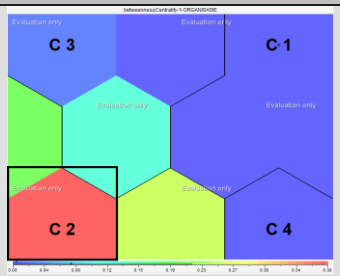
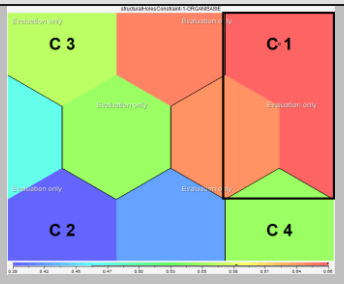
The SNA metric data shown in Table 6.4 on its own conveys relatively little information regarding possible risks that the nodes may pose within the network, and requires more in-depth analysis in order to be meaningful. By using these data as attributes for the nodes and then clustering these nodes using the SOM technique, however, it is possible to obtain useable information from the data without in-depth analysis. To demonstrate this, consider the SOM and discussion shown in Table 6.6, which was generated using only the data shown in Table 6.4. The nodes situated in each of the four clusters are shown in Table 6.5. The CIA columns in Table 6.6 show which of the confidentiality, integrity, or availability aspects of information security are associated with each of the metrics. The rationale for each of these metrics being associated with particular CIA risks is discussed in Chapter 4.

TABLE 6.5: NODES CONTAINED IN EACH OF THE FOUR CLUSTERS

Cluster	Nodes in cluster
C1	NODE6, NODE7, NODE8, NODE11
C2	NODE4, NODE5
C3	NODE1, NODE2, NODE3, NODE9, NODE10
C4	NODE12

TABLE 6.6: SOM OF DATA IN TABLE 6.4, COLOURED USING EACH OF THE SIX METRICS USED

Metric used to colour SOM	CIA			Coloured SOM	Comments
	C	I	A		
TOTAL DEGREE CENTRALITY (TDC)		X	X		The nodes in cluster C2 have a much higher value for TDC than the other nodes in the network. This suggests that these nodes present a risk to integrity and availability in the network. Cluster C4 contains the lowest TDC values, which suggests that the only node in C4, namely NODE12, presents a minimal risk to integrity and availability.

EIGENVECTOR CENTRALITY (EiC)	X	X	X		Cluster C2 contains the highest EiC values, which are associated with a general increased risk. This increased risk is due to the leadership qualities these nodes possess. C3 also contains nodes with elevated EiC measures, which suggests that the cluster may contain those nodes that are a high risk, but not yet as high as the nodes in C2. As with TDC, the node in C4 presents the smallest risk to the network.
ECCENTRICITY CENTRALITY (ECC)			X		The increased ECC values in clusters C4 and C1 suggest that there are nodes in the network that are somewhat isolated. These nodes pose a risk to availability, as their isolation may negatively impact the flow of information in the network. The nodes in clusters C2 and C3, by contrast, have a very low measure of ECC and therefore pose minimal risk to availability in this regard.
CLOSENESS CENTRALITY (CC)	X				Both clusters C2 and C4 contain nodes with high CC values. Of particular interest is cluster C4, which also contains nodes with high ECC values and low TDC values. This suggests that NODE12 is somewhat of an irregular presence in the network, as the node is isolated and has little direct "friendships", but has access to a significant amount of information. This node therefore poses a significant risk to confidentiality.
BETWEENNESS CENTRALITY (BC)	X	X			The elevated BC values in C2 suggest that some of the nodes in C2 act as "shortcuts" for information flow within the network. As these nodes have the potential to modify data and control information flow between nodes, these nodes present a risk to both confidentiality and integrity. These are the same nodes that have high TDC and EiC nodes, which suggests that the well-connected, emergent leaders in the network pose a significant risk to information in the network.
STRUCTURAL HOLES CONSTRAINT (SHC)		X	X		Cluster C1 contains nodes that have a high measure of SHC. This suggests that these nodes may not be well connected enough to function adequately. This suggests that there is a communication problem within the network, which may negatively impact the integrity and availability of information in the network.

The brief discussion presented in Table 6.6 shows how much simpler it is to evaluate risk metric data using SOM when compared to in-depth statistical analysis. Furthermore, as the attributes of the nodes determine the clusters, the technique can be used to easily monitor how the risk in the network changes over time.

6.3. IMPROVING INFORMATION SECURITY AWARENESS USING SNA

In this section, a method that can be used to develop cost-effective security awareness programmes using SNA is proposed. This method, developed for this study, was presented at the WISE conference in Poland in 2018, and the article was published as a chapter in Volume 531 of the IFIP Advances in Information and Communication Technology series. The pre-publication manuscript version of the article, cited as Serfontein *et al.* (2018), is presented in Appendix A.

In an article by Dang-Pham *et al.* (2017a), a method was discussed whereby SNA was used in a Vietnamese organisation to identify individuals that would be able to act as information security champions. These individuals, once identified, received information security training so that their influence could help shape the information security culture of the organisation. In Serfontein *et al.* (2018), this method is referred to as the Dang-Pham Awareness (DPA) method.

The information security awareness technique described in this section is broadly based on the DPA-method, but differs from it in two crucial ways:

- The DPA-method uses formal networks exclusively, whereas the technique described here can be used for both formal- and informal networks; and
- This method can be used to develop information security programmes that specifically target the awareness level of a group, which should help prevent security fatigue, which is a form of mental fatigue that is caused by over-exposure to information security knowledge, from developing. It therefore differs from the DPA-method in that it does not identify individuals to train, but rather identifies ideal locations to target with information security programmes.

The proposed method follows a phased process, which is shown in Figure 6.7. The three stages of the method, named Preparation, Network Construction, and Evaluation and Implementation, broadly follow a standard process structure. During the first phase, namely Preparation, a clear plan is developed that will be used during the second phase to collect all of the data needed to develop a targeted information security awareness programme. In order to accomplish this, the phase deals with the issue of bordering by identifying target groups. Once the target groups have been identified, the appropriate data collection techniques can be selected. As the goal is also to prevent the development of information security fatigue, the overall level of information security awareness in the target group is estimated. This estimation can be made either by making use of formal techniques developed for this exact purpose, such as the one presented in Kruger and Kearney (2006), or by using the group's field of study and an educated guess. In either event the group's

existing knowledge is taken into consideration and respected, which should help mitigate information security fatigue.

The second phase, named Network Construction, focusses on using the plan developed during the first phase to collect the relevant data, and then using that data to construct a social network that can be used to identify key nodes. These nodes can be anything from an influential individual, to a shared resource or a location. The output of the second phase is however not a list of individuals to target, but only a social network and a number of selected metrics. The metrics should be chosen based on the group being targeted. Certain groups may respond negatively if one of its emergent leaders, i.e. someone with a high eigenvector centrality, is suddenly drafted to help improve the information security awareness of the group. Selecting the correct metrics may require both intimate knowledge of the targeted group's culture, and some trial-and-error experimentation may be required before the correct measures are identified.

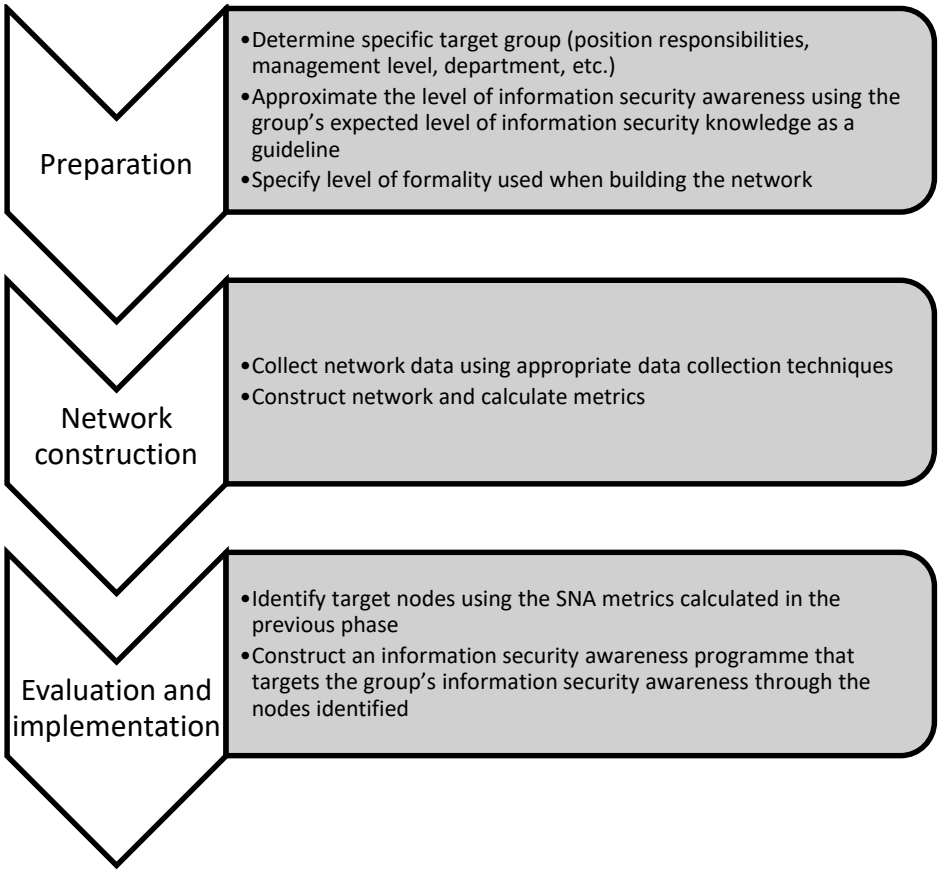


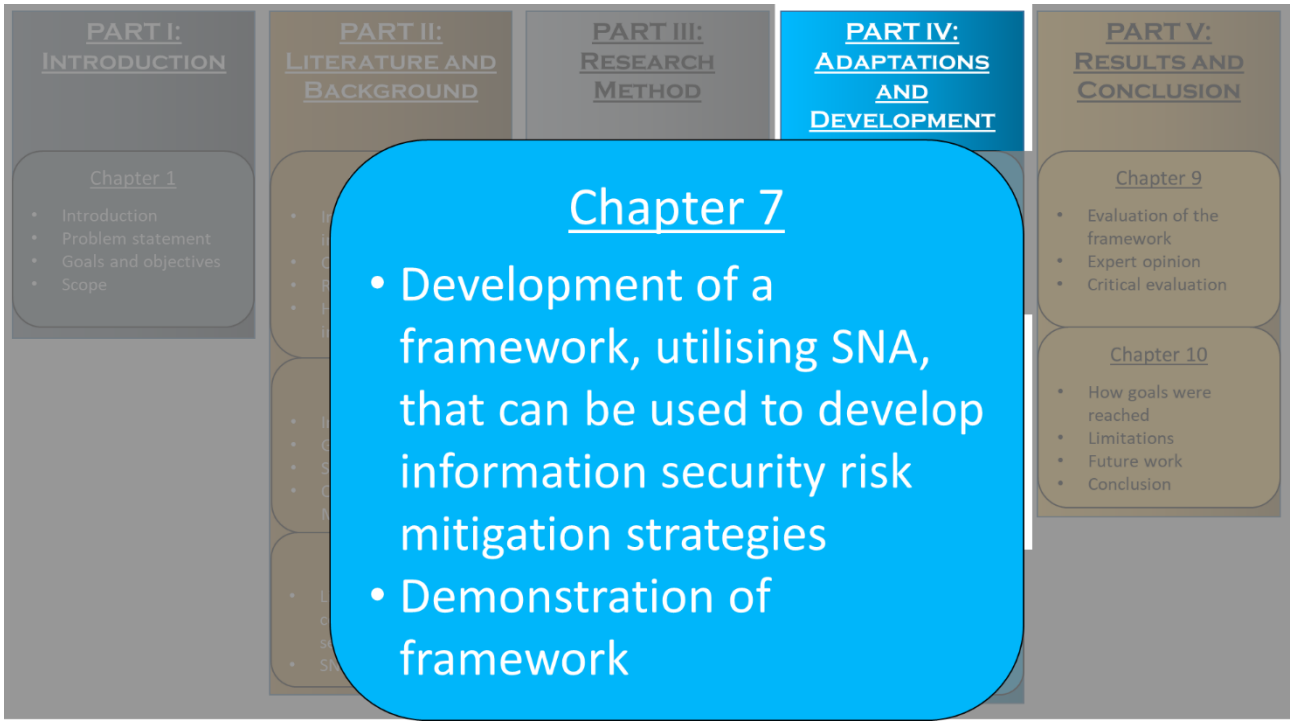
FIGURE 6.7: PROCESS FOR DEVELOPING A TARGETED SECURITY AWARENESS PROGRAMME USING SNA

The final Evaluation and Implementation phase deals specifically with developing and implementing awareness programmes using the SNA data obtained during the previous

phase. With the metrics for the network now available, the best suited nodes can be selected. These nodes will ideally represent entities such as locations that either many members of the group frequent, resources, such as posters in important locations that may impact influential members, or other strategic resources that can be used to naturally disseminate information through the network. Once these nodes have been identified, the method used to target them should also be identified. Finally, all of the information about the network, including the targeted nodes and the methods used to target them, should be used to develop an information security awareness programme. This programme, because of its targeted nature, may be more effective and less expensive to implement than programmes developed using more traditional planning techniques.

6.4. CHAPTER SUMMARY

In this chapter, the adaptation of a number of methods for use with SNA were discussed. The chapter started by presenting a network optimisation model and discussed how it can be adapted to make it useable with social networks. This was followed by a description of how SOMs work, and how they can be used to evaluate risk in a social network. The chapter concluded with a description of a technique that could be used to improve security awareness programmes. In the next chapter, a framework is presented that implements the first two of these techniques, namely the network optimisation and SOM techniques.



CHAPTER 7: A NOVEL FRAMEWORK FOR ADDRESSING INFORMATION SECURITY RISK USING SNA

CHAPTER HIGHLIGHTS:

- What is the structure of the framework that implements SNA in order to develop risk mitigation strategies?
- How is the optimisation technique discussed in Chapter 6 incorporated into the framework?
- Where is the SOM analysis technique used in the framework?
- Can the framework be used to evaluate risk in both small and large social networks?
- How can risk mitigation strategies be developed using the framework?

7 A NOVEL FRAMEWORK FOR ADDRESSING INFORMATION SECURITY RISK USING SNA

In this chapter, a framework for a novel method is introduced that proposes risk management steps by applying network optimisation to a social network. First, a description of the framework is provided, which focusses on its various phases and network optimisation approaches. Then an illustrative example is provided to demonstrate how the framework can be applied to data. Finally, the chapter concludes with an in-depth application of the method to a large real-world dataset to demonstrate the scalability of the method. This concluding section also describes the specific data-collection methods used and the nature of the collected data.

7.1. METHOD FRAMEWORK

In this section, the development of the framework for the method is discussed in detail. The motivation for choosing to present the method as a framework, rather than as a rigid method, rests in the adaptability a framework has compared to a method. Furthermore, as it can easily be adapted to incorporate new methods and techniques, a framework allows for a greater measure of malleability. As each organisation has its own unique network and cultural features, it also allows for a greater measure of specialisation: depending on the network and the metrics used, different network optimisation methods may be more effective for different organisations. In the interest of clarity, the overall structure of the framework is shown in Figure 7.1.

The framework is divided into five phases. The first phase is concerned with collecting data and developing social networks that can be used to evaluate information security risks. In the second phase the SNA metrics are calculated, and the SOM method discussed in Chapter 6 is used to provide an initial evaluation of the metrics data. The third phase focusses on optimising the networks using the risk metrics calculated in the previous phase. In the penultimate phase, the results of the optimisation process are used to develop risk mitigation strategies. The final phase deals with the implementation of the strategies developed in the fourth phase. The fifth phase also deals with monitoring the network, so that any changes to the network can be detected as needed. This monitoring process makes use of the Statistical Control Chart (SCC) method described in Chapter 3.

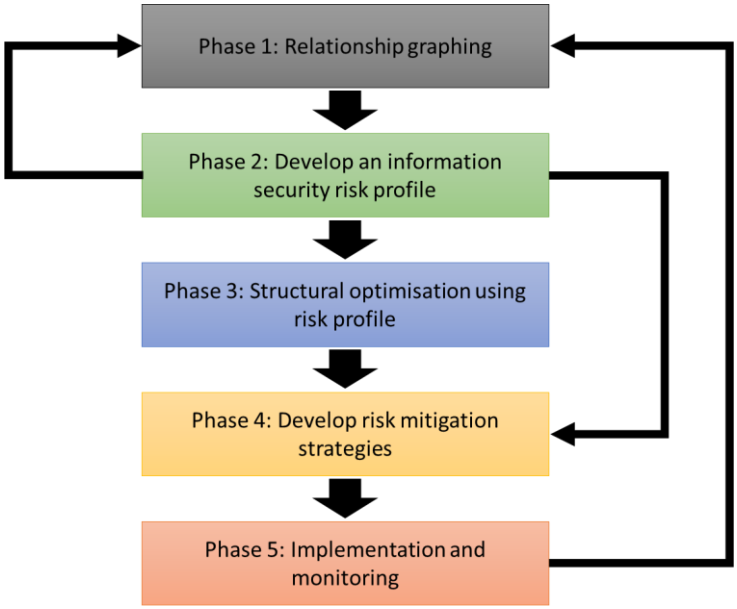


FIGURE 7.1: OVERALL STRUCTURE OF FRAMEWORK

While all five phases are included in the framework, it is important to clarify that the focus in this study is predominantly on the first three phases. As discussed in Chapter 1, one of the aims of this study is to demonstrate how SNA can be used to inform the development of risk mitigation strategies. Because the development of the strategies themselves can potentially be an incredibly complex process, and the strategies will likely differ from organisation to organisation, the scope in this study is focussed on how SNA can be used to identify which nodes and relationships should be targeted by the risk mitigation strategies. Phases 4 and 5 are, nevertheless, still discussed in detail, as they are important to real-world implementations of the novel framework presented in this chapter.

Each of the five phases, namely Relationship Graphing, Develop an Information Security Risk Profile, Structural Optimisation Using Risk Profile, Develop Risk Mitigation Strategies, and Implementation and Monitoring, will now be discussed in turn.

7.1.1. Phase 1: Relationship Graphing

The ultimate objective of a method based on this framework is to address information security risks using SNA. This is done by developing risk mitigation strategies, which are then implemented, and the effects monitored. As with any such method, whereby the goal is to improve an existing situation, current data about the situation is needed. Normally, such a phase would be called “data collection”. In order to emphasise that this phase deals with representing various social networks using data, however, the phrase “Relationship Graphing” was chosen as the phase title. This also summarises the purpose of the first

phase: **collect data that can be used to graph the relationships in the organisation and use the data to produce social network graphs.**

As discussed in Chapter 3, one of the most important steps in collecting SNA data is to border the relationships. This is a necessary process that involves selecting specific relationships that are relevant to the problems being evaluated. As there are many ways that can be used to border an organisation's social network, an incremental approach is proposed. By using an incremental approach, the relationships can be described using various attributes, and these attributes can be ordered based on priority. However, regardless of how the relationship attributes are prioritised, the first stage of the bordering process should involve choosing between **formal** and **informal** relationships.

A formal relationship, as used in this context, is a relationship that primarily exists by virtue of a formal organisational structure, and is used to construct a **formal network**. The relationship between a manager and an employee, for example, would be formal in nature. The reporting structure of an organisation, in similar fashion, would also be formal. Because of the nature of these relationships, it is highly unlikely that any individual in an organisation would be isolated in a formal network - if there are any isolated individuals, this would indicate a problem with the organisation itself.

Informal relationships, by contrast, exist as a result of personal interaction between two entities. When an employee leaves an organisation, his relationship with his former manager would no longer be formal. If the two were friendly and remain in contact, an informal relationship would however still exist between the two. An informal relationship therefore does not primarily exist as a result of any formal structure, and can exist between any two members of an organisation. The implication of this is that a formal network, constructed using formal relationships, and an **informal network**, constructed using informal relationships, will likely be different for the same organisation. Additionally, it is highly probable that the data for a formal network will be easier to collect, as formal organisational information can be used. Collecting data on informal networks typically require much more intrusive data collection methods, such as questionnaires and email analysis, and there is a risk that the organisation's members and culture will not be welcoming to such methods.

The selection of a formal or an informal network can therefore have a profound impact on the process as a whole. Depending on the unique properties of the organisation itself, it is possible that only one of the two types of network will be usable. A small organisation with only one manager, for example, will most likely produce a simple formal network that is unlikely to yield any insights. In larger organisations, however, the formal network may provide sufficient data to properly evaluate the information security risk of the organisation.

In order to choose between using formal or informal networks, the characteristics of the organisation being evaluated need to be considered. Typically, the decision will depend on

aspects such as the culture of the organisation, its structure, and its policies with regard to privacy. It may, for example, be difficult to collect sufficient enough data for an informal network if there is a persistent culture of non-compliance amongst its members. In Table 7.1, a number of these characteristics are compared based on how they impact on the decision to choose between formal and informal networks. It should be noted that an organisation may have characteristics that make both formal- and informal networks equally appropriate; in these instances, data for both network types can be collected and used in parallel.

TABLE 7.1: COMPARISON OF ORGANISATIONAL CHARACTERISTICS THAT ARE APPROPRIATE FOR THE SELECTION OF FORMAL AND INFORMAL NETWORKS

Characteristics	Appropriate for Formal Networks	Appropriate for Informal Networks
Size of Organisation	Medium to Large organisations	Any size
Culture	Non-cooperative, adverse to any activities requiring additional effort, questionnaire avoidant	Cooperative, willing to help with managerial data collection, poor managerial recordkeeping wherein management fails to keep track of structural changes
Approach to privacy	Privacy-neutral or privacy oriented	Not focussed on privacy, or organisation considers all communication during working hours to be a part of the organisation's data
Organisation stability	Organisation is relatively stable in that its structure does not change regularly	Organisation's structure changes constantly, or there is rapid employee turnover
Resource sharing	Resources are shared based on a clear organisational policy, or access is strictly regulated	Resources are shared in an informal fashion, or there are no access controls in place

There is a risk involved in selecting the wrong type of network for a particular organisation. If, for example, a formal network is selected, the risk posed by a hypothetical mid-level manager with a high number of personal, informal relationships may be overlooked. Similarly, if an informal network that is evaluated using questionnaires is selected, and there is a culture of non-compliance amongst the members of the organisation, the resultant network data may be incomplete, highly inaccurate, or both. For this reason, it is important for the selected network type to match the organisation.

Once the choice has been made between using formal- or informal networks (or both, depending on the situation), the specific methods that will be used to collect the necessary data can be selected. For formal networks, the reporting structure of the organisation, also called an **organogram**, may be sufficient. Other formal relationships, such as participation in interdepartmental projects and formal access to resources, may also be included. For informal networks, more direct data collection methods are needed. The methods may

include, but are not limited to, questionnaires, email analysis, interviews, observations, and social media analysis. How resources are shared should also be taken into account: if there are resources that are shared, the manner in which they are shared and amongst whom they are shared should be taken into consideration. If there are any tasks that certain members participate in together, these tasks should also be evaluated.

After the data collection methods have been selected, they should be used to determine the various relationships in the organisation. Some of these techniques will necessarily produce either a formal or an informal network. An organogram, for instance, necessarily produces a formal network. Other techniques, such as questionnaires, can be used to collect either formal or informal network data. The data, once collected, can now be used to produce the **relationship graph** (or graphs, if both formal and informal networks are to be used) that will be used in the following phase. An overview of Phase 1 is shown graphically in Figure 7.2.

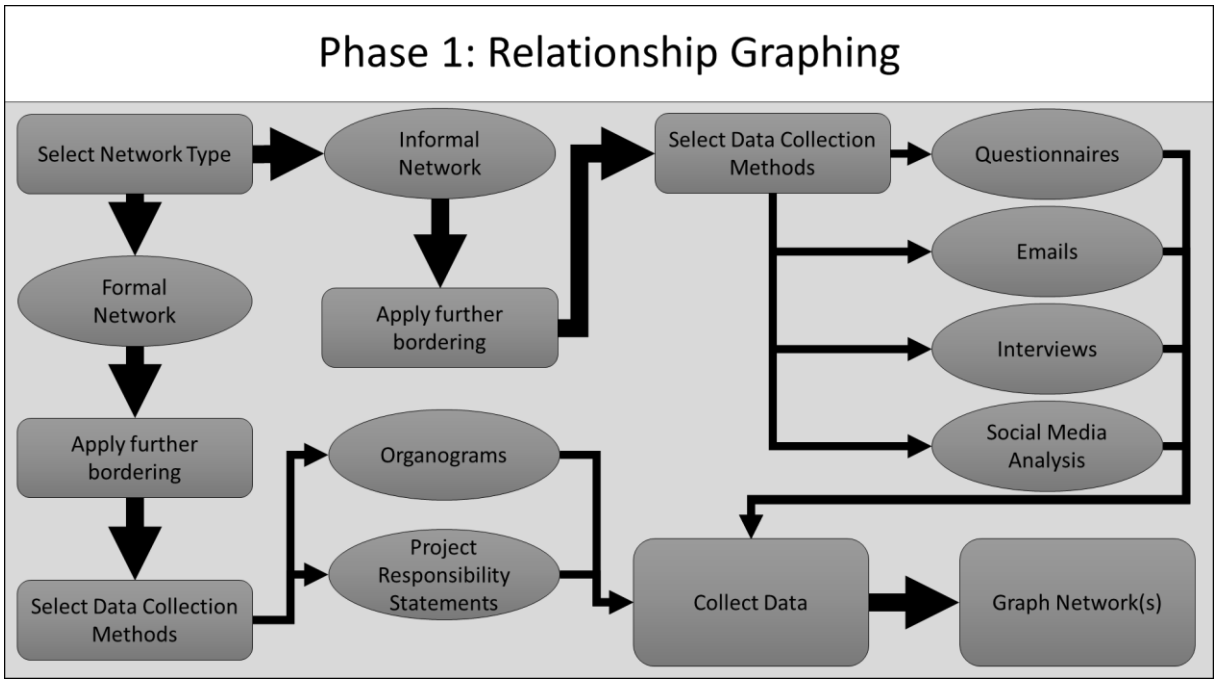


FIGURE 7.2: OVERVIEW OF PHASE 1

Phase 1, in conclusion, starts with selecting the network type. The network can be either formal or informal, or both formal and informal networks can be selected if there is a possibility that different types of risk in the organisation may be affected by the type of relationship. Then, after the choice has been made between either formal, informal, or both, the specific types of data that have to be collected are identified and appropriate data collection techniques are selected. The final output of this phase is a graph of the selected network constructed using the data collected.

7.1.2. Phase 2: Develop an Information Security Risk Profile

The purpose of Phase 1 is to produce a graph, or graphs, that represent the social network of an organisation. In Phase 2, the graphs are used to produce an information security risk profile. In order to develop this risk profile, various SNA metrics are used to determine CIA attributes, as discussed in Chapter 4. Additionally, by applying the SOM clustering method described in Chapter 6, it may be possible to identify a number of nodes that pose a risk due to the similarity in their attributes. The various steps that can be followed to develop the risk profile will now be discussed.

The importance of Phase 2 cannot be overstated. As shown in Chapter 4, it is possible to determine the information security risk posed by a node by using its SNA metrics. If these metrics are not calculated properly, or used incorrectly, it is possible that a high-risk node may be ignored, or even made into a higher risk. Additionally, if the risk profile is not produced properly, it becomes next to impossible to develop meaningful, useful, and impactful risk mitigation strategies during the later phases. It is also important that the SOM technique is used correctly during this phase, as the technique can be used to identify nodes that, while they do not have a high level of risk based on their SNA metrics, still have characteristics that nodes with high levels of risk do have. The SOM technique can therefore be used to identify nodes that may become a risk in future. By identifying these nodes beforehand, the risk they may pose can be addressed before it is realised.

Another important aspect of this phase is that the quality of the data collection methods selected in Phase 1 can be appraised. During Phase 2, as the data is evaluated using the selected SNA metrics and the SOM technique, it should be possible to identify potential shortcomings in the collected data. If, for example, the total degree centrality of a person who is a known gossip in an informal network is calculated as being very low, then the quality and appropriateness of the data collection methods selected in Phase 1 can be questioned, and Phase 1 can be repeated before possibly damaging actions are implemented in later phases.

In order to develop a risk profile for the network, a number of things need to be taken into consideration. The first of these considerations is whether or not the risk profile needs to include the entire organisation. If not, there are a number of different techniques that can be used to divide the network into component parts. One of the methods that can be used to separate the network into component parts is community detection. Some of the various community detection techniques that can be used are discussed in Chapter 3. Alternatively, certain departments or management layers can be excluded entirely. Nodes may also be included or excluded based on certain inherent attributes. If the entire network is to be evaluated as a single unit, then no separation is needed.

Once the sections of the network that have to be evaluated have been selected, the SNA metrics for each node in the selected sections can be calculated. The metrics as described in Chapter 4 are used for this purpose. It should be noted, however, that any meaningful SNA metric can be used. The SNA metrics are then used to determine the risk profile for each of the nodes. The second consideration is that it is possible that a node may be a risk in one aspect, but not a risk overall. It is therefore important to calculate a single value that can be used to describe a node's risk within the context of the network. In order to do this, all of the metrics should be used to determine a singular risk value. Whilst the SNA metrics as calculated will still be used as input data for the SOM method, the risk value can also be used to monitor how the overall risk of a node changes over time. Without the use of specific monitoring techniques, such as the SCC technique implemented in Phase 5, use of the calculated risk values to monitor the change in risk would likely be for more informal uses, such as checking to see if a node has undergone a change, no matter how insignificant. The primary value of the SOM technique at this stage is to help determine if the risk is high enough to warrant further analysis. Due to the graphical nature of SOMs, a quick visual evaluation of a risk profile SOM should be sufficient in some cases to determine if further steps are justified. It is possible, however, that poor bordering decisions may help discard crucial information. As such, if the SOM analysis does not highlight any at-risk nodes, then the network data and bordering steps should be revisited in order to verify that the collected data is indeed correct. If there really are no risks in the network, then Phase 2 need not be revisited, and Phase 3 may be skipped.

One method of combining the metrics for a node into a unitary risk value for said node is to apply a weighted summation. This provides for a computationally simple way to combine the metrics into a single value, and also allows for each implementation to be adjusted according to various specifications. The equation for such a weighted summation would be

$$Risk\ value\ of\ node\ i = C_i \sum_{m=1}^{k} w_m v_{mi} \tag{7.1}$$

where w_m is the weight of metric m, v_{mi} is the value of metric m for node i, k is the total number of metrics used, and C_i is an importance, or criticality, weight for node i. This criticality weight can be used to give special attention to certain nodes, or exclude other nodes entirely.

The above equation does have one major drawback, however, in that it does not account for the fact that most metrics may differ greatly in the range of values they have. In order to address this shortcoming, the metrics should be normalised. Normalisation, in this context, involves scaling all of the metric values so that they each have a value between 0 and 1. The specific normalisation method used is linear feature scaling (Aksoy & Haralick, 2001). This

method is used because it respects the bounds of each measure individually and allows for summations to be used in calculating the overall risk of a network. Feature scaling normalises the values of a metric based on the metric's upper and lower bounds, which are given as the minimum and maximum values obtained for any node in the network. The equation for calculating the normalised value x' for any value x is

$$x' = \frac{x - \min(X)}{\max(X) - \min(X)} \tag{7.2}$$

where X is the set containing x, $\min(X)$ is the smallest value in X, and $\max(X)$ is the largest. For a social network, X would contain all of the values for all of the nodes for a particular metric. When equation 7.1 is modified to implement this form of normalisation, equation 7.3 is obtained:

$$Normalised\ risk\ value\ of\ node\ i = C_i \sum_{m=1}^{k} w_m \left(\frac{v_{mi} - \min(V_m)}{\max(V_m) - \min(V_m)} \right) \tag{7.3}$$

where V_m is the collection of all values for all nodes for metric m.

There are a significant number of methods that can be used to determine the weights used in the above equation, if the weights are used at all. One way is to experimentally assign weights to each of the metrics until the results match the known risks for extant nodes. This process would however likely be highly resource intensive, which would only make it feasible in experimental environments, or in environments where these risks are already known. Another option is to use known multi-criteria techniques, such as the Analytical Hierarchy Process (AHP) (Bentley & Whitten, 2007). It is also possible to assign weights based on the perceived importance of particular nodes or metrics. If, for instance, one node or group of nodes is considered inherently more important than other nodes, the weights can be used to give priority to those nodes. In a similar fashion, if it can be shown or reasoned that certain measures are a greater risk than their values would suggest, the metric weights can be used to potentially address these risks first. Because of the wide range of options available when assigning weights to the metrics, no one method will be clearly defined as being part of the framework, and the specific method chosen is left to the discretion of the user. This further contributes to the malleability of the framework, which improves its usability in diverse circumstances. Regardless of the method used to calculate the weights, the ultimate objective is to produce a risk value that adequately describes the risk of each node. The overall structure of Phase 2 is shown graphically in Figure 7.3.

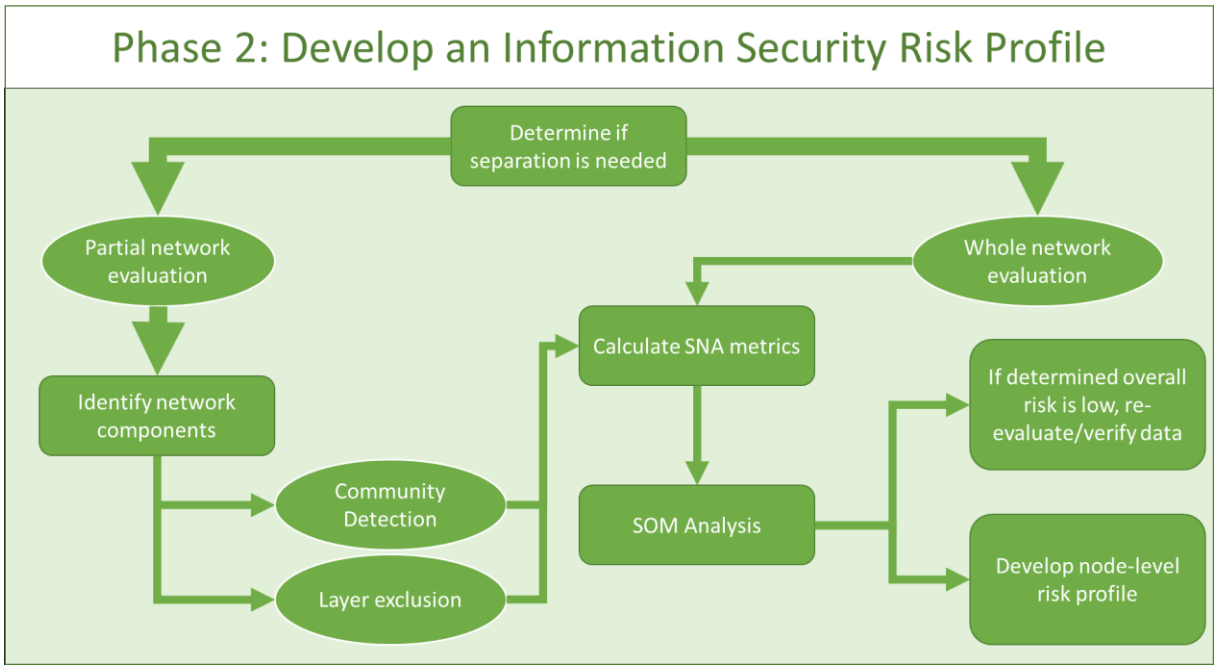


FIGURE 7.3: OVERVIEW OF PHASE 2

Phase 2, in summary, provides two sources of information. The first, the SOM analysis, can be used to quickly evaluate the current state of the organisation and determine if the observed risks warrant further use of the method. Alternatively, if there is a strong belief that the risk is higher than the data would indicate, the SOM analysis can be used to justify recollecting the data at an earlier stage, thus improving the efficiency of the method. The second source of information, namely the node-level risk profile, is used in further steps to evaluate the risk in the organisation.

7.1.3. Phase 3: Structural Optimisation Using Risk Profile

Upon the conclusion of Phase 2, a node-level risk profile of the organisation being evaluated is obtained. This risk profile takes on the form of a tabulated list wherein the various metrics are listed for each of the nodes, along with a risk value for each of the nodes. The purpose of Phase 3 is to use the developed risk profile to identify possible relationships that, when either added or removed, should reduce the overall risk in the network.

The value of Phase 3 lies in the structural optimisation technique that is used to identify relationships that, had they existed, would have reduced the overall risk in the network. By identifying these relationships in this phase, it becomes possible to develop targeted risk mitigation strategies that aim to reduce the overall risk in the network. If this phase is not implemented correctly, however, it is likely that the risk mitigation strategies developed will

target the wrong nodes. If this happens, it may lead to a reduction in trust in the process, or even an increase in risk as high-risk nodes may be given even more power. It could also cause a level of frustration if, for example, a large number of individuals who already know each other are invited to a function designed to introduce strangers to one another.

The first problem to address lies in the concept of "overall risk". As the output of Phase 2 is a node-level risk profile, and not a network level one, the first step is to clearly define how an overall risk level would be determined. As with the node-level risk values described in Phase 2, a good way to determine the overall risk of the network would be to use a weighted summation of the risk values of the nodes in the network. One advantage of following this approach is that Equation 7.3 can be used as the basis for the calculation. As Equation 7.3 already includes weights that can be used to adjust the risk value of each individual node, the derivation of the equation for overall risk is very straightforward:

$$Overall\ risk\ value\ of\ network = \sum_{i=1}^{N}\left[C_i \sum_{m=1}^{k} w_m \left(\frac{v_{mi} - \min(V_m)}{\max(V_m) - \min(V_m)}\right)\right] \qquad (7.4)$$

where N is the total number of nodes in the network, and all of the remaining variables match those described for Equations 7.1, 7.2, and 7.3.

It is important, at this point, to take into consideration that it will be impossible to completely eliminate all of the risks in the network. As a result of this, a threshold value should be determined once the overall risk value has been calculated for the first time. The determination of the risk threshold is delayed up to this point for two reasons:

1. It reduces the likelihood that the risk threshold will be determined arbitrarily beforehand, which may make the threshold value unrealistically low or unnecessarily high.
2. The risk threshold is made relevant to the overall risk value of the network.

Ideally, in order to ensure that the second of these reasons is maintained, the threshold should be calculated from the initial overall risk value. One way is to define the risk threshold as a percentage value of the initial overall risk; doing so would also make reporting more meaningful, as the risk threshold can be described as being a goal wherein the risk is reduced by $x\%$. Deciding on the specific risk threshold, be it a constant value or as a percentage of the calculated overall risk, can be done in any number of ways. One way involves management setting a threshold value. This will likely be the method used initially, as it presents management with the opportunity to test the method using simple-to-define goals and outcomes. An alternative method, which may become increasingly necessary over time as a result of under- or overoptimistic management goals, makes use of historical data

to determine an "ideal" risk threshold. Such a threshold is one that is realistic, reasonable, and achievable, and results in an acceptable reduction in risk. The historical data can be obtained from previous implementations of the method, or from literature sources describing similar situations and by how much the risk was reduced.

With both the overall risk and risk threshold values determined, the network can now be evaluated for structural improvements. In order to identify these potential improvements, the network optimisation technique described in Chapter 6 can be applied to the network. This forms part of an iterative process that is continued until the required risk threshold is reached, or no more improvements can be made without destabilising the network. The overall structure of Phase 3 is shown in Figure 7.4.

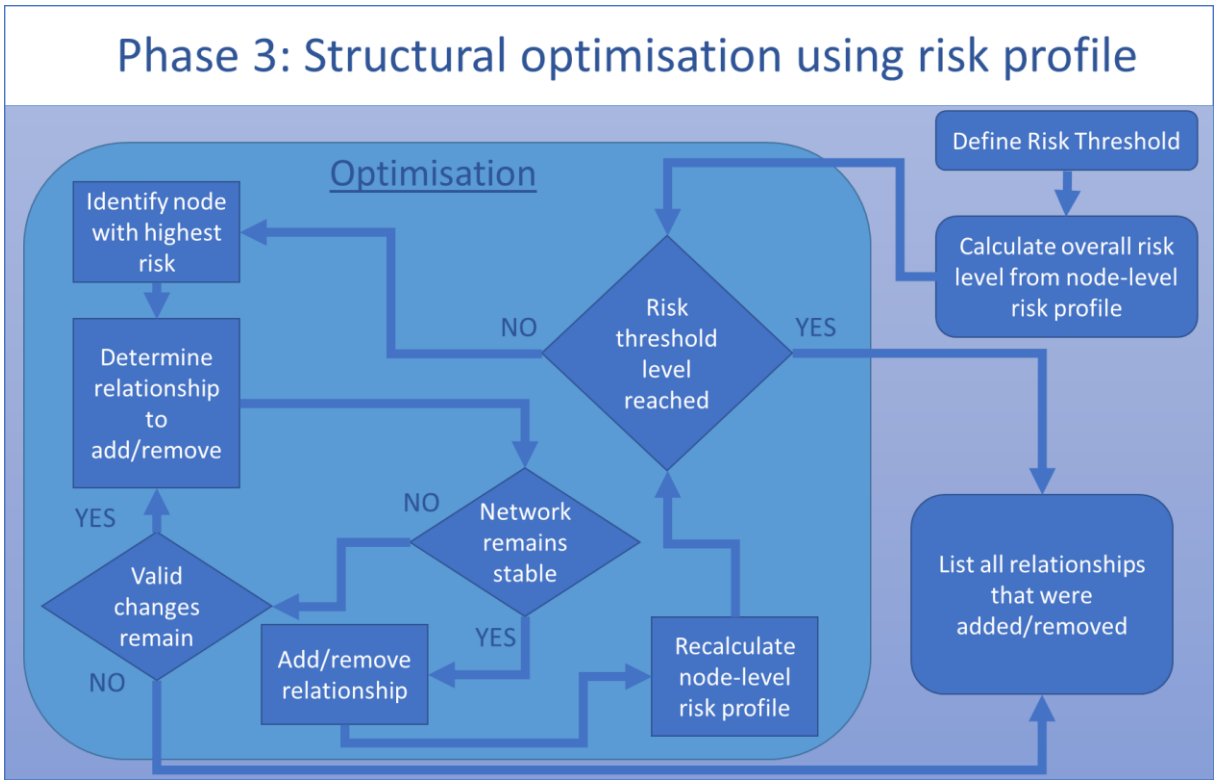


FIGURE 7.4: OVERVIEW OF PHASE 3

The core purpose of Phase 3 is to utilise the node-level risk profile developed during Phase 2 to ascertain where there is room for structural improvements in the network. Once these possible improvements have been identified, they can be used to develop risk mitigation strategies that may improve the overall risk in the organisation. A further side-effect of Phase 3 is that solutions to structural problems, such as border spanners, may be presented as a result.

7.1.4. Phase 4: Develop Risk Mitigation Strategies

Unlike Phases 1 – 3, which can be automated to a greater extent, Phase 4 requires some creative input. Once Phase 3 has been completed, a list of relationships that can be added or removed to improve the overall risk in the network becomes available. This list may contain extant relationships that need to be removed, such as those that give one node far greater access to confidential information than is necessary, or relationships that should be there, such as those between managers. In order to implement the changes these new relationships require, a number of methods can be used. A non-exhaustive list with regard to people includes such actions as promoting cooperation between project groups, reassigning personnel, hosting seminars and conferences, and organising teambuilding events. When resources are involved, an informative programme can be developed to help familiarise people with the targeted resources, office managers can be appointed to improve access to the resources, and the policies regarding resources can be improved, simplified and made readily accessible. If the targeted nodes are tasks, the people involved in the various tasks can be put into contact with one another. Alternatively, the location for the tasks' execution can be shared, or even the resources they have in common. The specific methods that can be used will be influenced heavily by the culture and policies of the organisation itself, which is why a creative touch is needed. Depending on the node identified and the methods chosen, the decision could even be made to accept the risk a particular node poses.

Once the specific nodes have been identified, and the methods have been chosen, the last step in Phase 4 should be to develop a risk mitigation plan (RMP). The RMP should identify the nodes, as well the methods used to target them. The plan also needs to list the methods according to a priority level determined using the current risk the associated nodes pose, based on the risk profile developed during Phase 2 and the overall risk determined at the beginning of Phase 3. The ideal RMP therefore contains a prioritised list of events, or actions, and the nodes involved. An RMP is developed so that the implementation phase can follow a clearly structured process. The RMP is also of value from a managerial perspective, as it can be used to budget for the various events and the plan can be assessed and approved by management before the implementation phase commences. The existence of an RMP should also help to secure the support of management, as the RMP condenses Phases 1 to 4 into a straight-forward document that effectively lists problems, solutions, and the possible advantages gained from implementing the solutions.

To reiterate: Phase 4 requires the creative input of an individual that understands the organisation's culture and policies well enough to suggest effective risk strategies. The purpose of the previous optimisation steps is to identify the people, tasks, and resources that should be targeted by the strategies to improve the overall risk experienced by the organisation.

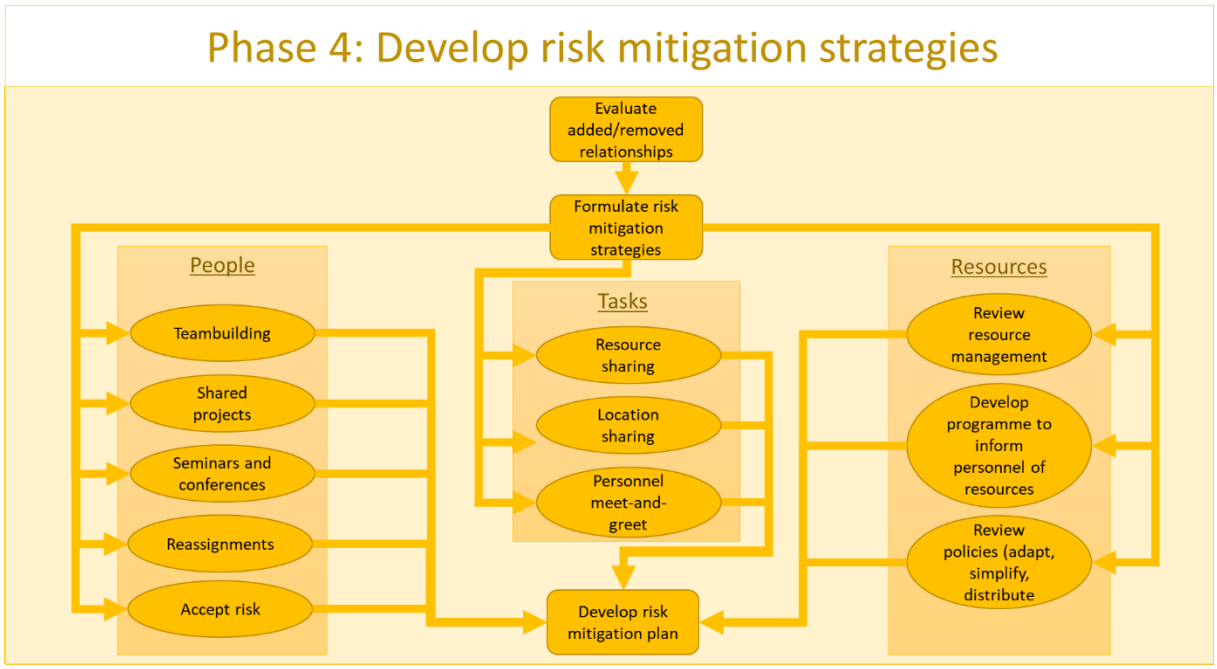


FIGURE 7.5: OVERVIEW OF PHASE 4

The purpose of Phase 4, in summary, is to use the structural improvements identified in Phase 3 to develop risk mitigation strategies. This will likely be a complex phase in practice, as a number of external factors need to be taken into account when selecting a risk mitigation strategy. These factors include the cost of implementing the strategy, the impact of the organisation's culture, time constraints, the locality of the nodes involved, etc. It is therefore possible that each time Phase 4 is reached, a different approach is used to select a risk mitigation strategy.

A final matter to consider with Phase 4 is that the process used to select the various risk mitigation strategies is not implicitly included in the framework. As a result, it is possible to use any technique or process that identifies ideal risk mitigation strategies without the overall process being affected, and each of these selection techniques can be evaluated in any number of ways. The specific method and measure used to determine how optimal the choice of risk mitigation strategy is will depend on the type of nodes targeted, as well as the type of strategy selected.

7.1.5. Phase 5: Implementation and Monitoring

During Phase 5, the RMP developed over the course of Phases 1 – 4 is implemented and the changes to the network is monitored. This is done so that the effects of the various chosen methods can be evaluated based not only on their efficacy, but also on their impact. If a

method based on this framework is implemented over a long term, it should prove valuable to take note of which methods are more or less effective given the circumstances and the people involved. The implementation part of Phase 5 therefore has two important functions: the first is to implement the RMP, whilst the second is to learn from the implementation process in order to improve future iterations.

The monitoring process, on the other hand, may make use of the Statistical Control Chart (SCC) method described in Chapter 3. This method makes use of SCCs to monitor the changes to certain specific SNA metrics in order to determine when, and if, certain statistically relevant changes have occurred in the network. While this may not seem immediately useful to formal networks that are based on organograms, significant changes in informal networks, that are more susceptible to seemingly random changes, are much easier to detect using the SCC method. This does however not mean that the SCC method is useless with formal networks. In the event of automated network monitoring, where a system is used to monitor the formal structure of the organisation for example, the SCC method remains useful as it can still be used to detect significant changes without having to monitor relationships individually. One example would be of project responsibility lists: the SCC method can be used to detect when certain targeted project managers begin to cooperate on shared projects. An alternative to the SCC method is to track the changes in the overall risk in the network over time. This is done by repeating the risk profile steps in Phases 2 and 3 after an action has been taken to modify the network, and then observing how and if the risk values change. This may potentially be simpler, but less instructive.

The starting point for the implementation of the RMP should be the selection of the event or action that has the highest priority. If possible, a number of events or actions can be selected that can run simultaneously. The events that are selected should however aim to target different groups, so that the efficacy of a single method on a group can be determined. Once the method and the associated group has been identified, a micro-network containing only the involved participants should be constructed using the data collected in Phase 1. This network will be used to calculate the normal SNA metrics used for the SCCs. This is done so that that minor, natural changes in the rest of the network do not affect the monitoring process.

After the action has been taken, or the event held, the micro-network is monitored over time by collecting new data focussed on only the micro-network's members. As soon as a significant change is detected, a report is compiled detailing the time taken for the change to take effect and how the change affected the network. This process is then repeated for each of the items in the RMP until the plan has either been implemented completely, or an administrative decision is made to reformulate the plan by returning to Phase 1. The overall structure of Phase 5 is shown in Figure 7.6. Regardless of how Phase 5 is concluded, the ideal implementation of a process based on this framework would involve a continual process: following the completion of Phase 5, Phase 1 is revisited, followed by Phases 2 – 5.

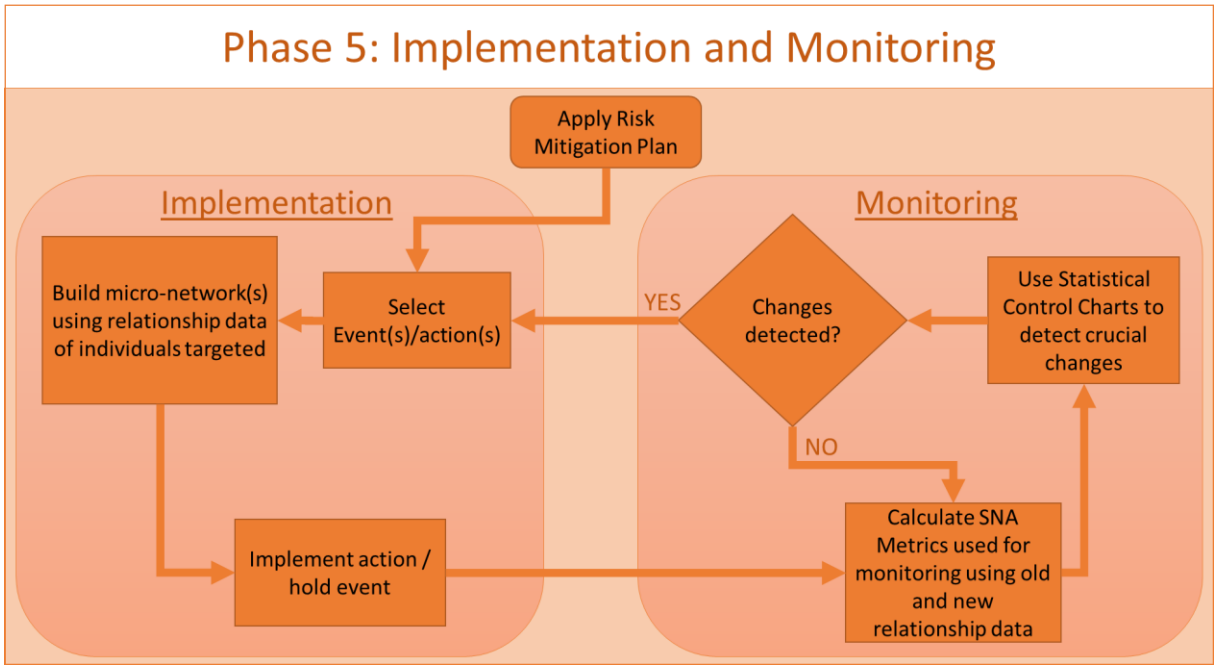


FIGURE 7.6: OVERVIEW OF PHASE 5

In conclusion, the entire framework progression is as follows: during Phase 1, data is collected that can be used to graph the relationships of an organisation. During Phase 2, the collected data is used to construct social networks that are then used to develop node-level risk profiles. Phase 3 then uses the node-level risk profiles to identify relationships that can be added or removed to address the risks identified in each node's risk profile. In Phase 4, the relationships identified in Phase 3 are used to produce a Risk Mitigation Plan that aims to either create or remove relationships within the context of the borders established during Phase 1. The final phase, Phase 5, features the implementation of the RMP. As the RMP is implemented, the organisation is monitored to ascertain how, and if, the desired changes take effect.

7.2. ILLUSTRATIVE EXAMPLE

In order to demonstrate how the framework can be applied to real-world data, an illustrative example using data collected from a real-world department at a university in South Africa will be presented in this section. Due to various logistical limitations and time constraints, only the first four phases will be applied to the data. The working of Phase 5, namely the Implementation and Monitoring phase, will be demonstrated using hypothetical scenarios based on the data collected. The data is referred to as the University Department (UD) data for the remainder of this section.

7.2.1. PHASE 1

Recall from section 7.1.1 that Phase 1 deals with the selection of the network type and data collection methods, as well as the actual collection of the network data using the selected methods. In the interest of completeness and clarity, the overview of the phase, first shown in Figure 7.2, is repeated in Figure 7.7.

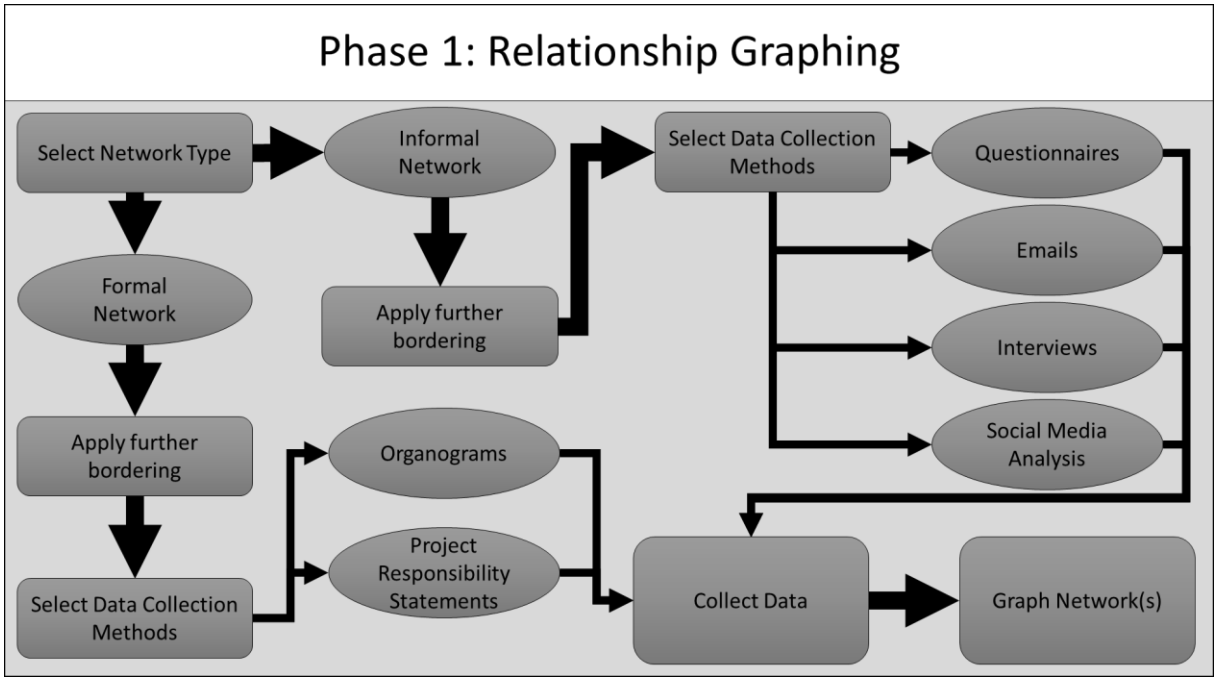


FIGURE 7.7: OVERVIEW OF PHASE 1

Select network type: The starting point of this phase is the selection of the type of network data to collect, which can be formal or informal, or both. For the UD network, **a formal network was selected**. The decision to select a formal network is informed primarily by two reasons. The first reason is that a significant amount of usable formal data is known to be available, and therefore the framework can be used without the need to collect informal data. The second reason is that the collection of informal network data would have necessitated the use of questionnaires. As there is a possibility that the people in the department would not respond to questionnaires in a timely manner, or indeed at all, the decision was made to not use informal network data at all.

It should be noted that the selection of a formal network type in this instance is informed by the impact the available data collection methods could have on the process. A certain amount of foresight was subsequently needed in order to expedite the process. This foresight may not necessarily be available, however: in certain complex organisations the

choice between formal and informal networks may depend on the additional bordering steps taken in the next step. If it should happen that one of the network types cannot realistically be used because infeasible data collection methods would be needed, the network type can be abandoned in favour of the other type.

Apply further bordering: The first step taken in bordering the UD network was to select formal relationships exclusively. The second step involves selecting which relationships to add. The UD, like many departments with office space, has a number of resources that are shared between the members of the department. These resources, such as a photocopier and kitchen, is excluded from the consideration, specifically because the decision was made to only evaluate relationships between individuals. For this same reason, certain shared tasks, such as administration, are also excluded.

Select data collection methods: With a formal network type with only individual relationships included, the selection of a data collection method depends on the formal data that was believed to be readily available. The data, which describes the formal relationships between study supervisors, lecturers, and post-graduate students, can be collected using statements provided by the various supervisors, as well as official class lists. As such, the data collection methods that are selected are the use of declared supervisor responsibilities with regard to post-graduate students, and the formal class lists. The statements are then **collected** from each of the members in the department, whilst the class lists are obtained from the university's administrative system. Honours-, masters- and doctoral level students are included, as well as those members of the department that do not have any post-graduate students.

Graph network(s): In order to develop the network graph, the data from the supervisor responsibility statements are aggregated into a single relationship matrix. This matrix is then used to produce the relationship graphs that are used in the subsequent phases. In order to ensure that the data is loaded correctly, the freeware suite Gephi is used to produce a Pajek data file from an adjacency matrix, which was then loaded into ORA-Lite. Gephi was used to load the data initially as its data modification features are more user friendly than ORA-Lite's, and the output format of the file increases the reliability of the loaded data. Two graphical representations of the graph of the network, developed using ORA-Lite, are shown in Figure 7.8 and Figure 7.9. Both of these graphs are shown to demonstrate how the readability of a visualised network can be improved merely by using a layout technique that sorts the nodes according to a particular metric. A circular layout, such as the one shown in Figure 7.9, also makes it easier to highlight specific changes in the network.

Summary of Phase 1: The outcome of Phase 1 is a network graph that can be used with SNA methods to evaluate the state of information security risk in an organisation. For this illustration of the framework, the organisation chosen was a department at a South African university, specifically with regard to the academic staff and post-graduate students. The details of the particular relationships were collected using class lists and supervisor

responsibility statements, and consolidated into a single adjacency matrix. A value of 1 was given for each instance of a relationship; if a relationship occurred in more than one way (e.g. a student is in a class presented by their study supervisor), the weight of the relationship was increased by 1 for every occurrence of that relationship.

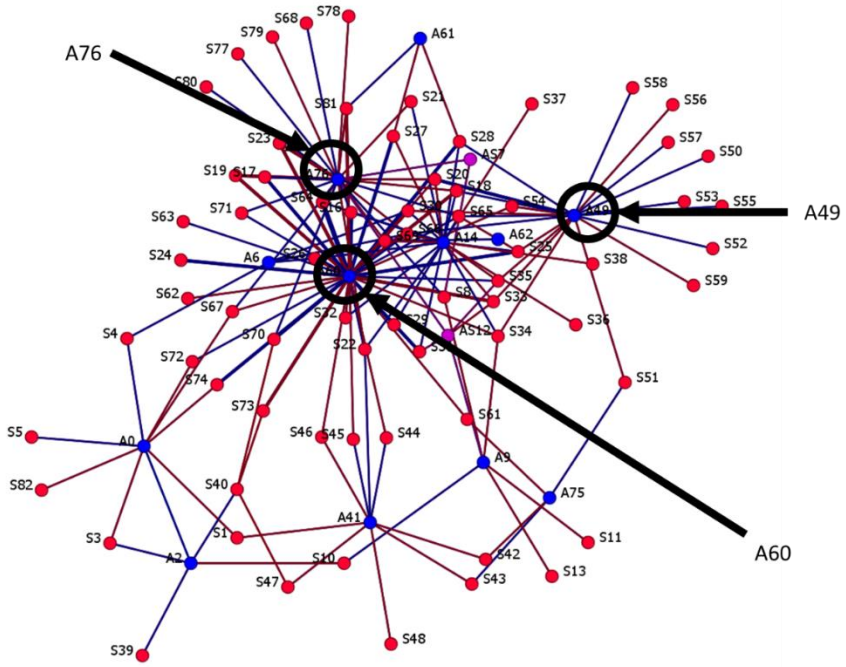


FIGURE 7.8: UD NETWORK WITHOUT ANY LAYOUT CHANGES. STUDENTS ARE COLOURED RED AND ACADEMIC STAFF BLUE. THE NODES IN PURPLE ARE POST GRADUATE STUDENTS IN ACADEMIC STAFF POSITIONS

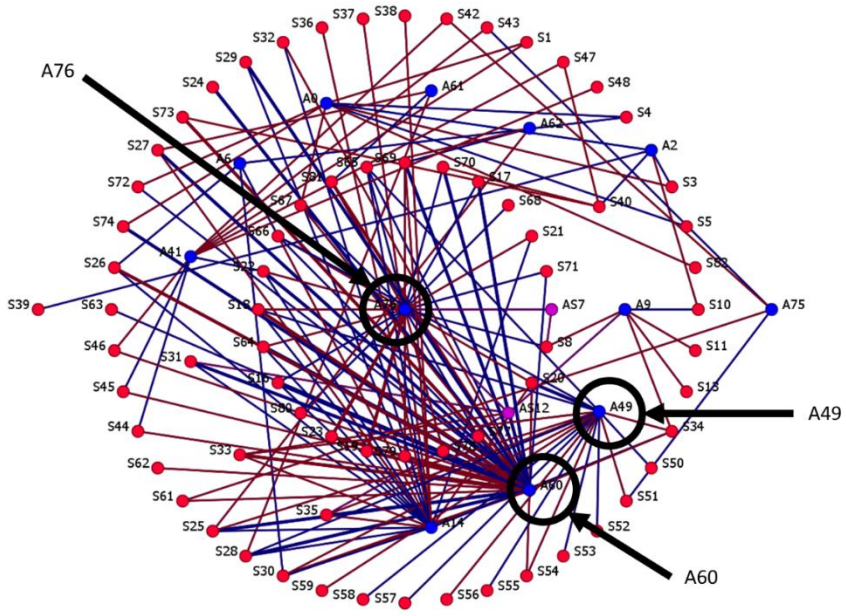


FIGURE 7.9: UD NETWORK WITH A CIRCULAR LAYOUT. THE NODES WITH THE HIGHEST BETWEENNESS ARE PLACED IN THE CENTRE

The data shown in Figure 7.8 and Figure 7.9 were anonymised before processing with Gephi was started. The meaning of the codes used is as follows:

- A node with a code that begins with “S” represents a student;
- “A” indicates a member of the academic staff; and
- Any node that contains both A and S is a member of the academic staff that is also pursuing a post-graduate qualification.

The network has 84 nodes, of which 12 are academic, 70 are students, and 2 fall into both categories. As shown in Figure 7.9, three members of the academic staff (A60, A49, and A76) had stronger relationships to the students. One of these nodes, namely A60, presented two subjects, and therefore any student who has both classes also has a stronger relationship. Following the completion of Phase 1, a graph of the network is now available. The following phase, which involves the development of a risk profile, can now commence.

7.2.2. PHASE 2

Once Phase 1 has been completed successfully, and a relationship graph has been obtained as a result, the focus can now shift to developing an information security risk profile. In the interest of ease, the overview of Phase 2 first shown in Figure 7.3 is presented again in Figure 7.10.

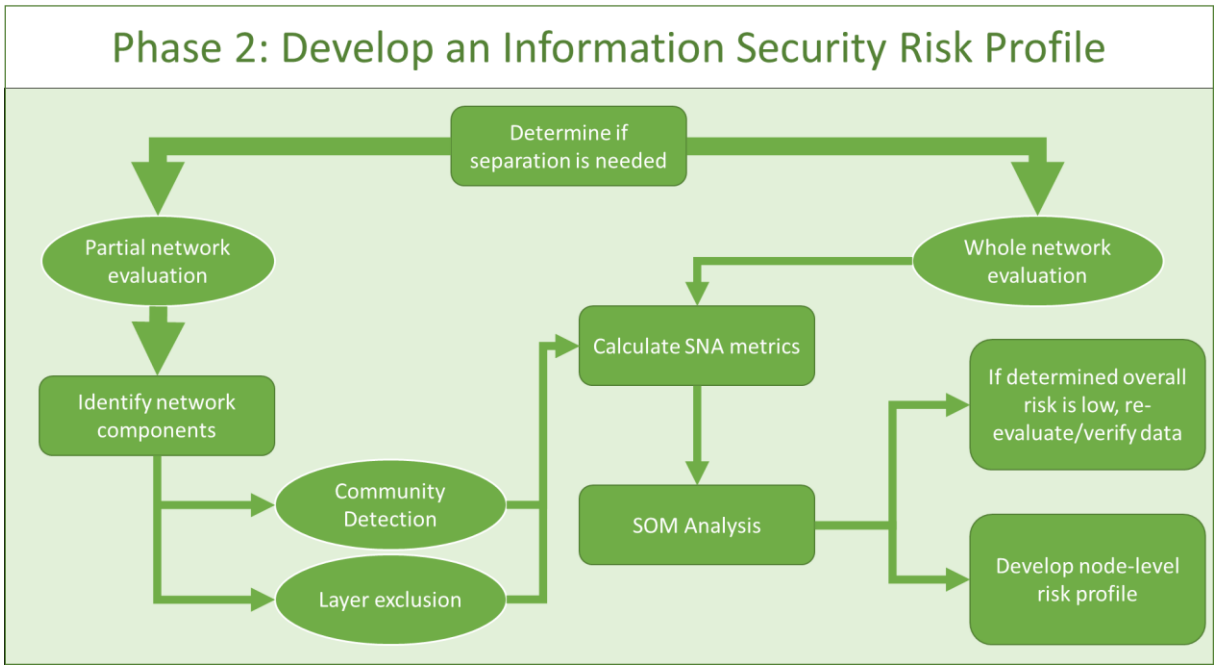


FIGURE 7.10: OVERVIEW OF PHASE 2

An important factor to keep in mind is that this phase has two possible outcomes. The first outcome, which may occur if there was a problem with the data gathering process, or if the true overall risk is low, is that the risk profile shows that there are no risks and that the remainder of the method is therefore unnecessary. In this instance, the data is verified by checking the process followed during Phase 1. If there are no inherent problems in the data gathering techniques, or how they were used, then the method ends at the conclusion of this phase. The second outcome is that risk that is identified in the network is presented as a node-level risk profile. This profile will then be used in further phases to develop risk mitigation strategies.

Determine if separation is needed/desired: The UD network likely features a number of communities, such as class groups, students studying under the same supervisor, etc. The network, due to its relatively small size, and the large number of identifiable subnetworks, is evaluated as a single network without any separation.

Calculate SNA metrics: The SNA metrics for this network are calculated using ORA-Lite. Seven metrics were selected, specifically betweenness centrality (BC), closeness centrality (CC), eccentricity centrality (EcC), eigenvector centrality (EiC), structural holes constraint (SHC), and total degree centrality (TDC). The resulting metrics are saved to a file, and all of the unscaled values are removed.

SOM Analysis: The maps for the SOM analysis is generated using Viscovery's SOMine software suite. The metrics for each node are added as attributes without any changes to the default settings. The file contains attributes for the 84 nodes in the network, so the SOM was set to use 84 nodes. The resulting maps are shown in Figure 7.11 - Figure 7.17.

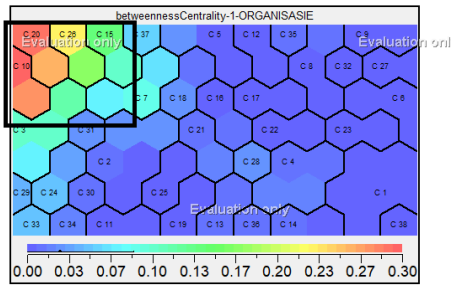


FIGURE 7.11: SOM BETWEENNESS CENTRALITY (BC)

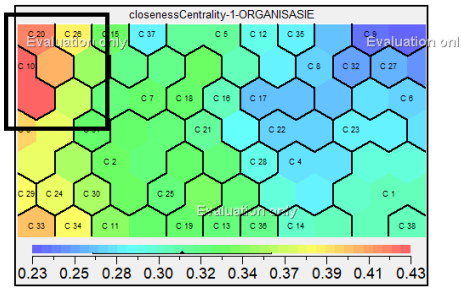


FIGURE 7.12: SOM CLOSENESS CENTRALITY (CC)

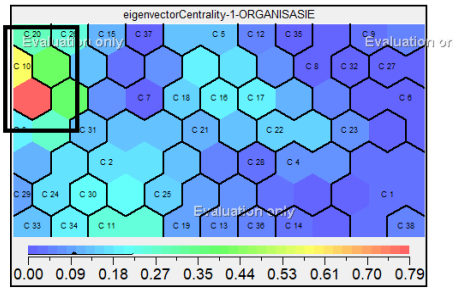


FIGURE 7.13: SOM EIGENVECTOR CENTRALITY (EIC)

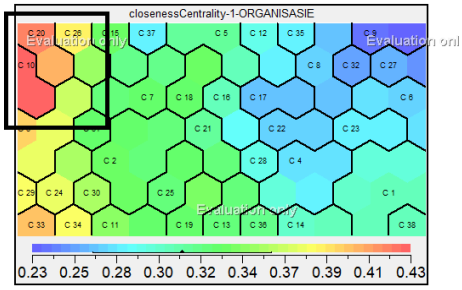


FIGURE 7.14: SOM CLOSENESS CENTRALITY (CC)

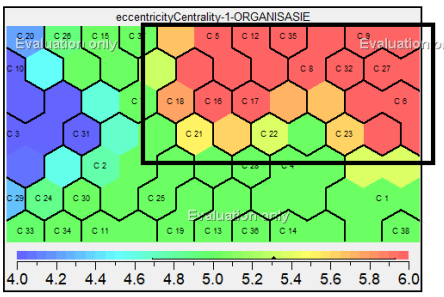


FIGURE 7.15: SOM ECCENTRICITY CENTRALITY (ECC)

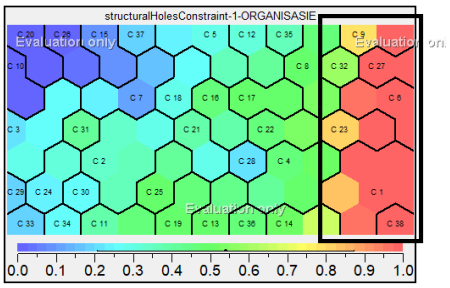


FIGURE 7.16: SOM STRUCTURAL HOLES CONSTRAINT (SHC)

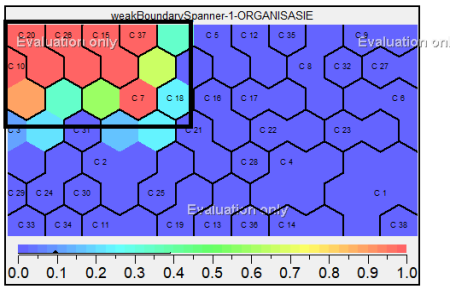


FIGURE 7.17: SOM BOUNDARY SPANNER (BS)

The SOMs presented above show that a number of risk "hotspots" are present in the network. These "hotspots", identified by the red colouration of a cluster, suggest that the remainder of the method may produce usable risk mitigation strategies that could improve the overall risk in the network. At this stage, the SOMs are not used for more in-depth analysis; they can however be used at a later stage to monitor and evaluate the effectiveness of the risk mitigation strategies by providing a point of comparison.

Develop node-level risk profile: As the SOM analysis demonstrated a certain level of risk in the network, a node-level risk profile that can be used in further phases is needed. The risk profile is developed by using the same metrics file generated using ORA-Lite during the *Calculate SNA metrics* step. The following data are added to the file: the minimum, maximum, and weight value for each metric, as well as the criticality, risk value, and weighted risk value for each of the nodes. This file now contains the node-level risk profile that will be used in subsequent phases. A part of this profile is presented in Table 7.2, and the full profile can be found in Appendix B.

TABLE 7.2: PORTION OF PHASE 2 RISK PROFILE

Node	BC	CC	EcC	EiC	SHC	TDC	BS	Criticality of Node	Node Risk value
DN7	0	0.309	5	0.028	0.714	0.024	0	2	5.787
N8	0.012	0.317	5	0.029	0.481	0.036	0	1	2.516
				.			.		
N37	0	0.283	6	0.026	1	0.012	0	1	4.339
N38	0	0.283	6	0.026	1	0.012	0	1	5.787
Metric MIN	0	0.213	4	0.005	0.031	0.012	0		
Metric MAX	0.303	0.439	6	0.88	1	0.446	1		
Metric Weight	1	1	2	1	2	1	1		

A number of weights are also assigned to certain nodes and metrics. As the supervisors are considered to be of greater importance in the network, every node with a supervisor designation is given a criticality weight of 2. As far as the metrics are concerned, both EcC and SHC are given a weight of 2. The motivation behind this decision is that EcC and SHC measures have the largest red areas on the analysis SOM, and should therefore be given priority.

7.2.3. PHASE 3

After the conclusion of Phases 1 and 2, a network graph and a node level risk profile, which is partially shown in Table 7.2, is available. In Phase 3, the structural optimisation technique discussed in Chapter 6 is used to identify relationships that may help to mitigate some of the risk in the network. The procedure for Phase 3, which was previously shown in Figure 7.4, is shown again in Figure 7.18.

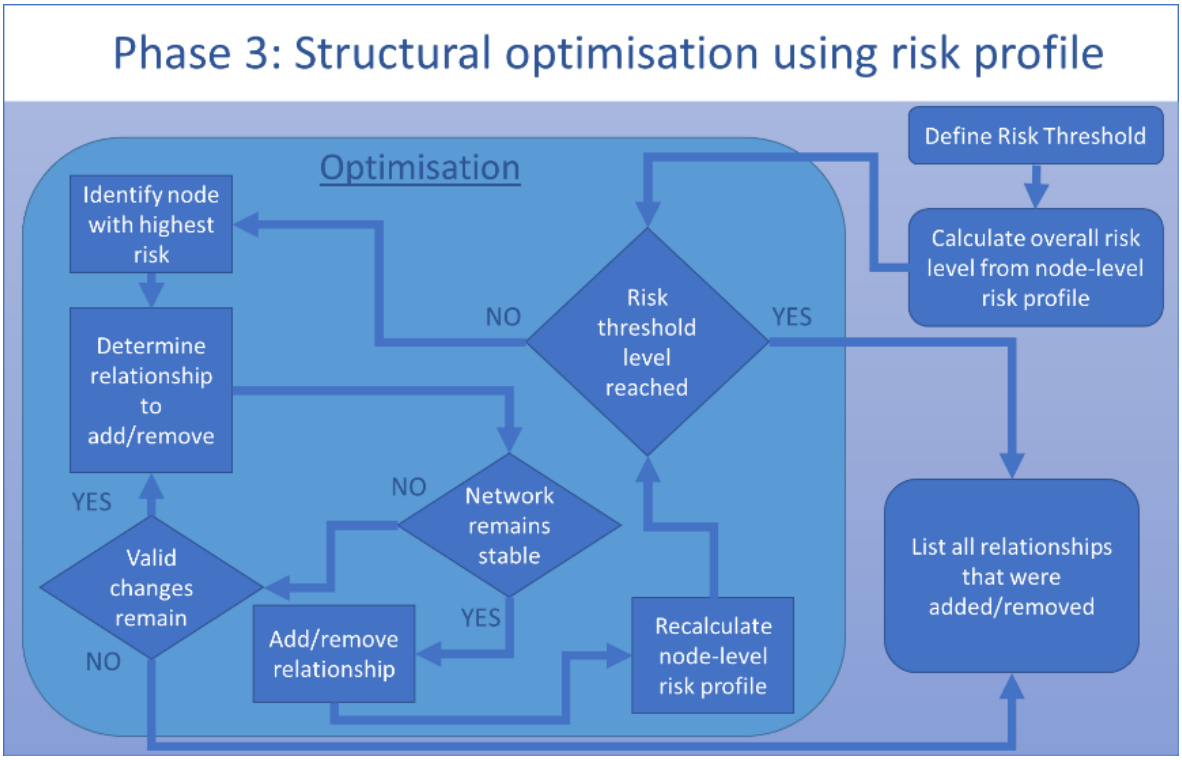


FIGURE 7.18: OVERVIEW OF PHASE 3

Define risk threshold: For the purposes of this illustration, a risk reduction of 20% of the overall risk is selected. This reduction is chosen arbitrarily in this instance as there are no policies or mandates that can be used to inform the decision.

Calculate overall risk level from node-level risk profile: Using the per node risk values in the risk profile (Table 7.2) as input variables for Equation 7.4, the overall risk in the network was calculated. By adding the weighted risk values of all the nodes together (as per Equation 7.4), an overall risk value of 313.77 is obtained. With the 20% reduction threshold decided on in the previous step, this means that the value for the risk threshold level is 251.

Structure optimisation process: The structure optimisation step is highly influenced by the nature of the relationships in the network. The nature of the department means that no relationships can be removed, primarily for two reasons: firstly, a number of the relationships exist because certain members of the academic staff are responsible for classes taken by a number of students. As the selection of subjects is left to the discretion of the students, it is not feasible to remove the relationships that exist by virtue of a student taking a class. The second reason is that the relationship between a post-graduate student and a study supervisor is fixed for the duration of the post-graduate course. The implication, therefore, is that the only measures that can be modified are those that are decreased when the relationships a node have is increased. Only two of the selected measures can be influenced mathematically by adding relationships to a node, and those metrics are EcC and SHC. The process used to find the new relationships, first introduced in Chapter 6, is:

1. Select the node with the highest risk.
2. Determine whether EcC or SHC has the higher value for that node.
3. Select the node with the lowest value for the metric selected in the previous step. If there are multiple options, select the node with the lowest overall risk. if a relationship between the high-risk node and the selected node already exists, select the next node that matches the criteria (i.e. lowest risk and lowest value for metric).
4. Use ORA-Lite to add the identified relationship to the network.
5. Recalculate the metrics.
6. Determine if the risk threshold has been reached.
7. Repeat steps 1 to 6 until the risk threshold has been reached.

A total of 13 new relationships are identified before the risk threshold is reached. The overall risk in the network, following the addition of these 13 new relationships, is calculated as 250.85. This represents a reduction in overall risk of 20.05%. The network, with the new links added, is shown in Figure 7.19.

While the inherent nature of the original network, where relationships stay in place until nodes are removed, means that relationships themselves cannot be removed, the same is not true for all networks. In network that does allow for the removal of links, such as in corporate management structures where employees can be reassigned at management's discretion, the optimisation process followed here would be slightly different. If relationships can both be added and removed, then there are no restrictions on the metrics that can be used to select the low risk nodes. The remainder of the process remains the same.

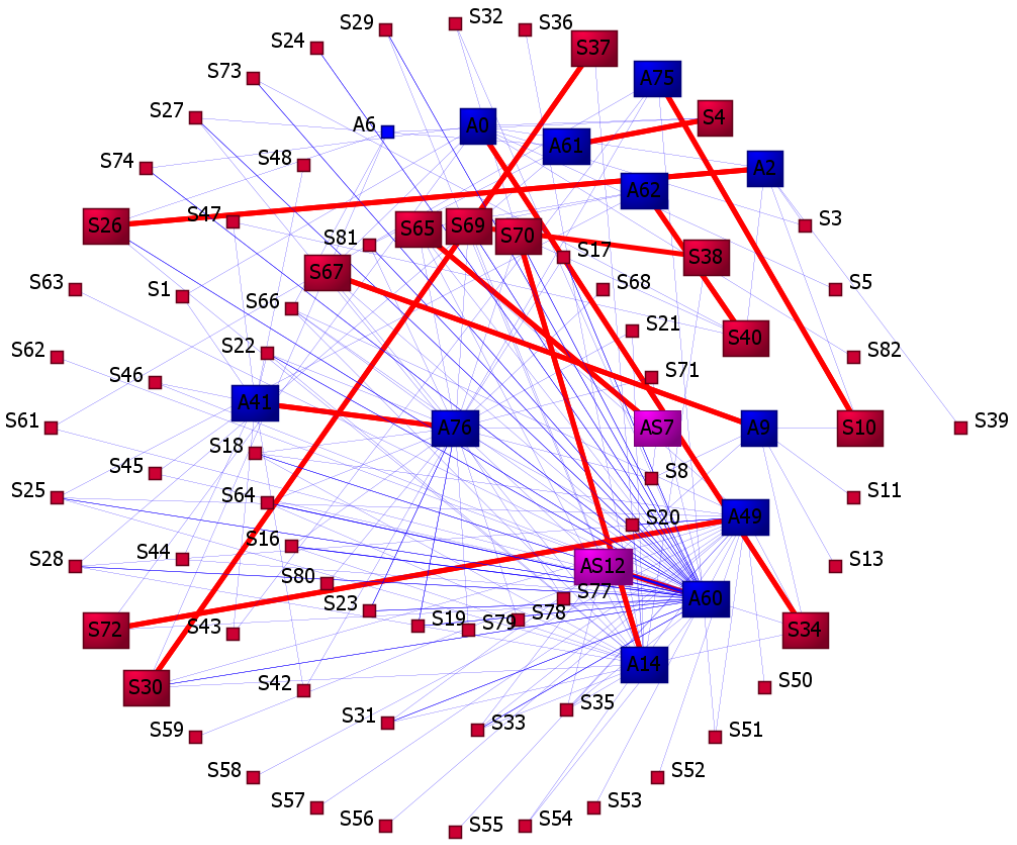


FIGURE 7.19: UD NETWORK AFTER STRUCTURAL OPTIMISATION. THE NODES THAT GAINED RELATIONSHIPS ARE SHOWN WITH A LARGE RECTANGULAR FORM; THE RELATIONSHIPS THAT WERE ADDED ARE HIGHLIGHTED. THIS GRAPH DOES NOT SHOW WEIGHTED LINKS.

The algorithm can also be used, in an adapted manner, to determine how relationships have to change if a node is suddenly removed. Consider, for instance, the possibility that a member of the academic staff in the UD network retires, becomes seriously ill, or dies. In this situation, the relationships with this node have to be changed, as supervisor-, class- and administrative responsibilities have to be reassigned. In such a situation, all of the nodes with a relationship to the now absent node are taken as the total collection of at-risk nodes, and no other nodes are considered. The algorithm is then applied in a mostly familiar way:

1. From among those nodes that have a relationship with the absent node, select the one with the highest risk.
2. Identify the metric that contributes the most to the risk value of the node selected in step 1.
3. Identify an appropriate node, e.g. a member of the academic staff, which has the lowest value for the metric identified in step 2.
4. Create a link between the node selected in 1 and the node identified in step 3. By connecting a low risk node to a high-risk node in this manner, the overall risk in the network should be kept to a minimum.
5. Remove the link between the node selected in step 1 and the absent node.

6. Repeat steps 1 to 5 for all of the nodes that had a relationship with the absent node, **but use the newer network with the new links (both added and removed) included and recalculate the metrics every time.**

Once this process has been repeated for all of the nodes that had a link to the absent node, the network with the new nodes should have a lower amount of risk, in addition to the responsibilities being reassigned in a balanced manner.

Compile list of all relationships that were added: The 13 identified relationships are shown in Table 7.3. As per the optimisation algorithm introduced in Chapter 6, a node is only considered once during the course of the optimisation process. There are therefore 26 nodes that are affected by the 13 new relationships. For each of the 26 nodes that were identified, the selection metrics EcC and SHC are shown in Table 7.3, as well as the risk value for each of the nodes. While only EcC and SHC are shown, the risk value is calculated using all of the metrics as input variables for Equation 7.3.

TABLE 7.3: NEW RELATIONSHIPS CREATED IN UD NETWORK. THE NEW RELATIONSHIP EDGES ARE INDICATED USING THEIR SOURCE NODE (HIGH RISK) AND DESTINATION NODE (LOW RISK)

Relationships found during structural optimisation							
Identified high risk node	**Source Node metrics before addition of new links**			**New relationship (Source -> Destination)**	**Source Node metrics following addition of new links**		
	EcC	***SHC***	**Risk**		***EcC***	***SHC***	**Risk**
A60	***4***	*0.031*	*9.954*	*A60 -> AS12*	*4*	*0.031*	*9.415*
A14	***5***	*0.043*	*8.960*	*A14 -> S70*	*4*	*0.047*	*7.333*
A49	***5***	*0.052*	*8.466*	*A49 -> S72*	*4*	*0.05*	*6.786*
A76	***4***	*0.047*	*7.935*	*A76 -> A41*	*4*	*0.048*	*8.077*
A0	***5***	*0.132*	*6.601*	*A0 -> S34*	*4*	*0.116*	*4.127*
A9	***5***	*0.167*	*6.483*	*A9 -> S67*	*4*	*0.143*	*4.436*
A62	***6***	*0.333*	*6.204*	*A62 -> S40*	*5*	*1*	*6.23*
A61	***6***	*0.333*	*6.120*	*A61 -> S4*	*4*	*0.5*	*1.496*
A2	***5***	*0.259*	*5.968*	*A2 -> S26*	*5*	*0.207*	*5.078*
AS7	***5***	*0.714*	*5.787*	*AS7 -> S65*	*5*	*0.529*	*5.51*
A75	***6***	*0.25*	*5.710*	*A75 -> S10*	*5*	*0.2*	*3.777*
S37	***6***	*1*	*4.339*	*S37 -> S30*	*5*	*0.612*	*2.665*
S38	***6***	*1*	*4.339*	*S38 -> S69*	*5*	*0.612*	*2.657*

Table 7.3 shows only those metrics that were used to select the nodes. As a total of seven metrics are used to calculate the risk values, the metrics shown in Table 7.3 may be the same after the addition of the new relationship. Also of note is that two nodes, namely A76 and A62, experience a marginal increase in risk with the new relationships. The increase experienced by these nodes is not a direct cause for concern, as the goal of the process is to reduce the overall risk in the network. In this instance, as the marginal increase in risk is nevertheless accompanied by an overall decrease in risk, it is allowable.

This list is used in the next phase to develop risk mitigation strategies. It is important to note that there is potential for further risk mitigation in the network, however, due to the established risk threshold, these risks may have to be addressed during a later iteration of the method.

7.2.4. PHASE 4

Phase 4 involves the development of risk mitigation strategies using the relationships list that was developed during the previous phase. The development of these strategies requires a certain degree of creative thinking, as certain relationship building techniques may not be appropriate, or indeed feasible, in some situations. It is important to note that, in an ideal situation, the strategies would be developed by personnel familiar with the corporate culture and knowledge of applicable "link-building" methods, i.e. those methods that can be used in practice to create relationships between nodes. As the skills and techniques required to develop such strategies effectively lie outside the scope of this study, and any strategies proposed here will likely be unsuitable for most applications, only the overall process is discussed. The overview of Phase 4's process is shown in Figure 7.5.

Evaluate added relationships: The relationships that are identified in Phase 3 can be placed into one of three groups: supervisor to supervisor, supervisor to student, or student to student. As the nodes targeted are all people, strategies that aim to create new relationships between people should be considered. If one of the targeted nodes had been a resource instead, such as a photocopier, strategies that attempt to connect nodes to that photocopier may be considered. Similarly, if a node is targeted that represents a task, then the strategies would have to be appropriate to the situation.

Formulate risk mitigation strategies: For the UD network, the nodes that are targeted in this phase all represent people. This means that strategies such as teambuilding events and shared projects, can be considered. In real-world situations, due to a number of factors, such as cost for example, management may place certain limitations on, or specify requirements for, the strategies that can be selected. This increases the complexity of the process, as an increasingly significant number of factors may have to be taken into account when selecting strategies.

In the event that the targeted nodes include tasks, the strategies that can be used differ from those used to create relationships between people, but may still be subject to similar restrictions or requirements. Consider, for instance, a task that has to be connected to a person. One of the strategies that can be employed is to assign that task to the person. This may, however, not be feasible if the person is already assigned to a large number of other tasks. In order to select the correct strategy, it is therefore clear that a working knowledge

of the network is needed. Additionally, a strategy that works on one occasion may not work a second time, and identifying which strategies will work can be a complex process.

The third type of node that can be targeted represent resources. Connections to these nodes, much like those discussed earlier, require a working knowledge of how resources are used in the network in order to develop effective strategies. If a resource has certain access limitations, such as locked storage rooms or password protected photocopiers, it is important to know why those limitations are in place. If a strategy is proposed that would violate policy if it were to be implemented, for example, the strategy is not feasible. Furthermore, if a node does not have access to a resource due to certain historic reasons specific to that node, then more creative solutions may be required in order to create a new relationship between the person and the resource.

Develop risk mitigation plan: Once all of the necessary, and possibly feasible, strategies have been selected, they should be consolidated into a single Risk Mitigation Plan (RMP), which can be presented to management if needed. The RMP is then used in the final phase to actively create or remove the relationships that should lower the overall risk in the network.

In conclusion: the purpose of Phase 4 is to develop risk mitigation strategies using the relationships identified in Phase 3 as a guide. This phase requires a creative hand in order to be implemented effectively. As previously stated, it is vital that the individual(s) that develop the risk mitigation strategies are familiar with the nuances, culture, and nature of the network for which the mitigation strategies are developed. Furthermore, the ideal person(s) would also have the skills or training needed to develop effective risk mitigation strategies.

7.2.5. PHASE 5

The final phase deals with the implementation of the strategies developed during Phase 4, as well as the monitoring process used to verify that the desired changes occur. No workable strategies specific to the organisation were presented during Phase 4, as the development of such strategies is a complex affair that is likely to differ from organisation to organisation. As a result of this complexity, as well as the likelihood that most organisations will approach it in vastly differing manners and subsequently develop unique solutions, the development of such strategies lie outside the scope of this study. Therefore, a demonstration using the network data is presented to show how the network may have changed, if strategies meant to create relationships in the network had been developed, and these strategies were to be implemented. In this demonstration all of the changes suggested during Phase 3 are incorporated, i.e. the demonstration functions as if all of the

identified relationships end up coming into existence. The overall structure of Phase 5 is shown in Figure 7.6.

Apply risk mitigation plan: At the conclusion of Phase 4, in practice, a risk mitigation plan (RMP) would be available that describes the strategies that can be used to create the new relationships in the network. This RMP would have the intended goal of creating a total of 13 new relationships in the network. These new relationships involve 26 separate nodes, specifically the 26 nodes named in Table 7.3. In this instance, the impact of all 13 of these relationships coming into existence over the course of several weeks is shown. In order to illustrate the creation of the new relationships in a more natural manner, the relationships shown in Table 7.3 are randomly placed into one of five groups, and each of these groups of relationships is then individually added to the original network first introduced in Section 7.2.1. The intention is to demonstrate how the network changes over time, if all of the strategies are implemented simultaneously and new network data is collected once every few weeks, or months. This illustration is therefore intended to feature a network that experiences significant changes between observations. Each of these observations, i.e. points at which new total degree centrality (TDC) values are calculated, are referred to as "observations". After the addition of all of the relationships in a group, the TDC for the nodes in the network, with the new relationships added, is calculated. The TDC for all of the nodes are then added together, to obtain a total TDC value for the entire network. These TDC values are then used as inputs for the Statistical Control Chart monitoring method.

Select events and actions: In practice, either management or the entities responsible for implementing the strategies proposed in the RMP, may choose to only select certain events or actions for implementation at any given time. This selection of specific strategies may be as a result of financial considerations, as certain strategies may be very costly to implement, or in the interest of determining which strategies are the most effective given a certain group of targeted nodes. Here, the impact of simultaneously implementing the strategies that are intended to create all of the new relationships is shown.

Build micro-networks using relationship data of individuals targeted: In order to properly monitor the events that are meant to create the new relationships, and ensure that these new relationships are indeed created, a micro-network that shows all of the extant relationships between the 26 targeted nodes is created. By creating such a micro-network, any natural fluctuations in the rest of the network is eliminated from the monitoring process. This allows for the nodes that are targeted by the strategies to be monitored with a degree of isolation, which should also help to study natural trends amongst the targeted nodes over time. As this illustration shows what the implementation process looks like when all 13 of the new relationships are targeted for creation simultaneously, all 26 of the targeted nodes, as well as their relationships to one another, are included in a single micro-network. If the implementation process had featured the selection of only a subset of the relationships, for instance to determine how effective a specific strategy is, then the micro-

network would only have contained the nodes that would be affected by the creation of the new relationships targeted by the selected subset of strategies. The initial relationships between the nodes in the micro-network are those relationships that are already present prior to the implementation process. The resultant micro-network, which is smaller than the full network as it only contains 26 nodes, should be easier to monitor and evaluate. Furthermore, as the monitoring process evaluates the network over time, the use of this micro-network should reduce or even eliminate the impact that any possible changes in the rest of the network may have on the 26 nodes that are targeted by the RMP, increasing the effectiveness of the SCC monitoring method. The micro-network is shown graphically in Figure 7.20.

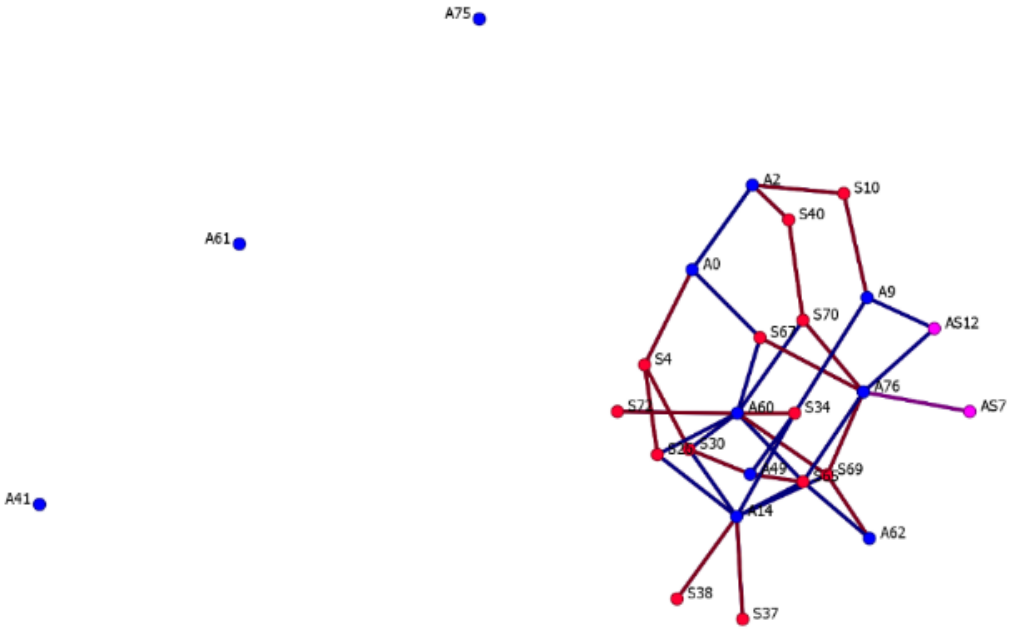


FIGURE 7.20: MICRO-NETWORK CONTAINING NODES IDENTIFIED IN PHASE 3. THE RELATIONSHIPS BETWEEN THESE NODES ARE INCLUDED. STUDENTS ARE COLOURED RED, ACADEMIC STAFF BLUE, AND THE NODES IN PURPLE ARE POST GRADUATE STUDENTS IN ACADEMIC STAFF POSITIONS.

Implement actions: In this instance, the creation of all of the new relationships will be attempted simultaneously. This also means that the micro-network containing all 26 targeted nodes can be monitored as a singular entity. If different phases had been proposed to target a subset of nodes, then each subset would have been monitored as part of a separate micro-network.

MONITORING: The monitoring sub-phase will likely encompass a continuous process conducted over several weeks or months. In order to monitor the effects of various actions over time, new data is collected every several weeks in order to keep track of changes in the

network. The micro-network is re-graphed every three weeks over the course of five months, for a total of six observations.

Calculate SNA metrics used for monitoring: The purpose of the monitoring phase is to detect significant changes in the network. In larger networks, there is a possibility that even the relationships between the nodes in the micro-networks change and fluctuate over time. In order to detect when there are significant changes in the network, statistical control charts (SCCs) are used. As demonstrated by McCulloh (2009), the natural fluctuations in a network can be statistically differentiated from sudden and unexpected changes, or "shocks", as the natural fluctuations tend to follow a statistically regular pattern. This means that, when SCCs are used, the natural fluctuations are those that lie within a statistical control area, and shocks are those that lie outside the area. The bounds for this statistical control area are given by the equation

$$Threshold\ values = \bar{p}\ \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\ \ ,where \tag{7.5}$$

$$\bar{p} = \frac{total\ value\ of\ deviations}{total\ of\ all\ metric\ values}\ \ , \tag{7.6}$$

and n is the average value per observation. The value for the lower threshold is 0 if the calculated value is negative. In this demonstration the total degree centrality metric is used to detect the changes. McCulloh, however, demonstrates how network denisity and the cumulative sum of betweenness centrality (BC) over time can be used to detect changes in a network. In practice, it may be possible to select the SNA metric used to monitor changes in the network based on the nature of the relationships that need to be created. Certain metrics, such as the cumulative sum of BC, may be more appropriate than TDC to detect changes in networks when relationships can both be added and removed.

The type of SCC used here, as mentioned previously, requires at least two sets of values, where one set represents the norm and the other the deviation from the norm. When this type of SCC is applied to quality control processes, the one set contains the number of units produced, whilst the other contains the number of defective units produced. When applied to social networks, where there are no defective or normal units, the change in network metrics is used in lieu of a "defective value", and the total network metrics are used as the "production" values. In a normal network, the interaction between these two values is such that they are statistically controlled, i.e. all of the deviation values fall within the control area. A SCC of a network in a natural state of fluctuation will therefore look like the one in Chart 7.1. As the number of available observations for a network increases, the accuracy of

the SCC method may also increase, due to the availability of a larger number of non-shock network states.

As the intention with the application of the SCC method is to detect if, and when, the risk reducing changes occur, the SCC method may not be appropriate for all networks or network data. In these cases, where the network is subject to a certain amount of constant, significant change, the SOM technique may be more appropriate to monitor change in the network. As the SOM technique clusters nodes based on a wide range of metrics, it should be possible to determine if a node is still a high risk by determining if the other nodes that are in the same cluster are also high-risk nodes. If the composition of the cluster changes significantly, it could indicate that the fundamental structure of the network changed, even in light of its unstable, dynamic nature. Even so, if the network is too unstable over a significant period of time, it may be impossible to meaningfully monitor the network over that period of time. If the intention is to attempt to stabilise such an unstable network by creating risk reducing links, then the monitoring phase is only used to determine when, and if, the network stabilises. Once the network has stabilised, the SCC method can again be used to detect targeted changes, which should begin to present as shocks on an SCC.

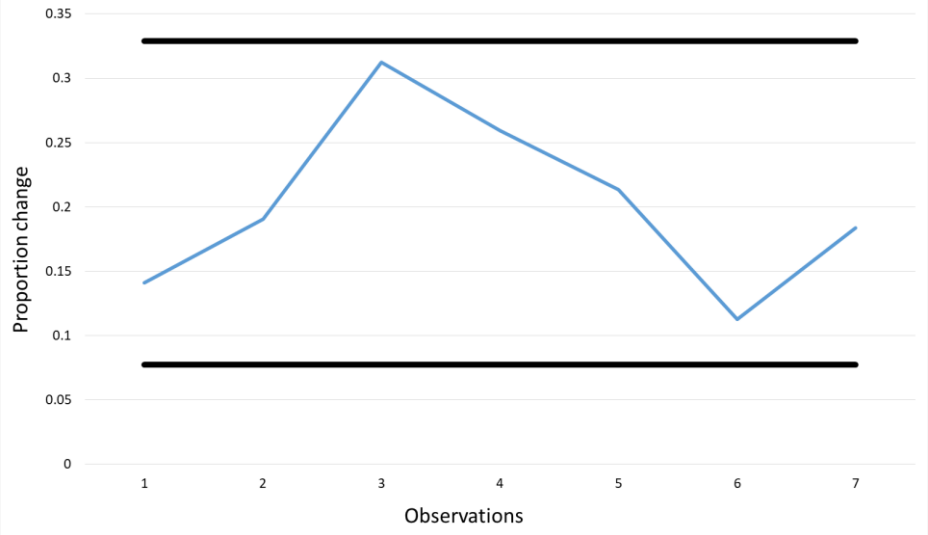


CHART 7.1: EXAMPLE SCC OF A NETWORK IN A NATURAL STATE OF FLUX, WITH NO SHOCKS.

Because no relationships are removed and the aim of the monitoring process is to detect new and significant relationships, the change in the sum of the total degree centrality ($\sum \mathrm{TDC}$) of all the nodes in the micro-network is used to measure overall network change. The $\sum \mathrm{TDC}$ is first calculated for the original network, which is used as the first observation. For the remainder of the observations, groups of relationships are created in the network between each of the observations, which means that the network is "shocked" between each observation. The $\sum \mathrm{TDC}$ is then recalculated after each observation, in order to

determine the measurable state of the network at that point in time. A change in the $\sum \text{TDC}$ since the previous observation is then calculated, and this change is used as the observation's deviation value. In order to ensure that the values are scaled to the network every time, the proportion change in the $\sum \text{TDC}$ is calculated. This proportion change in $\sum \text{TDC}$ can be calculated as:

$$Proportion\ change = \frac{|\sum TDC_2 - \sum TDC_1|}{\sum TDC_2} \tag{7.7}$$

where $\sum TDC_2$ is the sum total of the TDC of all nodes in the current observation, and $\sum TDC_1$ is the sum total of the TDC of all nodes in the previous observation. The $\sum \text{TDC}$ for the original micro-network, and the six simulated observations, is shown in Table 7.4.

TABLE 7.4: MONITORING DATA FOR THE UD MICRO-NETWORK DURING THE COURSE OF THE 5TH PHASE

Observation	$\sum TDC$	Proportion change in $\sum TDC$ since previous observation
Original	74	0
1st	74	0
2nd	80	0.075
3rd	162	0.506
4th	88	0.457
5th	190	0.537
Final	198	0.042

Note that, in this instance, only the creation of the 13 targeted relationships occurred, and no natural fluctuations were observed. In a real-world network it may happen that there is constant, gradual change as relationships grow stronger and weaker, or relationships come into existence or are removed. This fluctuation in the network may also be impacted by the techniques used to collect the network data, as certain techniques, such as questionnaires, may provide network data that, while representative of a node's state at a specific point in time, may not be an accurate representation of the node's state in general.

Use statistical control charts to detect crucial changes: The use of statistical control charts (SCCs) aids in the process of separating significant changes from natural fluctuations in the network. Therefore, in an ideal situation, network data would be available for periods before the intended changes to the network are implemented, so that the statistical patterns of the organisation's network can be better identified. This data, which would describe the network and network changes over the course of months or years before

monitoring is started, was not available for the UD network. This is due to the organisation itself not being monitored from a SNA perspective at any point prior to the UD network, first presented in Section 7.2.1, being graphed. In this instance, as such long-term network data are not available, historical, natural trends in network fluctuations cannot be identified easily. The SCC for the UD micro-network, based on the data presented in Table 7.4, is shown in Chart 7.2. This SCC presents the proportion change in $\sum \mathrm{TDC}$ on the *y*-axis for each of the observations shown in Table 7.4. By applying Equations 7.5 and 7.6 to the data in Table 7.4, the statistical control borders, i.e. the thresholds that separate natural fluctuations from potential shocks, were calculated as being at 0.189 and 0.439.

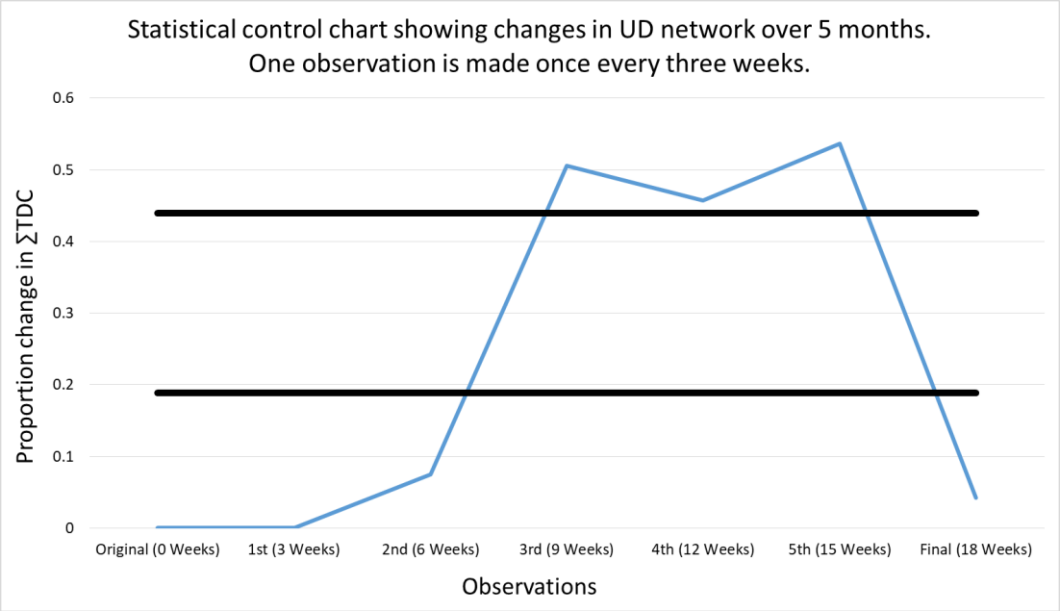


CHART 7.2: SCC FOR UD MICRO-NETWORK

Chart 7.2 shows that all of the changes that occurred in the UD micro-network are shocks, as none of the observed changes in $\sum \mathrm{TDC}$ are between the upper and lower bounds. The chart further indicates that there was no change in the network for the first three weeks, i.e. between the "Original" and "1st" observations, but that this was followed by significant changes during the following six weeks, i.e. between the "1st" and "3rd" observations. This suggests that the UD network is very stable, and that it is likely to have a lower bound SCC value of close to 0, due to the complete absence of changes during this period. Finally, during the final three weeks, there was a significant downward trend in changes to the network, as the final relationships were created. The order in which the relationships were added to the network in this demonstration does not match the order wherein they were identified in Phase 3 (Table 7.3), as it is unlikely that a real-world network in similar conditions would produce new relationship in such a specified order. If each relationship had been targeted individually and consecutively, this may however have been possible.

However, as it would take a very long time and have significant cost implications to add the relationships one by one, it is unlikely that a real-world implementation would involve creating the relationships individually.

The SCC also shows that there was consistent, significant change in the network starting with the second observation. If the trend had fallen back into the control area, or below the lower bound, it may have indicated that there was stagnation in the implementation process. This would, in turn, have prompted investigation into why the process had become stagnant.

Changes detected: In total, five significant changes (i.e. outside of the bounds of the SCC) were detected in the network. There were no changes detected during the "Original" and 1st observations so, while these observations do lie outside of the control area, they are not considered relevant shocks in this instance. As no natural fluctuations were observed, only the 13 shock relationships were added to the network, and there were changes between all but the first and second observations, this matches the expected outcome. To put it another way: one would expect, based on the work done by McCulloh (2009), that, if only shock relationships are added to a network, there would be no observed points within the statistical control borders of a SCC that are created using the changes in that network as data.

Using only the SCC in Chart 7.2, it is not possible to distinguish between an outcome wherein all of the relationships are created, and an outcome wherein only some of the relationships are created and the implementation progress becomes stagnant. In order to ascertain if all of the desired relationships have been created, the network graph should be investigated. This investigation is made easier by the use of a micro-network, as the smaller micro-network can potentially be evaluated graphically. This reduces the work that needs to be done in order to verify that all of the desired relationship have been created. If the SCC indicates that no shock relationships are added to the network over a significant period of time, i.e. the observed values fall within the statistical control borders, and some of the relationships still need to be created, then there may be a problem with the actions used to create the new relationships. In such a situation, the RMP should be reconsidered, and potentially revised, for the relevant relationships.

7.2.6. Summary of Illustrative Example

In Section 7.2 an application of the novel framework is applied to a small real-world network. The data for the network was obtained from a department at a university, and describes the relationships between the academic staff and students. This network is referred to as the UD network. In Phase 1, the data for the network is collected using formal

class lists, as well as statements from study supervisors. The data is then used to graph a network containing 84 nodes, of which 12 are academic personnel, 70 are students, and 2 are both students and academic personnel. In Phase 2, the network metrics for the UD network are calculated, and a SOM analysis is used to determine if there are risks in the network. A cursory investigation of the risk profile shows that certain nodes, such as A60 and A14, pose a high risk. A number of areas on the SOM have high values for the risk nodes, indicating that there are nodes in the network that pose a possible risk. The SOM is also used to identify metrics that may benefit from greater attention, i.e. to determine if there are metrics that should receive a greater weight. The metrics eccentricity centrality and structural holes constraint are given a greater weight due to their overall high levels in the network. The members of the academic staff are also given higher criticality weights than the students. At the conclusion of Phase 2, a node-level risk profile is available that contains both the metric- and risk values for each node.

In Phase 3, the node-level risk profile is used to identify 13 new relationships that would reduce the overall risk in the network by 20% if they were to come into existence. These 13 new relationships are identified using the optimisation method first discussed in Chapter 6. In Phase 4, the various approaches that can be used to develop risk mitigation strategies are discussed, as the development of specific strategies is a very complex process unique to each organisation. Some options, such as shared projects and teambuilding, are presented as examples of strategies that can be used to create new relationships. In the final Phase, a demonstration is presented to show how the changes in the network would be monitored in a real-world implementation. The addition of the 13 new relationships is shown over the course of 5 months, and the SCC method is used to detect when the new relationships are added. While no shocks are detected during the first three weeks, the network changes significantly from there on.

In conclusion, all five of the phases of the framework are illustrated using the UD network. With long-term, real-world applications of the framework, this process is likely to be iterative, and the various phases can be applied to different sub-networks at different times, depending on the specific network and the minutia of the particular implementation.

7.3. LARGE DATASET EXAMPLE

The network used to demonstrate the framework in Section 7.2 contains 84 nodes. While a network of this size is ideally suited to demonstrate how the framework functions when applied to data, it does not inherently indicate how scalable the technique is. In order to show that the framework can indeed be scaled to larger networks, consider the network shown in Figure 7.21. This network, which is one of the datasets provided by the Centre for

Computational Analysis of Social and Organisational Systems (CASOS) at Carnegie Melon University, was developed as part of the NATO Trident Juncture exercise in 2016. The network shows the social media response on Twitter to Trident Juncture. This network contains 1363 nodes, of which 611 are individuals, 571 are Tweets, 132 are topics, and 49 are locations. The network was selected because it is a large, natural, real-world network that contains a number of structures and communities similar to what is expected in a large organisation's informal network. The communities are more clearly seen in the 3D version shown in Figure 7.22. Sadly, due to the size of the network, Figure 7.22 does not contain enough detail to identify individual nodes with specificity, and only serves to show that there may be communities in the network.

As seen in Figure 7.22, there are a seemingly large number of disconnected dyads and triads in the network, which greatly increases the overall boundary spanner (BS) metric for the network. This may have to be taken into account by giving the BS metric a reduced weight, as a high BS value is typical in the network and may not be indicative of increased risk.

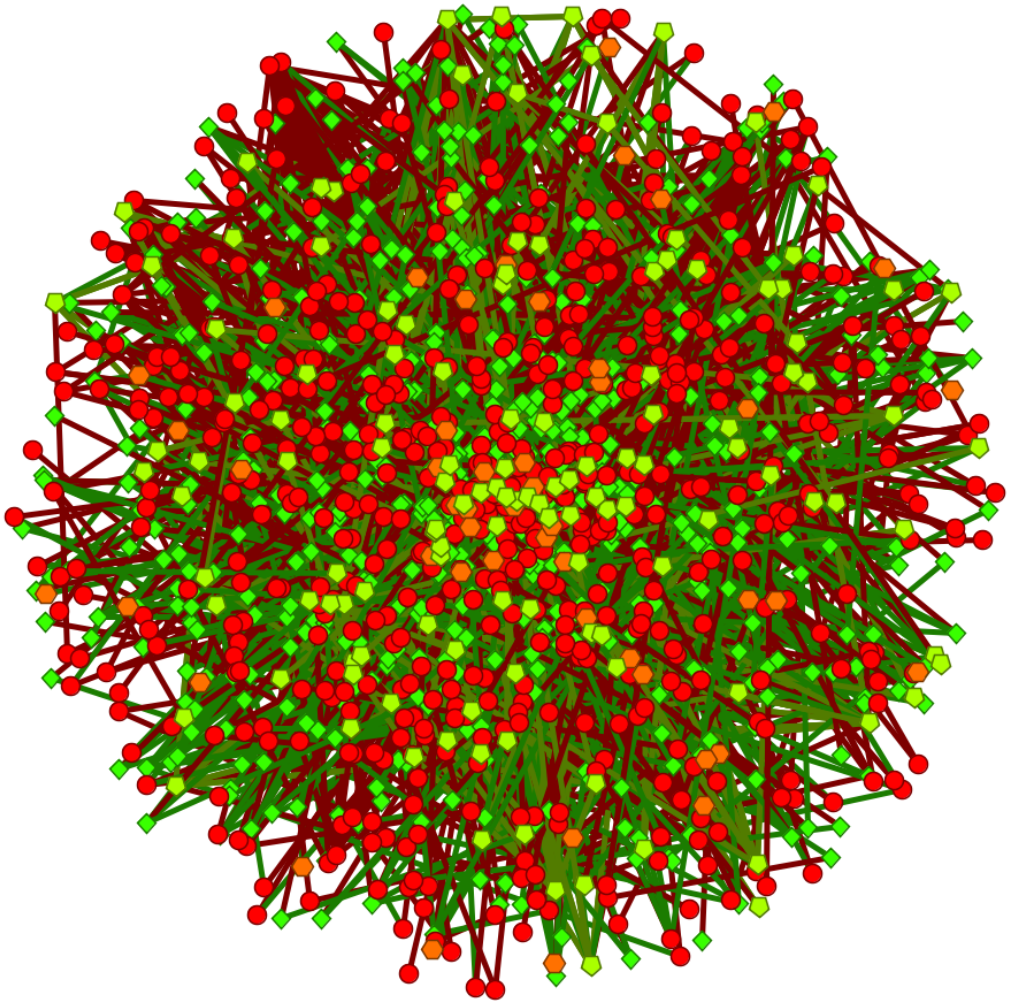


FIGURE 7.21: TRIDENT JUNCTURE TWITTER NETWORK (FRANKENSTEIN *ET AL.*, 2016). THE CIRCULAR RED NODES ARE INDIVIDUALS, THE GREEN DAMONDS ARE TWEETS, THE LIGHT GREEN PENTAGONS ARE TOPICS, AND THE ORANGE HAXAGONS ARE LOCATIONS

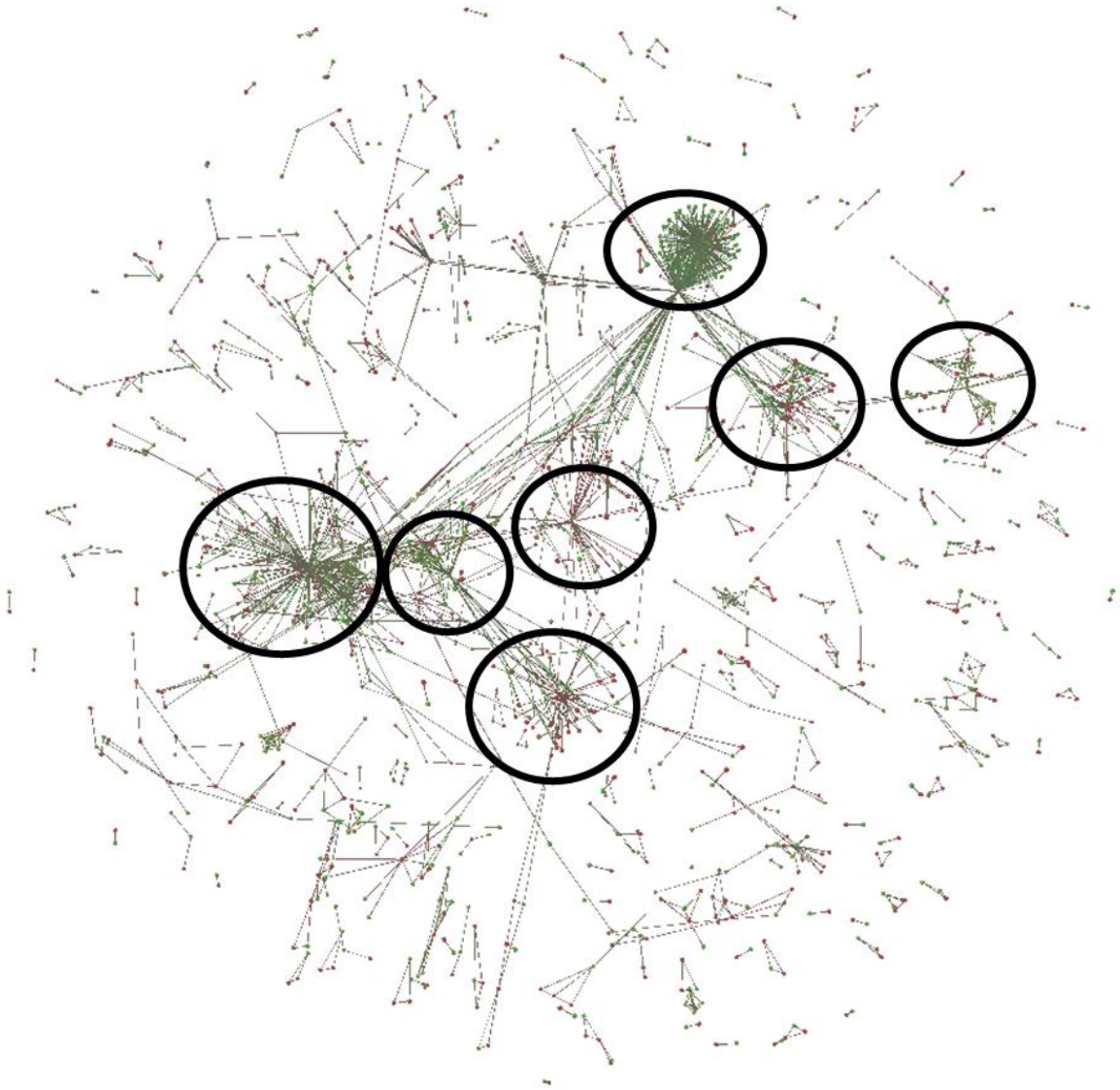
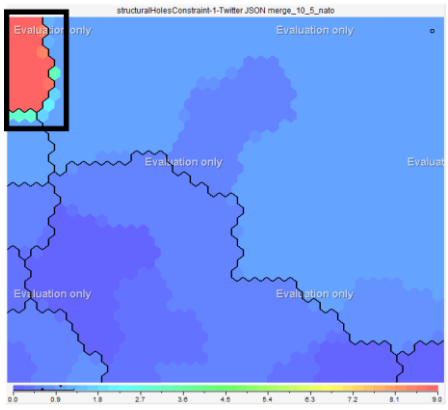
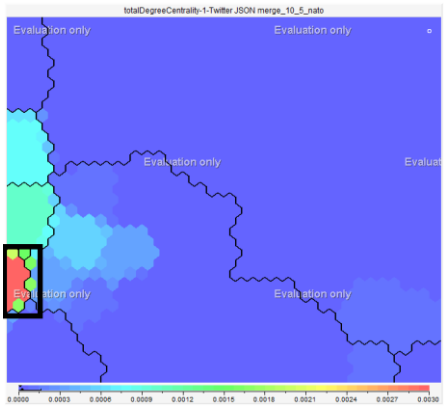


FIGURE 7.22: ZOOMED OUT 3D VERSION OF THE TRIDENT JUNCTURE TWITTER NETWORK, SHOWING THE POSSIBLE EXISTENCE OF AT LEAST 7 COMMUNITIES

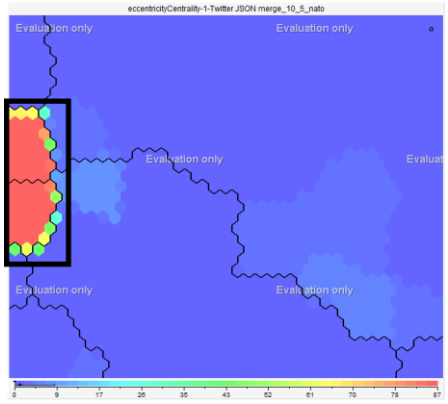
Finally, in order to ascertain whether or not there are risks in the network, consider the SOM shown in Figure 7.23. The SOM clearly shows that there is a measure of risk in the network that may be resolved through optimisation. This is seen in the fact that, for every metric, there is a hotspot affecting the nodes in at least one cluster. Additionally, as the network contains nodes that not only have high measures of eigenvector centrality (Figure 7.23(e)), and total degree centrality (Figure 7.23(b)), but also have high eccentricity centrality measures (Figure 7.23(c)), there is a significant chance that certain individuals in the network may have access to disproportionately high amounts of information, which may warrant further investigation. Figure 7.23(f) shows the large number of nodes that are boundary spanners. As these nodes, based on the network, are mostly disjointed, the nodes in the four clusters with correspondingly high boundary spanner values should be approached with this in mind.



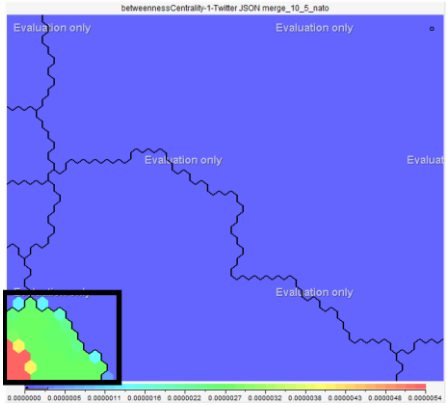
(A): STRUCTURAL HOLES CONSTRAINT



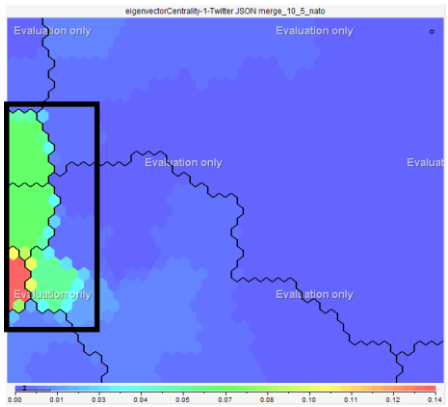
(B): TOTAL DEGREE CENTRALITY



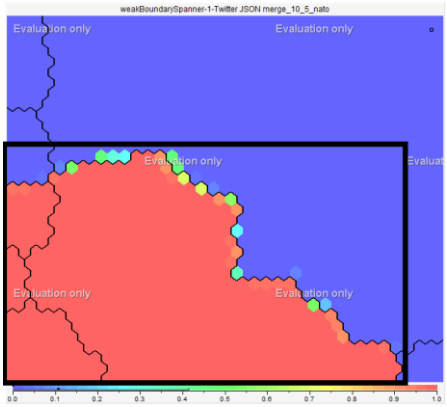
(C): ECCENTRICITY CENTRALITY



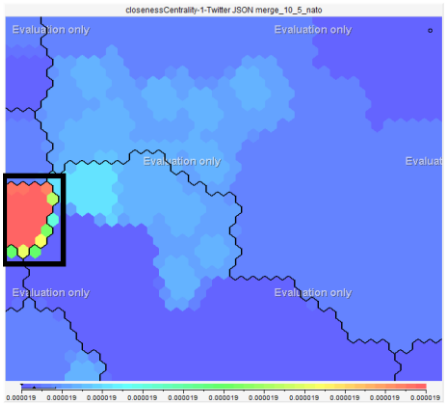
(D): BETWEENNESS CENTRALITY



(E): EIGENVECTOR CENTRALITY



(F): WEAK BOUNDARY SPANNER



(G): CLOSENESS CENTRALITY

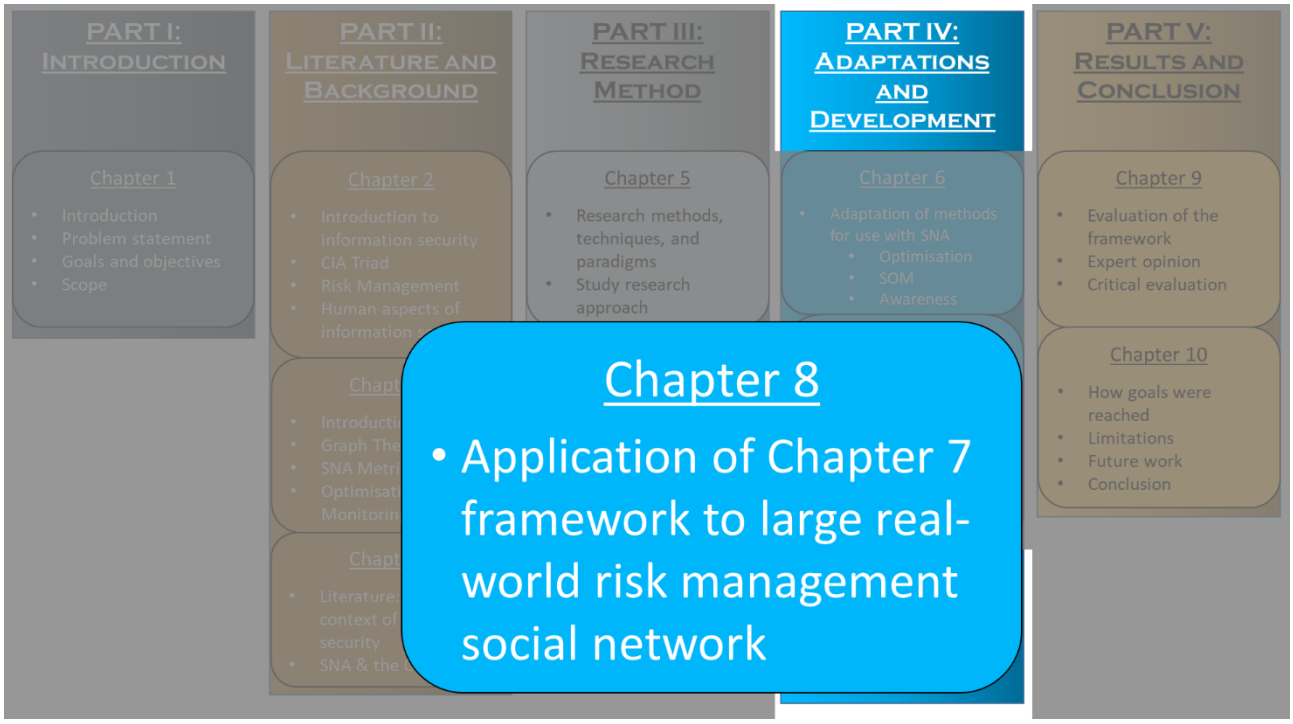
FIGURE 7.23 (A) TO (G): SOM FOR THE TRIDENT JUNCTURE TWITTER NETWORK, COLOURED USING EACH OF THE 7 SELECTED METRICS

The remainder of the framework should not be affected by the size of the network, as the risk reducing relationships are all identified one at a time. The only impact the greater size of the network may have is an increase in the amount of iterations needed to sufficiently reduce the risk in the network. This, by implication, means that larger networks may require the creation or deletion of a much larger number of relationships in order to address the risk in the network.

7.4. CHAPTER SUMMARY

In this chapter, a novel framework was presented that can be used to potentially lower the overall information security risk in a social network. This framework, which is primarily based on the principles of SNA, consists of five phases. In the first phase, data is collected to graph a network. The second phase is concerned with calculating and identifying possible risks, whereas the third phase is focussed on identifying relationships that can be added or removed to lessen the identified risks. The final two phases deal with the specifics of real-world applications, specifically the development of organisation specific risk mitigation strategies and the implementation of those strategies. Two examples, used to demonstrate the framework using real-world data, were also presented.

This framework is perhaps the most significant of the contributions made in this study. As such, a more in-depth validation is appropriate. In the next chapter, the framework is applied in great detail to a real-world risk management network, followed in Chapter 9 by an evaluation of the results.



CHAPTER 8: RISK MANAGEMENT NETWORK: ANALYSIS AND OPTIMISATION

CHAPTER HIGHLIGHTS:

- How was the data for the real-world network collected?
- What were the results of the graphical risk evaluation using SOM?
- What were the results of the optimisation?
- How would the risk improve if strategies were to be developed to implement the proposed changes?
- How was risk addressed with regard to the CIA Triad?

8 RISK MANAGEMENT NETWORK: ANALYSIS AND OPTIMISATION

In Chapter 7, a novel framework that can be used to develop risk mitigation strategies is introduced, and the working of the framework illustrated using a network built with real-world data. In this chapter, the framework is applied to a much larger real-world network that describes the relationships between various entities in a risk management network. This is done to demonstrate the practical applicability of the framework to large real-world networks. The network used in this chapter is much more relevant to real-world applications than some of those introduced previously, as it has no disconnected dyads or triads. This makes it much more representative of formal real-world networks.

The focus of this study is to investigate how Social Network Analysis (SNA) can be used in the context of information security risk management. Specifically, this includes the development of a novel method, utilising SNA, which can be used to inform the development of risk mitigation strategies. The work presented in this chapter focusses mainly on an implementation of the first three phases of this framework, as the final two phases deal with the development and implementation of organisation-specific strategies. As every organisation will require different, and likely unique, strategies, the development and implementation of these strategies are considered to be outside the scope of this study and are not discussed in depth. This chapter starts with a description of the data used to build the framework, focussing on its nature and origin. The network data is then processed using the first three phases of the framework. Finally, the results are presented, and a number of theoretical strategies are briefly discussed.

8.1. OVERVIEW OF DATA AND METHOD OF COLLECTION (PHASE 1)

The network presented in this chapter is generated using data collected from a Corporate Risk Report (CRR) of a large firm. The CRR was made available on condition that any data obtained from it is kept confidential. As such, the data used in this chapter is in an anonymised form. This means that all of the nodes are given generalised names and the links between the nodes are not provided for each and every node. The network is formal, as the relationships of the network do not describe any “social”, or informal, interactions. Furthermore, as the CRR has a specific scope, the network is naturally bordered to only those nodes that are relevant to the management of risks at the firm. In short, the network is formal, bordered, and the data were obtained from a CRR. This correlates to the first six steps of Phase 1.

The network data obtained from the CRR details the relationship between six different kinds of nodes. In total, there are 695 nodes in the network, linked to one another by means of 2962 distinct relationships. The nature of the relationships between the nodes is shown graphically in Figure 8.1. The six different types of nodes will now be introduced individually.

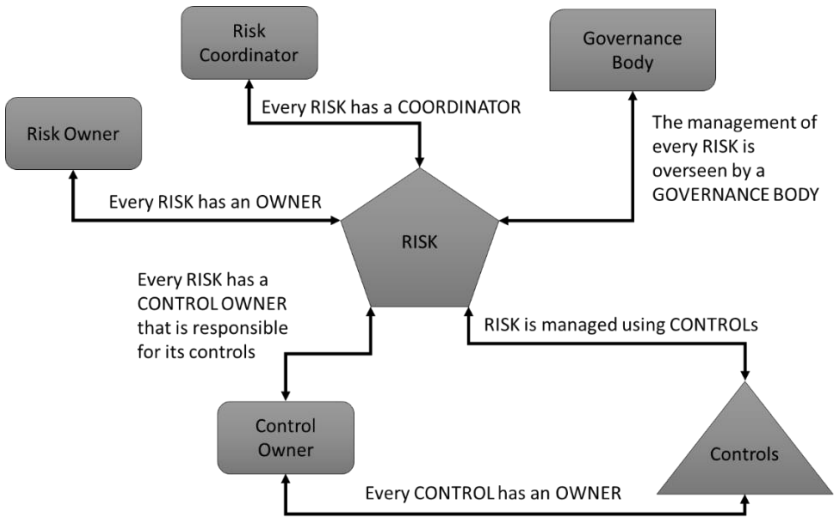


FIGURE 8.1: RELATIONSHIP BETWEEN THE VARIOUS TYPES OF NODES IN THE CRR NETWORK

Risks: These nodes represent the real-world risks that the firm has to manage. The risks range from physical risks, such as fires, to more intangible risks, such as damage to the corporate image. A total of 26 risks are mentioned in the CRR. Each risk has at least one risk owner, a risk coordinator, and any number of controls.

Controls: The controls are put in place to manage the real-world risks. There are 612 controls in total, with a varying number assigned to each of the risks. All of the controls are also associated with a control owner. The controls are used to manage the risks in a number of ways, and range from quality control processes, to business management systems and training and competency programmes.

Risk owners: Six of the nodes represent those individuals who are ultimately responsible for developing strategies to mitigate the risks. They are therefore also responsible for allocating risk coordinators and controls to the risks.

Control owners: Every control is managed by a control owner, who in turn is responsible for ensuring that the controls are implemented. There are 26 control owners in the network.

Risk coordinators: A risk coordinator is an individual, typically a manager, who understands the risk management process. The coordinator ensures that interrelated risks, and their associated processes, are properly coordinated amongst one another. There are 13 risk coordinators included in the network.

Governance bodies: Every risk is overseen by a governance body, which is typically a committee appointed to oversee a specific subset of risks. The task of a governance committee is to ensure that the risks are properly managed, and to appoint risk owners as needed. A total of 12 governing bodies are relevant to the network.

While some of the nodes in the network have exclusive associations with one another, most of the risk owners and –coordinators are responsible for more than one risk. Similarly, some of the risks are owned and coordinated by the same node, and certain governance bodies oversee multiple risks. It is also possible for a control to be used to manage multiple risks, and for a similar control used to manage a separate risk, to also have a separate control owner. In those cases where one node is connected to another more than once, the strength of the relationships between the nodes is increased.

One way in which the graphical evaluation of a network can be made simpler is to use network layout techniques. These techniques use node attributes, such as metrics, to determine the location of each node, thereby including metric values in the graphical representation of the network. The raw CRR network, with no layout applied, is shown in Figure 8.2. In order to more clearly show the complexity of the network, a circular layout was also applied to the network. The network, with the circular layout, is shown in Figure 8.3. A zoomed-in portion of the network following application of the layout is shown in Figure 8.3.

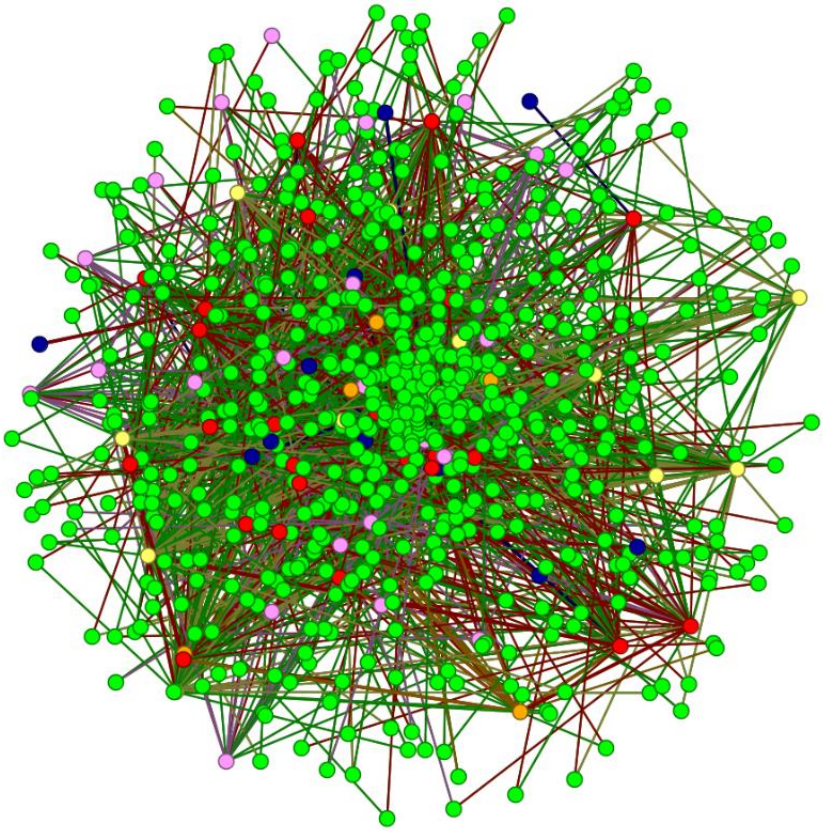


FIGURE 8.2: CRR NETWORK SHOWING RELATIONSHIPS BETWEEN RISKS (RED), RISK CONTROLS (GREEN), RISK COORDINATORS (YELLOW), RISK OWNERS (ORANGE), CONTROL OWNERS (PINK) AND THE GOVERNANCE BODIES (BLUE).

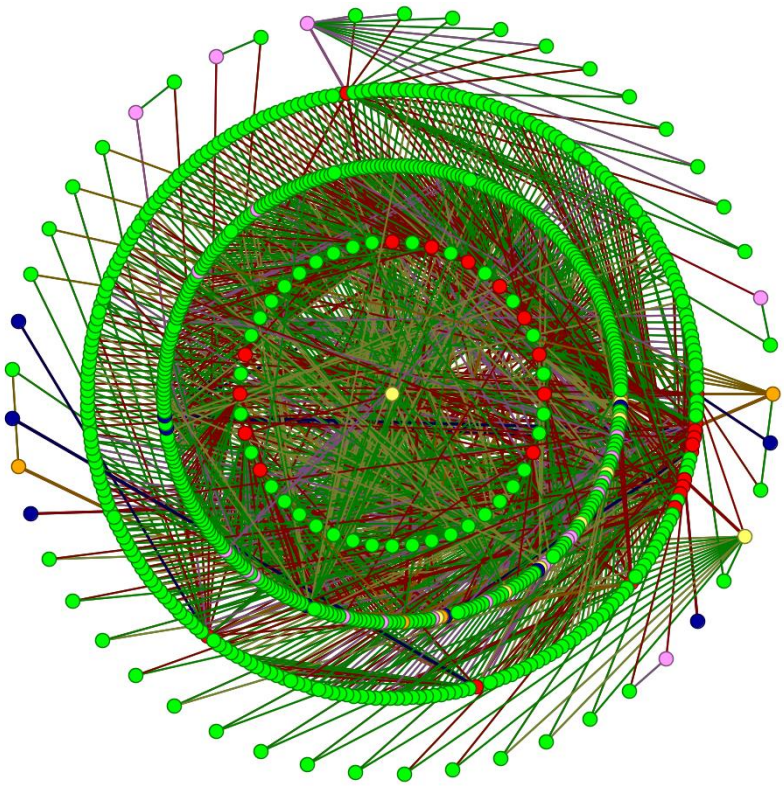


FIGURE 8.3: CRR NETWORK WITH A CIRCULAR LAYOUT. NODES WITH THE HIGHEST BETWEENNESS ARE PLACED IN THE CENTRE.

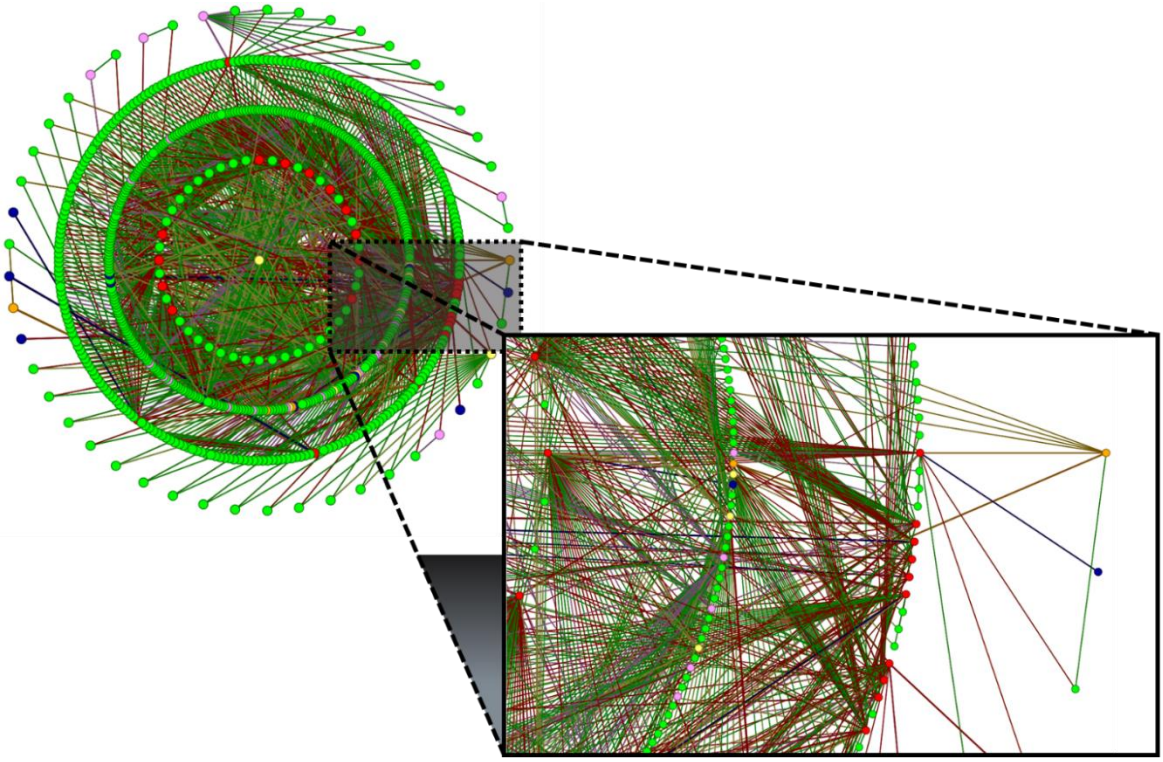


FIGURE 8.4: CRR NETWORK WITH A CIRCULAR LAYOUT, WITH A ZOOMED-IN PORTION OF THE NETWORK SHOWN

While the network shown in Figure 8.2 provides very little useful information on its own, the graphical layout used to produce Figure 8.3 does expose one noteworthy node. This node is a risk coordinator (RISK_COORDINATOR_8) situated in the centre of the network, indicating that it has the highest betweenness. As betweenness is an indication of how often a node is found on the shortest path between other nodes, this suggests that the node may be a shortcut, or "go-to" node. It also suggests that the risk coordinator may be in danger of suffering fatigue, due to the large amount of information that potentially flows through it. The node should therefore be re-evaluated following the optimisation phase to ensure that the risk is resolved, if it does indeed pose a significant risk.

With the network and network data obtained, Phase 1 is complete and Phase 2 can commence.

8.2. GRAPHICAL EVALUATION (PHASE 2)

The ultimate goal of the second phase is to develop an appropriate node-level risk profile that can be used to identify high-risk nodes, as well as to identify relationships that can be added or removed to lessen those risks. In this section, Phase 2 will be applied in full to the CRR network. This involves three steps: firstly, the appropriate SNA metrics should be calculated either for the full network, or individual subsets thereof. Secondly, a Self-Organising Map (SOM) analysis is conducted on the network to determine if there are indeed risks in the network, and re-evaluate the data collection process if no risks are found. The SOM analysis is also used to identify nodes and/or metrics that increase the overall risk in the network, and subsequently give them higher criticality weights. Lastly, the SNA metrics and the identified (or selected) weights are used to develop a node-level risk profile. Each of these three steps will now be discussed in turn.

8.2.1. ANALYSIS OF CRR NETWORK GRAPH

For the analysis of the network, seven SNA metrics are used. All seven of these metrics were discussed in greater detail in Chapters 3 and 4, so the reasoning behind the selection of these particular metrics for the application in this chapter will now be explained briefly.

Total Degree Centrality (TDC): TDC is used to identify nodes that are highly influential in the network and are generally responsible for the dissemination of a considerable amount of information in the network. Nodes that have a high measure of this metric pose a greater

risk due to this influence, as they can potentially manipulate the information in the network, or delay its availability.

Eigenvector Centrality (EiC): Nodes with a high EiC measure have connections to nodes that are themselves well connected. This means that these nodes tend to act as informal leaders in the network. As such a leader can have either a positive or a negative effect on the other nodes in the network, the risk they may or may not pose should be considered along with other metric values.

Eccentricity Centrality (ECC): The ECC metric for a node indicates how "far away" the node is from the other nodes in the network. This measure can be used to identify nodes that are somehow isolated in the network. If the measure is used in conjunction with other metrics, it can be used to identify nodes that may have greater access to information than they should.

Closeness Centrality (CC): This metric can be used to identify those nodes that have access to the greatest amount of information in the network. Whereas TDC can be used to identify those nodes that serve as the most likely origin of information, CC can be used to determine if a node is a good source of information. Nodes with a high CC value are also more likely to cause a significant amount of damage if they are compromised through information gathering attacks, such as social engineering or phishing.

Betweenness Centrality (BC): BC is an indication of how often a node is on the shortest path between other nodes. As this measure increases, so too does the power the node has over the information that flows through it. Additionally, as such nodes tend to act as the best messenger for other nodes, they are at risk of being overworked. This may compromise the quality of the information that flows through such a node.

Structural Holes Constraint (SHC): If a node has a high SHC value, it may be limited in its ability to function efficiently. Because of this, a node with a high SHC measure is considered a risk as these nodes may have access to crucial information, but may not be able to communicate it effectively. Alternatively, the nodes may need information that does not arrive in a timely manner.

Boundary Spanner (BS): The BS metric is used to identify nodes that are the exclusive commonality between two otherwise separate parts of the network. If a node with a high BS value is removed, the network will split into two completely isolated networks. This means that nodes with a high BS value pose a risk in two ways: firstly, they pose a risk to the stability of the network, and secondly they may have complete control over the information that flows through them.

The seven selected metrics are used both for the SOM analysis, as well as for calculating the risk values for each of the nodes. They are therefore the basis for the risk profile that is obtained at the end of the second phase.

8.2.2. SOM Analysis

In order to verify that there are risks present in the network, a SOM (Self-Organising Map) analysis is used. This analysis can also be used to potentially identify nodes and metrics that should be given a higher priority in the risk profile, especially if there are metrics or nodes that clearly pose a higher risk that is not reflected in the risk profile.

The SOM is obtained by using the SOM algorithm, as described in Chapter 6, to give each of the nodes a geographical position based on the metric values for each node. All seven of the selected metrics are used to describe each node. This means that one "base" map is obtained, with all seven metrics having been used to determine each node's position on the map. The algorithm produced a "base" SOM of the CRR network that contains five clusters, as shown in Figure 8.5. Each cluster contains nodes that, based on their attributes, are similar in some way, and also different in some way from nodes in the other clusters. A summary of the contents of each cluster is provided in Table 8.1.

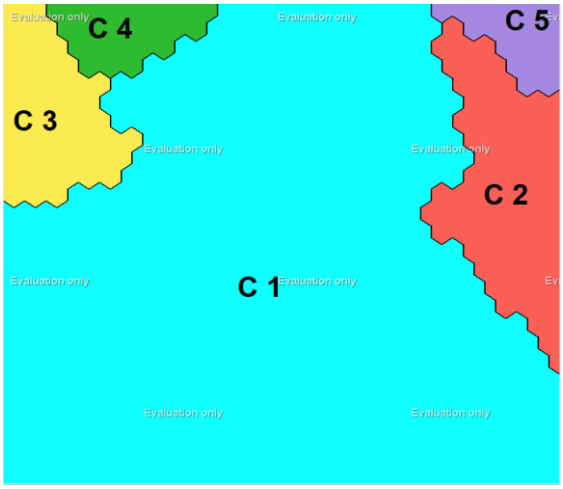
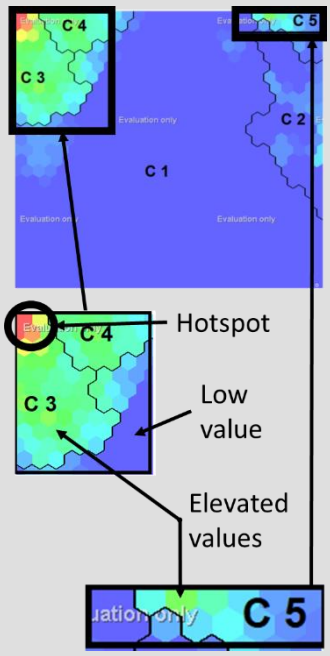
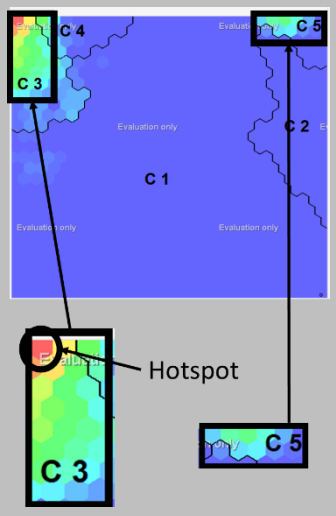


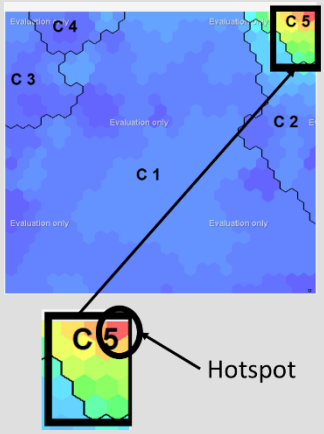
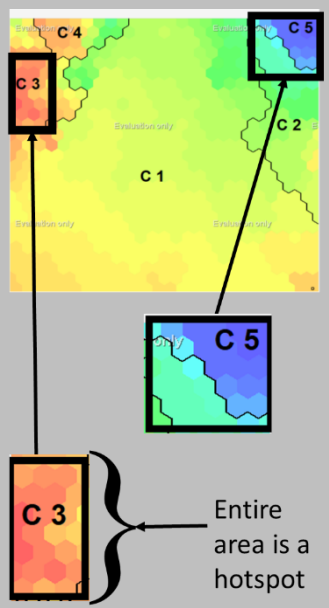
FIGURE 8.5: SOM CLUSTERS

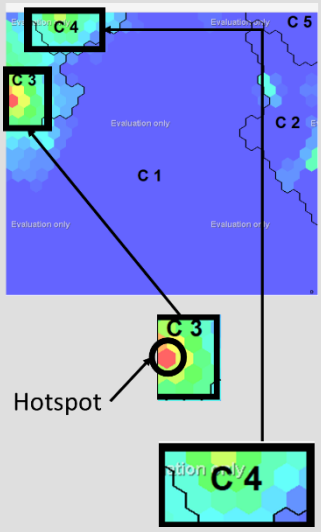
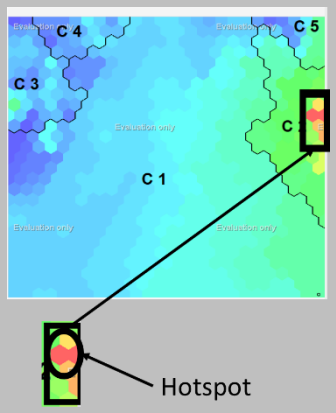
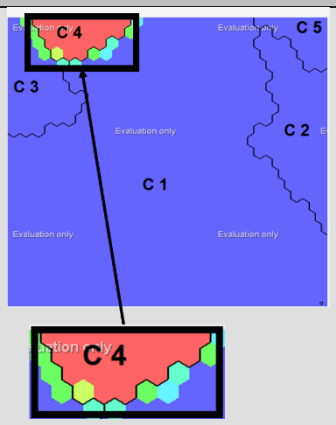
TABLE 8.1: NUMBER OF EACH TYPE OF NODE IN EACH CLUSTER. TO ENHANCE READABILITY, THE CELLS WITH A VALUE OF 0 HAVE BEEN FILLED IN USING BLACK.

Cluster	Number of nodes in cluster						
	Risks	Controls	Risk Coordinators	Risk Owners	Control Owners	Governance Bodies	TOTAL
C1	2	546	1	3	23		575
C2		66	5	2	2		75
C3	13		7	1	1		22
C4	11						11
C5						12	12
TOTAL	26	612	13	6	26	12	695

Each of the metrics that are used to produce the SOM can be utilised to colour the SOM in different ways. This aids in identifying trends and patterns in the data, and also highlights high risk areas that should be investigated. Each of these coloured SOMs will now be briefly discussed in turn. The CIA (Confidentiality, Integrity, and Availability) column shows the security risk attribute associated with each of the different metrics. The rationale behind the association of a metric to particular CIA risks was discussed in Chapter 4.

Metric used to colour SOM	CIA			Coloured SOM	Comments
	C	I	A		
TOTAL DEGREE CENTRALITY (TDC)		X	X		There is a single hotspot in cluster C3, which suggests that there is at least one node in the cluster that poses a risk. An evaluation of the values for the nodes suggests that nodes RISK OWNER 6 and RISK COORDINATOR 10 are the problematic nodes in this region. The overall increased level of total degree centrality in the remainder of clusters C3 and C4 indicate that the nodes in these clusters may have a higher influence than nodes in other clusters. Cluster C4 contains 11 of the risk nodes, whereas C3 contains 13. As the network centres on these risks, this is not necessarily a problem for those particular nodes. The elevated total degree centrality values in C5 are noteworthy, as this node contains 12 of the governance bodies. As these nodes are responsible for oversight, it is interesting that some of them have higher total degree centrality values than others. This intimates that some of the governance bodies oversee a greater number of risks than the rest, and therefore have greater influence in the network.
EIGENVECTOR CENTRALITY (EiC)	X	X	X		The same hotspot identified using total degree centrality is also present for eigenvector centrality. This is an indication of emergent leadership qualities, which suggests that one possible explanation for the integrity and availability risk posed by nodes RISK OWNER 6 and RISK COORDINATOR 10 is that they act as informal leaders in the network. These nodes are regarded as a general risk due to their power in the network. A section of cluster C5 also shows an increase in risk. The section corresponds to a similar measure of heightened risk identified using total degree centrality, as indicated by the similarity of the area and location. This suggests that some of the governance bodies, which are situated in C5, also pose a risk due to their more active leadership positions in the network.

					Unlike the areas identified in C3 and C5, C4 has a very low overall eigenvector centrality measure, which implies that the integrity and availability risk posed by the nodes in this cluster may not be related to emergent leadership qualities, but are more likely caused by other factors. An investigation of the betweenness centrality and BS measures indicate that these nodes may tend to act as intermediaries, thereby increasing the risk they pose to integrity and availability.
ECCENTRICITY CENTRALITY (ECC)			X		There is a single hotspot present in cluster C5 with regard to eccentricity centrality. This measure can be used to identify nodes that are somehow "far away" from the rest of the network, thereby making them eccentric. The nodes in C5 are all governance bodies, which intimates that these oversight committees are not as involved in the network as the other nodes. While this may not be a problem, the heightened risk to availability might be. If the role of these governance bodies is understood as being crucial to ensuring that the appropriate risks are managed effectively, then they should have ready access to relevant information. While the nodes themselves may not necessarily pose a risk, the lack of relationships involving them may indicate a problem with availability in the network.
CLOSENESS CENTRALITY (CC)	X				There are two areas of note with regard to closeness centrality. The first area, shown as a hotspot in C3, suggests that some of the nodes in C3 may have access to an amount of information greater than the norm. As C3 contains a number of risk nodes, as well as risk coordinators and risk owners, this may not necessarily indicate a problem. However, as these nodes do pose an increased risk to confidentiality, due to their increased vulnerability to attacks that aim to obtain information, they should nevertheless receive increased focus to ensure that they are protected against such attacks. The second area, which is in C5, supports an earlier conclusion that the governance bodies in C5 are isolated from the network. As most of the nodes in this cluster likely do not have access to significant amounts of information, they pose a much smaller information leak risk than the nodes in C3. C1, which contains the greatest number of nodes, is mostly uniform in the risk its nodes pose with regard to closeness centrality.

BETWEENNESS CENTRALITY (BC)	X	X			Two areas of interest are highlighted by the betweenness centrality measures. Both of these areas overlap with areas that have been previously identified as containing either high risk or at-risk nodes, specifically using total degree centrality, eigenvector centrality and closeness centrality. Betweenness centrality is an indication of how often a node is on the shortest path between other nodes, and can be used to identify nodes that act as intermediaries, either by necessity, or as a matter of convenience. These nodes facilitate the flow of information between other nodes and have access to an increased amount of information as a result. While they may not be responsible for exclusive communication between the other nodes, the nodes are in a position where they have an increased amount of power over the integrity of the information transferred between them. This strengthens the previous conclusion that these particular nodes pose a risk to integrity in the network. RISK COORDINATOR 8 has the highest betweenness centrality value of the nodes in C3, suggesting that this node may also be overworked.
STRUCTURAL HOLES CONSTRAINT (SHC)		X	X		Structural holes constraint is used to determine if there are nodes in the network that may be limited in their capacity as a result of insufficient connections to other nodes. When a node is constrained in this manner, it may have a negative impact on the integrity and availability of the information the node has access to. There is an overall heightened structural holes constraint in C2, with one clear hotspot. This hotspot, based on the structural holes constraint of the nodes in C2, contains RISK COORDINATORs 5 and 6. This constraint may have far-reaching implications, as it could have an impact on how effectively the risks these coordinators are responsible for are managed.
BOUNDARY SPANNER (BS)		X	X		Boundary spanners are nodes that act as exclusive links between parts of the network. This means that boundary spanner nodes pose risks in two ways. Firstly, as a boundary spanner node's removal will completely isolate a part of the network, it poses a threat to the stability of the network. Secondly, as such a node is an exclusive connection between other nodes, it is in a position to alter the information that passes between two sections of the network, thereby compromising its integrity. Any delays with information passing through such a node may also negatively affect the availability of information. While the risks that the risk nodes represent are

					managed using controls, which may almost eliminate the chances of a risk being realised, the nodes themselves can only be removed from the network if the risk itself can no longer be realised in any way. Therefore, so long as a risk is managed using controls, the node that represents that risk remains in the network. Cluster C4 only contains risk nodes and, as the removal of these nodes is unlikely, the high boundary spanner values should not be a problem. Furthermore, as these nodes do not represent individuals, it is also unlikely that information will be altered by the node. The focus in this instance should therefore be to ensure that those nodes that neighbour the nodes in C4 are monitored, so as to ensure that no part of the network becomes isolated. If this should happen, which could occur if a risk is resolved completely and no longer needs to be managed, all of the controls, owners and coordinators assigned to that risk either need to be removed, or reassigned.

Of the five clusters, the ones containing the largest number of nodes, i.e. C1 and C2, seem to have the lowest overall risk. The nodes in C1, in particular, only tend to have elevated values for a metric when that metric's value is higher for most of the network. C2, in a similar fashion, only has an increased risk caused by structural holes constraint values. This implies that most nodes may not pose a heightened risk, and that the overall risk in the network could be lowered significantly by reducing the risk posed by a few select nodes. It is also important to note that, while the risk can be reduced, it will unlikely be removed entirely.

Summary of SOM Analysis

Seven metrics were used to conduct the SOM analysis. This analysis suggests that cluster C1, which contains the largest number of nodes, also contains the nodes that pose the least risk in the network. While C1 does contain nodes with elevated values for both closeness centrality and structural holes constraint, the heightened values are not outside the norm for most of the network. This intimates that the nodes do not pose an inherently greater risk than most other nodes in the network. The same is also true for cluster C2, with the exception of structural holes constraint. This cluster is the only cluster that contains a hotspot with elevated structural holes constraint values. Specifically, the hotspot contains the nodes RISK COORDINATOR 5 and RISK COORDINATOR 6. As structural holes constraint indicate that a node may be incapable of fulfilling its function adequately due to missing links to important nodes, this indicates that these two nodes may be exposed to an

increased risk to availability. Furthermore, because these nodes may not have adequate access to the information in the network, they are also a threat to integrity.

C3, in contrast to C1 and C2, contains hotspots for four of the seven metrics. This means that the nodes in this cluster, especially those that are found in hotspot regions more than once, pose a significant risk in the network. In particular, RISK OWNER 6 and RISK COORDINATOR 10 are found in the hotspot region for both total degree centrality and eigenvector centrality. The implication of this observation is that these two nodes have much stronger leadership qualities than other nodes in the network, and may be risks to confidentiality, integrity, and availability, due to the influence they have. RISK COORDINATOR 8, which has the single highest betweenness centrality value in the network, also poses a risk as it acts as intermediary for a significant number of nodes. Apart from the increased power the nodes have to affect the integrity and confidentiality of the information that flows through it, the node is also in danger of being overworked. If this should happen, and the node is eventually lost, it may have a significant negative impact on the network as a whole.

The boundary spanners in this network are effectively concentrated in cluster C4, which contains the nodes representing risks. As the likelihood of the risk nodes being removed is unlikely, and these nodes do not represent people, the risk posed by the nodes in this regard is minimal. These nodes should nevertheless be monitored so that, in case they are removed, the network is modified accordingly.

The final cluster, C5, contains a single hotspot for eccentricity centrality, but also has elevated values for total degree centrality, eigenvector centrality, and structural holes constraint. The cluster also has the lowest values for closeness centrality, and contains all 12 of the governance bodies exclusively. Based on the values for these metrics, it can be interpreted that the governance bodies are mainly peripheral in the network, despite having minor leadership roles and influence. Considering the role these nodes play in the risk management process, it may not be a problem; however, there is a risk that they may not have access to all the necessary information all of the time.

8.2.3. Node-Level Risk Profile

Once the metrics have been obtained, most of the data needed to develop the risk profile is available. The only information that is still needed deals with the importance levels of each node and metric, i.e. their risk weight values. By weighing certain nodes and metrics, the risk value for each node can be adjusted to more closely match its expected real-world risk. It is important to consider the importance, or criticality, of certain nodes, as not all nodes may be equally important. Similarly, if there are nodes or metrics that should be removed

from consideration completely, the selection of the correct weights can be used to accomplish this. Due to the large number of methods that can be used to determine these weights, which range from arbitrary selection to empirical determination and historical impact, no specific techniques are mentioned in the framework. The weighting technique that is used, if weights are assigned at all, can therefore be chosen and adapted according to the requirements of a specific application. As shown using the SOM analysis, there are a number of nodes and metrics in this instance that could be weighted to improve the risk profile.

8.2.3.1. Determining Risk Weight Values

One method that can be used to determine the weights for the nodes and metrics implements the results of the SOM analysis. By using these results, the weights can be selected based on the attributes and features of the data itself, instead of subjective rationales. This approach, which is used to determine the weights for the risk profile of the CRR network, involves two steps:

Step 1: Each of the **metrics** are investigated using the SOM. The purpose of this step is to determine the role each metric plays in the final risk value for each node, and how well this matches the expected real-world risk. Each metric is therefore considered based on its meaning, and whether or not that meaning matches the nature of the nodes themselves. If a metric is found to be particularly meaningful, it can be given a greater weight. However, if the contribution to the risk of a node by a metric does not match the real-world risk posed by that node, then that metric can be weighted to reduce its impact, or remove it completely.

Step 2: The **nodes** with a heightened risk are identified next. These nodes are predominantly found in the hotspot areas of the SOM, but may also be located in areas that have an abnormally low value for a metric. The identification of such higher risk nodes therefore still requires a degree of inference. Each of the high-risk nodes are evaluated using the SOM hotspots and the risk values for the nodes. The results of this evaluation are then used to determine if the risk value for a node should be weighted and, if such is the case, what the weight should be.

In order to minimise the subjectivity when determining weight values, weight scales can be used. These scales allow for the selection of a weight based on certain attributes, where an attribute is given a value on a scale. The scale that is used to determine the weights for the metrics was developed in such a way that the SOM for a metric can be used to determine its risk level. The risk levels are broadly adapted from other weight scales (Hancock, 2015; Gardoni & Murphy, 2014; Taylor, 2013). The range of the scale was selected such that a

metric that does not contribute significantly to the overall risk, or does not contribute at all, has a weight of 0, and a metric that indicates a catastrophic risk has a value of 3. Each of the levels are separated by a 0.25 difference in weight, starting with 0.75 for the lowest contributing level. These values were selected because all of the metric values are normalised to a value between 0 and 1 and, if the weight values are too large, they may skew the risk profile for certain nodes. The selection of the 0.25 interval is based on the assumption that the increase in risk value is linear. The weight scale for the metric values is presented in Table 8.2. This table also shows selection guidelines based on the results of the SOM analysis. The selection guidelines are summarised in Table 8.3, which show the number of the various traits, such as hotspots, that are needed for a metric to be described using a specific risk level. In order for a metric to indicate a **major** risk, for example, the SOM coloured according to that particular metric needs to show exactly one area that contains a hotspot, and more than one area with elevated values. In addition to this, the SOM for this particular metric does not show areas of spot elevation, i.e. the areas that are elevated are heightened in such a way that they appear as uniform, contiguous areas. Finally, for a major risk, the hotspot for this metric must be in the same area where exactly one other metric also has a hotspot. The size and shape of the hotspots do not need to match, but there has to be a significant overlap. The contents of Table 8.3 can be similarly applied to identify the weight value for each of the metrics.

TABLE 8.2: WEIGHT SCALE FOR METRICS.

Risk Level	Weight Value	SOM Criteria
Insignificant	**0**	Metric does not contribute meaningfully to risk evaluation
Negligible	**0.75**	Metric has no clear hotspots
Minimal	**1**	Metric shows a single hotspot for an area. The area does not have a hotspot for any other metrics. There are no areas of elevation outside the hotspot area.
Slight	**1.25**	Metric shows a single hotspot for an area. The area does not have a hotspot for any other metrics. There is a single area of elevation.
Minor	**1.5**	Metric shows a single hotspot for an area. The area does not have a hotspot for any other metrics. There are multiple areas of elevation.
Moderate	**1.75**	Metric shows a single hotspot for an area. The area does not have a hotspot for any other metrics, but there are areas with "spot" elevation[2].
Significant	**2**	Metric shows a single hotspot for an area. The area has a hotspot for one other metric. There are no areas of elevation outside the hotspot area.
Serious	**2.25**	Metric shows a single hotspot for an area. The area has a hotspot for one other metric. There is a single area of elevation.
Major	**2.5**	Metric shows a single hotspot for an area. The area has a hotspot for one other metric. There are multiple areas of elevation.
Critical	**2.75**	Metric shows a single hotspot for an area. The area has a hotspot for one other metric, and there are areas with "spot" elevation.
Catastrophic	**3**	Metric shows multiple hotspots in areas that are hotspots for at least two other metrics

[2] One of the selection criteria mentioned is "spot elevation", which refers to areas that have locally high values, but do not have values high enough to classify as hotspots.

TABLE 8.3: SUMMARY OF SELF-ORGANISING MAP (SOM) WEIGHTING CRITERIA

Risk Level	Weight Value	Hotspots	Elevated areas	Spot elevated areas	Hotspots in other metrics
Negligible	0.75	0	0	0	0
Minimal	1	1	0	0	0
Slight	1.25	1	1	0	0
Minor	1.5	1	>1	0	0
Moderate	1.75	1	>1	>=1	0
Significant	2	1	0	0	1
Serious	2.15	1	1	0	1
Major	2.5	1	>1	0	1
Critical	2.75	1	>1	>=1	1
Catastrophic	3	>1	>=0	>=0	>1

The weights for the nodes are selected in a similar fashion. As with the metric values, a weight scale can be used to determine the values for the nodes. It is important to clarify, however, that one reason for the inclusion of node weights in the framework is to allow for the risk profile to be adapted to better fit an organisation. The node weights, for example, can be used to emphasise specific nodes that do not seem to pose significant risks, yet are known to be problematic. Alternatively, these weights can be used to reduce the risk values of nodes that are known to be at risk, but which could not function adequately if their situations are altered drastically. In this instance, a weight scale is used to assign weights to nodes based on the perceived level of risk they pose as a result of their nature. As individual people tend to pose a greater risk than non-human actors, they are given a higher criticality weight. Groups of people, in contrast, are considered to be a smaller risk than individuals, as the chances of an individual, acting in isolation, and causing a risk to be realised is considered greater than a group of people causing damage. While groups may pose a smaller risk than individuals, they are still people, and therefore still pose a risk that is greater than non-human actors. The weight scale used here is shown in Table 8.4. This scale was developed using the Analytic Hierarchy Process (AHP) (Serfontein, 2016; Taylor, 2013), which is a multi-criteria decision making technique that uses pairwise comparisons to determine the relative importance of each entity being compared. Each of the pairwise comparisons are quantified by use of a preference scale, which assigns a value between 1 and 9 depending on the preference for one comparison attribute over another. If one entity or attribute is extremely more important than the one it is compared to, for example, a value of 9 is given to that particular comparison. The individual person was considered to be strongly more important compared to the non-human entity and moderately more important compared to a group. A group, on the other hand, was considered to be moderately to strongly more important compared to a non-human entity.

TABLE 8.4: WEIGHT SCALE FOR NODES

Node type	Weight Value
Non-human	0.097
Group of People	0.284
Individual person	0.619

While this process can be used to determine weights that allow for the risk profile to better match the real-world risks, it may be infeasible to use this method if the risk profile contains a large number of metrics, or if the network contains a large number of nodes. In an ideal situation the weights would be determined by evaluating the network and risk profiles over the course of months or years, which may produce weights that are "fine-tuned" to the network itself. The process presented here is intended as a starting point for networks where these weight values are unknown. As no known weight values for the CRR network are available, this procedure will now be applied in order to determine the weights for the risk profile. Due to the large number of nodes in the network, the risk levels for the nodes will be determined using their "type" attribute, i.e. whether they are a person, group, or risk. In contrast, only seven metrics are used, so each of these metrics can be considered individually.

Assigning node weights

The only consideration when assigning criticality weights to nodes in this case is the nature of those nodes. The CRR network contains six types of nodes, of which three types represent individual people, one represents groups of people, and two represent non-human entities. As it is more likely that an individual would pose a greater risk, the types of nodes that describe people will be assigned a weight of 0.619, as per the scale in Table 8.4. The single group type, namely the governance bodies, will be assigned a weight of 0.284. The remainder of the nodes, i.e. the control- and risk nodes, are non-human entities that are considered to be lower risk and are assigned a weight of 0.097.

Assigning metric weights

The first metric to be considered is the boundary spanner metric. As the only boundary spanners in the network are risk-type nodes, and these nodes will likely not cause problems, the impact of this metric should be reduced to better reflect its true risk. Due to the fact that only the risk-type nodes have an elevated value for this metric, and that the boundary spanner values indicate risks that are unlikely to be realised with these risk nodes, the weight of the metric should be selected such as to reduce the risk contribution of the boundary spanner value. Stated differently, the weight value should be smaller than 1 to ensure that the boundary spanner metric, when multiplied by the weight value, contributes less to the risk value of the node than only the boundary spanner value would. Considering that the risks associated with elevated boundary spanner values are to integrity and availability, and that the risk-type nodes themselves are unlikely to cause a realisation of these risks, the weight is selected as 0. This removes the metric from the risk equation entirely.

As discussed, the SOM for each of the metrics is used to determine their weight values. The SOMs are shown in Section 8.2.2. Of the remainder of the metrics, the two that are considered next are total degree centrality and eigenvector centrality. The SOMs for these two metrics not only show a hotspot in the same area, but are very similar overall. This suggests that these two metrics will have the same weight value. In addition to the hotspot, there are two areas of significant elevation. Both of these observations suggest that total degree centrality and eigenvector centrality indicate areas of **major** risk, and are given a weight of 2.5, as per the criteria given in Table 8.2.

Betweenness centrality is heightened in similar places on the SOM as eigenvector centrality, with the exception that the hotspot is located in a different place and is not heightened in C4. With the approach used to select the weight value for eigenvector centrality and total degree centrality taken into consideration, the weight for betweenness centrality should similarly be chosen using the weight scale shown in Table 8.2. As the hotspot for betweenness centrality is located in an area where closeness centrality also has a hotspot, and spot elevation is present in cluster C4, the weight value for betweenness centrality is 2.75, as determined by its **critical** risk level.

While closeness centrality is heightened for most of the network, with the exception of the governance bodies, there are multiple areas that are heightened more than other areas. This, combined with an area that contains a hotspot for betweenness centrality as well, suggests that closeness centrality has a **major** impact on the perceived risk, and should be given a weight of 2.5, as per Table 8.2.

Eccentricity centrality, in contrast to most of the metrics discussed previously, has a single hotspot in an area where no other metric has hotspots. This, coupled with the lack of elevated areas outside of the hotspot area, indicate that eccentricity centrality has a **minimal** contribution, and should be given a weight value of 1 according to Table 8.2. Structural holes similarly have a hotspot in an area where no other metric has a hotspot, but is elevated for most of the map. The elevation is found in multiple areas, which suggest that the structural holes constraint metric shows a **minor** contribution to risk, and should be given a weight value of 1.5. The weights selected for the various node types are shown in Table 8.5, and the weights for all of the metrics are shown in Table 8.6.

TABLE 8.5: RISK PROFILE WEIGHTS FOR NODE TYPES

Node Type	Weight
Risk coordinators	0.619
Risk owners	0.619
Control owners	0.619
Governance bodies	0.284
Controls	0.096
Risks	0.096

TABLE 8.6: RISK PROFILE WEIGHTS FOR METRICS

Metric	Weight
Boundary spanner	0
Total degree centrality	2.25
Eigenvector centrality	2.25
Betweenness centrality	2.75
Closeness centrality	2.5
Eccentricity centrality	1
Structural holes constraint	1.5

The approach employed to obtain these weights may not be appropriate for all networks, as each network may require a unique approach in order to find weights that are appropriate. This is also true for a network where historical data are available, as such data could greatly simplify the process. With the weight values determined for all of the nodes and metrics, the node-level risk profile can almost be completed. All that remains is the calculation of the risk value for each node, as well as the overall risk value. Before revisiting how these risk values are determined, however, two alternative techniques that can be used to determine the weight values are briefly discussed first.

8.2.3.2. DETERMINING RISK WEIGHT VALUES – ALTERNATIVE APPROACHES

Two alternatives to the weighing approach employed are introduced in this section. These two techniques, referred to as historical approximation and the AHP, can both be used to determine metric and nodes weights. However, as the usability and appropriateness of these techniques are dependent upon the nature of the network being evaluated, they may not be suited to certain networks or organisations. It is therefore important to reiterate that any technique can be used to determine the weight values, and that the framework does not specify a weighting technique.

Historical approximation: The first alternative technique discussed uses historical data collected from the network in order to determine the most effective weights. This can be accomplished in any number of ways, from selecting weights and never changing them, to running simulations using the data. The simulations could, for example, be used to identify the weights that most closely match the observed historical risks. This should help to produce a risk profile that matches the real-world risks even more closely. Another approach is to use the historical data with neural techniques, such as the SOM, in order to find weight values that accurately describe the observed impact of each node and metric. In order to use such a neural technique for this purpose it may be necessary to specially adapt the technique in order to obtain useable weight values.

Analytic Hierarchy Process (AHP): The AHP is a multi-criteria decision making technique that makes use of pairwise comparisons in order to calculate weights for each of the objects

being compared (Tadisina *et al.*, 1991). The comparisons are done using a preference level scale, such as the one shown in Table 8.7. Unfortunately, because of the large number of pairwise comparisons that would have to be done in order to calculate the weights, the AHP will likely be unsuitable for use with large networks. One option to reduce the amount of work needed and make the technique useable with large networks, is to select a specific number of metrics and nodes, and only apply the AHP to those nodes and metrics that were specially selected. The downside to this approach is that it may not produce weights that are appropriate for all of the nodes or, indeed, all of the metrics. The use of the AHP should therefore be carefully considered and compared to the alternatives, especially with large networks.

TABLE 8.7: PREFERENCE SCALE FOR PAIRWISE COMPARISONS, ADAPTED FROM (TAYLOR, 2013)

Preference Level	Numerical Value
Equally Important	1
Equally to Moderately Important	2
Moderately Important	3
Moderately to Strongly Important	4
Strongly Important	5
Strongly to Very Strongly Important	6
Very Strongly Important	7
Very Strongly to Extremely Important	8
Extremely Important	9

In addition to the two techniques discussed, there may be other methods that can be used to determine the appropriate weights. These techniques are therefore not intended to represent an exhaustive list, but merely an indication of the kinds of techniques that could potentially be used to calculate the weights.

8.2.3.3. NODE-LEVEL RISK PROFILE, INCLUDING WEIGHTS

The second phase of the framework, which focusses on developing an information security risk profile, has two main objectives. The first is to use the network data obtained during the first phase to produce a node-level risk profile, which can then be used during subsequent phases in order to develop risk mitigation strategies. For this purpose, equations 7.3 and 7.4 (Chapter 7), which are reshown as equations 8.1 and 8.2 respectively, are used. These equations are used to calculate both the risk values for each node, and the overall risk value for the network:

$$Normalised\ risk\ value\ of\ node\ i = C_i \sum_{m=1}^{k} w_m \left(\frac{v_{mi} - \min(V_m)}{\max(V_m) - \min(V_m)} \right) \qquad (8.1)$$

$$Overall\ risk\ value\ of\ network = \sum_{i=1}^{N}\left[C_i \sum_{m=1}^{k} w_m \left(\frac{v_{mi} - \min(V_m)}{\max(V_m) - \min(V_m)}\right)\right] \quad (8.2)$$

where w_m is the weight of metric m, v_{mi} is the value of metric m for node i, k is the total number of metrics used, C_i is the criticality weight for node i, V_m is the collection of all values for all nodes for metric m, and N is the total number of nodes in the network. When equation 8.2 is applied to the data in the node-level risk profile for the CRR network, an overall risk value of 238.223 is obtained.

The second objective is to conduct an initial assessment of the risks in the network, using the calculated metric values and the SOM technique. The purpose of this assessment is simultaneously to identify nodes that may pose a significant risk, so that it can be verified that their risks are addressed, and to ensure that the data collection process used in the first phase was adequate and appropriate. This was accomplished, as the SOM analysis clearly identified nodes and groups of nodes that pose a risk, or are at risk, in the network.

At the conclusion of the second phase, all of the metrics have been calculated for each of the nodes, and the weights for various nodes and metrics have been determined. These values are all that is needed to produce a node-level risk profile that can be used in later phases. A small portion of the node-level risk profile is presented in Table 8.8, and the complete risk profile can be found in Appendix C.

TABLE 8.8: EXTRACT OF NODE-LEVEL RISK PROFILE. THE TABLE SHOWS THE NORMALISED VALUES FOR BETWEENNESS CENTRALITY (BC), CLOSENESS CENTRALITY (CC), ECCENTRICITY CENTRALITY (ECC), EIGENVECTOR CENTRALITY (EIC), STRUCTURAL HOLES CONSTRAINT (SHC), AND BOUNDARY SPANNER (BS). THE WEIGHT AND RISK VALUES FOR THESE NODES ARE ALSO SHOWN, AS WELL AS THE METRIC WEIGHTS AND THE TOTAL RISK VALUE.

Node-Title	BC	CC	EcC	EiC	SHC	TDC	BS	Weight of Node	Risk of Node
RISK 1	0.085	0.526	0.086	0	0.249	0.329	1	0.097	0.267
RISK OWNER 1	0.005	0.485	0.114	0.001	0.293	0.105	0	0.619	1.249
RISK COORDINATOR 1	0.339	0.749	0.029	0	0.685	0.217	0	0.619	2.692
GOV BODY 1	0	0.052	0.629	0	0.416	0.054	0	0.284	0.427
CONTROL 1	0.002	0.454	0.114	0	0.356	0	0	0.097	0.173
. . .									
	METRIC WEIGHTS							TOTAL RISK:	238.223
	2.75	2.5	1	2.25	1.5	2.25	0		

With the data verification complete and the weights assigned, a complete node-level risk profile is available. This profile is now used in the third phase of the framework in order to

identify relationships that, if either added or removed, may result in a decrease in overall risk in the network.

8.3. OPTIMISATION (PHASE 3)

The optimisation phase of the framework can be applied to the network data once the node-level risk profile has been obtained. The first task to be completed in Phase 3 is the selection of a measure that can be used to determine when the risk has been sufficiently reduced, or an adequate number of nodes have been evaluated. In the previous chapters, two such methods were discussed. With the first method, every single node is evaluated until a point is reached where the overall risk can no longer be reduced. The second method involves selecting a risk reduction threshold. This threshold can be obtained in a number of ways, from arbitrarily selecting a percentage reduction in overall risk value, to using management determined reductions and historically determined reduction rates. In order to demonstrate the flexibility of the framework in selecting this threshold, a third option will now be introduced.

The threshold value, in addition to the two methods mentioned, can also be determined using statistical methods. One approach is to use basic statistical metrics, such as the average and standard deviation, to determine when a risk value is abnormally high, and then use the smallest of the abnormally high values as the threshold value. In this instance, the threshold is applied to the risk values of the individual nodes, and the process is continued until there are no nodes left that have an initial risk value above the threshold. Another method is to use Tukey's rule (Swanepoel *et al.*, 2008) to identify outliers, and then use the value of the smallest high-risk outlier as the threshold value. Only the high-risk outliers are considered, as the goal is to reduce the overall risk in the network. Tukey's rule states that an outlier is 1.5 times greater or smaller than the interquartile range of a set of values. The interquartile range is determined by first finding the median of the set, and then finding the values that lie halfway between the median and both the minimum and maximum values in the set. The interquartile range is then calculated by subtracting the smaller of these values from the larger one.

In order to determine the risk threshold, the z-score for the node risk values are used. The z-score, or standard score, is the distance between a value in a set and the statistical mean, in terms of the standard deviation of the set (Glen, 2019; Cheadle *et al.*, 2003). This means that a z-score of 1 indicates a value that is exactly the value of the standard deviation greater than the mean. The equation for calculating the z-score for a node risk value is

$$z_i = \frac{x_i - \mu}{\sigma} \tag{8.3}$$

where z_i is the z-score for node i, x_i is the risk value for node i, μ is the average, or mean, of the risk values of all the nodes, and σ is the standard deviation of the risk values.

The z-score can be used to evaluate values in terms of normal statistical distributions. In general, if a node has a risk value z-score greater than 2, then it has a risk that is not only above average, but is statistically high as well (Swanepoel *et al.*, 2008). As such, a z-score of 2 will be used as a threshold value during the optimisation phase in this chapter, i.e. only the nodes with a risk value z-score of 2 or greater will be considered. There are a total of 30 nodes in the network that satisfies this requirement, and the risk values for these nodes are shown in Table 8.9, as well as the calculated z-score for each of the risk values.

TABLE 8.9: EXTRACT OF NODE-LEVEL RISK PROFILE, SHOWING ALL THE NODES WITH A RISK Z-SCORE GREATER THAN 2. THE TABLE SHOWS THE NORMALISED VALUES FOR BETWEENNESS CENTRALITY (BC), CLOSENESS CENTRALITY (CC), ECCENTRICITY CENTRALITY (ECC), EIGENVECTOR CENTRALITY (EIC), STRUCTURAL HOLES CONSTRAINT (SHC), AND BOUNDARY SPANNER (BS). THE WEIGHT AND RISK VALUES FOR THESE NODES ARE ALSO SHOWN.

Node-Title	BC	CC	EcC	EiC	SHC	TDC	BS	Weight of Node	Risk of Node	Z-Score
RISK_COORDINATOR_10	0.87	0.745	1	1	0	1	0	0.619	6.038	8.962
RISK_COORDINATOR_13	0.954	0.835	1	0.864	0.006	0.754	0	0.619	5.794	8.572
CTRL_OWN_18	0.991	0.922	0	0.743	0.008	0.7	0	0.619	5.131	7.513
RISK_COORDINATOR_11	0.898	1	0	0.759	0.017	0.5	0	0.619	4.845	7.056
RISK_COORDINATOR_8	1	0.935	0	0.633	0.018	0.523	0	0.619	4.776	6.945
RISK_COORDINATOR_2	0.981	0.874	0	0.486	0.027	0.415	0	0.619	4.302	6.188
RISK_COORDINATOR_7	0.685	0.935	0	0.652	0.026	0.4	0	0.619	4.102	5.868
RISK_COORDINATOR_12	0.463	0.701	1	0.529	0.014	0.615	0	0.619	4.098	5.862
RISK_COORDINATOR_4	0.37	0.688	1	0.36	0.056	0.246	0	0.619	3.21	4.443
CTRL_OWN_16	0.194	0.684	1	0.478	0.053	0.262	0	0.619	3.088	4.248
RISK_OWNER_6	0.324	0.866	0	0.526	0.053	0.192	0	0.619	2.941	4.013
CTRL_OWN_17	0.139	0.606	1	0.286	0.057	0.269	0	0.619	2.619	3.498
CTRL_OWN_11	0.25	0.792	0	0.401	0.064	0.215	0	0.619	2.569	3.418
CTRL_OWN_8	0.185	0.645	1	0.248	0.117	0.123	0	0.619	2.557	3.399
CTRL_OWN_3	0.148	0.619	1	0.219	0.081	0.2	0	0.619	2.488	3.289
CTRL_OWN_21	0.093	0.545	1	0.224	0.069	0.246	0	0.619	2.339	3.051
CTRL_OWN_20	0.259	0.706	0	0.245	0.057	0.262	0	0.619	2.292	2.976
RISK_COORDINATOR_1	0.528	0.554	0	0.082	0.067	0.246	0	0.619	2.275	2.949
CTRL_OWN_19	0.037	0.537	1	0.24	0.134	0.115	0	0.619	2.132	2.72
RISK_OWNER_2	0.037	0.602	1	0.173	0.144	0.085	0	0.619	2.107	2.68
CTRL_OWN_5	0	0.39	1	0.034	0.819	0.015	0	0.619	2.051	2.591
CTRL_OWN_26	0	0.385	1	0.039	0.819	0.015	0	0.619	2.05	2.589
CTRL_OWN_23	0	0.342	1	0.045	0.821	0.015	0	0.619	1.994	2.5
RISK_COORDINATOR_5	0.176	0.589	0	0.184	0.061	0.269	0	0.619	1.899	2.348
CTRL_OWN_22	0.007	0.459	1	0.152	0.257	0.062	0	0.619	1.878	2.314
RISK_OWNER_4	0	0.268	1	0.023	0.824	0.015	0	0.619	1.852	2.273
CTRL_OWN_1	0.102	0.429	1	0.039	0.259	0.069	0	0.619	1.847	2.265
CTRL_OWN_9	0	0.199	1	0.005	0.827	0.015	0	0.619	1.723	2.067
CTRL_OWN_24	0	0.199	1	0.005	0.827	0.015	0	0.619	1.723	2.067
CTRL_OWN_12	0	0.199	1	0.005	0.827	0.015	0	0.619	1.723	2.067

It is noteworthy that all of the nodes shown in Table 8.9, i.e. those nodes that have an above average risk value, represent individuals, which is also why the weight value for all of the nodes are the same. This means that none of the risk, control, or governance body nodes have an above average risk value. One explanation for this is that these individuals have the vast majority of the authority in the network. Additionally, most of them are associated with multiple risks and controls, and therefore have access to considerable information. Lastly, they represent a very small portion of the network: of the 695 nodes in the network, only 45 are individuals. This suggests that this small portion of the network is situated very centrally and, as a result, is highly influential. This naturally increases the risk associated with these nodes, as any threat to any of them will likely have a significant impact on the network as a whole.

With the node-level risk profile completed and the threshold determined, the optimisation step can commence. Unlike the networks used to illustrate the optimisation method in previous chapters, this network features links that can both be added and removed. However, in order to ensure that the essential functioning of the network itself is not compromised, for example by a risk no longer having a risk coordinator, the following rules have to be followed when identifying links to add or remove:

- No node may be isolated from the network;
- The basic structure shown in Figure 8.1 has to be maintained, e.g. the link between a risk owner and a risk can only be severed if that risk has multiple risk owners; and
- Governance bodies provide oversight and should therefore not be connected to controls directly.

The remainder of the optimisation process is exactly the same as described in the previous chapters. Firstly, the node with the highest risk value (*R*) is identified, then the metric for that node that has the highest value (*M*). The metric *M* is then investigated: if the metric's value can be lowered by removing a link, then the links between the selected node *R* and each of its neighbours are considered. In such a case, if appropriate, the link to the neighbour with the highest value for the identified metric *M* is removed. If the metric can be improved by adding a link, however, the node with the lowest value for the identified metric is identified and a link is created between that node, and node R. This process is followed with the rules mentioned above taken into consideration.

As 30 nodes were identified that lie above the threshold level, 30 steps are needed in the optimisation process. The data used to identify the risk reducing links are shown in Table 8.10. This table illustrates the progression of the optimisation phase, by first listing the data of the high-risk node. In each row containing a high-risk node's data, for example RISK_COORDINATOR_10, the metric that contributes the most to that node's risk value is highlighted in yellow. In the following row, the details for the node that matches the

optimisation requirements are shown. This is done for all 30 high risk nodes. In those cases where a node could not be found that satisfies the optimisation requirements for that node, a red row with the text “NO VALID NODES” is shown. The links that are added are shown in Figure 8.7, and the links that are removed are shown in Figure 8.8.

TABLE 8.10: DATA USED TO IDENTIFY NEW RELATIONSHIPS DURING OPTIMISATION PROCESS. CELLS HIGHLIGHTED IN RED INDICATE THE HIGHEST REMAINING Z-SCORE AT THE START OF THE STEP. CELLS HIGHLIGHTED IN YELLOW ARE THE METRICS THAT CONTRIBUTE THE MOST TO A NODE'S RISK VALUE. “REMOVE” INDICATES THAT LINKS HAVE TO BE REMOVED IN ORDER TO REDUCE THE METRIC, WHEREAS “ADD” INDICATES THAT A METRIC VALUE WILL BE REDUCED IF LINKS ARE ADDED.

Node-Title	Risk of Node	Z-Score	BC (REMOVE)	CC (REMOVE)	SHC (ADD)
STEP 1					
RISK_COORDINATOR_10	6.038	8.962	***2.393***	1.863	0
RISK_25	0.637	0.331	1.477	2.035	0.03
STEP 2					
RISK_COORDINATOR_13	5.794	8.572	***2.624***	2.088	0.009
RISK_22	0.738	0.492	1.986	2.315	0.024
STEP 3					
CTRL_OWN_18	5.131	7.513	***2.725***	2.305	0.012
RISK_8	0.702	0.435	2.217	2.153	0.026
STEP 4					
RISK_COORDINATOR_11	4.845	7.056	2.47	***2.5***	0.026
RISK_19	0.621	0.305	1.35	2.23	0.044
STEP 5					
RISK_COORDINATOR_8	4.776	6.945	***2.75***	2.338	0.027
RISK_2	0.693	0.421	2.318	2.165	0.029
STEP 6					
RISK_COORDINATOR_2	4.302	6.188	***2.698***	2.185	0.041
RISK_5	0.609	0.286	2.217	1.98	0.047
STEP 7					
RISK_COORDINATOR_7	4.102	5.868	1.884	***2.338***	0.039
RISK_15	0.658	0.365	1.68	2.1	0.033
STEP 8					
RISK_COORDINATOR_12	4.098	5.862	1.273	***1.753***	0.021
RISK_26	0.664	0.374	1.656	2.078	0.027
STEP 9					
RISK_COORDINATOR_4	3.21	4.443	1.018	***1.72***	0.084
RISK_20	0.607	0.283	1.298	2.1	0.039
STEP 10					
CTRL_OWN_16	3.088	4.248	0.534	***1.71***	0.08
RISK_11	0.572	0.227	1.196	1.948	0.044
STEP 11					
RISK_OWNER_6	2.941	4.013	0.891	***2.165***	0.08
RISK_21	0.587	0.251	1.298	1.938	0.042
STEP 12					
CTRL_OWN_17	2.619	3.498	0.382	***1.515***	0.086
RISK_13	0.492	0.099	1.07	1.883	0.068
STEP 13					
CTRL_OWN_11	2.569	3.418	0.688	***1.98***	0.096
RISK_10	0.548	0.189	1.196	1.818	0.044
STEP 14					
CTRL_OWN_8	2.557	3.399	0.509	***1.613***	0.176
RISK_16	0.489	0.095	0.916	1.83	0.06
STEP 15					
CTRL_OWN_3	2.488	3.289	0.407	***1.548***	0.122
RISK_17	0.461	0.05	0.842	1.743	0.06
STEP 16					
CTRL_OWN_21	2.339	3.051	0.256	***1.363***	0.104
RISK_12	0.482	0.083	0.891	1.98	0.072
STEP 17					
CTRL_OWN_20	2.292	2.976	0.712	***1.765***	0.086
RISK_18	0.442	0.019	0.712	1.655	0.06

Node-Title	Risk of Node	Z-Score	BC (REMOVE)	CC (REMOVE)	SHC (ADD)
STEP 18					
RISK_COORDINATOR_1	2.275	2.949	***1.452***	1.385	0.101
RISK_6	0.414	-0.025	1.603	1.298	0.083
STEP 19					
CTRL_OWN_19	2.132	2.72	0.102	***1.343***	0.201
NO VALID NODES					
STEP 20					
RISK_OWNER_2	2.107	2.68	0.102	***1.505***	0.216
NO VALID NODES					
STEP 21					
CTRL_OWN_5	2.051	2.591	0	0.975	***1.229***
RISK_9	0.539	0.174	2.318	1.733	0.044
STEP 22					
CTRL_OWN_26	2.05	2.589	0	0.963	***1.229***
RISK_7	0.431	0.002	0.789	1.603	0.059
STEP 23					
CTRL_OWN_23	1.994	2.5	0	0.855	***1.232***
RISK_4	0.384	-0.073	1.221	1.31	0.077
STEP 24					
RISK_COORDINATOR_5	1.899	2.348	0.484	***1.473***	0.092
NO VALID NODES					
STEP 25					
CTRL_OWN_22	1.878	2.314	0.019	***1.148***	0.386
NO VALID NODES					
STEP 26					
RISK_OWNER_4	1.852	2.273	0	0.67	***1.236***
GOV_BODY_5	0.853	0.676	0	0.498	1.5
STEP 27					
CTRL_OWN_1	1.847	2.265	0.281	***1.073***	0.389
RISK_1	0.233	-0.315	0.407	0.758	0.09
STEP 28					
CTRL_OWN_9	1.723	2.067	0	0.498	***1.241***
RISK_3	0.279	-0.241	1.196	0.498	0.108
STEP 29					
CTRL_OWN_24	1.723	2.067	0	0.498	***1.241***
RISK_23	0.406	-0.038	0.611	1.798	0.111
STEP 30					
CTRL_OWN_12	1.723	2.067	0	0.498	***1.241***
RISK_24	0.355	-0.12	0.305	1.655	0.122

To illustrate how the optimisation is done, consider the data for Step 3, as shown in Table 8.10:

STEP 3					
CTRL_OWN_18	5.131	7.513	***2.725***	2.305	0.012
RISK_8	0.702	0.435	2.217	2.153	0.026

Each of the steps progress individually for each of the nodes, as first shown in Algorithm 6.1. With each step, of the nodes that were not considered in previous steps, the one with the highest risk is identified. In Steps 1 and 2, the nodes RISK_COORDINATOR_10 and RISK_COORDINATOR_13 were considered. In Step 3, the node CTRL_OWN_18 was identified as the node with the highest risk, with a risk value of 5.131 and a risk z-score of 7.513. Once the node has been selected, all of the metric values are considered and the metric with the highest value is identified. In this instance, the metric with the highest value is betweenness centrality, with a value of 2.725. This value is greater than the values for closeness centrality (2.305) and structural holes constraint (0.012). The metric with the highest value, in this

case betweenness centrality, is used to determine how the rest of the optimisation for the selected node will proceed.

The betweenness centrality metric for a node is determined by calculating how many times the node is part of the shortest route between other nodes. As such, the metric value is reduced by removing links, as adding new links will likely increase the value of the metric for the node. When a link has to be removed, all of the high-risk node's neighbours are considered, and the link to the neighbour with the highest value for the selected metric is removed, if such a removal is appropriate. Stated differently, all of CTRL_OWN_18's neighbours are investigated, and the one with the highest betweenness centrality is selected. If this node has not been considered in previous steps, and removing the link to this node does not violate the rules, then the link between CTRL_OWN_18 and the neighbour is removed. As shown in Table 8.10, out of all of CTRL_OWN_18's neighbours, RISK_8 satisfied these requirements. The node has a betweenness centrality metric value of 2.217, and has a risk value of 0.702.

In those cases where the highest metric value is reduced by adding links, as is the case for Steps 21, 22, and 23, where the structural holes constraint metric has the greatest value, then all of the nodes that are not neighbours of the high risk node are investigated. Out of all these nodes, the node with the lowest value for identified metric is selected. If a new link to this node is appropriate, then the link is created. For CTRL_OWN_5, in Step 21, the new link is created to RISK_9, which has a structural holes constraint value of 0.044. In Step 22, CTRL_OWN_26 is connected to RISK_7 and, in Step 23, CTRL_OWN_23 is connected to RISK_4.

There are four cases, shown in Table 8.10, where adding or removing an identified link is not appropriate. In Step 24, for instance, the metric closeness centrality had the greatest value. This metric is reduced by removing links, which is why all of its neighbours are considered. Unfortunately, none of the links could be removed without violating the structural rules of the network and, subsequently, no links could be identified for removal. Figures 8.6 – 8.8 show the original network and all of the links that were added and removed. These links are not shown on a network with a circular layout, as the position of the nodes change based on the metric used. Figure 8.9 shows how the optimisation impacts the network by applying a circular layout technique that uses betweenness centrality to determine a node's position. With this technique, the node with the highest betweenness centrality is placed in the centre, while the remaining nodes are placed further away as their respective betweenness centrality values decrease.

It is interesting to note, when graphically inspecting these circular graphs, that the risk coordinator that used to occupy the central position, i.e. the position with the highest betweenness centrality, was replaced by a control owner. Additionally, the outermost ring, which contains those nodes with the lowest betweenness centrality values and therefore pose a lower risk, has grown to contain a significantly greater number of nodes. This shows

that the optimisation process has had a ripple effect on the network which, while it may be slight with regard to the overall risk value, has positively impacted a larger number of nodes than those evaluated.

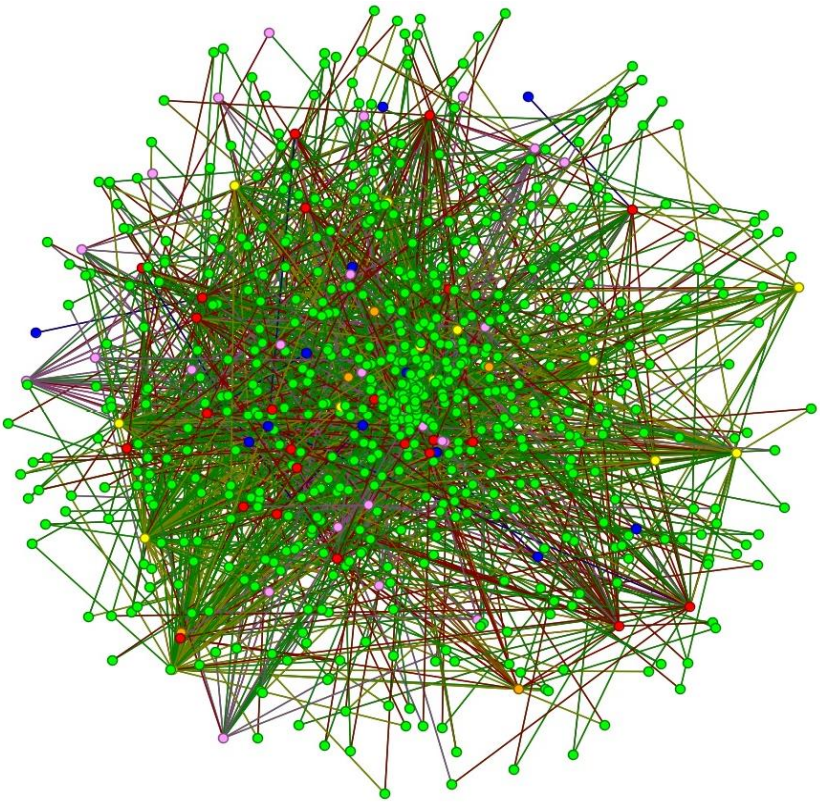


FIGURE 8.6: CRR NETWORK PRIOR TO OPTIMISATION

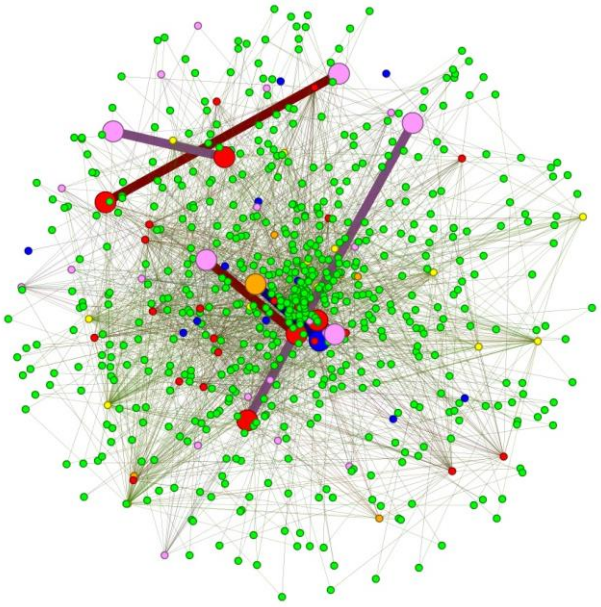


FIGURE 8.7: LINKS THAT WERE ADDED TO CRR NETWORK DURING OPTIMISATION

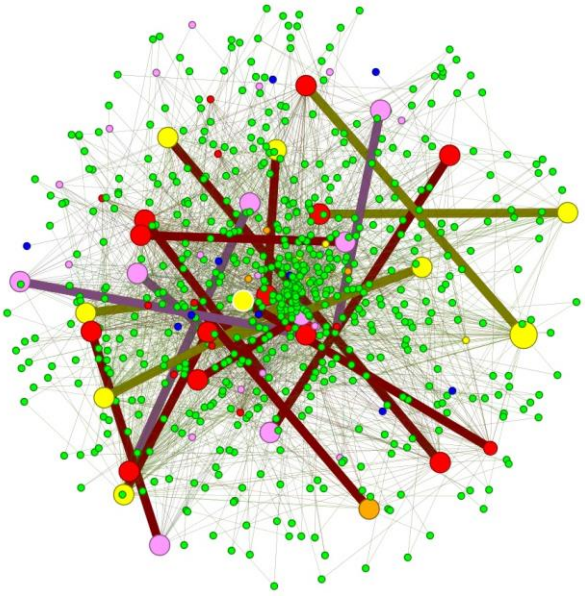


FIGURE 8.8: LINKS THAT WERE REMOVED FROM CRR NETWORK DURING OPTIMISATION

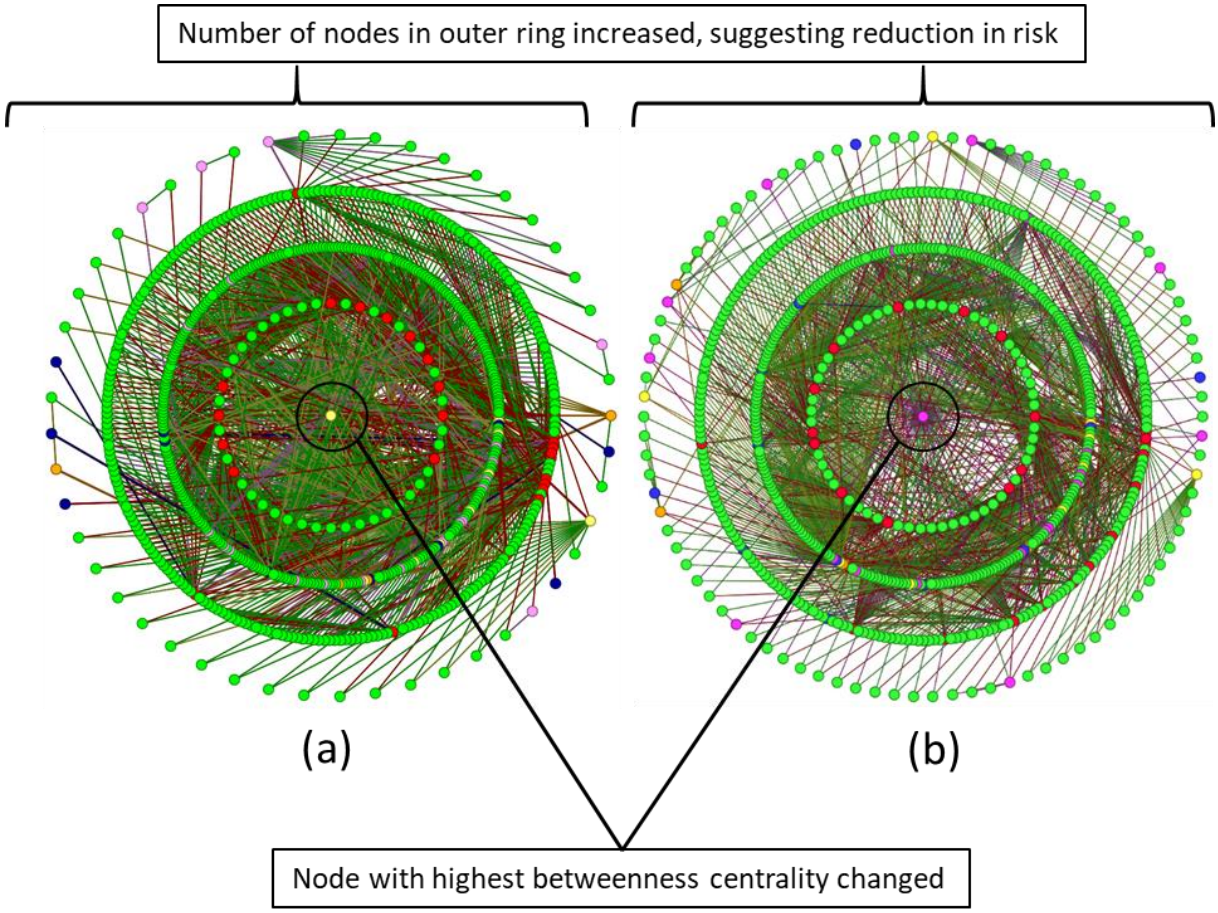


FIGURE 8.9: CRR NETWORK BEFORE (a) AND AFTER (b) OPTIMISATION, WITH CIRCULAR LAYOUT APPLIED. NODE WITH HIGHEST BETWEENNESS CENTRALITY IS PLACED IN THE CENTRE.

Reflection on optimisation process

When the nodes involved in the optimisation, as shown in Table 8.10, are considered, it is notable that, with the exception of RISK_14, all of the risk nodes are involved. This is likely due to a number of factors, such as the risk-central structure of the network, and the occurrence of individual entities taking on multiple roles, e.g. a risk owner also acting as a risk coordinator. It should be emphasised that, in those cases where a link is removed, no links are considered that may violate the structure of the network. This means that, in every instance where a link between a risk and a risk coordinator or owner is removed, that risk was already connected to another coordinator or owner. The implication of this is that most, if not all, of the risk nodes are associated with individuals that have multiple roles for different risks, i.e. a person that acts as the risk owner for one may be a risk coordinator for another. As a result, there may be exploitable redundancies in the network, and the removal of these links should help streamline the responsibilities of the individuals by not assigning too many different roles to each individual.

The inclusion of most of the risk nodes in the optimisation process further suggests that, while they may not be among the nodes that pose the highest risk in the network, most of them are connected to high-risk nodes. This implies that the network that exists to manage the risks poses a greater threat than the risks themselves, at least as far as the network itself is concerned. This is a noteworthy observation, as it shows that the risk in the network can be reduced by optimising the structure of the network, and fine-tuning the responsibilities of the individuals involved, without needing to introduce any major structural changes. Subsequently, any mitigation strategies developed to address the risks in the network should focus on reducing the number of roles or responsibilities individuals have, and in so doing reduce the risk that each high-risk individual poses. This is also true for cases where new relationships have to be created: by creating a new relationship between a risk and a control owner, for example, the control owner becomes better informed of issues that affect a different part of the network. Ideally, by incorporating this information into his management strategy, it may be possible to find more cost-effective controls. It may also be possible to enhance the effectiveness of the controls by utilising knowledge of the greater network. Finally, it is interesting to note that all of the nodes that receive new links following the optimisation, as shown in Table 8.10, had a structural holes constraint value as the greatest metric value. As this metric is an indication of structural inefficiencies that may limit a node's ability to function, it suggests that there are minor structural problems in the network that should be addressed. By connecting these nodes that are somewhat restricted by structural inefficiencies to nodes that are not constrained, new paths to necessary information are opened. This should improve the efficiency of information flow within the network and make the controls more effective and reactive.

In total, the optimisation process identified 19 links to remove and 7 links to add. Of these 26 links, 9 removed a link between a risk coordinator and a risk, 9 removed a link between a control owner and a risk, 6 created new links between control owners and risks, 1 created a link between a risk owner and a governance body, and 1 removed a link between a risk owner and a risk. These details are summarised in Table 8.11.

TABLE 8.11: SUMMARY OF LINKS THAT WERE ADDED AND REMOVED

Total number of links added/removed	26
Total number of links removed	19
Number of links removed between a RISK COORDINATOR and a RISK	9
Number of links removed between a CONTROL OWNER and a RISK	9
Total number of links created	7
Number of links created between a CONTROL OWNER and a RISK	6
Number of links removed between a RISK OWNER and a RISK	1
Number of links created between a RISK OWNER and a GOVERNANCE BODY	1

There were also 4 nodes that could have benefitted from having links removed, but there were no valid links to remove. As discussed, this suggests that there are potentially unnecessary redundancies in the network, such as one risk having multiple risk coordinators

and owners. It also shows that 6 out of the 26 (23%) control owners may be constrained in their ability to function, and the best solution to this problem is to create new links to other risks. This will, by virtue of mutual association, give them access to contacts and relationships they would not have otherwise. The one link that is created between a risk owner and a governance body also suggests that certain risk owners may benefit from greater interaction with the oversight bodies. This is not necessarily true for all risk owners, but the risk owner considered is shown to be constrained by structural holes.

A number of observations can be made based on the reduction in overall risk value, and the section of the network that was investigated:

- Prior to optimisation, the risk value, calculated using Equation 8.2, was 298.739. Upon conclusion of the optimisation phase, this value had been reduced to 290.411. A calculation of the percentage change relative to the original value shows that this 8.328 reduction equals a 2.8% drop in the overall risk value. Furthermore, this risk reduction was achieved by only considering 30 of the 695 nodes, or 4.3% of all the nodes in the network.
- It is also important to clarify that this network contains a small number of individuals, 71 in total, and 612 controls. Accounting for the 12 governance bodies, this means that 88% of the network consists of controls. These controls all have a mostly contiguous risk value, which is significantly lower than the risk values associated with the individuals. This suggests that the network, as it relates to the controls, is properly structured and poses no significant intrinsic structural risks. The implication of this is that the primary risks are found in 11.9% of the network. The 4.3% of the network that was evaluated therefore still accounts for an investigation of 36% of the higher risk nodes of the network.
- The remainder of this high-risk section of the network was not considered due to the use of a threshold technique that only allowed for nodes with a risk z-score of greater than 2 to be considered.

In conclusion, the seemingly slight reduction in risk is significant, especially considering the restrictions placed on the optimisation process by the selected threshold technique.

8.4. FINAL REFLECTION

As illustrated in Figure 8.10, the process followed in this chapter encompasses the first three phases of the framework introduced in Chapter 7. Briefly stated, the Corporate Risk Report (CRR) used as the source of the data was investigated in order to determine how the various entities involved in the risk management system were connected to one another. The names of all these entities were then anonymised, an adjacency list was created, and finally

a graph of the relationships was created. This graph was used to calculate the SNA metrics for the network, which were evaluated using the SOM technique first introduced in Chapter 6. The weights for the various nodes and metrics were then assigned, for the nodes according to attributes, and for the metrics using the results of the SOM analysis. The node-level risk profile was finalised by adding the weights, and calculating the risk values.

The final phase made use of the optimisation technique, first described in Chapter 6, to identify 19 links that would improve the overall risk if they were to be removed, and 7 links that would be of benefit if they came into existence.

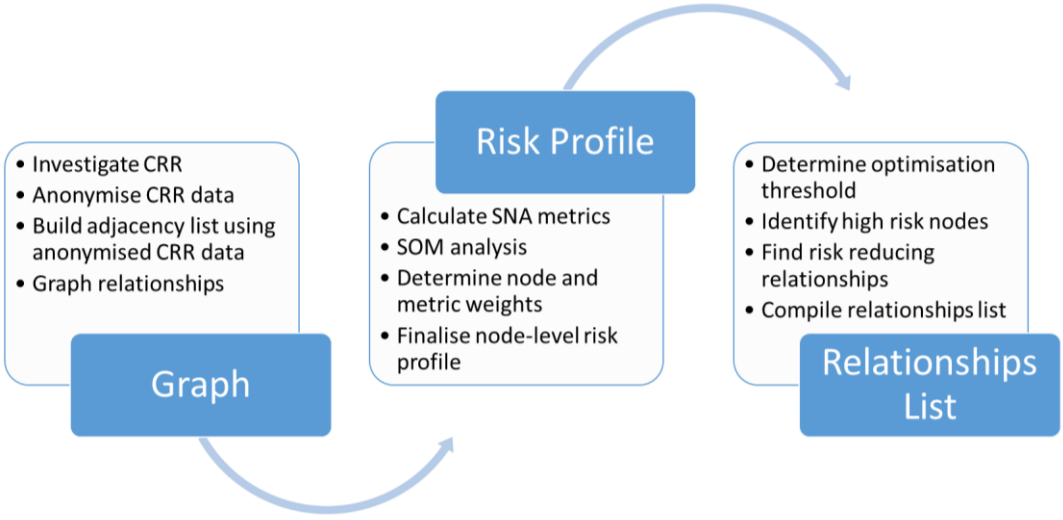


FIGURE 8.10: OVERALL PROGRESSION OF FIRST THREE PHASES

Each of the nodes can be evaluated individually to determine how the information security risk is affected with regard to confidentiality, integrity and availability (CIA). In order to illustrate the impact that could potentially be observed for specific nodes should these changes be implemented, and how they would impact the risk to confidentiality, integrity, and availability (CIA), consider the changes in the metrics for the node RISK_COORDINATOR_8 as an example. The changes to the metrics for this node are shown in Figure 8.11. Recall from Chapter 4 that each of the SNA metrics can be used to identify potential risks to each aspect of the CIA triad. With RISK_COORDINATOR_8, the metrics prior to optimisation suggest that the node is potentially at risk with regard to its confidentiality, integrity, and availability. This observation is made based on the fact that, with the exception of eccentricity centrality and the structural holes constraint value, all of the metrics have above average elevation. As summarised in Figure 8.11, the impact of the optimisation is a reduction in risk to all three members of the CIA triad. While the increase for the structural holes constraint does suggest a slight increase in risk to both integrity and

availability, this risk is mitigated by the reduction of the other metrics that also impact integrity and availability. It is important to reiterate at this point that the ultimate goal is not necessarily to eliminate all of the CIA risk present in the network, but to reduce and manage it. Therefore, while RISK_COORDINATOR_8 does show a slight increase in the value for the structural holes constraint metric, it is acceptable as the increased value is still below average.

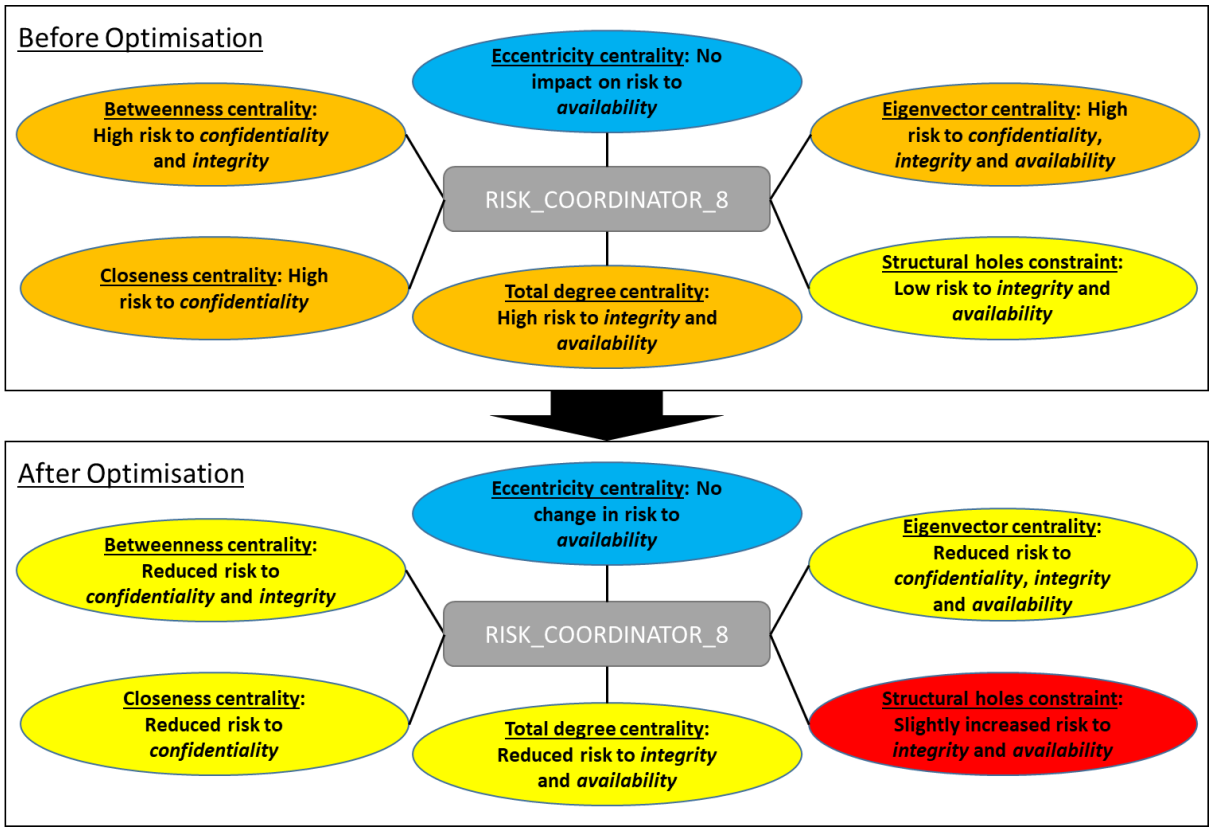


FIGURE 8.11: CHANGES TO THE METRICS FOR RISK COORDINATOR 8, AND HOW THESE CHANGES IMPACT CIA RISKS. BLUE INDICATES THAT A METRIC HAS NO IMPACT ON THE RISK VALUE OF THE NODE, ORANGE INDICATES AN ELEVATED RISK, YELLOW INDICATES A LOW RISK, AND RED INDICATES THAT THE VALUE FOR THE METRIC HAS INCREASED AS A RESULT OF THE OPTIMISATION.

Another noteworthy observation that can be made concerns the interaction between confidentiality, integrity, and availability. These three aspects of the CIA triad generally have to be balanced, e.g. a significant increase in confidentiality may lead to, or require, a significant decrease in availability. To demonstrate how this interaction can be observed in the changes to the metrics, consider the changes for CTRL_OWN_5, as shown in Figure 8.12. The pre-optimisation metrics indicate that the node has elevated levels of risk for all three members of the CIA triad. During the optimisation phase, a link is identified that, when added, would reduce the eccentricity centrality value of CTRL_OWN_5 to zero, potentially causing a significant reduction in the risk to availability for the node. Additionally, this new link would also cause a reduction in the value of the structural holes constraint metric,

indicating a slight reduction in risk to integrity as well. The remainder of the metrics, however, indicate an overall increase in the risk to confidentiality for the node, as none of the metrics associated with confidentiality are reduced. This suggests that the significant reduction in risk to availability causes a slight increase in risk to confidentiality, which is to be expected as the three aspects of the CIA triad need to be balanced. This increase in risk to confidentiality is therefore due to the trade-off between confidentiality, integrity and integrity that is a natural part of the process that maintains the balance between the three aspects.

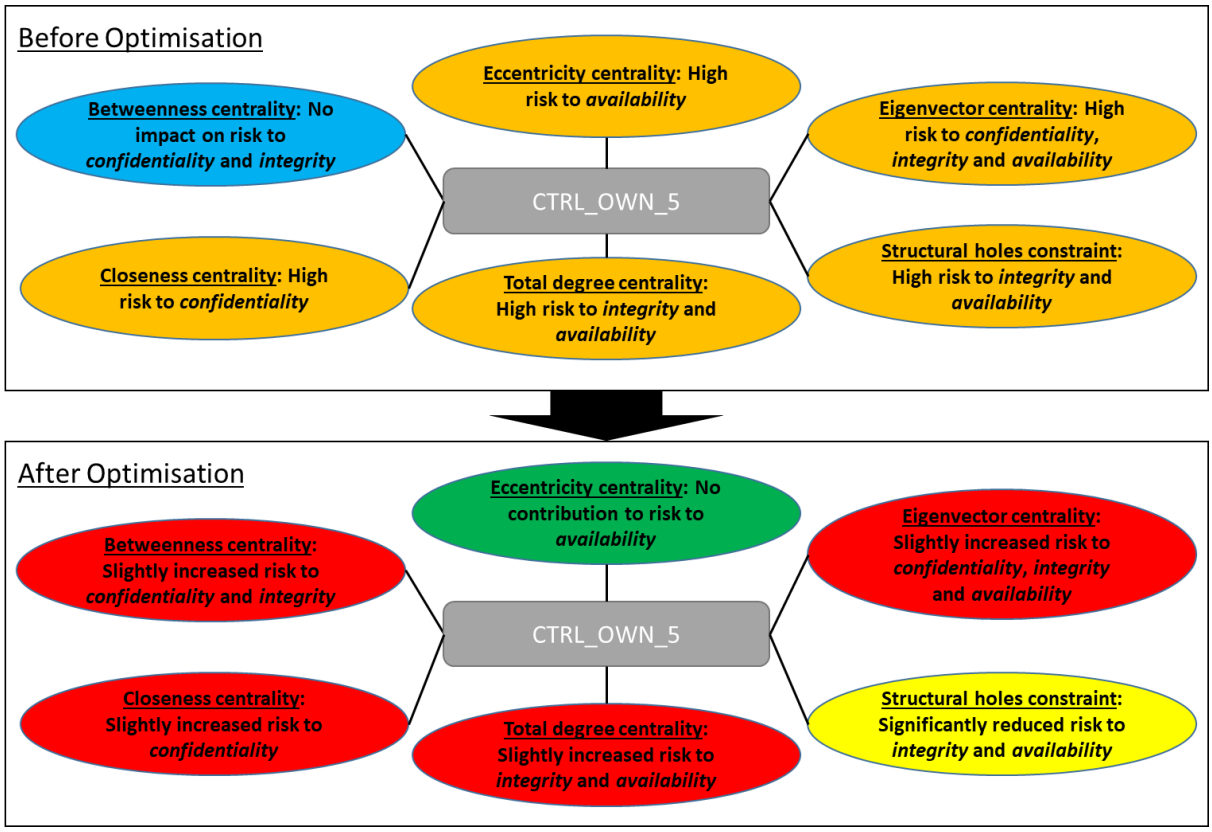


FIGURE 8.12: CHANGES TO THE METRICS FOR CONTROL OWNER 5, AND HOW THESE CHANGES IMPACT CIA RISKS. BLUE INDICATES THAT A METRIC HAS NO IMPACT ON THE RISK VALUE OF THE NODE, GREEN INDICATES THAT THE VALUE OF THE METRIC HAS BEEN REDUCED TO ZERO, ORANGE INDICATES AN ELEVATED RISK, YELLOW INDICATES A LOW RISK, AND RED INDICATES THAT THE VALUE FOR THE METRIC HAS INCREASED AS A RESULT OF THE OPTIMISATION.

In summary, the goal of reducing the overall risk to confidentiality, integrity, and availability could be achieved by implementing strategies that create or remove the 26 identified links. If these changes are affected in the real-world network, the greatest improvement over the long term should be to availability, followed closely by confidentiality and integrity. This means that a more substantial improvement to integrity is sacrificed in the interest of a more significant improvement to availability and confidentiality. It is important to emphasise, however, that all three aspects should improve, as the slightly smaller

improvement to integrity and confidentiality, when compared to availability, is nevertheless an improvement over the risk that exists in the original network. These changes to the CIA risk in the network are expected due to the way in which the metric values changed following optimisation.

8.5. CHAPTER SUMMARY

In this chapter, the framework, first presented in Chapter 7, was applied to a real-world risk management network. The source of the data, namely a confidential Corporate Risk Report, was discussed, the overall structure of the network was introduced, and the network was graphed. The graph of this network was then used to calculate the values for various SNA metrics, and the SOM analysis technique introduced in Chapter 6 was utilised in order to evaluate the data, as well as to determine weights for the various metrics. A node-level risk profile was then developed using the metric values, and this profile was evaluated in order to identify certain links that should either be added or removed in order to improve the overall risk of the network. The implications of the nodes associated with these links were also discussed briefly, and the impact in terms of confidentiality, integrity, and availability was highlighted.

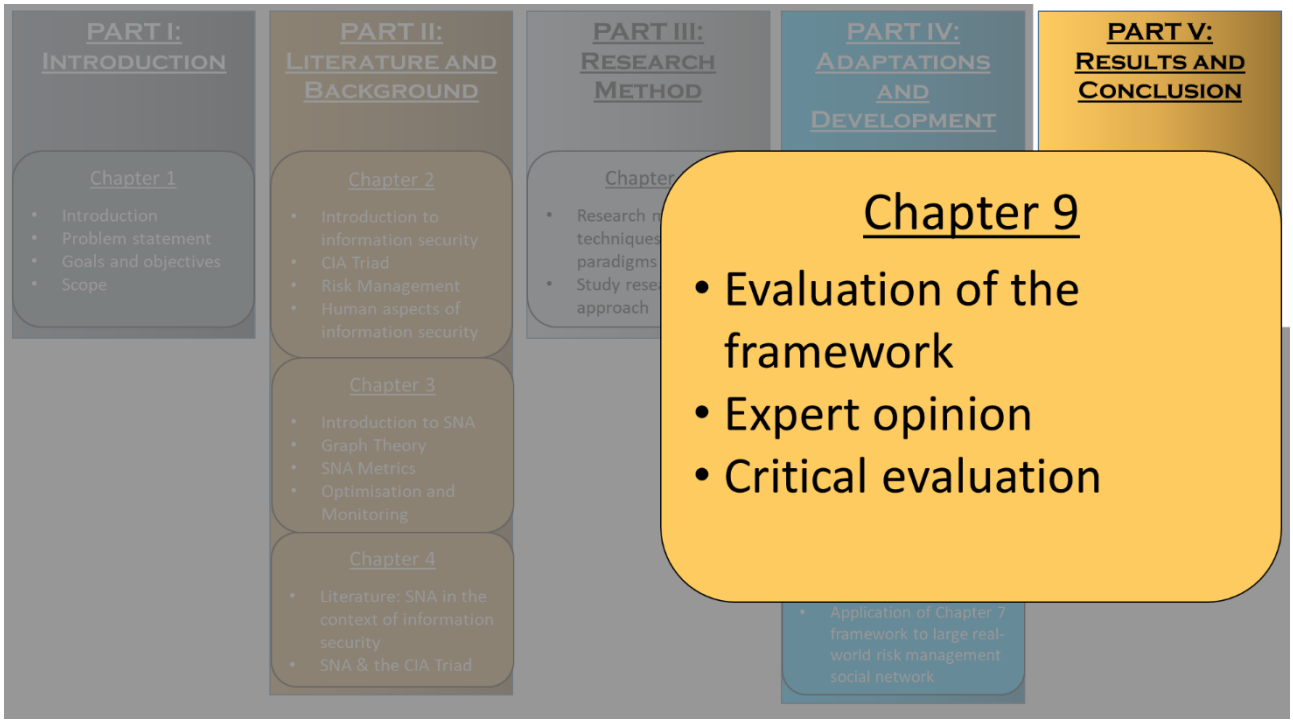
This chapter concludes Part IV of the study. In the first chapter of Part V, an evaluation of this application of the framework to a Corporate Risk Report is presented in order to demonstrate its validity.

PART V

Results & Conclusion

"Quality is never an accident. It is always the result of intelligent effort"

- John Ruskin



CHAPTER 9: EVALUATION OF THE FRAMEWORK

CHAPTER HIGHLIGHTS:

- In which ways are the framework evaluated?
- What was the feedback that was received regarding the framework?
- Can the framework be used to manage risk?
- Is the framework a valuable contribution to the field of information security risk management?

9 EVALUATION OF THE FRAMEWORK

The focus of Part IV is to investigate how certain methods can be adapted for use with Social Network Analysis (SNA), as well as to introduce a novel framework that implements those methods. This framework is intended to be used as a risk management tool, and is subsequently focussed on evaluating risk using SNA metrics. In Part V, which is also the final part of the study, the results of the research are evaluated and critiqued. This part contains two chapters, of which the first is focussed on evaluating the work presented in Part IV, and the final chapter evaluates and concludes the study as a whole.

In this chapter, two primary methods are used to critically evaluate the value of the framework and the adapted methods. The first method involves presenting the framework, and its application as shown in Chapter 8, to an expert in the field of risk management, so that the expert can evaluate the framework and provide feedback. The second method makes use of a number of critiques, formed as questions, in order to critically evaluate the framework. The framework and adapted methods, as shown in previous chapters, are applicable to real-world data and situations. The goal with the evaluations in this chapter is to investigate the worth and utility of the framework in a broader context, and in so doing establish its contributory worth.

9.1. EXPERT FEEDBACK

The first kind of evaluation used to determine the worth of the proposed novel framework involves feedback, both positive and critical, from an expert in risk management. In order to evaluate the framework in this manner, both the framework and its application to real-world data, as shown in Chapter 8, was presented to the expert. He was then asked a number of questions that were aimed at determining the real-world practical worth of the framework. The expert has over 35 year of experience in the Finance, Mining, and Utility and Infrastructure sectors, where he held senior managerial positions relating to Information Technology, Internal Auditing, and Risk and Assurance Management. He is the current Risk and Assurance Manager of the Risk and Assurance branch of a large corporation that has over 3000 employees. In this section, his feedback on the framework is discussed.

The expert, firstly, confirmed that the framework as presented would definitely make a positive contribution to any corporate risk management system, as it is capable of providing valuable insights that might not be evident from a simple review of the data. He also highlighted the value of applying the CIA approach to the CRR level of corporate risk

management. Such an approach, he explained, brings a different perspective beyond traditional risk management techniques. This may help to identify risks in the risk management system, and could therefore be used to improve such a system.

When probed about the results obtained from the evaluation presented in Chapter 8, and how useable the results were considered to be, some of the difficulties in using CRR data in this manner were highlighted. The primary issue is that CRR data, by nature, is in a continuous state of evolution and is therefore not truly mature. The data, as used in this study to test and develop the framework, is consequently best described as a snapshot of the system, taken at a specific point in time. It was further confirmed that the framework should work especially well with non-evolutionary data, such as risk management system snapshots and formal reporting structures. This shows how versatile the framework is, as it can be applied to a wide variety of data sources outside of CRRs as well.

The comments also highlighted the importance of the criticality weights, as the risks, controls, control owners, risk owners, risk coordinators, and governance bodies are not equal. Each of these entities would, in practice, have to be weighted individually in order to accurately represent their importance. The importance and worth of controls, for example, are determined by how effective each of them are, as well as how critical they are. It is therefore important that a procedure should be available that can produce a criticality weight that is representative of the true importance of each of the entities. Because of this, a singular method that can be used to determine the criticality weights for a number of entities at once is unlikely to be feasible or reliable. This is not necessarily a shortcoming of the framework, however, as the circumstances for each entity is in a constant state of flux. Determining the importance of these entities is consequently part of an ongoing and evolving process. It was subsequently emphasised that when a concept, such as a control measure, is used in different contexts, care should be taken to ensure that it is appropriate in its own context. In these cases, the opinion is that the method by which weights are assigned to various entities should be refined to ensure that the weights obtained are valid based on their unique circumstances. Stated differently, the criticality weights should be determined in such in a way that they can adapt quickly to any changes that may occur in the network. The framework allows for such changes to be implemented easily, as the weights can be changed immediately if any changes to the context of risks, controls, etc. are detected.

The opinion was expressed that the framework is an excellent foundation for identifying areas where an organisation can improve their risk management systems. Specifically, because it can identify which entities the organisation relies upon the most, it can be used to improve the robustness of risk management systems. If, for example, a control can be identified that appears to be more important than other controls, steps could be taken to ensure that the identified control is as robust as possible. The expert was not aware that

this particular approach had been considered previously, and that it represents a novel way in which the framework can be applied.

It was also remarked that management oversight and corporate knowledge would have to be integral to any implementation of this framework to real-world risk management systems. This supports the position that Phase 4 of the framework (Develop risk mitigation strategies), in particular, will likely differ significantly from organisation to organisation in the way that it is approached. It was also mentioned that it would be advantageous to optimise the framework for each organisation, as this would further increase its effectiveness and produce even more reliable results. This should be neither a significant nor an insurmountable task, as the continuous adjustments and fine tuning of the weights used by the framework could be implemented into any existing, continuous risk management processes. When all of these points are taken into consideration, the framework would be an asset to any integrated and mature risk function, according to the expert.

One suggestion on how the framework can be improved relates to its seemingly simplified view of risk management. Specifically, the framework has a strong focus on managing risks by "reducing" them. In practice, the primary aim of risk management is to allow for optimal decision making when it comes to risk. As a result, most corporations focus on finding a level of risk that it considers to be acceptable. This generally involves risk-reward scenarios, where rewards are balanced with the risks that come with them. This balancing act is often accomplished through the effective and efficient use of resources. The suggestion, subsequently, is that the risk-reward principle should be kept in mind when the framework is implemented, with a smaller focus on "reducing" risk, and more on managing it.

In summary, the expert was of the opinion that the framework would be an asset to any risk management system. It is capable of providing valuable insights, and can also be used to extract additional information from CRR data that could be used to improve the robustness of existing risk management systems.

9.2. CRITICAL EVALUATION

In this section, the framework is critiqued using selected aspects identified in the literature, from relevant sources such as Yue *et al.* (2007), Armstrong and McCulloh (2010), Whitman and Mattord (2011), and Wangen (2017). The points of critique are presented as questions, and the remainder of the section focusses on answering them constructively. The ultimate purpose of these questions is to provide a critical evaluation of the framework that can be used to objectively assess its worth. The evaluation questions are as follows:

1. Based on the literature introduced in Chapter 2, does the framework describe a risk management process? Is it subsequently reasonable to use the framework in a risk management context?
2. A significant novelty that separates the framework from existing risk management techniques is the use of SNA. Does the framework actually implement SNA, rather than just graph theory? Is the way in which SNA is applied appropriate within the context of the framework?
3. Can the framework be used to manage information security risks?
4. Considering the significant impact that security culture can have on attempts to manage information security risks, does the framework allow for information security culture to be taken into consideration?
5. Information security risks can have a number of causes, such as physical vulnerabilities, human interaction, and poorly structured information systems. Can the framework be used to address any of these risks? If not, which risks can it address?
6. It is entirely possible that the framework could be implemented in situations where other, more established risk management methods and techniques are already in place. Is the framework compatible with these methods? Is integration with these existing methods possible in theory?
7. There are a significant number of risk management techniques and approaches already in use. Is it therefore reasonable to expect that the framework could contribute meaningfully to information security risk management?
8. The framework allows for the use of Self-Organising Maps (SOMs) to investigate information security risk in the network. Is this use of SOMs appropriate? Does it produce useable results?

Each of these questions will now be addressed individually.

Question 1: *Based on the literature introduced in Chapter 2, does the framework describe a risk management process? Is it subsequently reasonable to use the framework in a risk management context?*

The proposed framework, as introduced in Chapter 7, consists of five phases. These phases, called *relationship graphing*, *develop an information security risk profile*, *structural optimisation using risk profile*, *develop risk mitigation strategies*, and *implementation and monitoring* respectively, each address one primary task. The overall structure of the framework is reshown in Figure 9.1.

Recall from Chapter 2 that the structure of a risk management process involves three steps, namely risk identification, risk analysis, and risk control. In the interest of completeness, this process, as introduced in Chapter 2, is reshown in Figure 9.2. The implication is that, in order

for the framework to be valid as a risk management tool, it has to address all three of these steps.

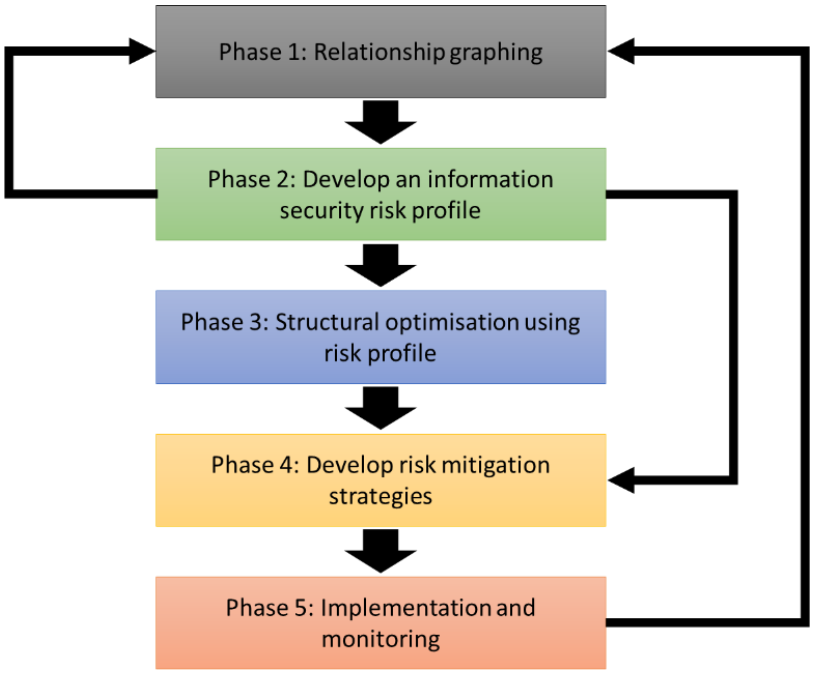


FIGURE 9.1: OVERALL STRUCTURE OF THE NOVEL FRAMEWORK

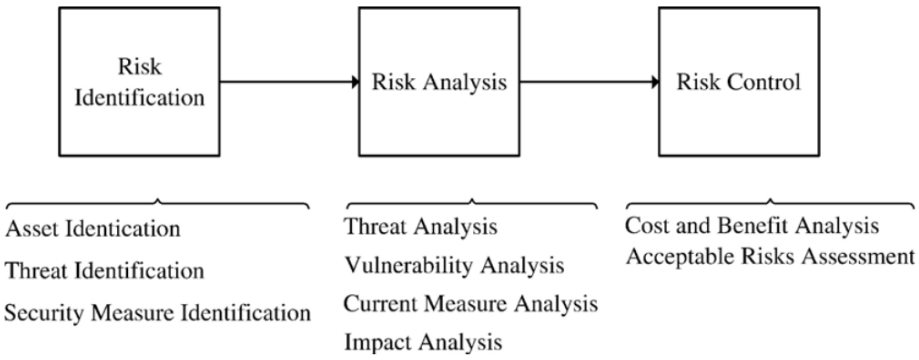


FIGURE 9.2 RISK MANAGEMENT PROCESS (YUE *ET AL.*, 2007)

The first risk management task, namely risk identification, typically involves identifying risk assets, threats, and existing security measures. In the proposed framework, this task is accomplished over the course of the first two phases. In the first phase, certain bordering criteria are selected that are used to restrict which nodes and relationships are included in the network being evaluated. An evaluation of all the possible types of nodes and relationships in the network is therefore necessary, and can be used to exclude certain

nodes and relationships. This helps to identify those assets, known as risk assets, which are considered worth protecting, as nodes and relationships that are deemed worthless can be excluded through bordering. Furthermore, by graphing the network using data that can be obtained for the nodes and relationships most, if not all, of the extant security measures can be identified and accounted for. The threats are effectively identified during the second phase, following the completion of the node-level risk profile. The framework therefore clearly addresses the risk identification task of the risk management process.

The second task in the risk management process, specifically risk analysis, is primarily addressed in the third phase of the framework, but certain aspects of this task are also addressed in the fourth phase. In particular, the third phase deals firstly with identifying those nodes with the highest risk, which corresponds to vulnerability analysis. Secondly, possible ways to address the risk are also investigate in this phase. This in turn ties into impact analysis. The impact of implementing certain strategies is considered in the fourth phase, when these strategies are developed and selected. Ultimately though, while the framework does indeed identify risks and provide for a way to quantify those risks, it cannot be used to identify threats that exist outside the network. It is therefore crucial to emphasise that the framework exists to address risk in the network; other risk management techniques are required for risk emanating from outside the network.

The final task, risk control, is accomplished in the fourth and fifth phases of the framework. In particular, the fourth phase deals with the development and selection of appropriate risk mitigation strategies, whereas the fifth phase controls and monitors the implementation of those strategies. As it would be counterproductive to implement strategies that would cost more than the realisation of the risks they have to mitigate, the development of strategies in the fourth phase addresses both cost-benefit analysis and acceptable risk assessment. The risk control task is therefore also clearly integrated into the framework.

Finally, it is clear that these tasks are not merely present in the framework, but they follow the same basic pattern. In phases one and two, the risk identification task is completed. Based on the result of the risk identification step, the risks are analysed in phases three and four. Finally, the risks are controlled in phases four and five. Therefore, it is clear that the framework does follow a risk management process and can subsequently be considered appropriate for use in a risk management context.

Question 2: *A significant novelty that separates the framework from existing risk management techniques is the use of SNA. Does the framework actually implement SNA, rather than just graph theory? Is the way in which SNA is applied appropriate within the context of the framework?*

In order to address this question, it should first be clarified how SNA differs from graph theory. Recall from Chapter 3 that SNA is used to evaluate interactions using graph theory. This means that, unlike graph theory which investigates networks using quantitative methods, SNA uses the results of graph theory calculations in order to make inferences. Stated differently, graph theory is used, for example, to determine the eigenvector centrality value of a node, whereas SNA uses that value in order to make determinations about a node's emergent leadership qualities. For an implementation of graph theory to be described as SNA, it is therefore critical that the meaning behind a graph theory metric is used to evaluate a node. In this, the framework clearly implements SNA rather than just graph theory, as the meaning associated with each of the selected metrics is continually considered. As for the appropriateness of the way in which SNA is applied in the framework, consider the type of networks being investigated. These networks all contain multiple entities that are linked, through relationships, to one another. In addition, SNA can be used to find and identify relationships that may not be obvious or observable under normal circumstances. All of these relationships can then be investigated for potential information security risks in terms of the CIA triad. Even if the nodes themselves are not human, there is still a clear relationship between them. In this way, the networks can be considered social in nature, and investigating the traits the nodes have using SNA is valid.

Question 3: *Can the framework be used to manage information security risks?*

As previously demonstrated, the framework can be applied in a risk management context. The question remains, however, if it is appropriate for managing information security risks specifically. In order to address this, consider the associations that are made between the three members of the CIA triad, namely confidentiality, integrity, and availability, and selected SNA metrics in Chapter 4. As shown, it is possible to describe the information security risk of a node using selected SNA metrics. Closeness centrality, for example, can be used to identify the level of risk to confidentiality for a particular node. Considering the fact that the framework can be exploited to manage risks, and that it incorporates SNA metrics that can describe information security risks, the framework can indeed be implemented to manage information security risks specifically.

Question 4: *Considering the significant impact that security culture can have on attempts to manage information security risks, does the framework allow for information security culture to be taken into consideration?*

One of the advantages to developing a framework, rather than a rigid method, is that the minutiae of the framework can be modified to fit the organisation it is applied to. With regard to information security culture, the specific implementation of the framework can be

adjusted in multiple ways in order to take it into account. The first way is to identify and use risk mitigation strategies that take the information security culture of the organisation into consideration. This is one of the reasons why the development of risk mitigation strategies is outside the scope of this study: the efficacy of a strategy depends significantly on the organisation it is applied to, and is therefore also affected by the security culture of the organisation.

Another way in which the information security culture of an organisation can be accounted for in the framework is found in the methodology used to determine weights for the nodes and metrics in the node-level risk profile. If, for instance, a specific group of people are known to engage in openly risky behaviour, their corresponding nodes in the node-level risk profile can be weighted accordingly. In a similar fashion, if a person is known to be extremely responsible and exceptionally cautious, then that node can be assigned a weight that reduces its risk value accordingly. Due to these attributes, it is clear that the framework can indeed be used in a way that accounts for the information security culture of an organisation.

Question 5: *Information security risks can have a number of causes, such as physical vulnerabilities, human interaction, and poorly structured information systems. Can the framework be used to address any of these risks? If not, which risks can it address?*

When applied correctly, and provided appropriate data are used, the framework can potentially be used to identify a wide variety of information security risks. Specifically, if an organisation, or even a system, can be represented as a graph, the framework can be applied to identify and manage potential information security risks in that organisation. There are limitations to the types of risks that can be identified, of course. In particular, the framework requires mostly complete information about the nodes and relationships of the network it is used with, and cannot be used to identify risks that lie outside the network. For example, it may be possible to identify those individuals that would be ideal targets for ransomware attacks, but the nature and timing of the attack would be difficult to determine. One of the best applications of the framework is to manage risks to information flow and quality, i.e. confidentiality, integrity, and availability.

Question 6: *It is entirely possible that the framework could be implemented in situations where other, more established risk management methods and techniques are already in place. Is the framework compatible with these methods? Is integration with these existing methods possible in theory?*

In Chapter 2, a number of risk management techniques, such as CORAS and ISRAM, are introduced. All of these techniques can potentially be used in conjunction with, and even

integrated into, the framework. This is mainly due to the flexibility built into the framework. If, for instance, the risk level of a node does not match the risk obtained using other techniques, the weight values for the nodes and metrics can be adjusted to better reflect its true risk. If the chances of a risk being realised is especially small, the values for the nodes affected can be adjusted accordingly. Lastly, if a high-risk node should be removed from the network completely, the framework can be used to inform the best way to reassign the node's responsibilities. The framework can therefore be used in conjunction with other risk management techniques, and integration with these techniques is indeed possible.

Question 7: *There are a significant number of risk management techniques and approaches already in use. Is it therefore reasonable to expect that the framework could contribute meaningfully to information security risk management?*

The core premise of the proposed framework is that SNA can be used to develop risk mitigation strategies that manage information security risks by adding and removing relationships in a network. Alternatively, depending on the strategies selected, new nodes could even be added to improve the risk levels in the network. Unlike other risk management approaches, this framework therefore attempts to introduce improvements in more subtle ways. Consider, for example, how the risks are managed with regard to at-risk individuals. One approach is to introduce stricter security policies aimed at the at-risk individuals, or more intrusive awareness programmes. While both of these approaches are valid in the correct context, the framework can be used to reduce the risk by subtly adjusting the structure of the underlying social network. As a result of this, it is reasonable to expect that the framework will contribute meaningfully, as it aids the risk management process by targeting causes that other methods do not address directly.

Question 8: *The framework allows for the use of Self-Organising Maps (SOMs) to investigate information security risk in the network. Is this use of SOMs appropriate? Does it produce useable results?*

SOM is a neural technique that uses the attribute data for entities to cluster them into regions on a two-dimensional map, which is similar to how the human brain sorts objects into groups. In doing so, the SOM allows for the entities to be evaluated graphically in a way that feels more natural than alternative visualisation techniques. In the framework, SOMs can be used to cluster and visualise the nodes in the network by using the same SNA metrics that are used to evaluate the risks for these nodes. As shown previously, there are metrics that can be associated with specific CIA risks. This means that, by using these same metrics to generate a SOM, the risk in the network can be visualised in a way that is relatively easy to evaluate graphically. In addition, the SOM can also be used to identify nodes that have a

similar risk profile despite being completely different. This is made possible by the clustering that is done by the SOM technique. All of this means that SOM can be used to evaluate and monitor risk in a network, as changes can be identified graphically. This makes the technique useful in managing risk. Therefore, not only is the use of SOMs in this manner appropriate, it produces results that are both useable and valuable.

9.3. CONCLUDING REMARKS

The proposed novel framework, first introduced in Chapter 7, was evaluated in this chapter using two separate approaches. The first approach, which relied on a review by an expert, highlighted singular points of interest. The most significant of these is that the framework will likely have to be optimised for each specific organisation that implements it in order to produce the best results. However, even without these adjustments, the framework is certain to be a valuable asset to any risk management system. The results of the expert evaluation are considered to be mainly positive, which confirms that the framework is useable as a corporate risk management technique. Based on the expert's feedback, it is also clear that the framework is flexible enough for incorporation into existing risk management systems. This observation is based on the fact that risk management operates in a continuously changing environment, and the framework was designed with appropriate levels of adaptability.

One comment, namely that the framework seems to convey a simplified view of risk management in some cases, should be considered. While it is indeed the case that the framework is mostly presented as a technique that can be used to mitigate risks, its ultimate use in practice would be highly influenced by the approaches followed in Phase 1 to collect network data, and the way in which strategies are developed during Phase 4. This opinion, while completely valid, does therefore not address a fundamental shortcoming. As such, it does not diminish the utility of the framework, but rather strengthens the position that it has to be implemented properly, and that the weights and weighting techniques used have to take the context of each entity into consideration.

The second evaluation approach relied on a critical evaluation of the framework. This critical evaluation was done by identifying a number of crucial points from the literature, and then using those points to consider the worth of the framework. During the course of this evaluation, several crucial points were confirmed about the framework:

- The framework adheres to the principles of proper risk management processes;
- It implements SNA in an effective and relevant way;
- Information security risks can be managed using the framework; and

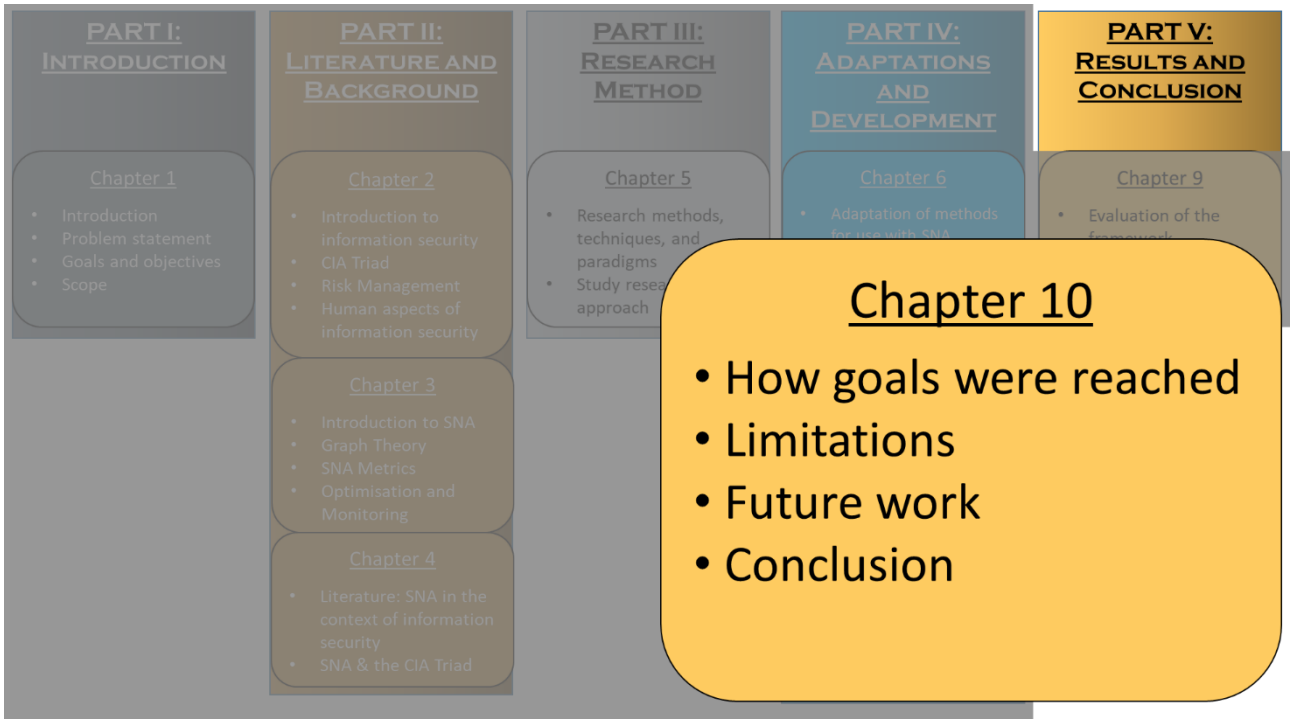
- Applying the framework to existing risk management systems without causing problems is feasible.

When considered together, these points show that the proposed framework satisfies the requirements for an information security risk management technique. Furthermore, when considered alongside the comments from the risk management expert, these points confirm that the framework is not only useable in a risk management context, but also that it will make a positive contribution to any risk management system that implements it. Lastly, the framework has shown that it is capable of producing insight into CRR data that more traditional risk management techniques do not provide. This demonstrates that, even if it is not implemented exactly as presented in this study, the framework has the flexibility to be useful in a vast number of risk management systems.

9.4. CHAPTER SUMMARY

In this chapter, the proposed framework that was presented in Chapter 7, and applied to real-world data in Chapter 8, was evaluated. The evaluation was done using feedback that was received from an expert in the field of risk management, as well as a critical evaluation. The framework was shown to be both useable and valuable to real-world risk management systems.

In the next chapter, the synopsis of the entire study is revisited and its contributions are considered. Additionally, the limitations are clarified and areas for possible research are presented.



CHAPTER 10: SUMMARY AND CONCLUSION

CHAPTER HIGHLIGHTS:

- What is the synopsis of the study?
- What are the contributions of the study?
- What are the limitations of the study?
- What future work can be done based on this study?

10 SUMMARY AND CONCLUSION

In this, the final chapter, the study is concluded. The chapter starts with a synopsis of the study, in which the contents of each of the chapters are discussed briefly. This is followed by a description of how the goals and objectives, first introduced in Chapter 1, were achieved. The contributions of this study are then described, followed by a consideration of the limitations. The chapter concludes with a brief discussion on possible future work that could be done, based on this study.

10.1. SYNOPSIS OF THE STUDY

The study contains ten chapters, which are separated into five parts. The first part (one chapter) serves as the introduction to the study, and introduces the research problem, objectives, and overall structure. The second part (three chapters) focusses on literature and background, specifically relevant literature in the fields of information technology and social network analysis. In the third part (one chapter) the research methodology is discussed, and the structure of the research process is presented. A number of techniques and adaptations are discussed in the fourth part (three chapters), including a novel framework that implements the adapted techniques. The framework is also applied to selected networks and its functioning and validity is demonstrated. In the fifth part (two chapters), the framework is critically evaluated, and the study is concluded. The contributions, limitations, and future work are also highlighted in the fifth part.

PART I

In Chapter 1, the premise of the study was introduced, and some of the problems with existing information security techniques were discussed. Specifically, it was highlighted that certain approaches used to address the human aspect of information security may not necessarily be sufficient. Social Network Analysis (SNA) was then proposed as a technique that could potentially be used to address these shortcomings, with selected references to how SNA has been used in the past. The formal research question, namely "*Can social network analysis be used to develop risk management strategies in an information security context?*" was then presented. In order to answer this research question, nine objectives were specified. These objectives were:

1. **Conduct** a review of the literature, and use the information and insight gained from it to compile informative sections on the following topics:
    (a) Information security, specifically focussing on its core principles, various approaches to addressing the human aspects, and risk management strategies;
    (b) Network theory, including network metrics and visualisation techniques, with special attention given to appropriate sections of graph theory and its relevance to SNA; and
    (c) Past instances of SNA being used to address information security risk.
2. **Adapt** relevant existing methods so that they can be used in conjunction with SNA in the context of this study;
3. **Develop** a novel method framework that incorporates SNA that can be used to evaluate information security risk in an organisation and propose improvements that will address those risks;
4. **Demonstrate** how the novel method can be applied to real-world data;
5. **Investigate** the viability of applying Self-Organising Maps (SOMs) as a visualisation- and data processing technique to the novel method;
6. **Demonstrate** how the method, utilising SOMs, can be applied to real-world data;
7. **Critically evaluate** the method and its applications, and provide an overview of its advantages and potential;
8. **Identify** possible shortcomings in the method and propose how future research may improve the method;
9. **Suggest** sensible and related future work that may follow from this study.

The remainder of the study focussed on achieving these objectives. Each of the five parts, with the possible exception of Part III, were aimed at addressing these goals. Part II dealt with the literature, while Part III introduced the research methodology and approach. In Part IV various techniques and adaptations were discussed, and a novel framework was developed and demonstrated. Part V focussed on the evaluation of the framework, as well as concluding remarks. Each of the remaining parts, and how the objectives were reached, will now be discussed briefly.

PART II

The second part of the study was primarily concerned with the literature that served as background for the study. In Chapter 2, the main concepts of information security were introduced. A number of relevant topics were discussed, such as the CIA triad (confidentiality, integrity, and availability), various risk management approaches, and the human aspect of information security. Chapter 2 therefore served to satisfy Objective 1(a),

as it contains a review of the literature which was **conducted** to gain insight surrounding information security topics.

The second chapter in Part II, Chapter 3, introduced SNA and its associated principles. The chapter started by introducing the fundamental ideas behind SNA, and then switched focus to introduce selected, and relevant, topics from the field of graph theory. Various graph theory terms, such as edges, nodes, walks, and cliques were introduced. Selected examples of visualisation techniques were also shown, and Self-Organising Maps (SOM) were proposed as a possible method for visualising social networks. A number of selected SNA metrics were presented and discussed, as well as techniques that could be used to identify possible communities in social networks. The chapter concluded with the introduction of Statistical Control Charts as a monitoring method that can be used to detect changes in social networks. Chapter 3 served to address Objective 1(b), as the literature study that was **conducted** on the background of SNA and related topics provided significant insight.

The final chapter in Part II, i.e. Chapter 4, combined the principles discussed in Chapters 2 and 3, and investigated the interaction between SNA and information security. The first part of the chapter dealt with investigating and reviewing existing research on the use of SNA with information security. The various ways in which SNA metrics can be used to identify potential CIA risks were then introduced. A simple example was used to explain this interaction further. The chapter concluded with tabulated lists of additional relevant literature sources. By **conducting** a review of the extant literature, this chapter provided insight into how SNA could be used to investigate information security risks. This chapter therefore satisfied Objective 1(c).

PART III

The third part of the study was concerned with the research methodology, methods, and techniques. The topics were discussed in Chapter 5, which started by introducing the Research Onion model. The remainder of the chapter was structured according to this model, and subsequently introduced a large number of concepts, such as research philosophies, methodological choices, and research strategies. The chapter concluded with a discussion of the approach employed in the study, and closed with an introduction of the overall structure and process followed in the study. Part III is a crucial part for a number of reasons. The first is to ensure that the study adheres to accepted scientific principles, and to establish a proper scientific research process. Furthermore, by clearly identifying the research philosophy, methods, time horizon, etc., the method used to report on the research can be improved. Finally, by properly planning the research according to established scientific principles, the chances of an important piece of evidence being overlooked are minimised.

PART IV

The penultimate part contains the majority of the novel contributions, and subsequently the chapters in this part addressed the majority of the objectives. The first chapter in this part, i.e. Chapter 6, focussed on existing techniques and how they can be **adapted** for use with SNA. The chapter centred around three techniques in particular. The first was a network optimisation technique that utilised SNA metrics. The original technique is aimed at physical networks, but was adapted for use with social networks. The second technique is an application of SOM, whereby the SNA metric data for a network was used to generate a map that visualises potential information security risks in the network. The third, and final, technique discussed in this chapter utilises SNA to identify optimal distribution points when developing security awareness programs. In this chapter, Objective 2 was reached, and Objective 5 was partially achieved.

The next chapter, i.e. Chapter 7, introduced a novel framework that can not only be used to identify nodes in a social network that have a high information security risk, but also develop risk mitigation strategies to deal with them. The first part of the chapter discussed the **development** of the framework, while the second part **demonstrated** the functioning of the framework using two distinct networks. The first demonstration used network data collected from a university department, while the second demonstration made use of data for a Twitter network that was publically available. In this chapter, Objective 3 was reached, and Objective 4 was partially addressed.

Chapter 8, which concludes Part IV, contains an application of the framework to a corporate risk management network. The purpose of this application was to **demonstrate** how the framework functions when applied to a large real-world network. The chapter started by introducing the risk management network, and how the data used to graph the network were obtained. The network was graphed and visualised, and the selected metrics were calculated for the network. The metrics were then used to develop a SOM, which was investigated graphically. Objectives 5 and 6 were achieved completely following this **investigation**, as the SOM technique was fully **demonstrated** using real-world data. The remainder of the chapter presented the results following the use of the optimisation technique presented in Chapter 6, and concluded with a discussion of the impact the identified changes could have on the levels of information security risk in the network. Objective 4 was therefore fully reached in this chapter, as the novel framework was applied to real-world data, **demonstrating** its validity.

PART V

The final part dealt with the results of the study, in particular the outcomes of the application of the framework to the risk management network. Chapter 9, which was the

first chapter in this part, started with a discussion of the comments received from an expert in the field of risk management. The second part of the chapter presented a **critical evaluation** of the novel framework and the adapted techniques. In the final chapter, Chapter 10, possible limitations were **identified** and future work was **suggested**. This chapter also contains a summary of the research contributions and objectives. In Part V, Objectives 7, 8, and 9 were therefore achieved.

10.2. CONTRIBUTIONS

The primary contribution of this study is the development of a framework that implements SNA in order to develop information security risk mitigation strategies. The framework contributes to risk management systems by utilising SNA to identify potential risk areas in the structure of a network, and then identifies ways in which the risks can potentially be reduced. Both formal and informal network types can be investigated, which makes it possible to apply the framework in a wide variety of situations. Furthermore, because the framework does not absolutely specify which metrics and techniques should be used, it can be improved and fine-tuned over time, making it a valuable management tool.

During the development of the framework, a number of methods were identified that can be used with SNA if adjusted appropriately. These methods were all modified for use with SNA, and the adaptations are therefore considered contributions:

- The network optimisation technique presented in Chapter 6 was originally developed to help optimise physical computer networks. However, because the technique used graph metrics to identify optimisations, it was mostly appropriate for use with social networks. This method nevertheless still required adaptation, as social networks have inherent properties that make them more difficult to alter than physical networks. The technique was subsequently adapted to allow for these properties, and the optimisation technique as used for the remainder of the study was obtained. This adapted technique differs from the original to a significant enough degree to be considered novel.
- Prior to this study, SOMs had already been used to visualise social networks. The application used in this study, however, did not visualise the network itself, but rather the SNA metrics associated with particular information security risks. This novel application allows for SOMs to be used to investigate information security risk in a social network in a graphical manner, which is especially valuable on managerial level. Also, because SOMs cluster data, they can be used to identify similarities that would have remained hidden otherwise. Another application of SOMs presented in this study is their use in determining weight values for SNA

metrics. Using SOMs in this manner allows for metric weights to be determined in a non-subjective way, which improves the reliability of the optimisation process.

- While the primary focus of the study is on developing a framework that can be used to identify and address information security risks in a social network, this is not the only way in which these risks can be addressed. A more traditional method for improving information security culture involves the use of security awareness programmes. By applying SNA to the planning process involved when developing these programmes, the overall efficacy of the programme can potentially be improved, whilst also reducing costs. This is also a novel application of SNA that can be used to improve overall awareness, and subsequently reduce risk.

The study also succeeded in developing informative sections on information security and SNA that would be valuable to those unfamiliar with these topics. These sections, i.e. Chapters 2 and 3, are presented as novel contributions, as the composition of the information contained in these sections, as well as the style of presentation, is distinct from other sources.

A further contribution was made by not only developing a novel framework, but demonstrating its utility by applying it to a real-world network with a large number of nodes. In doing so, not only was it shown that the novel framework can be applied to networks of the size expected in a corporate environment, but that the framework itself is scalable to a network of any size.

Another contribution was made with regard to the interaction between SNA and the CIA triad. Prior to this study, SNA had been used to identify organisational risks, but had never been used in conjunction with SOM to identify specific CIA risks. By considering the meaning of various SNA metrics, and then using those meanings to identify CIA risks, a new approach to investigating information security risk was introduced. This principle, whereby the meaning of the SNA metrics is used to evaluate information security risks, is therefore also a valuable contribution.

The final contribution presented is the identification of relevant future work. By investigating the research and discerning areas that could lead to future research, it is possible to provide guidance on how to further expand the fields of SNA, information security risk management, and SOM applications.

10.3. LIMITATIONS AND FUTURE WORK

Due to the nature of the research method and the scope declared in Chapter 1, this study has the following limitations:

- The timeframe for this study meant that the framework could only be used to evaluate the network of a single real-world organisation. The different ways in which different risk managers would apply the technique to their specific organisations could therefore not be investigated.
- Ideally the implementation and impact of the application of the framework to real-world organisations would be investigated over a period of multiple years. Unfortunately, a longitudinal study of sufficient scale could not be completed within the limited timeframe of this study.

In part, due to these limitations, a number of areas are available for future work. Firstly, a research project could be initiated wherein the framework is applied in a real-world network over a significant period of time. This should help identify possible areas for improvement, and will confirm the utility of the framework. Another avenue for research is to investigate methods to improve on the weighting of the nodes and metrics contained in the node-level risk profiles. This will likely also be a longitudinal study, as identifying the best weighting technique will require significant amounts of long-term risk data.

10.4. CHAPTER SUMMARY

In this chapter, the study was concluded. The chapter started with a synopsis of the study, whereby each of the parts and chapters were discussed. The various objectives identified at the beginning, and how they were achieved, were also highlighted. The contributions made by this study were then presented, followed by a brief overview of the limitations of the study, and possible future work.

A APPENDIX A: PUBLISHED ARTICLES

During the course of this study, two academic papers were produced and successfully published. Both these papers were presented at the *International Federation for Information Processing (IFIP) World Conference on Information Security Education (WISE)*, and were subsequently published as book chapters in *Information Security Education. Education in Proactive Information Security*, which is part of the Springer *IFIP Advances in Information and Communication Technology* series. Both these articles therefore have an associated ISBN and have been indexed in Scopus.

The first paper, entitled *The feasibility of raising information security awareness in an academic environment using SNA (Serfontein et al., 2018)*, was presented at WISE 11 in 2018, and subsequently published in the same year. In this paper an application of Social Network Analysis (SNA) to security awareness programme planning is presented. The final publication for this paper is available at Springer via https://doi.org/10.1007/978-3-319-99734-6_6.

The second paper, entitled *Identifying Information Security Risks in a Social Network Using Self-organising Maps (Serfontein et al., 2019),* was presented at WISE 12 in 2019 and published in the same year. This paper presents the Self-Organising Maps (SOM) information security risk evaluation technique that was first discussed in Chapter 6, and was then applied to the CRR network in Chapter 8. . The final publication for the second paper is available at Springer via https://doi.org/10.1007/978-3-030-23451-5_9

The feasibility of raising information security awareness in an academic environment using SNA

Rudi Serfontein[1[0000-0002-0428-6494]], Lynette Drevin[2[0000-0001-9370-8216]], Hennie Kruger[3[0000-0001-8514-4422]]

North-West University, Potchefstroom, South Africa
[1]rudi.serfontein@nwu.ac.za
[2]lynette.drevin@nwu.ac.za
[3]hennie.kruger@nwu.ac.za

Abstract. The human aspect is one of the key success factors in information security (InfoSec). Its impact on InfoSec is so significant that multiple studies have shown that a balanced approach combining technology and security awareness is needed in order to maintain the integrity of an organisation's security. At present, one of the methods most often used to address InfoSec awareness is to develop security awareness programmes that can be used to educate its users within an organisation. This method has several drawbacks; however, as such programmes might not be comprehensive enough, or quick enough to address newer threats. It can furthermore lead to the users developing InfoSec fatigue, which renders most attempts at improving security awareness pointless. These problems are compounded by non-traditional organisational structures, such as those found in educational institutions, where both students and staff should be made aware of information security risks on a regular basis. In order to address the potential information security awareness problem at educational institutions, this paper investigates the feasibility of using Social Network Analysis (SNA) to improve existing security awareness programmes. Following a brief introduction to SNA, two illustrative examples are offered to show that SNA presents a viable option to improve programmes for raising information security awareness in an academic environment, by allowing for the effective selection of ideal target locations.

Keywords: Social network analysis, Security awareness, Security fatigue

1. Introduction

In the field of information security, one of the primary success factors is the human aspect [1]. Past research has shown that a balanced approach in which both technological and social aspects are addressed is crucial to maintaining information security [2-4]. Despite repeated campaigns to educate users regarding information security, however, a significant number of users still engage in risky online behaviour [5] and are still considered the weakest link in information security [6]. Among the many places that can be negatively impacted by a lack of information security awareness, few are as vulnerable as universities. This stems from the fact that university networks need to be accessible to a wide variety of people, such as students, faculty members, administrative staff, and visitors [4]. With the massive

number and types of people that need to be able to access a university network, it is only reasonable to assume that a significant number of users will act in a way that compromises both the security of the university and their own personal security. One of the best known traditional methods of addressing this risk and educating users is security awareness programmes [7-9]. There are, however, a number of significant drawbacks to these awareness programmes, e.g. the awareness programmes might not be comprehensive enough [10], they might not address new threats quickly enough when the risks change continuously [11], and the programmes rely upon the users to consciously decide to comply with information security principles [12]. A significant amount of research is focused on attempting to address these shortcomings [13]. Another factor that may impact negatively on security awareness training is security fatigue. Security fatigue is a specific form of mental fatigue, which is a well-known phenomenon in psychology that describes the feeling a person has during or after prolonged periods of cognitive activity [14, 15]. Security fatigue is experienced by users when they are bombarded with information security knowledge to such a degree that they become overburdened with the information and may choose to abandon all conscious efforts to adhere to the security principles as explained during the course of the awareness programmes [16].

Given the importance of the human aspect in information security and the potential problems with broad security awareness programmes, an adaptive approach is proposed. In this paper, the feasibility of using Social Network Analysis (SNA) as a technique to positively influence information security awareness programmes, specifically those that are targeted at an academic environment, will be discussed. SNA is a method used to graphically represent a social organisation, such as a community or business, in such a way that the social interactions can be studied quantitatively [17]. The technique is suitable for use in environments where certain risks, including those risks associated with information security, are present, and has been used in the past to, among others:

- Identify core members and organisations within terrorist groups [18]; and
- Identify hierarchies in criminal Dark Web forums [19].

In addition to the studies mentioned above, SNA has also been used to enhance the information security of an organisation. The work done by Dang-Pham, Pittayachawan and Bruno [20] is of particular interest to this study as it serves to demonstrate the validity of the method discussed here. In the study done by Dang-Pham, Pittayachawan and Bruno, SNA was used to identify individuals who would be able to serve as information security champions. These individuals were then trained in information security so that their influence would help to shape the workplace culture with regard to information security. Because of the importance of this method, it will be referred to as the DPA-method (Dang-Pham Awareness) in the remainder of the paper. SNA has also been used in different studies to identify individuals who pose an organisational risk. By calculating the relative SNA metrics for the various nodes, individuals who may pose a risk due to their position in the network can be identified [21, 22].

The purpose of this paper is to address the information security awareness shortcomings that may exist in university classes and faculties by employing an SNA

approach. As the method can be applied to target important individuals and locations using both formal and informal social structures, it should prove useful when developing targeted awareness programmes that can be used to inform staff and students alike. Once these central individuals and locations have been identified, security awareness programmes using classic awareness items such as posters, pens, brochures, discussions, etc. can be used to inform people about security issues and thereby improve security awareness [23]. The purpose of the method proposed in this paper is therefore not to revolutionise traditional security awareness programmes, but merely to provide a way to improve their effectiveness and coverage in situations where security education and –training would not be feasible, and full-scale awareness programmes may be prohibitively expensive, or cause unwanted fatigue.

The remainder of this paper is organised as follows. In the next section, introductory background information with regard to some SNA metrics is provided. This is followed in Section 3 with the discussion of the proposed method, and two illustrative examples. A discussion of the findings is presented in Section 4, and in Section 5 the paper is concluded.

2. Background

2.1 Social Network Analysis

Any social organisation can be considered to be a series of interconnected networks, and as such standard graph modelling can be used to represent them. In such a network, nodes can be used to represent entities, such as people, knowledge, tasks or resources, whereas arcs can be used to represent the relationships that exist between them.

SNA allows for the quantitative analysis of a social organisation through graph theory, and various metrics can be calculated in order to analyse a network. Although a large number of metrics exists (a count of the work done by Clemente, Martins and Mendes [24] shows 28 metrics, whereas the help section of the ORA-Lite software suite names almost 200), only four basic metrics that are used in the illustrative examples will be briefly introduced. The discussion of the four SNA metrics is based on the work done in [21]. Comprehensive discussions of a large number of metrics can be found in a number of sources, such as [24], [25], and [26].

Degree Centrality. The degree centrality measure is concerned with an individual node and more importantly the particular node's position within the network [21, 27]. A node's ability to influence a particular network is governed by its position within the network, and this in turn is referred to as the node's centrality measure [21]. There are a number of different types of centrality, but the core principle of centrality is that a node that is located more centrally, i.e. has more specific connection types than other nodes, will have a greater specific influence on the network as a whole. One of the quantitative measures used to describe the influence of such a node is referred to as its total degree centrality, and is calculated by using several node properties, such

as the number of connections leading into the node, the number of connections leading out of the node, and the sum of the aforementioned connections [21]. A node with a high total degree centrality would be an excellent target for security awareness training, as any information injected into the network at this point is likely to propagate to the rest of the network in some way.

Closeness centrality. Closeness centrality is calculated by determining all the geodesic distances (i.e. the shortest distances) to all other nodes within the network [21], and takes all indirect connections to other nodes that a node possesses, together with all direct connections, into account. A node that has a high closeness centrality value is considered to be a good source of information, whereas nodes with a high degree centrality value aids in the diffusion of information throughout the entire network. This means that analysis of the nodes with the greatest closeness centrality values should provide the best information with regard to the information in the network, and would therefore mitigate the need for full node-by-node network analysis.

Betweenness centrality. When examining interactions between two non-adjacent nodes, the nodes that lie on the paths connecting the two nodes have some control over the interaction between the two nodes [28]. The betweenness centrality measure is a representation of the number of times that a particular node finds itself on the geodesic path of other nodes within the entire network [21]. This measure is reflective of the number of indirect nodes that are connected to a particular node. Thus, a node that has a high betweenness centrality measure would also be a good candidate to use to distribute knowledge and information throughout the network, as these types of node are exclusive, limited sources of information for parts of the network. There is, however, a downside to using such a node: a node with a high betweenness measure is at risk of being overburdened, as such a node would spend a portion, if not all, of its time facilitating interactions between other nodes.

A node that finds itself as an intermediary in an information exchange relationship between two nodes is also considered to be in a position of power, as any information exchanged between the two nodes has to go through the intermediary. The intermediary has a unique position of power in this instance, as it can determine not only the fidelity of the information being exchanged, but also whether information is exchanged at all. Thus, as the number of nodes that relies on such an intermediary increases, so too does the relative power the intermediary node possesses.

Eigenvector centrality. Eigenvector centrality measures the extent to which a particular node is connected to other nodes that are considered to be highly connected or are of some particular importance [21]. Nodes that have a high eigenvector centrality value are important to note since they are considered to possess emergent leadership properties [29]. Nodes with a high eigenvector centrality are therefore also considered good targets for security awareness, as they tend to take on the roles of early adopters.

2.2 Network formality

The formality of a network within the context of this paper is a measure of how formal the relationships that are used to construct a social network are. A highly formal network will utilize formal relationships, such as reporting structures, while a less formal network will make use of what is known as informal information systems (IIS). These systems are of particular interest, as they are found in every organisation and present one of the many places where SNA can be applied. IIS are special types of information system that represent the so-called "grapevine" of an organisation [30]. IIS are characterised by their lack of formal structure, their questionable reliability and their possible incompatibility with formal information systems. Unfortunately, due to their ability to collect a significantly greater subset of data, IIS are often crucial to business processes [31, 32]. It is important to take note of these types of information system, as they can have a profound impact on the flow of information within an organisation and must therefore be considered when developing a method that relies on the characteristics of a social network to improve security awareness within an organisation. Depending on the organisation, it may be necessary to target the social networks associated with IIS, rather than those networks associated with its formal structures, in order to obtain the desired results with regard to information security awareness. In an academic environment, for example, it is important to target both the more formal networks that include relationships, such as reporting structures and teaching responsibilities, and the less formal networks, such as those that include social relationships between students.

3. Method

The methodology employed in this paper broadly follows the DPA-method, with a number of notable exceptions:

- The DPA-method uses formal networks constructed from an organisation's hierarchy, whereas the method proposed here targets both formal and informal social networks;
- The method proposed in this study specifically targets personnel and students in an academic setup, such as a university, rather than an organisation; and
- The information security awareness programmes developed using the method in this study can be used to structure a programme that ideally targets the expected awareness level of a group, whereas in the DPA-method a number of influential employees are fully trained in information security awareness.

The proposed method is executed in three primary phases: Preparation, Network Construction, and Evaluation and Implementation. The basic process of the method is shown in **Fig.** .

The first phase, namely the **Preparation** phase, focusses on developing a clear and congruent approach to implementing the method. During this phase, a number of issues crucial to obtaining useful SNA data are addressed. The first of these issues deals with properly "bordering" the group the awareness programme is to target. In an academic environment, bordering may include aspects, such as field of study, the

faculty they belong to, their lecturers, etc. This phase also focusses on determining the scope and formality of the networks that will be used.

The **Network Construction** phase is primarily focussed on collecting and processing the network data needed to identify the target individuals. This phase focuses on selecting data collection methods that can be used to construct social networks. These methods may include questionnaires, email-scanning, class-list processing, etc. if a more informal network was selected. Otherwise, formal organisational structures, such as reporting hierarchies can be used, which negate the necessity of using intrusive techniques, such as questionnaires and email-scanning. Once the members of the group have been identified and the nature of the relationships between them has been established, the social network can be constructed. This, along with the calculation of the metrics, is ideally done using software. In this phase, the impact of selecting a more formal or a less formal network will also become clear. Should the impact of the network formality be too great in a negative sense, the Preparation phase should be repeated in order to either negate or mitigate the impact.

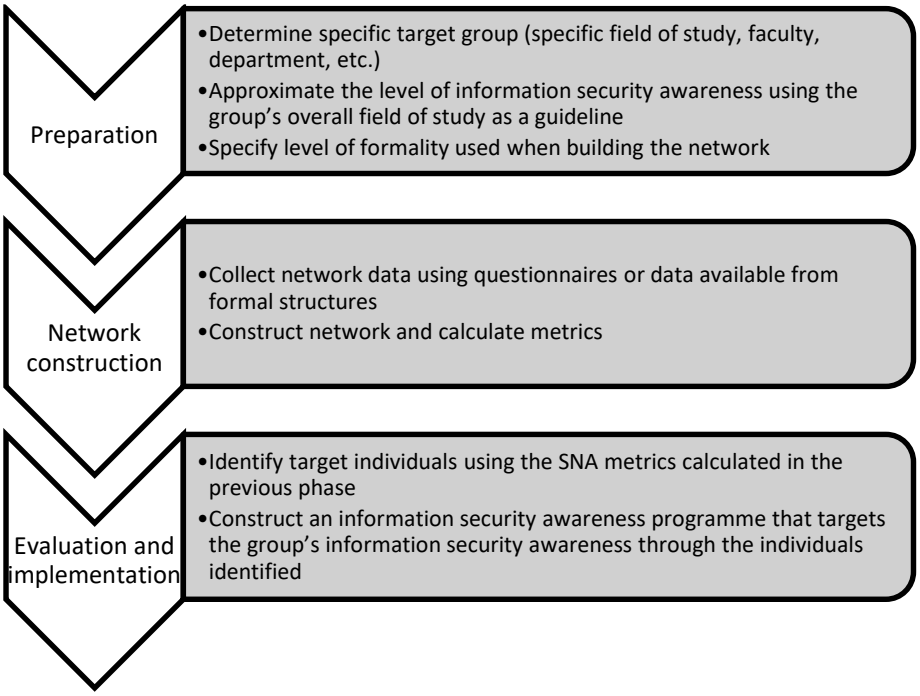


Fig. 1**.** Process of the proposed method, showing progression through the three phases

During the final **Evaluation and Implementation** phase, the data from the previous two phases are used to determine both the contents of the awareness programme and its intended targets. The specifics of the awareness programme's contents will likely differ from case to case, as the programme should be adapted to the targeted individual, as well as the group in general.

3.1 Illustrative examples

To illustrate the feasibility of the proposed approach, two practical experiments were conducted. In the first experiment an informal social network construction approach was used, whereas a formal social network was utilised in the second experiment.

Case study 1. During the Preparation phase a target group of 25 post-graduate students was chosen. An informal social network construction approach was decided upon, as there were no significant formal connections amongst the students apart from attending the same class. In the Network construction phase data were obtained from the students. The following social question was posed to the students:

Suppose the computer security group is invited to a function by the industry and everyone shows up. The venue is properly decorated and a number of round tables have been prepared, with exactly one chair for each of the students. If you could make the decision, who would you prefer to have on your right- and left-hand side at the table?

The response rate was 68%, which was deemed adequate for demonstration purposes.

Respondents were given the option of choosing from a list of names that correspond to the students registered for the class. The data obtained were analysed using ORA-Lite [33], which was also used to construct the network. The four measures discussed in Section 2 were calculated and are used in the next phase to determine candidates for disseminating security awareness information through the network. The network obtained is presented in **Fig. 2**.

During the final Evaluation and Implementation phase, the calculated measures were evaluated to identify candidates that should be targeted. Results indicated that node LR has the highest betweenness centrality at 0.028, eigenvector centrality at 0.322, and total-degree centrality at 0.174, while node CP has the highest closeness centrality at 0.055. These values indicate that the best singular candidate to target would be node LR. An evaluation of the network shown in **Fig. 2**, however, shows that selecting only node LR will not be entirely effective as there are three distinct, unconnected networks. Therefore, in order to expose the entire network, nodes AG, which is visually the centre of subnetwork B, and node LS in subnetwork C, which has a betweenness centrality of 0.01, an eigenvector centrality of 0.156, and a total-degree centrality of 0.109, should also be targeted.

Case study 2. For Case Study 2 a formal network construction approach was chosen. The relationships between the personnel at a Computer Science department at a South African university and their formal post-graduate students were used. Where duplicate connections were found, for instance where one student received guidance from more than one member of the department, the weight of the existing connection was increased to indicate a closer relationship. The same three phases used in Case Study 1 were used and the network shown in **Fig. 3** was obtained. The data were anonymised, and the node names were chosen to differentiate between students and staff. All node names that contain a D represent staff and all nodes that contain an N

represent students. From **Fig. 3** as well as the metrics calculated from this network, it is clear that nodes D60, D49, D14 and D76 represent the most connected and influential members in this network. Node D60 in particular has the highest value in all four metrics, which indicates that this person is not only an emergent leader within the network, but is also an influencer. This makes sense as this node is a member of the academic staff who has a large number of students that also receives guidance from other members of staff. Node D76 is also a good target as the node has the second-highest total-degree centrality value. The node does, however, have a significantly lower eigenvector centrality value, and the reduced leadership influence may impact the efficacy of using this node as a target. The ideal situation would involve all four of these individuals, namely D60, D49, D14 and D76, being targeted in information security awareness programmes. As these four individuals are likely to have regular meetings or discussions, any information passed to them should disseminate through the network relatively quickly and naturally.

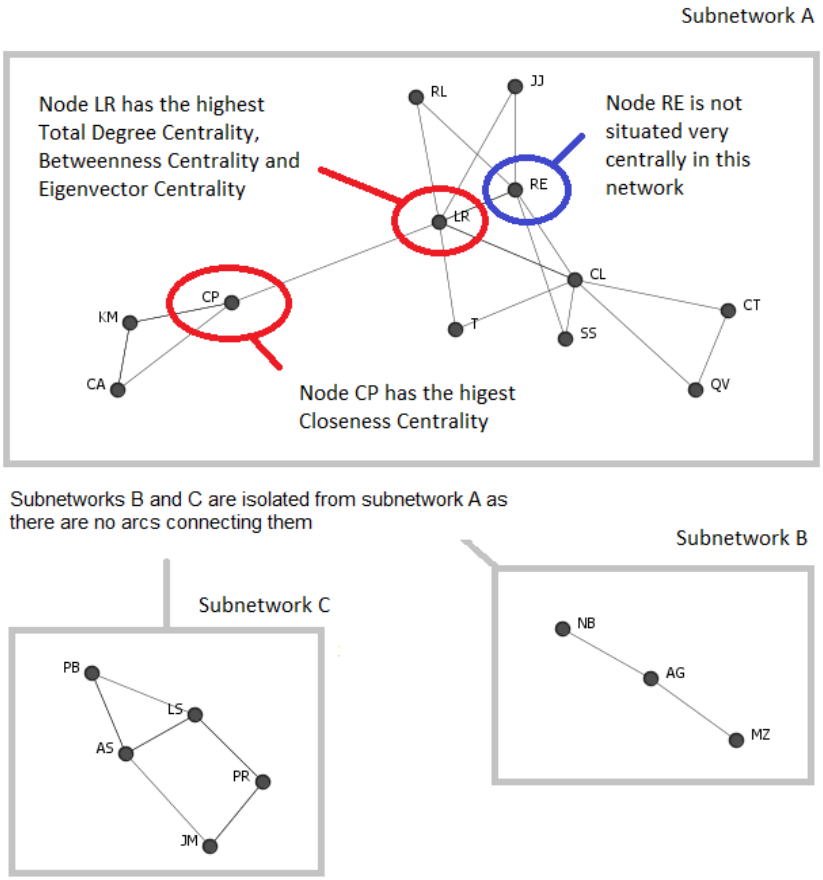


Fig. 2. Social network based on the informal social question

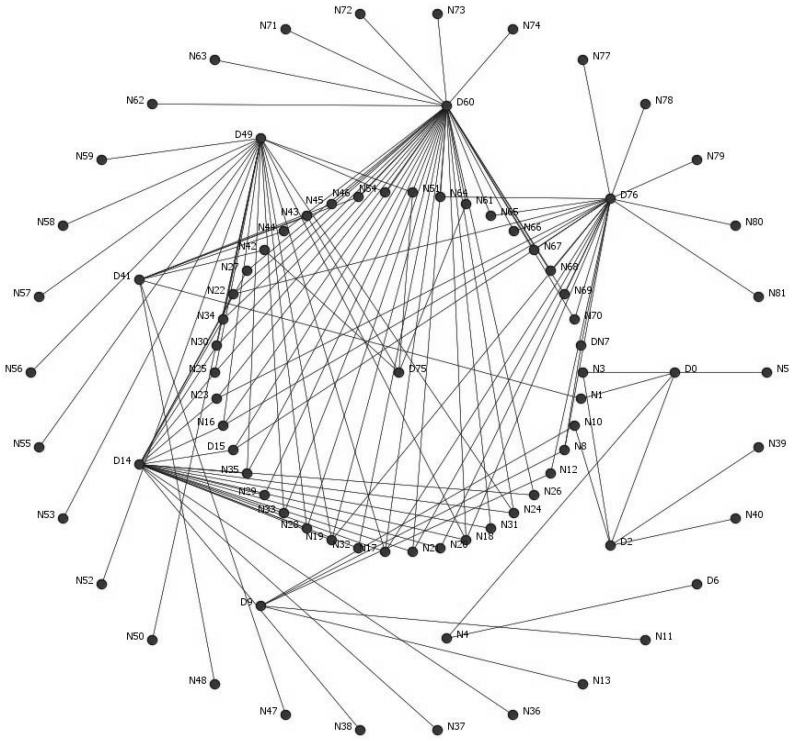


Fig. 3. Formal network of the Computer Science lecturers and post-graduate students

4. Discussion

The organisational structure of universities generally differs from more traditional organisations in that an academic organisation incorporates a large number of students, which are generally not part of a formal management structure. Universities also differ from other organisational structures in that the various departments at a university are quite often isolated from one another. While it is likely that an administration department may have contact with all the various faculties, it is generally quite rare for separate faculties to have frequent contact with one another. These aspects of an academic organisation make it difficult to target both the staff and the students at a university. While possible in theory, in practice it is difficult to provide information security training to both students and staff, as there is no simple way to organise events of such a size. In addition to the logistical difficulties, neither students nor staff like attending awareness training sessions, especially if there is a perception that no new information will be provided. These problems are further compounded by the financial limitations most universities have to implement in order to remain solvent. Security awareness training for a whole university will likely require significant funds. Such an investment, as far as most universities are concerned, offers too little in return.

The two case studies presented in this paper demonstrate that SNA is a feasible alternative to large scale security awareness programmes in an academic environment, due to a number of reasons. The first of these reasons is that both formal and informal techniques and relationships can be used to construct social networks, which means that there is no absolute dependency on specific structures. This is an advantage in academic environments where a comprehensive formal structure may be limited or non-existent. Another reason is that a handful of individuals can be identified for targeted awareness training. This significantly reduces the cost and, as the topics of the awareness programmes can be selected to correspond to the individual's level of information security knowledge, the chances of fatigue are also drastically reduced. A further advantage is that security awareness can be addressed less formally and more consistently: as new threats are identified, the various targeted individuals can be informed with minimal cost and effort. These individuals will also have a known level of information security knowledge, which will make a continuous programme more effective. SNA is also a relatively simple method to implement, as software packages that implement it do not require overly complex data in order to produce results. This makes the technique relatively easy to implement and use. A final advantage is that any number of networks can be constructed concurrently in order to target a large group. If say, for example, two departments have no contact with one another and their internal organisational structures are too distinct, a network can be constructed for each department using bordering techniques that are appropriate to each department.

5. Conclusion

Information security awareness programmes have to be implemented and used with great care in order to be effective. In more traditional organisations, formal awareness programmes are generally used to address information security awareness shortcomings. In academic organisations, where formal structures do not necessarily include all the members of the organisation, such as students, it is often much more difficult to conduct effective security awareness training. In this paper, in an attempt to address some of the problems of conducting security awareness training in an academic environment, the feasibility of using SNA to develop targeted awareness programmes was investigated. Two illustrative examples, one using formal structures and the other informal relationships, were presented to demonstrate that SNA is a feasible alternative to formal awareness programmes in an academic environment. The contribution of this study is that the suggested approach, that may be easier and faster to use, and reduce certain limitations, such as costs, fatigue, and the inclusion of information that is inappropriate for the target audience, is indeed feasible. Future work will include the use of more extensive tests, such as the use of larger sample groups and the monitoring of information security levels, to demonstrate the usability of the presented method. These studies will also show how effective, both in terms of cost and coverage, the proposed method is when compared to untargeted, traditional awareness programmes.

References

1. R. Shillair, S.R. Cotten, H.S. Tsai, S. Alhabash, R. LaRose and N.J. Rifon, "Online safety begins with you and me: Convincing Internet users to protect themselves," *Computers in Human Behavior*, vol. 48, 2015, pp. 199-207.
2. Parsons, A. McCormac, M. Butavicius, M. Pattinson and C. Jerram, "Determining employee awareness using the Human Aspects of Information Security Questionnaire (HAIS-Q)," *Computers & Security*, vol. 42, 2014, pp. 165-176.
3. Z.A. Soomro, M.H. Shah and J. Ahmed, "Information security management needs more holistic approach: A literature review," *International Journal of Information Management*, vol. 36, no. 2, 2016, pp. 215-225.
4. Y. Rezgui and A. Marks, "Information security awareness in higher education: An exploratory study," *Computers & Security*, vol. 27, no. 7–8, 2008, pp. 241-253; DOI http://dx.doi.org/10.1016/j.cose.2008.07.008.
5. Z.S. Byrne, K.J. Dvorak, J.M. Peters, I. Ray, A. Howe and D. Sanchez, "From the user's perspective: Perceptions of risk relative to benefit associated with using the Internet," *Computers in Human Behavior*, vol. 59, 2016, pp. 456-468.
6. N.A.G. Arachchilage and S. Love, "Security awareness of computer users: A phishing threat avoidance perspective," *Computers in Human Behavior*, vol. 38, 2014, pp. 304-312.
7. F.A. Aloul, "The Need for Effective Information Security Awareness," *Journal of Advances in Information Technology*, vol. 3, no. 3, 2012, pp. 176-183.
8. C.C. Chen, B.D. Medlin and R.S. Shaw, "A cross-cultural investigation of situational information security awareness programs," *Information Management & Computer Security*, vol. 16, no. 4, 2008, pp. 360-376.
9. M.E. Thomson and R. von Solms, "Information security awareness: educating your users effectively," *Information Management & Computer Security*, vol. 6, no. 4, 1998, pp. 167-173.
10. M.T. Siponen, "A conceptual foundation for organizational information security awareness," *Information Management & Computer Security*, vol. 8, no. 1, 2000, pp. 31-41.
11. H.A. Kruger and W.D. Kearney, "A prototype for assessing information security awareness," *Computers & Security*, vol. 25, no. 4, 2006, pp. 289-296.
12. B. Ng, A. Kankanhalli and Y. Xu, "Studying users' computer security behavior: A health belief perspective," *Decision Support Systems*, vol. 46, no. 4, 2009, pp. 815-825.
13. A. Tsohou, M. Karyda and S. Kokolakis, "Analysing the role of cognitive and cultural biases in the internalization of information security policies: recommendations for information security awareness programs," *Computers & Security*, vol. 52, 2015, pp. 128-141.
14. M.A.S. Boksem and M. Tops, "Mental fatigue: Costs and benefits," *Brain Research Reviews*, vol. 59, no. 1, 2008, pp. 125-139; DOI https://doi.org/10.1016/j.brainresrev.2008.07.001.
15. D. van der Linden, M. Frese and T.F. Meijman, "Mental fatigue and the control of cognitive processes: effects on perseveration and planning," *Acta Psychologica*,

vol. 113, no. 1, 2003, pp. 45-65; DOI https://doi.org/10.1016/S0001-6918(02)00150-6.
16. S. Furnell and K.-L. Thomson, “Recognising and addressing ‘security fatigue’,” *Computer Fraud & Security*, vol. 2009, no. 11, 2009, pp. 7-11; DOI https://doi.org/10.1016/S1361-3723(09)70139-3.
17. J. Scott and P.J. Carrington, *The SAGE Handbook of Social Network Analysis*, SAGE Publications, 2011.
18. J. Fu, D. Sun, J. Chai, J. Xiao and S. Wang, “The "six-element" analysis method for the research on the characteristics of terrorist activities,” *Ann Oper Res*, vol. 234, 2015, pp. 17-35.
19. E. Philips, J. Nurse, M. Goldsmith and S. Creese, “Applying social network analysis to security,” *Working Papers of the Sustainable Society Network*, 2015, pp. 11-27.
20. D. Dang-Pham, S. Pittayachawan and V. Bruno, “Applications of social network analysis in behavioural information security research: concepts and empirical analysis,” *Computers & Security*, vol. 68, 2017, pp. 1-15.
21. H.L. Armstrong and I. McCulloh, “Organizational Risk using Network Analysis,” *Proc. South African Information Security Multi-Conference*, 2010.
22. H. Armstrong, C. Armstrong and I. McCulloh, “A Course Applying Network Analysis to Organizational Risk in Information Security,” 2010.
23. M.E. Whitman and H.J. Mattord, *Principles of Information Security*, Cengage Learning, 2011.
24. F.M. Clemente, F.M.L. Martins and R.S. Mendes, *Social network analysis applied to team sports analysis*, Springer, 2016.
25. S. Brin and L. Page, “The anatomy of a large-scale hypertextual web search engine,” *Computer networks and ISDN systems*, vol. 30, no. 1-7, 1998, pp. 107-117.
26. L.C. Freeman, D. Roeder and R.R. Mulholland, “Centrality in social networks: II. Experimental results,” *Social networks*, vol. 2, no. 2, 1979, pp. 119-141.
27. R.A. Hanneman and M. Riddle, *Introduction to social network methods*, University of California, 2005.
28. S. Wasserman and K. Faust, *Social network analysis: Methods and applications*, Cambridge University Press, 1994.
29. S.P. Borgatti, “Centrality and network flow,” *Social Networks*, no. 27, 2005, pp. 55-71.
30. D.K. Clancy and F. Collins, “Informal accounting information systems: Some tentative findings,” *Accounting, Organizations and Society*, vol. 4, no. 1-2, 1979, pp. 21-30.
31. S. MacDonald, “Informal information flow and strategy in the international firm,” *International Journal of Technology Management*, vol. 11, no. 1-2, 1996, pp. 219-232.
32. R. Duncombe and R. Heeks, “Enterprise across the digital divide: information systems and rural microenterprise in Botswana,” *Journal of International Development*, vol. 14, no. 1, 2002, pp. 61-74.
33. CASOS, “ORA-Lite,” 2018; www.casos.cs.cmu.edu/projects/ora.

Identifying information security risks in a social network using self-organising maps

Rudi Serfontein[1[0000-0002-0428-6494]], Hennie Kruger[2[0000-0001-8514-4422]], Lynette Drevin[3[0000-0001-9370-8216]]

North-West University, Potchefstroom, South Africa
[1]rudi.serfontein@nwu.ac.za
[2]hennie.kruger@nwu.ac.za
[3]lynette.drevin@nwu.ac.za

Abstract. Managing information security risks in an organisation is one of the most important tasks an organisation has. Unfortunately, due to the complexity of most organisational systems, identifying information security risks can be difficult. One way to identify possible risks in an organisation is to make use of Social Network Analysis (SNA). While they can be used to identify risks, the metrics calculated using SNA are often numerous and daunting to managers unfamiliar with SNA. Furthermore, as the data in this form tend to be uncomfortable to process, educating managers about risks in their organisation can be quite difficult. Also, as these metrics often require quantitative processing in order to be useful, SNA on its own is not always an attractive method to use to identify risks in an organisation. In this paper the use of self-organising maps to identify possible information security risks in an organisation is investigated. Risk data were obtained from an organisation that deals in risk management, which were used to build a social network. A number of metrics associated with risk were calculated from the network, and these metrics were used to cluster the various entities using a self-organising map. Certain entities that pose a possible information security risk were identified. The results suggest that it may be viable to use self-organising maps, in concord with SNA, to more easily identify risks in an organisation using visual methods.

Keywords: Self-Organising Maps · Social Network Analysis · Information Security.

1 Introduction

Information security risk management is one of the most crucial parts of information security and should be one of the most important actions taken by organisations [1]. Unfortunately, due to the relative complexity of most organisational systems, identifying information security risks that are inherent to people using the systems, and making managers aware of them, is often quite difficult. One of the methods proposed in recent years to address such risks involve the use of Social Network Analysis (SNA) [2-4]. SNA is a method that can be used to evaluate an organisation, for instance a community or business, in such a way that social interactions can be

studied quantitatively, rather than qualitatively [5]. It does however have a significant drawback in that large networks, when visualised, may have so many nodes and arcs that the network is visually incomprehensible. In order to address this drawback, a number of studies have employed techniques that alter significant nodes and edges of a visualised network in order to draw attention to certain aspects. Some of these techniques include differentiating the colour of nodes and edges [6], using differing sized nodes to correspond to certain metrics [7], and using labels of various sizes [3]. A somewhat more novel technique makes use of Self Organising Maps (SOMs) to directly visualise network data [8]. A SOM is an effective technique that can be used not only to visualise high-dimensional data, but to visualise it in such a way that the result can act as both a similarity graph and a clustering diagram [9]. The SOM technique can be used to identify similar nodes within a social network, even in the presence of seemingly contradicting attributes, and present this data in a way that managers are quickly informed of risks in the network. SOMs have also been used in information security research to propose improvements to intrusion detection methods [10], and as a method for analysing information security behavioural data [11, 12]. While the approach suggested by Boulet, Jouve, Rossi and Villa [8] does allow for social networks to be visualised as SOMs, it has a shortcoming in that the SOMs generated can not necessarily be used in a way that is relevant to the process of identifying possible information security risks. This is mainly as a result of the fact that, in order to identify risks within a social network, a number of metrics calculated from the network data is used rather than the raw data itself.

In this paper the feasibility of using existing SOM techniques to inform managers of possible risks in an organisation is discussed. The value of such an application is twofold; firstly, by using a visualisation method that reduces the amount of data that is visualised, the often confusing graphs produced by traditional SNA visualisation techniques can be replaced with SOMs that are easier to process graphically. Furthermore, as SOM algorithms produce maps that naturally display data of interest, analysis and evaluation of the data should no longer require the visualisation results to be adapted (node enlargement, colouration, etc.) in order to be meaningful. Secondly, as SOMs organise similar data into clusters, their application should make it easier to inform a manager of groups of similar at-risk entities. This is thanks to the clustering done by the SOM algorithms, as, due to the relationship between certain SNA metrics and the CIA triad (Confidentiality, Integrity and Availability), as will be discussed in Section 2, clustering the nodes according to these metrics allows an evaluator to quickly identify similar problematic nodes. Furthermore, because the SOM algorithm uses the calculated SNA metrics as attributes to determine the clusters, the clusters themselves can be used to infer similarities that may not be readily transparent from the available data. Another advantage is that managers can be informed and educated of possible risks in an organisation early on, which may aid in developing effective awareness, education, and training programs. The graphical nature of SOM may also make it a useful tool for training inexperienced risk managers, and can potentially aid in identifying standard trends and patterns.

In the remainder of the article the background, research methodology, and results, will be discussed respectively. The background discussion will focus on SNA in the

context of information security, as well as SOMs. The discussion of the method will focus on both the application of the techniques, and the data collection phase. The paper will then conclude with a discussion of the results and implications.

2 Background

The primary theme of this paper is the use of Social Network Analysis (SNA) metrics as inputs for a Self-Organising Map (SOM), which should aid in evaluating risk in an organisation. In order to demonstrate how this can be done, five SNA metrics will be discussed briefly. While there are dozens of SNA metrics that can potentially be used, the five discussed here were chosen based on their established relationships with risk in the literature. The section will start with a description of SOMs, followed by the evaluation of the selected SNA metrics in relation to risk.

2.1 Self-Organising Maps (SOM)

The self-organising map (SOM) is a neural network technique that can be used to visualise and evaluate high-dimensional data [13]. The SOM technique uses given data to produce a self-organising neural network wherein the data points are clustered into topographical regions [14]. This visualisation technique has a wide range of known applications, from evaluating comparable biological adaptations [15] and improving optimisation algorithms [16, 17], to clustering data for problem-solving purposes [14, 18]. One of the greatest advantages SOM has over other high-dimension visualisation techniques is that it produces a two-dimensional topographical map that can be evaluated and interpreted without any special knowledge or skills. In addition to clustering known data points, depending on the data, the technique can also be used in vector quantisation, and as a regression modelling technique [13]. All these methods can arguably be used to obtain valuable information about data, but in the context of this paper only the clustering function of SOM will be considered. The algorithm for developing a SOM [19] is shown below.

Input: Dataset *N*
Output: A topographical map *M* containing the data from *N*, sorted into topographical areas

Variables:
w_{ij} – Weight vector describing topographical area ij; either randomised or defined at start

x – An input vector contained in *N*
α – Learning rate that is a slowly decreasing function of time

```
while Stop condition is false
    For each x in N
        For each vector j
            Compute D(j) = ∑_i (w_ij − x_i)^2
        end
```

```
                    Find index J such that D(J) is a minimum
                    For all units j in a topographical area J, and for all i:
                      |     Compute w_ij(new) = w_ij(old) + α[x_i − w_ij(old)]
                    end
            end
            Update α
            Reduce radius of topographical area at specified times
            Test Stop condition
    end
    return M
```

This algorithm produces one map with all of the entities sorted into clusters. Certain software suites, such as Viscovery SOMine [20], provide additional information by colouring the same map using values from different attributes.

As stated, SOMs can be used to cluster high-dimensional data on a two-dimensional map, producing a result that can be interpreted easily without training. This makes the technique especially valuable to those in managerial positions, as these individuals may not have the time to study large reports and data sets in detail, and may also hold true for the outcomes of SNA based studies – especially if the resulting network is particularly large or complex. By calculating the SNA metrics, as discussed in the next section, and applying SOM to the resulting data set, the risks posed by certain individuals or groups can be determined and presented visually in a way that is easy to process and interpret. Additionally, a number of at-risk individuals may be identified that would not necessarily have been evident through the use of more traditional visualisation techniques such as bar-graphs. It is possible for an individual to have all the traits of a high-risk individual and not be an obvious risk from the data itself. In these instances, a clustering technique such as SOM can be used to identify individuals that have similar, possibly hidden, attributes. This makes it significantly easier to address certain information security risks, as larger scale programmes can be developed to target groups consisting of similar individuals.

In summary, SOM is a valuable technique to use in addition to SNA, as the clustering function of SOM can be used to infer invaluable information about information security risks if the correct and relevant SNA metrics are used.

2.2 **Social Network Analysis (SNA) in the context of Information Security**

One of the most well-known frameworks for information security is the CIA triad [21], which references Confidentiality, Integrity, and Availability. ***Confidentiality*** describes the access rights that users have to a piece of information. For example, a manager having confidential access to certain business data that his employees do not, or should not, have. One possible SNA metric that can be used to evaluate a risk to confidentiality is *closeness centrality*. Closeness centrality is calculated by determining all the shortest distances to all other nodes within the network [22], so a node with a high closeness centrality has a large number of close relationships to

other nodes in the network. Such a node may therefore have access to information that it should not have access to. Alternatively, if the node is an object or a resource, such as a shared computer or a photocopier, it could become a significant confidentiality risk if malware or untrustworthy maintenance personnel are involved.

Integrity describes not only how accurate any piece of information is but, by extension, how trustworthy it is. One of the SNA metrics that can be used to evaluate the risk a node poses to the integrity of the information in the network is *total degree centrality*. The total degree centrality measure is concerned with an individual node's position within the network [22, 23], and is determined by using the number of nodes leading into and out of a node. A node with a high total degree centrality is well connected within the network, and may have enough influence over other nodes to impact the integrity of the information passing through them. Consider, as an example, an office worker with a high total degree centrality that has to capture data for a corporate database. If this worker were to make a mistake in capturing the data, the integrity of the data that a large number of nodes rely on may be compromised.

The *betweenness centrality* measure can be used to identify nodes that are risks to both integrity and confidentiality. This measure is a representation of the number of times that a particular node is part of the shortest path between any two nodes in the network [24]. It is reflective of the number of indirect nodes that are connected to that node. To demonstrate the rationale behind using betweenness centrality as an indicator of risk to both integrity and confidentiality, consider a department with a "go-to" individual. This individual will likely have access to greater amounts of information than is ideal, and would be in a position to alter the information flowing through the network. Furthermore, as nodes with high betweenness measures tend to act as brokers, this individual may be seen as a trustworthy shortcut to obtain information in the network, which places it in a position to obtain greater amounts of information, as well as manipulate information as it flows through the network.

The final member of the triad, ***availability***, deals with the ability to access the data in a timely manner. Availability is often at odds with both confidentiality and integrity, as systems meant to protect availability and confidentiality often impact on the availability of the data. With regard to SNA, one of the metrics that may identify a high risk node in terms of availability is the one that identifies a node as a *boundary spanner*. A boundary spanner is a node that, if removed, will cause one part of the network to become completely isolated [25], thereby negatively impacting the availability of information in certain parts of the network.

The final SNA metric to be mentioned here is *eigenvector centrality*. Eigenvector centrality measures the extent to which a particular node is connected to highly connected nodes [22]. Nodes that have a high eigenvector measure are considered to possess emergent leadership properties [26] and may be considered a potential risk to confidentiality, integrity, and availability. For example, consider the impact an informal leader can have on information in a network. Confidential information may be shared with such a node as a result of the connections with highly connected nodes, whereas the integrity of the data in the network may be impacted by the additional knowledge the node obtains. Availability may also be impacted negatively if the emergent leader convinces other nodes to delay the flow of information, or if

the information is redirected through the network along suboptimal routes. A summary of the SNA metrics discussed, and how they relate to the CIA triad, is given in **Table 1**.

Table 1. SNA metrics in the context of Information Security

SNA Metric	CIA Triad			Rationale
	C	I	A	
Total degree centrality		X		Nodes with a high total degree centrality have influence in the network and are connected to a significant portion of the network.
Closeness centrality	X			Nodes with a high closeness centrality have access to a significant amount of information in the network.
Betweenness centrality	X	X		Nodes with a high betweenness centrality may be prone to information brokering and tampering.
Boundary spanner			X	If a boundary spanner node is removed, an entire section of the network becomes isolated
Eigenvector centrality	X	X	X	Nodes with a high eigenvector centrality are considered emergent leaders and, depending on their influence and attitude, may be a general risk, depending on the circumstances

It should be emphasised that, while this brief discussion focussed on each member of the CIA triad individually, confidentiality, integrity, and availability are all interconnected. It is possible, for instance, for a significant enough increase in confidentiality to result in a significant reduction in availability. This is also true for the relationship between confidentiality and integrity, and the relationship between integrity and availability. The goal is typically to find a balance between these three aspects that is appropriate to the particular situation. This interrelatedness should be kept in mind when evaluating risks using SNA, as well as when selecting controls to address these risks.

From this short discussion it is clear that SNA metrics can be used to identify and possibly evaluate risks in terms of the CIA triad. The use of SNA metrics as input data for a SOM is therefore appropriate, and its use in this manner may help to identify and visualise risks in an organisation. This should help to improve overall awareness of risks in the organisation, aid in training managers, and may even help to determine overall preventative measures. The application in the rest of the paper uses the five SNA metrics discussed in this section as input for a SOM.

3 Method

The study was conducted using data provided by a manager from a large company that deals with risk evaluation. The data are confidential, and were subsequently anonymised prior to publication. Using this data, a network was built that describes the relationship between various entities. This network is shown in **Fig. 1**.

The entities, or nodes, of the network include 26 real-world risks, 612 controls, 6 risk owners, 26 control owners, 13 risk coordinators and 12 governing bodies. The risks are those risks that the organisation has to manage, whilst the controls are those controls used to manage the risks. The risk- and control owners are ultimately responsible for the risks and controls respectively, whereas the risk coordinators ensure that the correct risks are managed using the appropriate controls. The governing bodies are responsible for determining which control is used with which risk. These bodies also determine what the probability of a risk occurring is, as well as the severity of such an occurrence. The network is undirected, as unidirectional relationships between entities such as risks and risk owners do not seem realistic.

The network data were processed using the software suite ORA-Lite [24], while the SOMs were generated using the Viscovery SOMine Suite [20]. The data for the SOM consists of the 5 SNA metrics for each node. In total, 695 nodes are contained in the network, and a total number of 1738 links exist between them. The focus of the network is on managing real-world risks, and it was subsequently processed in a risk centric way, i.e. the relationships between the nodes are based on similar relationships to particular risks. This means that the relationship between a risk coordinator and a control owner, for example, is described only in terms of their shared relationship to the same risk.

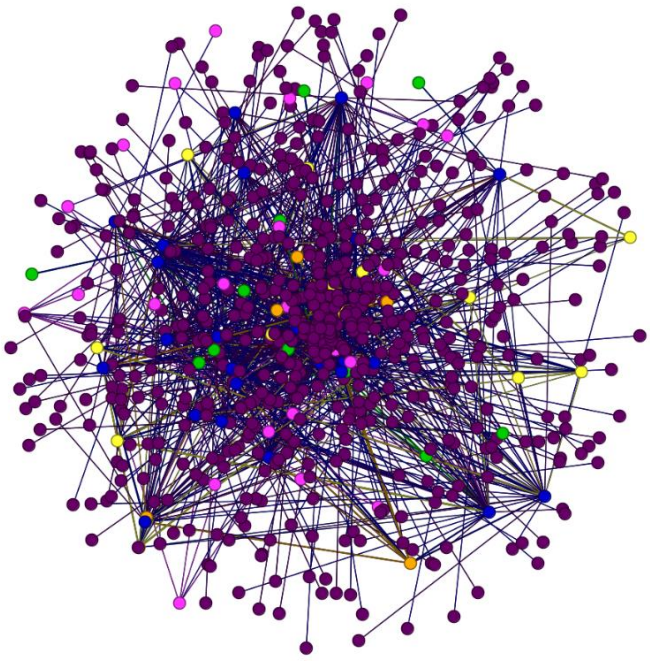


Fig. 1. Social network showing relationships between risks (blue), risk controls (purple), risk coordinators (yellow), risk owners (orange), control owners (pink) and the governing bodies (green) responsible for appointing the various role-players.

With networks of this size, the large number of metric values that are produced can be quite complex. In order to help evaluate such complex data in a simpler, graphical way, the SOM algorithm is used. The SOMs can be used to quickly identify problem areas, which should make it easier to evaluate the data. Additionally, as SOMs are graphical in nature, they can be applied iteratively to investigate how the risk in a network changes over time, or as certain controls are introduced that aim to manage those risks.

When applying this technique to training risk managers, one of the aims is to highlight certain trends or groups that may pose a natural risk in the network. In doing so, the risk manager should be better informed of the nuances of the network and may be able to introduce more effective risk mitigation measures than would have been possible otherwise. Consider, for example, a network, such as the one shown in **Fig. 1**, with risks, control coordinators, and controls. If a SOM is developed for the network, the manager should be able to readily identify groups of nodes that have similar risk profiles based on their clustering. If a certain grouping of controls and control coordinators are found in the same cluster, for instance, it may indicate that there is a problem with the way in which the controls are managed. Alternatively, if all of the control coordinators are grouped into one high risk cluster, it may be appropriate to introduce measures, such as policies, to address the risks posed by these nodes.

Another way in which the technique can be used in training is to monitor how the risk profile of certain clusters change when new controls are implemented to address the identified risks. As a SOM is graphical in nature, and the geographical structure of the map changes as the risk values for the nodes change, it should be possible to identify the changes in the network graphically. This is especially true for clusters that lose nodes, as the area of the map that the cluster occupies should be reduced. By using the SOMs as a graphical aid, the manager should be able to identify which approaches work best, and under which circumstances.

4 Results and discussion

The SOM algorithm produced a map with three regions, or clusters, when applied to the network data. This map is shown in **Fig. 2** to **Fig. 6**. Each figure shows the same map, but with different colourations. The colourations are used to show how the values of the five measures differ for various nodes. The clusters are the same in each image, as one map, built using all five SNA metrics as node attributes, was obtained.
In **Fig. 2,** where the boundary spanner measure is used to colour the map, the red colouration that covers most of Cluster C2 indicates that the nodes in C2 pose a significant possible risk, as the nodes in this cluster have a much higher boundary spanner value than the nodes in other clusters. A cursory evaluation of the cluster's data shows that C2 exclusively contains all of the nodes that represent the risks. It should be noted that, while the risks are all found in the same cluster for this network, this will not necessarily be the case for all networks. As the boundary spanner metric indicates that a node's removal will completely isolate a part of the network, this suggests that the current structure in the network includes nodes that will be isolated

if any of the risks are resolved. The network data itself shows that these nodes are primarily controls. If, for example, the risk “Corporate Brand” is completely resolved, i.e. if the company finds itself in a situation where there is no risk at all to the company brand, then the controls that exist to manage that risk, such as “Social Media strategy and protocols” and “Expert communications resources”, are no longer needed. In order to ensure that these controls are not kept in place unnecessarily, additional measures need to be implemented.

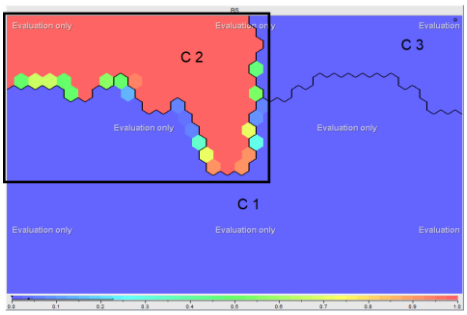


Fig. 2. SOM (boundary spanner)

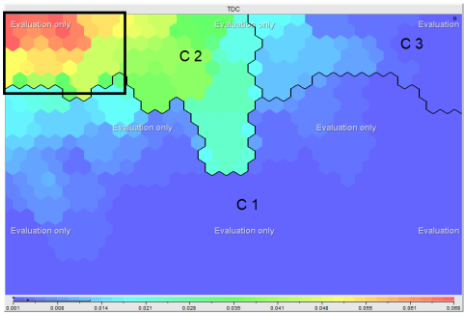


Fig. 3. SOM (total degree centrality)

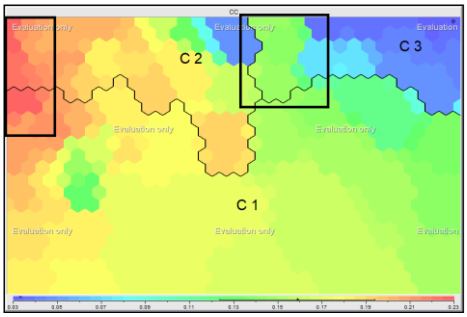


Fig. 4. SOM (closeness centrality)

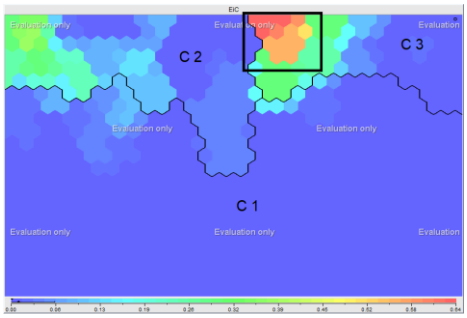


Fig. 5. SOM (eigenvector centrality)

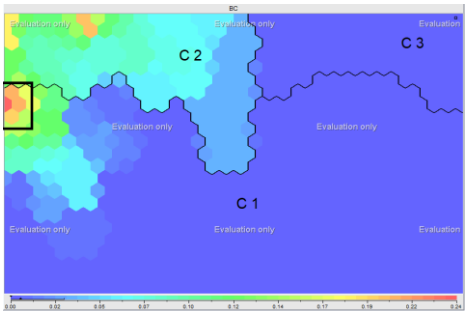


Fig. 6. SOM (betweenness centrality)

The colouration, based on the total degree centrality, used on the SOM in **Fig. 3** shows that there is an area in cluster C2 where the nodes have unusually high total degree centrality values, which is associated with a higher level of risk to integrity. Of the nodes in C2, there are six nodes, indicated by the red area in C2, that have a significantly higher amount of total degree centrality than the rest of the nodes. These nodes are the risks "Forest fires", "Environmental impacts", "Interruption to supply networks", "Waste Treatment Capacity", "Urban Resource Capacity", and "Rural Resource Capacity". All of these risks can have a significant impact if realised, which is why the influence they have is so substantial. This also means that any errors with regard to these risks, such as the risk of forest fires being over- or underemphasised, can have a significant impact on the information that is ultimately used in the network to manage other risks. If, for example, the integrity of the information with regard to the chances of a forest fire occurring is compromised, then there may not be enough water available to address the fire. If the fire affects any industrial assets, this may have environmental impacts, which in turn could negatively affect the company's corporate brand image. To protect the integrity of the information of these nodes, additional controls should be implemented.

An area with values much higher than the surrounding areas is shown to be present in cluster C3 in both **Fig. 4** and **Fig. 5**. This area, which is situated on the left hand side of cluster C3 where it borders Cluster C2, is coloured green in **Fig. 4** and green, orange, and red in **Fig. 5**. Additionally, a hotspot is present on the border between C1 and C2, as shown in **Fig. 4**. While the nodes in the hotspot area between C1 and C2 certainly pose a risk to confidentiality, the section of C3, where the nodes have much higher values for closeness centrality and eigenvector centrality than the rest of the nodes in the cluster, warrants further investigation. The higher closeness centrality measure of these nodes, shown in **Fig. 4**, suggest that they are a risk to confidentiality, whereas the very high eigenvector centrality measure shown in **Fig. 5** indicate that they are an overall risk. There are two nodes in particular that fall into this region, one being a risk owner that is responsible for 11 out of the 26 risks, and the other being a risk coordinator that is responsible for 6 risks. The remainder of C3 is low risk and, as both of these nodes are in the cluster, it intimates that the risk posed by this risk owner and risk coordinator could be alleviated by reducing the number of risks that they are responsible for. Some of the risks could be transferred to other risk owners and coordinators. Alternatively, the risks can be co-owned and co-coordinated with owners and coordinators that are responsible for a smaller number of risks.

With the exception of the small area of low risk in C3, **Fig. 4** shows that most of the nodes in the network have a high measure of closeness centrality. This suggests that controls should be in place to protect the confidentiality of the information in the network in general, as almost any one of the nodes could be responsible for compromising confidentiality.

Fig. 6 highlights the existence of a single hotspot that exists in cluster C1 with regard to betweenness centrality. The hotspot, which is situated on the left hand side of C1 and has a red colour, contains a single risk coordinator, which poses a risk to both the integrity and the confidentiality of information in the network. In order to resolve this risk, the dependence on the specific risk coordinator should be reduced.

The dependence of the risk coordinator can be reduced in one of at least two ways. The first method involves transferring some of the coordinator's responsibilities, such as risks, to another coordinator. This coordinator should preferably be located in cluster C1, as such a coordinator is likely to have a similar amount of power and influence. The second method that could be used is to employ an additional risk coordinator to take over some of the duties. This coordinator could also assume some of the duties of the risk coordinator in C3, thereby reducing the risk of two nodes simultaneously.

Based on all the SOMS shown in **Fig. 2** to **Fig. 6**, there are six risks, two risk coordinators, and a risk owner that pose potential risks to the overall security of information in this network. From this discussion, the advantage of using the SOM method to visualise SNA metric data is clear: by using SOMs to visualise SNA metric data, a relatively simple process can be followed in order to evaluate the risks in a network. The advantages are especially clear when compared to the process that would be needed in order to evaluate the risks in a network, such as the one shown in **Fig.** , or when using only raw data and statistical analysis. The use of SOMs in this manner gives managers the opportunity to evaluate risks graphically, as well as to compile risk discussion reports that do not require any prior knowledge of SOMs or SNA. This, in turn, could help improve the nature and quality of risk management, as a greater number of options and plans could develop as a result. Furthermore, SOMs allow for a way to systematically identify risks, and can also be used to monitor the progress and impact of risk mitigation strategies. Depending on the situation, it may be possible to identify positive or negative changes in the network almost instantaneously using this method. The central premise, i.e. that SOMs can be used to visualise SNA risk data, and in turn help educate managers of risks in their organisation in a quick and simple way, is therefore feasible.

5 Conclusion

Information security risk management is one of the most crucial parts of information security, but it is often complicated by the complexity of most organisational systems. In order to simplify the task of identifying risks in an organisation, SOM can be used to identify possible risks in an organisation by visualising SNA metric data of the organisation. A SOM, which clusters similar entities into geographically separate regions, is relatively simple to evaluate due to its graphical nature. When compared to other risk identification techniques that employ SNA metrics, which may require risk managers to process and evaluate large tables of numbers, the use of SOM may reduce the amount of work needed, as entities that pose a threat to the organisation can be identified with relative ease. Additionally, as a SOM is easier to evaluate, inexperienced risk managers may find the use of SOM less daunting than to use numerical data and statistical analysis. Finally, as a SOM provides an additional level of information that may not be readily apparent from the data, it could aid in educating risk managers of dangers in the organisation that may not be known, or obvious.

References

1. Wangen, G.: Information security risk assessment: A method comparison. Computer, vol. 50, no. 4, pp. 52-61 (2017).
2. Armstrong, H., Armstrong, C. and McCulloh, I.: A course applying network analysis to organizational risk in information security, In: South African Information Security Multi-Conference, pp. 204-214 (2010).
3. Dang-Pham, D., Pittayachawan, S. and Bruno, V.: Investigation into the formation of information security influence: Network analysis of an emerging organisation. Computers & Security, vol. 70, pp. 111-123 (2017).
4. Serfontein, R., Drevin, L. and Kruger, H.A.: The feasibility of raising information security awareness in an academic environment using SNA, In: IFIP World Conference on Information Security Education, pp. 69-80. Springer (2018).
5. Scott, J. and Carrington, P.J.: The SAGE handbook of social network analysis. SAGE publications (2011).
6. Tsui, E. and Liebowitz, J.: Linking social network analysis with the analytic hierarchy process for knowledge mapping in organizations. Journal of Knowledge Management, vol. 9, no. 1, pp. 76-86 (2005).
7. Dang-Pham, D., Pittayachawan, S. and Bruno, V.: Applying network analysis to investigate interpersonal influence of information security behaviours in the workplace. Information & Management, vol. 54, no. 5, pp. 625-637 (2017).
8. Boulet, R., Jouve, B., Rossi, F. and Villa, N.: Batch kernel SOM and related laplacian methods for social network analysis. Neurocomputing, vol. 71, no. 7, pp. 1257-1273 (2008).
9. Kohonen, T.: The self-organizing map. Neurocomputing, vol. 21, no. 1-3, pp. 1-6 (1998).
10. De la Hoz, E., De la Hoz, E., Ortiz, A., Ortega, J. and Prieto, B.: PCA filtering and probabilistic SOM for network intrusion detection. Neurocomputing, vol. 164, no. Supplement C, pp. 71-81 (2015).
11. Hunt, R. and Hill, S.: Using security logs to identify and manage user behaviour to enhance information security, In: 14th European Conference on Cyber Warfare and Security, pp. 111. Academic Conferences Limited (2015).
12. López, A.U., Mateo, F., Navío-Marco, J., Martínez-Martínez, J.M., Gómez-Sanchís, J., Vila-Francés, J. and Serrano-López, A.J.: Analysis of computer user behavior, security incidents and fraud using self-organizing maps. Computers & Security, vol. 83, pp. 38-51 (2019).
13. Bäck, T., Kok, J.N. and Rozenberg, G.: Handbook of natural computing. Springer (2012).
14. Pal, C., Hirayama, S., Narahari, S., Jeyabharath, M., Prakash, G. and Kulothungan, V.: An insight of world health organization (WHO) accident database by cluster analysis with self-organizing map (SOM). Traffic Injury Prevention, vol. 19, no. sup1, pp. S15-S20 (2018).
15. Nakayama, H., Sakamoto, T., Okegawa, Y., Kaminoyama, K., Fujie, M., Ichihashi, Y., Kurata, T., Motohashi, K., Al-Shehbaz, I., Sinha, N. and Kimura, S.: Comparative transcriptomics with self-organizing map reveals cryptic photosynthetic differences

between two accessions of north american lake cress. Scientific Reports, vol. 8, no. 1, pp. 3302 (2018).
16. Gu, F. and Cheung, Y.-M.: Self-organizing map-based weight design for decomposition-based many-objective evolutionary algorithm. IEEE Transactions on Evolutionary Computation, vol. 22, no. 2, pp. 211-225 (2018).
17. Kuo, R.J., Rizki, M., Zulvia, F.E. and Khasanah, A.U.: Integration of growing self-organizing map and bee colony optimization algorithm for part clustering. Computers & Industrial Engineering, vol. 120, pp. 251-265 (2018).
18. Lee, Y.: Using self-organizing map and clustering to investigate problem-solving patterns in the massive open online course: An exploratory study. Journal of Educational Computing Research, pp. 0735633117753364 (2018).
19. Fausett, L.V.: Fundamentals of neural networks: Architectures, algorithms, and applications. Prentice-Hall Englewood Cliffs (1994).
20. Viscovey SOMine. www.viscovery.net/somine, last accessed 10/02/2019.
21. Au, C.H., Fung, W.S. and Tses, A.: An investigation on the relationship between control self-assessment, cloud security, and cloud-related business performance-using partial least squares, In: Industrial Engineering and Engineering Management (IEEM), pp. 1879-1883. IEEE (2016).
22. Armstrong, H. and McCulloh, I.: Organizational risk using network analysis, In: South African Information Security Multi-Conference, pp. 132-141 (2010).
23. Hanneman, R.A. and Riddle, M.: Introduction to social network methods. University of California (2005).
24. ORA-Lite. www.casos.cs.cmu.edu/projects/ora, last accessed 24/04/2018.
25. Cormen, T.H., Leiserson, C.E., Rivest, R.L. and Stein, C.: Introduction to algorithms second edition. The MIT Press (2001).
26. Borgatti, S.P.: Centrality and network flow. Social networks, no. 27, pp. 55-71 (2005).

B Appendix B: Chapter 7 Appendix - Phase 2 Risk Profile

In this appendix the risk profile, as developed during Phase 2 of the UD example in Section 7.2.2, is presented. Table B.1 contains the details for each of the 84 nodes, as well as the risk values for each of the nodes. The "Risk value" column contains the risk value for each node, calculated using only the metrics and not weighted for criticality, whilst the "Weighted risk value" column contains the risk value for each node after the criticality weight has been taken into account. The final three rows of the table contain the minimum, maximum, and weight values for each of the metrics.

TABLE B.1: PHASE 2 RISK PROFILE

Node Title	BC	CC	EcC	EiC	SHC	TDC	BS	Criticality of Node	Risk value	Weighted risk value
AS7	0	0.309	5.000	0.028	0.714	0.024	0	2.000	2.893	5.787
A60	0.296	0.439	4.000	0.880	0.031	0.446	1.000	2.000	4.977	9.954
A14	0.212	0.393	5.000	0.307	0.043	0.277	1.000	2.000	4.480	8.960
A49	0.231	0.371	5.000	0.177	0.052	0.241	1.000	2.000	4.233	8.466
A76	0.303	0.430	4.000	0.296	0.047	0.289	1.000	2.000	3.967	7.935
A0	0.118	0.314	5.000	0.031	0.132	0.108	1.000	2.000	3.301	6.601
A41	0.101	0.325	5.000	0.040	0.111	0.108	1.000	2.000	3.260	6.520
A9	0.082	0.333	5.000	0.019	0.167	0.072	1.000	2.000	3.242	6.483
A62	0.001	0.296	6.000	0.047	0.333	0.036	0	2.000	3.102	6.204
A61	0.003	0.285	6.000	0.047	0.333	0.036	0	2.000	3.060	6.120
A2	0.045	0.269	5.000	0.006	0.259	0.060	1.000	2.000	2.984	5.968
A75	0.015	0.272	6.000	0.009	0.250	0.048	0	2.000	2.855	5.710
S36	0	0.283	6.000	0.026	1.000	0.012	0	1.000	4.339	4.339
S37	0	0.283	6.000	0.026	1.000	0.012	0	1.000	4.339	4.339
S38	0	0.283	6.000	0.026	1.000	0.012	0	1.000	4.339	4.339
S50	0	0.271	6.000	0.015	1.000	0.012	0	1.000	4.273	4.273
S52	0	0.271	6.000	0.015	1.000	0.012	0	1.000	4.273	4.273
S53	0	0.271	6.000	0.015	1.000	0.012	0	1.000	4.273	4.273
S55	0	0.271	6.000	0.015	1.000	0.012	0	1.000	4.273	4.273
S56	0	0.271	6.000	0.015	1.000	0.012	0	1.000	4.273	4.273
S57	0	0.271	6.000	0.015	1.000	0.012	0	1.000	4.273	4.273
S58	0	0.271	6.000	0.015	1.000	0.012	0	1.000	4.273	4.273
S59	0	0.271	6.000	0.015	1.000	0.012	0	1.000	4.273	4.273
S24	0	0.235	6.000	0.150	1.000	0.012	0	1.000	4.267	4.267
A6	0.021	0.319	5.000	0.045	0.250	0.048	0	2.000	2.124	4.248
S11	0	0.251	6.000	0.002	1.000	0.012	0	1.000	4.170	4.170
S13	0	0.251	6.000	0.002	1.000	0.012	0	1.000	4.170	4.170
S48	0	0.246	6.000	0.003	1.000	0.012	0	1.000	4.149	4.149
S5	0	0.240	6.000	0.003	1.000	0.012	0	1.000	4.122	4.122
S82	0	0.240	6.000	0.003	1.000	0.012	0	1.000	4.122	4.122
S39	0	0.213	6.000	0	1.000	0.012	0	1.000	4.000	4.000
S29	0	0.305	6.000	0.176	0.556	0.024	0	1.000	3.718	3.718
S3	0	0.247	6.000	0.003	0.669	0.024	0	1.000	3.498	3.498
S62	0	0.306	5.000	0.075	1.000	0.012	0	1.000	3.496	3.496
S63	0	0.306	5.000	0.075	1.000	0.012	0	1.000	3.496	3.496
S19	0	0.258	6.000	0.201	0.500	0.024	0	1.000	3.423	3.423
S23	0	0.258	6.000	0.201	0.500	0.024	0	1.000	3.423	3.423
S77	0	0.302	5.000	0.025	1.000	0.012	0	1.000	3.422	3.422
S78	0	0.302	5.000	0.025	1.000	0.012	0	1.000	3.422	3.422
S79	0	0.302	5.000	0.025	1.000	0.012	0	1.000	3.422	3.422

S80	0	0.302	5.000	0.025	1.000	0.012	0	1.000	3.422	3.422
S68	0	0.302	5.000	0.025	1.000	0.012	0	1.000	3.422	3.422
S28	0.015	0.337	6.000	0.195	0.280	0.048	0	1.000	3.416	3.416
S27	0.005	0.310	6.000	0.180	0.375	0.036	0	1.000	3.415	3.415
S51	0.017	0.286	6.000	0.016	0.500	0.024	0	1.000	3.392	3.392
S25	0.009	0.335	6.000	0.195	0.280	0.048	0	1.000	3.388	3.388
S1	0.014	0.268	6.000	0.006	0.500	0.024	0	1.000	3.292	3.292
S42	0.003	0.265	6.000	0.004	0.500	0.024	0	1.000	3.240	3.240
S43	0.003	0.265	6.000	0.004	0.500	0.024	0	1.000	3.240	3.240
S47	0.008	0.259	6.000	0.006	0.500	0.024	0	1.000	3.232	3.232
S17	0	0.317	5.000	0.175	0.556	0.024	0	1.000	2.770	2.770
S16	0.004	0.335	5.000	0.276	0.440	0.036	0	1.000	2.766	2.766
S64	0.004	0.335	5.000	0.276	0.440	0.036	0	1.000	2.766	2.766
S22	0.064	0.376	5.000	0.205	0.280	0.048	0	1.000	2.762	2.762
S66	0.067	0.415	5.000	0.145	0.200	0.060	0	1.000	2.739	2.739
S71	0.005	0.346	5.000	0.100	0.500	0.024	0	1.000	2.714	2.714
S18	0.025	0.386	5.000	0.217	0.280	0.048	0	1.000	2.691	2.691
S65	0.058	0.411	5.000	0.145	0.200	0.060	0	1.000	2.691	2.691
S54	0.010	0.339	5.000	0.090	0.500	0.024	0	1.000	2.688	2.688
S32	0.004	0.339	5.000	0.101	0.500	0.024	0	1.000	2.681	2.681
AS12	0.029	0.369	4.000	0.070	0.200	0.060	0	2.000	1.324	2.649
S21	0.004	0.335	5.000	0.051	0.500	0.024	0	1.000	2.606	2.606
S44	0.009	0.323	5.000	0.078	0.500	0.024	0	1.000	2.600	2.600
S45	0.009	0.323	5.000	0.078	0.500	0.024	0	1.000	2.600	2.600
S46	0.009	0.323	5.000	0.078	0.500	0.024	0	1.000	2.600	2.600
S20	0.013	0.388	5.000	0.157	0.280	0.048	0	1.000	2.592	2.592
S61	0.014	0.316	5.000	0.076	0.500	0.024	0	1.000	2.583	2.583
S35	0.025	0.393	5.000	0.122	0.250	0.048	0	1.000	2.552	2.552
S74	0	0.271	5.000	0.153	0.556	0.024	0	1.000	2.541	2.541
S8	0.012	0.317	5.000	0.029	0.481	0.036	0	1.000	2.516	2.516
S73	0.001	0.264	5.000	0.153	0.556	0.024	0	1.000	2.514	2.514
S33	0.007	0.361	5.000	0.197	0.280	0.048	0	1.000	2.498	2.498
S81	0.010	0.324	5.000	0.179	0.375	0.036	0	1.000	2.492	2.492
S31	0.001	0.329	5.000	0.182	0.375	0.036	0	1.000	2.488	2.488
S69	0.010	0.353	5.000	0.205	0.280	0.048	0	1.000	2.482	2.482
S30	0.013	0.346	5.000	0.195	0.280	0.048	0	1.000	2.449	2.449
S26	0.004	0.317	5.000	0.180	0.375	0.036	0	1.000	2.443	2.443
S10	0.023	0.278	5.000	0.002	0.500	0.024	0	1.000	2.361	2.361
S4	0.010	0.273	5.000	0.006	0.500	0.024	0	1.000	2.300	2.300
S40	0.026	0.284	5.000	0.029	0.250	0.048	0	1.000	1.967	1.967
S67	0.073	0.374	4.000	0.103	0.333	0.036	0	1.000	1.749	1.749
S34	0.078	0.393	4.000	0.118	0.250	0.048	0	1.000	1.722	1.722
S72	0.024	0.329	4.000	0.078	0.500	0.024	0	1.000	1.676	1.676
S70	0.027	0.332	4.000	0.178	0.375	0.036	0	1.000	1.583	1.583
Measure MIN	0	0.213	4.000	0476	0.031	0.012	0			
Measure MAX	0.303	0.439	6.000	0.88	1.000	0.446	1.000			
Weight	1.000	1.000	2.000	1.000	2.000	1.000	1.000			

C APPENDIX C: CHAPTER 8 APPENDIX - NODE-LEVEL RISK PROFILE

In this appendix the risk profile, as developed in Section 8.2.3 for the CRR network, is presented. For completeness, the metric weights are provided in Table C.1, along with a list of metric abbreviations used in the risk profile. Table C.2 contains the details for each of the 695 nodes, as well as the risk values for each of the nodes. The "Node Risk value" column contains the risk value for each node. All of the values are normalised. The rows in the risk profile are sorted according to the node risk value for each node, with the highest risk nodes being shown first.

TABLE C.1: RISK PROFILE WEIGHTS FOR METRICS

Metric		Weight
Boundary spanner	(BS)	0
Total degree centrality	(TDC)	2.25
Eigenvector centrality	(EiC)	2.25
Betweenness centrality	(BC)	2.75
Closeness centrality	(CC)	2.5
Eccentricity centrality	(ECC)	1
Structural holes constraint	(SHC)	1.5

TABLE C.2: NODE-LEVEL RISK PROFILE

Node-Title	BC	CC	ECC	EiC	SHC	TDC	BS	Criticality of Node	Node Risk Value
RISK COORDINATOR 10	0.870	0.745	1.000	1.000	0	1.000	0	0.619	6.038
RISK COORDINATOR 13	0.954	0.835	1.000	0.864	0.006	0.754	0	0.619	5.794
CTRL OWN 18	0.991	0.922	0	0.743	0.008	0.700	0	0.619	5.131
RISK COORDINATOR 11	0.898	1.000	0	0.759	0.017	0.500	0	0.619	4.845
RISK COORDINATOR 8	1.000	0.935	0	0.633	0.018	0.523	0	0.619	4.776
RISK COORDINATOR 2	0.981	0.874	0	0.486	0.027	0.415	0	0.619	4.302
RISK COORDINATOR 7	0.685	0.935	0	0.652	0.026	0.400	0	0.619	4.102
RISK COORDINATOR 12	0.463	0.701	1.000	0.529	0.014	0.615	0	0.619	4.098
RISK COORDINATOR 4	0.370	0.688	1.000	0.360	0.056	0.246	0	0.619	3.210
CTRL OWN 16	0.194	0.684	1.000	0.478	0.053	0.262	0	0.619	3.088
RISK OWNER 6	0.324	0.866	0	0.526	0.053	0.192	0	0.619	2.941
CTRL OWN 17	0.139	0.606	1.000	0.286	0.057	0.269	0	0.619	2.619
CTRL OWN 11	0.250	0.792	0	0.401	0.064	0.215	0	0.619	2.569
CTRL OWN 8	0.185	0.645	1.000	0.248	0.117	0.123	0	0.619	2.557
CTRL OWN 3	0.148	0.619	1.000	0.219	0.081	0.200	0	0.619	2.488
CTRL OWN 21	0.093	0.545	1.000	0.224	0.069	0.246	0	0.619	2.339
CTRL OWN 20	0.259	0.706	0	0.245	0.057	0.262	0	0.619	2.292
RISK COORDINATOR 1	0.528	0.554	0	0.082	0.067	0.246	0	0.619	2.275
CTRL OWN 19	0.037	0.537	1.000	0.240	0.134	0.115	0	0.619	2.132
RISK OWNER 2	0.037	0.602	1.000	0.173	0.144	0.085	0	0.619	2.107
CTRL OWN 5	0	0.390	1.000	0.034	0.819	0.015	0	0.619	2.051
CTRL OWN 26	0	0.385	1.000	0.039	0.819	0.015	0	0.619	2.050
CTRL OWN 23	0	0.342	1.000	0.045	0.821	0.015	0	0.619	1.994
RISK COORDINATOR 5	0.176	0.589	0	0.184	0.061	0.269	0	0.619	1.899
CTRL OWN 22	0.007	0.459	1.000	0.152	0.257	0.062	0	0.619	1.878
RISK OWNER 4	0	0.268	1.000	0.023	0.824	0.015	0	0.619	1.852
CTRL OWN 1	0.102	0.429	1.000	0.039	0.259	0.069	0	0.619	1.847
CTRL OWN 9	0	0.199	1.000	0.005	0.827	0.015	0	0.619	1.723
CTRL OWN 24	0	0.199	1.000	0.005	0.827	0.015	0	0.619	1.723
CTRL OWN 12	0	0.199	1.000	0.005	0.827	0.015	0	0.619	1.723

Node-Title	BC	CC	ECC	EiC	SHC	TDC	BS	Criticality of Node	Node Risk Value
CTRL OWN 14	0.111	0.610	0	0.144	0.135	0.115	0	0.619	1.619
CTRL OWN 25	0	0.299	0.500	0.018	0.821	0.015	0	0.619	1.580
RISK OWNER 3	0.269	0.485	0	0.045	0.204	0.069	0	0.619	1.557
CTRL OWN 15	0.065	0.554	0	0.122	0.188	0.092	0	0.619	1.441
CTRL OWN 4	0.139	0.519	0	0.063	0.158	0.115	0	0.619	1.434
CTRL OWN 7	0	0	1.000	0	0.831	0.015	0	0.619	1.411
RISK COORDINATOR 3	0.065	0.203	1.000	0.007	0.263	0.069	0	0.619	1.394
RISK COORDINATOR 6	0.028	0.216	1.000	0.010	0.108	0.177	0	0.619	1.362
RISK OWNER 1	0.019	0.221	1.000	0.007	0.260	0.069	0	0.619	1.341
CTRL OWN 13	0.037	0.368	0.500	0.021	0.297	0.062	0	0.619	1.333
RISK OWNER 5	0.093	0.325	0.500	0.026	0.208	0.085	0	0.619	1.318
CTRL OWN 2	0.046	0.329	0.500	0.021	0.341	0.046	0	0.619	1.307
RISK COORDINATOR 9	0.028	0.346	0.500	0.023	0.341	0.046	0	0.619	1.305
CTRL OWN 10	0.008	0.450	0	0.061	0.398	0.038	0	0.619	1.217
GOV BODY 12	0.028	0.606	1.000	0.280	0.103	0.085	0	0.284	1.013
GOV BODY 2	0	0.390	1.000	0.031	1.000	0	0	0.284	1.007
GOV BODY 7	0	0.385	1.000	0.037	1.000	0	0	0.284	1.007
GOV BODY 9	0	0.329	1.000	0.026	1.000	0	0	0.284	0.960
CTRL OWN 6	0.001	0.004	1.000	0.001	0.167	0.123	0	0.619	0.955
GOV BODY 6	0	0.268	1.000	0.021	1.000	0	0	0.284	0.914
GOV BODY 11	0.009	0.528	1.000	0.168	0.232	0.038	0	0.284	0.896
GOV BODY 4	0	0.199	1.000	0.005	1.000	0	0	0.284	0.854
GOV BODY 5	0	0.199	1.000	0.002	1.000	0	0	0.284	0.853
GOV BODY 10	0	0.160	1.000	0.005	1.000	0	0	0.284	0.827
GOV BODY 8	0	0.364	1.000	0.077	0.488	0.015	0	0.284	0.809
GOV BODY 1	0	0.069	1.000	0.001	1.000	0	0	0.284	0.760
RISK 22	0.722	0.926	0.500	0.724	0.016	0.515	0	0.097	0.738
RISK 8	0.806	0.861	0.500	0.516	0.017	0.523	1.000	0.097	0.702
RISK 2	0.843	0.866	0.500	0.449	0.019	0.500	1.000	0.097	0.693
RISK 26	0.602	0.831	0.500	0.647	0.018	0.500	0	0.097	0.664
RISK 15	0.611	0.840	0.500	0.631	0.022	0.469	0	0.097	0.658
GOV BODY 3	0.176	0.459	0	0.037	0.317	0.023	0	0.284	0.637
RISK 25	0.537	0.814	0.500	0.647	0.020	0.477	0	0.097	0.637
RISK 19	0.491	0.892	0.500	0.636	0.029	0.377	0	0.097	0.621
RISK 5	0.806	0.792	0.500	0.283	0.031	0.400	0	0.097	0.609
RISK 20	0.472	0.840	0.500	0.615	0.026	0.415	0	0.097	0.607
RISK 21	0.472	0.775	0.500	0.588	0.028	0.423	1.000	0.097	0.587
RISK 11	0.435	0.779	0.500	0.583	0.029	0.400	0	0.097	0.572
RISK 10	0.435	0.727	0.500	0.529	0.029	0.400	0	0.097	0.548
RISK 9	0.843	0.693	0	0.229	0.029	0.423	1.000	0.097	0.539
RISK 13	0.389	0.753	0.500	0.390	0.045	0.300	1.000	0.097	0.492
RISK 16	0.333	0.732	0.500	0.435	0.040	0.338	0	0.097	0.489
RISK 12	0.324	0.792	0.500	0.401	0.048	0.277	0	0.097	0.482
RISK 17	0.306	0.697	0.500	0.377	0.040	0.338	0	0.097	0.461
RISK 18	0.259	0.662	0.500	0.377	0.040	0.346	0	0.097	0.442
RISK 7	0.287	0.641	0.500	0.302	0.039	0.362	1.000	0.097	0.431
RISK 6	0.583	0.519	0.500	0.055	0.055	0.292	1.000	0.097	0.414
RISK 23	0.222	0.719	0.500	0.328	0.074	0.192	0	0.097	0.406
RISK 4	0.444	0.524	0.500	0.077	0.051	0.300	1.000	0.097	0.384
RISK 24	0.111	0.662	0.500	0.304	0.081	0.177	0	0.097	0.355
CONTROL 465	0	0.524	1.000	0.112	0.506	0.015	0	0.097	0.325
CONTROL 466	0	0.524	1.000	0.112	0.506	0.015	0	0.097	0.325
CONTROL 467	0	0.524	1.000	0.112	0.506	0.015	0	0.097	0.325
CONTROL 468	0	0.524	1.000	0.112	0.506	0.015	0	0.097	0.325
CONTROL 165	0	0.528	1.000	0.098	0.506	0.015	0	0.097	0.323
CONTROL 166	0	0.528	1.000	0.098	0.506	0.015	0	0.097	0.323
CONTROL 392	0	0.519	1.000	0.106	0.510	0.015	0	0.097	0.323
CONTROL 393	0	0.519	1.000	0.106	0.510	0.015	0	0.097	0.323
CONTROL 34	0	0.390	1.000	0.034	0.819	0.015	0	0.097	0.321
CONTROL 302	0	0.511	1.000	0.106	0.507	0.015	0	0.097	0.321
CONTROL 582	0	0.385	1.000	0.039	0.819	0.015	0	0.097	0.321
CONTROL 413	0	0.498	1.000	0.104	0.509	0.015	0	0.097	0.318
CONTROL 414	0	0.498	1.000	0.104	0.509	0.015	0	0.097	0.318
CONTROL 516	0	0.498	1.000	0.106	0.507	0.015	0	0.097	0.318
CONTROL 517	0	0.498	1.000	0.106	0.507	0.015	0	0.097	0.318
CONTROL 518	0	0.498	1.000	0.106	0.507	0.015	0	0.097	0.318
CONTROL 519	0	0.498	1.000	0.106	0.507	0.015	0	0.097	0.318
CONTROL 520	0	0.498	1.000	0.106	0.507	0.015	0	0.097	0.318
CONTROL 550	0	0.494	1.000	0.106	0.506	0.015	0	0.097	0.317
CONTROL 551	0	0.494	1.000	0.106	0.506	0.015	0	0.097	0.317
CONTROL 552	0	0.494	1.000	0.106	0.506	0.015	0	0.097	0.317
CONTROL 553	0	0.494	1.000	0.106	0.506	0.015	0	0.097	0.317

Node-Title	BC	CC	ECC	EiC	SHC	TDC	BS	Criticality of Node	Node Risk Value
CONTROL 372	0	0.476	1.000	0.114	0.508	0.015	0	0.097	0.314
CONTROL 373	0	0.476	1.000	0.114	0.508	0.015	0	0.097	0.314
CONTROL 374	0	0.476	1.000	0.114	0.508	0.015	0	0.097	0.314
CONTROL 375	0	0.476	1.000	0.114	0.508	0.015	0	0.097	0.314
CONTROL 376	0	0.476	1.000	0.114	0.508	0.015	0	0.097	0.314
CONTROL 377	0	0.476	1.000	0.114	0.508	0.015	0	0.097	0.314
CONTROL 378	0	0.476	1.000	0.114	0.508	0.015	0	0.097	0.314
CONTROL 438	0	0.485	1.000	0.104	0.508	0.015	0	0.097	0.314
CONTROL 439	0	0.485	1.000	0.104	0.508	0.015	0	0.097	0.314
CONTROL 440	0	0.485	1.000	0.104	0.508	0.015	0	0.097	0.314
CONTROL 441	0	0.485	1.000	0.104	0.508	0.015	0	0.097	0.314
CONTROL 460	0	0.494	1.000	0.085	0.520	0.015	0	0.097	0.314
CONTROL 461	0	0.494	1.000	0.085	0.520	0.015	0	0.097	0.314
CONTROL 422	0	0.342	1.000	0.045	0.821	0.015	0	0.097	0.312
CONTROL 459	0	0.481	1.000	0.077	0.522	0.015	0	0.097	0.310
CONTROL 490	0	0.455	1.000	0.117	0.505	0.015	0	0.097	0.310
CONTROL 491	0	0.455	1.000	0.117	0.505	0.015	0	0.097	0.310
CONTROL 492	0	0.455	1.000	0.117	0.505	0.015	0	0.097	0.310
CONTROL 493	0	0.455	1.000	0.117	0.505	0.015	0	0.097	0.310
CONTROL 494	0	0.455	1.000	0.117	0.505	0.015	0	0.097	0.310
CONTROL 495	0	0.455	1.000	0.117	0.505	0.015	0	0.097	0.310
CONTROL 496	0	0.455	1.000	0.117	0.505	0.015	0	0.097	0.310
CONTROL 497	0	0.455	1.000	0.117	0.505	0.015	0	0.097	0.310
CONTROL 498	0	0.455	1.000	0.117	0.505	0.015	0	0.097	0.310
CONTROL 499	0	0.455	1.000	0.117	0.505	0.015	0	0.097	0.310
CONTROL 500	0	0.455	1.000	0.117	0.505	0.015	0	0.097	0.310
CONTROL 501	0	0.455	1.000	0.117	0.505	0.015	0	0.097	0.310
CONTROL 502	0	0.455	1.000	0.117	0.505	0.015	0	0.097	0.310
CONTROL 503	0	0.455	1.000	0.117	0.505	0.015	0	0.097	0.310
CONTROL 216	0	0.455	1.000	0.112	0.507	0.015	0	0.097	0.309
CONTROL 217	0	0.455	1.000	0.112	0.507	0.015	0	0.097	0.309
CONTROL 218	0	0.455	1.000	0.112	0.507	0.015	0	0.097	0.309
CONTROL 219	0	0.455	1.000	0.112	0.507	0.015	0	0.097	0.309
CONTROL 220	0	0.455	1.000	0.112	0.507	0.015	0	0.097	0.309
CONTROL 221	0	0.455	1.000	0.112	0.507	0.015	0	0.097	0.309
CONTROL 222	0	0.455	1.000	0.112	0.507	0.015	0	0.097	0.309
CONTROL 223	0	0.455	1.000	0.112	0.507	0.015	0	0.097	0.309
CONTROL 224	0	0.455	1.000	0.112	0.507	0.015	0	0.097	0.309
CONTROL 225	0	0.455	1.000	0.112	0.507	0.015	0	0.097	0.309
CONTROL 226	0	0.455	1.000	0.112	0.507	0.015	0	0.097	0.309
CONTROL 227	0	0.455	1.000	0.112	0.507	0.015	0	0.097	0.309
CONTROL 228	0	0.455	1.000	0.112	0.507	0.015	0	0.097	0.309
CONTROL 272	0	0.455	1.000	0.114	0.505	0.015	0	0.097	0.309
CONTROL 273	0	0.455	1.000	0.114	0.505	0.015	0	0.097	0.309
CONTROL 274	0	0.455	1.000	0.114	0.505	0.015	0	0.097	0.309
CONTROL 275	0	0.455	1.000	0.114	0.505	0.015	0	0.097	0.309
CONTROL 276	0	0.455	1.000	0.114	0.505	0.015	0	0.097	0.309
CONTROL 277	0	0.455	1.000	0.114	0.505	0.015	0	0.097	0.309
CONTROL 278	0	0.455	1.000	0.114	0.505	0.015	0	0.097	0.309
CONTROL 279	0	0.455	1.000	0.114	0.505	0.015	0	0.097	0.309
CONTROL 280	0	0.455	1.000	0.114	0.505	0.015	0	0.097	0.309
CONTROL 281	0	0.455	1.000	0.114	0.505	0.015	0	0.097	0.309
CONTROL 282	0	0.455	1.000	0.114	0.505	0.015	0	0.097	0.309
CONTROL 283	0	0.455	1.000	0.114	0.505	0.015	0	0.097	0.309
CONTROL 284	0	0.455	1.000	0.114	0.505	0.015	0	0.097	0.309
CONTROL 458	0	0.472	1.000	0.069	0.543	0.015	0	0.097	0.309
CONTROL 477	0	0.468	1.000	0.085	0.525	0.015	0	0.097	0.309
CONTROL 478	0	0.468	1.000	0.085	0.525	0.015	0	0.097	0.309
CONTROL 325	0	0.468	1.000	0.090	0.512	0.015	0	0.097	0.308
CONTROL 557	0	0.459	1.000	0.098	0.513	0.015	0	0.097	0.308
CONTROL 558	0	0.459	1.000	0.098	0.513	0.015	0	0.097	0.308
CONTROL 381	0	0.476	1.000	0.069	0.526	0.015	0	0.097	0.307
CONTROL 443	0	0.468	1.000	0.088	0.508	0.015	0	0.097	0.307
CONTROL 444	0	0.468	1.000	0.088	0.508	0.015	0	0.097	0.307
CONTROL 445	0	0.468	1.000	0.088	0.508	0.015	0	0.097	0.307
CONTROL 446	0	0.468	1.000	0.088	0.508	0.015	0	0.097	0.307
CONTROL 447	0	0.468	1.000	0.088	0.508	0.015	0	0.097	0.307
CONTROL 448	0	0.468	1.000	0.088	0.508	0.015	0	0.097	0.307
CONTROL 336	0	0.463	1.000	0.088	0.512	0.015	0	0.097	0.306
CONTROL 337	0	0.463	1.000	0.088	0.512	0.015	0	0.097	0.306
CONTROL 338	0	0.463	1.000	0.088	0.512	0.015	0	0.097	0.306
CONTROL 339	0	0.463	1.000	0.088	0.512	0.015	0	0.097	0.306

Node-Title	BC	CC	ECC	EiC	SHC	TDC	BS	Criticality of Node	Node Risk Value
CONTROL 340	0	0.463	1.000	0.088	0.512	0.015	0	0.097	0.306
CONTROL 341	0	0.463	1.000	0.088	0.512	0.015	0	0.097	0.306
CONTROL 342	0	0.463	1.000	0.088	0.512	0.015	0	0.097	0.306
CONTROL 343	0	0.463	1.000	0.088	0.512	0.015	0	0.097	0.306
CONTROL 344	0	0.463	1.000	0.088	0.512	0.015	0	0.097	0.306
CONTROL 345	0	0.463	1.000	0.088	0.512	0.015	0	0.097	0.306
CONTROL 346	0	0.463	1.000	0.088	0.512	0.015	0	0.097	0.306
CONTROL 408	0	0.463	1.000	0.077	0.524	0.015	0	0.097	0.306
CONTROL 415	0	0.442	1.000	0.112	0.507	0.015	0	0.097	0.306
CONTROL 416	0	0.442	1.000	0.112	0.507	0.015	0	0.097	0.306
CONTROL 417	0	0.442	1.000	0.112	0.507	0.015	0	0.097	0.306
CONTROL 418	0	0.442	1.000	0.112	0.507	0.015	0	0.097	0.306
CONTROL 419	0	0.442	1.000	0.112	0.507	0.015	0	0.097	0.306
CONTROL 420	0	0.442	1.000	0.112	0.507	0.015	0	0.097	0.306
CONTROL 421	0	0.442	1.000	0.112	0.507	0.015	0	0.097	0.306
CONTROL 423	0	0.442	1.000	0.112	0.507	0.015	0	0.097	0.306
CONTROL 546	0	0.463	1.000	0.080	0.520	0.015	0	0.097	0.306
CONTROL 612	0	0.463	1.000	0.088	0.512	0.015	0	0.097	0.306
CONTROL 382	0	0.459	1.000	0.080	0.525	0.015	0	0.097	0.305
CONTROL 190	0	0.433	1.000	0.109	0.507	0.015	0	0.097	0.303
CONTROL 191	0	0.433	1.000	0.109	0.507	0.015	0	0.097	0.303
CONTROL 192	0	0.433	1.000	0.109	0.507	0.015	0	0.097	0.303
CONTROL 193	0	0.433	1.000	0.109	0.507	0.015	0	0.097	0.303
CONTROL 194	0	0.433	1.000	0.109	0.507	0.015	0	0.097	0.303
CONTROL 195	0	0.433	1.000	0.109	0.507	0.015	0	0.097	0.303
CONTROL 196	0	0.433	1.000	0.109	0.507	0.015	0	0.097	0.303
CONTROL 197	0	0.433	1.000	0.109	0.507	0.015	0	0.097	0.303
CONTROL 198	0	0.433	1.000	0.109	0.507	0.015	0	0.097	0.303
CONTROL 199	0	0.433	1.000	0.109	0.507	0.015	0	0.097	0.303
CONTROL 200	0	0.433	1.000	0.109	0.507	0.015	0	0.097	0.303
CONTROL 201	0	0.433	1.000	0.109	0.507	0.015	0	0.097	0.303
CONTROL 357	0	0.450	1.000	0.088	0.511	0.015	0	0.097	0.303
CONTROL 358	0	0.450	1.000	0.088	0.511	0.015	0	0.097	0.303
CONTROL 359	0	0.450	1.000	0.088	0.511	0.015	0	0.097	0.303
CONTROL 360	0	0.450	1.000	0.088	0.511	0.015	0	0.097	0.303
CONTROL 361	0	0.450	1.000	0.088	0.511	0.015	0	0.097	0.303
CONTROL 362	0	0.450	1.000	0.088	0.511	0.015	0	0.097	0.303
CONTROL 363	0	0.450	1.000	0.088	0.511	0.015	0	0.097	0.303
CONTROL 364	0	0.450	1.000	0.088	0.511	0.015	0	0.097	0.303
CONTROL 365	0	0.450	1.000	0.088	0.511	0.015	0	0.097	0.303
CONTROL 366	0	0.450	1.000	0.088	0.511	0.015	0	0.097	0.303
CONTROL 367	0	0.450	1.000	0.088	0.511	0.015	0	0.097	0.303
CONTROL 368	0	0.450	1.000	0.088	0.511	0.015	0	0.097	0.303
CONTROL 369	0	0.450	1.000	0.088	0.511	0.015	0	0.097	0.303
CONTROL 370	0	0.450	1.000	0.088	0.511	0.015	0	0.097	0.303
CONTROL 371	0	0.450	1.000	0.088	0.511	0.015	0	0.097	0.303
CONTROL 469	0	0.437	1.000	0.093	0.523	0.015	0	0.097	0.303
CONTROL 470	0	0.437	1.000	0.093	0.523	0.015	0	0.097	0.303
CONTROL 487	0	0.446	1.000	0.082	0.527	0.015	0	0.097	0.303
CONTROL 488	0	0.446	1.000	0.082	0.527	0.015	0	0.097	0.303
CONTROL 21	0	0.463	1.000	0.069	0.508	0.015	0	0.097	0.302
CONTROL 22	0	0.463	1.000	0.069	0.508	0.015	0	0.097	0.302
CONTROL 23	0	0.463	1.000	0.069	0.508	0.015	0	0.097	0.302
CONTROL 290	0	0.455	1.000	0.069	0.523	0.015	0	0.097	0.302
CONTROL 457	0	0.424	1.000	0.061	0.591	0.015	0	0.097	0.302
CONTROL 559	0	0.437	1.000	0.098	0.512	0.015	0	0.097	0.302
CONTROL 560	0	0.437	1.000	0.098	0.512	0.015	0	0.097	0.302
CONTROL 561	0	0.437	1.000	0.098	0.512	0.015	0	0.097	0.302
CONTROL 562	0	0.437	1.000	0.098	0.512	0.015	0	0.097	0.302
CONTROL 563	0	0.437	1.000	0.098	0.512	0.015	0	0.097	0.302
CONTROL 564	0	0.437	1.000	0.098	0.512	0.015	0	0.097	0.302
CONTROL 565	0	0.437	1.000	0.098	0.512	0.015	0	0.097	0.302
CONTROL 407	0	0.450	1.000	0.069	0.525	0.015	0	0.097	0.301
CONTROL 545	0	0.450	1.000	0.072	0.522	0.015	0	0.097	0.301
CONTROL 289	0	0.442	1.000	0.061	0.544	0.015	0	0.097	0.300
CONTROL 291	0	0.442	1.000	0.077	0.522	0.015	0	0.097	0.300
CONTROL 394	0	0.442	1.000	0.080	0.510	0.015	0	0.097	0.299
CONTROL 395	0	0.442	1.000	0.080	0.510	0.015	0	0.097	0.299
CONTROL 511	0	0.437	1.000	0.063	0.544	0.015	0	0.097	0.299
CONTROL 512	0	0.433	1.000	0.080	0.522	0.015	0	0.097	0.299
CONTROL 544	0	0.433	1.000	0.063	0.543	0.015	0	0.097	0.298
CONTROL 449	0	0.437	1.000	0.066	0.522	0.015	0	0.097	0.297

Node-Title	BC	CC	ECC	EiC	SHC	TDC	BS	Criticality of Node	Node Risk Value
CONTROL 522	0	0.433	1.000	0.082	0.508	0.015	0	0.097	0.297
CONTROL 523	0	0.433	1.000	0.082	0.508	0.015	0	0.097	0.297
CONTROL 524	0	0.433	1.000	0.082	0.508	0.015	0	0.097	0.297
CONTROL 525	0	0.433	1.000	0.082	0.508	0.015	0	0.097	0.297
CONTROL 526	0	0.433	1.000	0.082	0.508	0.015	0	0.097	0.297
CONTROL 527	0	0.433	1.000	0.082	0.508	0.015	0	0.097	0.297
CONTROL 528	0	0.433	1.000	0.082	0.508	0.015	0	0.097	0.297
CONTROL 529	0	0.433	1.000	0.082	0.508	0.015	0	0.097	0.297
CONTROL 530	0	0.433	1.000	0.082	0.508	0.015	0	0.097	0.297
CONTROL 531	0	0.433	1.000	0.082	0.508	0.015	0	0.097	0.297
CONTROL 532	0	0.433	1.000	0.082	0.508	0.015	0	0.097	0.297
CONTROL 533	0	0.433	1.000	0.082	0.508	0.015	0	0.097	0.297
CONTROL 534	0	0.433	1.000	0.082	0.508	0.015	0	0.097	0.297
CONTROL 570	0	0.437	1.000	0.066	0.522	0.015	0	0.097	0.297
CONTROL 599	0	0.437	1.000	0.066	0.522	0.015	0	0.097	0.297
CONTROL 600	0	0.437	1.000	0.066	0.522	0.015	0	0.097	0.297
CONTROL 610	0	0.437	1.000	0.066	0.522	0.015	0	0.097	0.297
CONTROL 326	0	0.416	1.000	0.096	0.510	0.015	0	0.097	0.296
CONTROL 327	0	0.416	1.000	0.096	0.510	0.015	0	0.097	0.296
CONTROL 328	0	0.416	1.000	0.096	0.510	0.015	0	0.097	0.296
CONTROL 19	0	0.424	1.000	0.042	0.567	0.015	0	0.097	0.295
CONTROL 20	0	0.424	1.000	0.042	0.567	0.015	0	0.097	0.295
CONTROL 235	0	0.424	1.000	0.074	0.524	0.015	0	0.097	0.295
CONTROL 236	0	0.424	1.000	0.074	0.524	0.015	0	0.097	0.295
CONTROL 237	0	0.424	1.000	0.074	0.524	0.015	0	0.097	0.295
CONTROL 380	0	0.416	1.000	0.061	0.551	0.015	0	0.097	0.295
CONTROL 387	0	0.429	1.000	0.066	0.524	0.015	0	0.097	0.295
CONTROL 388	0	0.429	1.000	0.066	0.524	0.015	0	0.097	0.295
CONTROL 389	0	0.429	1.000	0.066	0.524	0.015	0	0.097	0.295
CONTROL 390	0	0.429	1.000	0.066	0.524	0.015	0	0.097	0.295
CONTROL 28	0	0.442	1.000	0.047	0.527	0.015	0	0.097	0.294
CONTROL 250	0	0.433	1.000	0.053	0.531	0.015	0	0.097	0.294
CONTROL 406	0	0.394	1.000	0.053	0.594	0.015	0	0.097	0.294
CONTROL 433	0	0.420	1.000	0.074	0.523	0.015	0	0.097	0.294
CONTROL 434	0	0.420	1.000	0.074	0.523	0.015	0	0.097	0.294
CONTROL 543	0	0.390	1.000	0.055	0.591	0.015	0	0.097	0.293
CONTROL 204	0	0.411	1.000	0.072	0.524	0.015	0	0.097	0.292
CONTROL 205	0	0.411	1.000	0.072	0.524	0.015	0	0.097	0.292
CONTROL 206	0	0.411	1.000	0.072	0.524	0.015	0	0.097	0.292
CONTROL 471	0	0.420	1.000	0.061	0.527	0.015	0	0.097	0.292
CONTROL 504	0	0.416	1.000	0.061	0.528	0.015	0	0.097	0.291
CONTROL 505	0	0.416	1.000	0.061	0.528	0.015	0	0.097	0.291
CONTROL 29	0	0.403	1.000	0.034	0.578	0.015	0	0.097	0.290
CONTROL 30	0	0.403	1.000	0.034	0.578	0.015	0	0.097	0.290
CONTROL 123	0	0.268	1.000	0.023	0.824	0.015	0	0.097	0.290
CONTROL 287	0	0.398	1.000	0.061	0.548	0.015	0	0.097	0.290
CONTROL 288	0	0.398	1.000	0.061	0.548	0.015	0	0.097	0.290
CONTROL 472	0	0.416	1.000	0.047	0.540	0.015	0	0.097	0.290
CONTROL 473	0	0.416	1.000	0.047	0.540	0.015	0	0.097	0.290
CONTROL 510	0	0.394	1.000	0.063	0.548	0.015	0	0.097	0.289
CONTROL 295	0	0.407	1.000	0.063	0.520	0.015	0	0.097	0.288
CONTROL 296	0	0.407	1.000	0.063	0.520	0.015	0	0.097	0.288
CONTROL 297	0	0.407	1.000	0.063	0.520	0.015	0	0.097	0.288
CONTROL 298	0	0.407	1.000	0.063	0.520	0.015	0	0.097	0.288
CONTROL 299	0	0.407	1.000	0.063	0.520	0.015	0	0.097	0.288
CONTROL 300	0	0.407	1.000	0.063	0.520	0.015	0	0.097	0.288
CONTROL 301	0	0.407	1.000	0.063	0.520	0.015	0	0.097	0.288
CONTROL 396	0	0.407	1.000	0.058	0.525	0.015	0	0.097	0.288
CONTROL 567	0	0.407	1.000	0.058	0.525	0.015	0	0.097	0.288
CONTROL 568	0	0.407	1.000	0.058	0.525	0.015	0	0.097	0.288
CONTROL 569	0	0.407	1.000	0.058	0.525	0.015	0	0.097	0.288
CONTROL 598	0	0.407	1.000	0.058	0.525	0.015	0	0.097	0.288
CONTROL 456	0	0.571	0.500	0.104	0.510	0.015	0	0.097	0.287
CONTROL 234	0	0.385	1.000	0.058	0.550	0.015	0	0.097	0.286
CONTROL 303	0	0.394	1.000	0.069	0.513	0.015	0	0.097	0.286
CONTROL 304	0	0.394	1.000	0.069	0.513	0.015	0	0.097	0.286
CONTROL 305	0	0.394	1.000	0.069	0.513	0.015	0	0.097	0.286
CONTROL 306	0	0.394	1.000	0.069	0.513	0.015	0	0.097	0.286
CONTROL 307	0	0.394	1.000	0.069	0.513	0.015	0	0.097	0.286
CONTROL 308	0	0.394	1.000	0.069	0.513	0.015	0	0.097	0.286
CONTROL 309	0	0.394	1.000	0.069	0.513	0.015	0	0.097	0.286
CONTROL 310	0	0.394	1.000	0.069	0.513	0.015	0	0.097	0.286

Node-Title	BC	CC	ECC	EiC	SHC	TDC	BS	Criticality of Node	Node Risk Value
CONTROL 311	0	0.394	1.000	0.069	0.513	0.015	0	0.097	0.286
CONTROL 312	0	0.394	1.000	0.069	0.513	0.015	0	0.097	0.286
CONTROL 313	0	0.394	1.000	0.069	0.513	0.015	0	0.097	0.286
CONTROL 314	0	0.394	1.000	0.069	0.513	0.015	0	0.097	0.286
CONTROL 315	0	0.394	1.000	0.069	0.513	0.015	0	0.097	0.286
CONTROL 316	0	0.394	1.000	0.069	0.513	0.015	0	0.097	0.286
CONTROL 317	0	0.394	1.000	0.069	0.513	0.015	0	0.097	0.286
CONTROL 323	0	0.394	1.000	0.047	0.549	0.015	0	0.097	0.286
CONTROL 424	0	0.398	1.000	0.055	0.529	0.015	0	0.097	0.286
CONTROL 425	0	0.398	1.000	0.055	0.529	0.015	0	0.097	0.286
CONTROL 426	0	0.398	1.000	0.055	0.529	0.015	0	0.097	0.286
CONTROL 427	0	0.398	1.000	0.055	0.529	0.015	0	0.097	0.286
CONTROL 428	0	0.398	1.000	0.055	0.529	0.015	0	0.097	0.286
CONTROL 429	0	0.398	1.000	0.055	0.529	0.015	0	0.097	0.286
CONTROL 566	0	0.398	1.000	0.055	0.529	0.015	0	0.097	0.286
CONTROL 146	0	0.576	0.500	0.090	0.510	0.015	0	0.097	0.285
CONTROL 485	0	0.390	1.000	0.047	0.541	0.015	0	0.097	0.284
CONTROL 462	0	0.558	0.500	0.098	0.512	0.015	0	0.097	0.283
CONTROL 348	0	0.385	1.000	0.063	0.513	0.015	0	0.097	0.282
CONTROL 349	0	0.385	1.000	0.063	0.513	0.015	0	0.097	0.282
CONTROL 350	0	0.385	1.000	0.063	0.513	0.015	0	0.097	0.282
CONTROL 351	0	0.385	1.000	0.063	0.513	0.015	0	0.097	0.282
CONTROL 352	0	0.385	1.000	0.063	0.513	0.015	0	0.097	0.282
CONTROL 464	0	0.558	0.500	0.096	0.509	0.015	0	0.097	0.282
CONTROL 210	0	0.381	1.000	0.058	0.523	0.015	0	0.097	0.281
CONTROL 211	0	0.381	1.000	0.058	0.523	0.015	0	0.097	0.281
CONTROL 212	0	0.381	1.000	0.058	0.523	0.015	0	0.097	0.281
CONTROL 213	0	0.381	1.000	0.058	0.523	0.015	0	0.097	0.281
CONTROL 214	0	0.381	1.000	0.058	0.523	0.015	0	0.097	0.281
CONTROL 243	0	0.385	1.000	0.045	0.531	0.015	0	0.097	0.281
CONTROL 244	0	0.385	1.000	0.045	0.531	0.015	0	0.097	0.281
CONTROL 379	0	0.550	0.500	0.098	0.513	0.015	0	0.097	0.281
CONTROL 261	0	0.381	1.000	0.047	0.528	0.015	0	0.097	0.280
CONTROL 285	0	0.545	0.500	0.098	0.511	0.015	0	0.097	0.280
CONTROL 286	0	0.545	0.500	0.098	0.511	0.015	0	0.097	0.280
CONTROL 479	0	0.372	1.000	0.058	0.528	0.015	0	0.097	0.280
CONTROL 480	0	0.372	1.000	0.058	0.528	0.015	0	0.097	0.280
CONTROL 481	0	0.372	1.000	0.058	0.528	0.015	0	0.097	0.280
RISK 3	0.435	0.199	0.500	0.010	0.072	0.246	1.000	0.097	0.279
CONTROL 203	0	0.359	1.000	0.053	0.550	0.015	0	0.097	0.279
CONTROL 539	0	0.541	0.500	0.098	0.510	0.015	0	0.097	0.279
CONTROL 540	0	0.541	0.500	0.098	0.510	0.015	0	0.097	0.279
CONTROL 541	0	0.541	0.500	0.098	0.510	0.015	0	0.097	0.279
CONTROL 542	0	0.541	0.500	0.098	0.510	0.015	0	0.097	0.279
CONTROL 383	0	0.541	0.500	0.090	0.516	0.015	0	0.097	0.278
CONTROL 384	0	0.541	0.500	0.090	0.516	0.015	0	0.097	0.278
CONTROL 476	0	0.372	1.000	0.042	0.538	0.015	0	0.097	0.278
CONTROL 330	0	0.372	1.000	0.042	0.532	0.015	0	0.097	0.277
CONTROL 331	0	0.372	1.000	0.042	0.532	0.015	0	0.097	0.277
CONTROL 347	0	0.372	1.000	0.042	0.532	0.015	0	0.097	0.277
CONTROL 403	0	0.537	0.500	0.096	0.512	0.015	0	0.097	0.277
CONTROL 404	0	0.537	0.500	0.096	0.512	0.015	0	0.097	0.277
CONTROL 405	0	0.537	0.500	0.096	0.512	0.015	0	0.097	0.277
CONTROL 450	0	0.532	0.500	0.104	0.507	0.015	0	0.097	0.277
CONTROL 451	0	0.532	0.500	0.104	0.507	0.015	0	0.097	0.277
CONTROL 452	0	0.532	0.500	0.104	0.507	0.015	0	0.097	0.277
CONTROL 453	0	0.532	0.500	0.104	0.507	0.015	0	0.097	0.277
CONTROL 454	0	0.532	0.500	0.104	0.507	0.015	0	0.097	0.277
CONTROL 594	0	0.532	0.500	0.104	0.507	0.015	0	0.097	0.277
CONTROL 252	0	0.368	1.000	0.042	0.529	0.015	0	0.097	0.276
CONTROL 253	0	0.368	1.000	0.042	0.529	0.015	0	0.097	0.276
CONTROL 254	0	0.368	1.000	0.042	0.529	0.015	0	0.097	0.276
CONTROL 255	0	0.368	1.000	0.042	0.529	0.015	0	0.097	0.276
CONTROL 256	0	0.368	1.000	0.042	0.529	0.015	0	0.097	0.276
CONTROL 257	0	0.368	1.000	0.042	0.529	0.015	0	0.097	0.276
CONTROL 258	0	0.368	1.000	0.042	0.529	0.015	0	0.097	0.276
CONTROL 391	0	0.537	0.500	0.090	0.513	0.015	0	0.097	0.276
CONTROL 442	0	0.528	0.500	0.088	0.528	0.015	0	0.097	0.276
CONTROL 455	0	0.541	0.500	0.085	0.512	0.015	0	0.097	0.276
CONTROL 506	0	0.528	0.500	0.098	0.510	0.015	0	0.097	0.275
CONTROL 507	0	0.528	0.500	0.098	0.510	0.015	0	0.097	0.275
CONTROL 508	0	0.528	0.500	0.098	0.510	0.015	0	0.097	0.275

Node-Title	BC	CC	ECC	EiC	SHC	TDC	BS	Criticality of Node	Node Risk Value
CONTROL 509	0	0.528	0.500	0.098	0.510	0.015	0	0.097	0.275
CONTROL 44	0	0.541	0.500	0.077	0.510	0.015	0	0.097	0.274
CONTROL 45	0	0.541	0.500	0.077	0.510	0.015	0	0.097	0.274
CONTROL 292	0	0.528	0.500	0.090	0.513	0.015	0	0.097	0.274
CONTROL 335	0	0.359	1.000	0.047	0.526	0.015	0	0.097	0.274
CONTROL 409	0	0.528	0.500	0.090	0.515	0.015	0	0.097	0.274
CONTROL 410	0	0.528	0.500	0.090	0.515	0.015	0	0.097	0.274
CONTROL 412	0	0.528	0.500	0.088	0.512	0.015	0	0.097	0.274
CONTROL 547	0	0.528	0.500	0.090	0.513	0.015	0	0.097	0.274
CONTROL 548	0	0.528	0.500	0.090	0.510	0.015	0	0.097	0.274
CONTROL 142	0	0.528	0.500	0.088	0.506	0.015	0	0.097	0.273
CONTROL 143	0	0.528	0.500	0.088	0.506	0.015	0	0.097	0.273
CONTROL 144	0	0.528	0.500	0.088	0.506	0.015	0	0.097	0.273
CONTROL 149	0	0.532	0.500	0.082	0.509	0.015	0	0.097	0.273
CONTROL 150	0	0.532	0.500	0.082	0.509	0.015	0	0.097	0.273
CONTROL 151	0	0.532	0.500	0.082	0.509	0.015	0	0.097	0.273
CONTROL 152	0	0.532	0.500	0.082	0.509	0.015	0	0.097	0.273
CONTROL 153	0	0.532	0.500	0.082	0.509	0.015	0	0.097	0.273
CONTROL 154	0	0.532	0.500	0.082	0.509	0.015	0	0.097	0.273
CONTROL 155	0	0.532	0.500	0.082	0.509	0.015	0	0.097	0.273
CONTROL 156	0	0.532	0.500	0.082	0.509	0.015	0	0.097	0.273
CONTROL 157	0	0.532	0.500	0.082	0.509	0.015	0	0.097	0.273
CONTROL 158	0	0.532	0.500	0.082	0.509	0.015	0	0.097	0.273
CONTROL 159	0	0.532	0.500	0.082	0.509	0.015	0	0.097	0.273
CONTROL 160	0	0.532	0.500	0.082	0.509	0.015	0	0.097	0.273
CONTROL 161	0	0.532	0.500	0.082	0.509	0.015	0	0.097	0.273
CONTROL 162	0	0.532	0.500	0.082	0.509	0.015	0	0.097	0.273
CONTROL 163	0	0.532	0.500	0.082	0.509	0.015	0	0.097	0.273
CONTROL 164	0	0.532	0.500	0.082	0.509	0.015	0	0.097	0.273
CONTROL 430	0	0.519	0.500	0.096	0.512	0.015	0	0.097	0.273
CONTROL 431	0	0.519	0.500	0.096	0.512	0.015	0	0.097	0.273
CONTROL 432	0	0.519	0.500	0.096	0.512	0.015	0	0.097	0.273
CONTROL 513	0	0.524	0.500	0.090	0.513	0.015	0	0.097	0.273
CONTROL 593	0	0.528	0.500	0.088	0.506	0.015	0	0.097	0.273
CONTROL 31	0	0.532	0.500	0.077	0.513	0.015	0	0.097	0.272
CONTROL 233	0	0.515	0.500	0.096	0.513	0.015	0	0.097	0.272
CONTROL 535	0	0.515	0.500	0.098	0.507	0.015	0	0.097	0.272
CONTROL 536	0	0.515	0.500	0.098	0.507	0.015	0	0.097	0.272
CONTROL 537	0	0.515	0.500	0.098	0.507	0.015	0	0.097	0.272
CONTROL 538	0	0.515	0.500	0.098	0.507	0.015	0	0.097	0.272
CONTROL 69	0	0.359	1.000	0.031	0.529	0.015	0	0.097	0.271
CONTROL 70	0	0.359	1.000	0.031	0.529	0.015	0	0.097	0.271
CONTROL 71	0	0.359	1.000	0.031	0.529	0.015	0	0.097	0.271
CONTROL 72	0	0.359	1.000	0.031	0.529	0.015	0	0.097	0.271
CONTROL 73	0	0.359	1.000	0.031	0.529	0.015	0	0.097	0.271
CONTROL 229	0	0.515	0.500	0.093	0.510	0.015	0	0.097	0.271
CONTROL 230	0	0.515	0.500	0.093	0.510	0.015	0	0.097	0.271
CONTROL 231	0	0.515	0.500	0.093	0.510	0.015	0	0.097	0.271
CONTROL 232	0	0.515	0.500	0.093	0.510	0.015	0	0.097	0.271
CONTROL 249	0	0.519	0.500	0.082	0.519	0.015	0	0.097	0.271
CONTROL 321	0	0.519	0.500	0.085	0.515	0.015	0	0.097	0.271
CONTROL 322	0	0.519	0.500	0.085	0.515	0.015	0	0.097	0.271
CONTROL 385	0	0.524	0.500	0.080	0.516	0.015	0	0.097	0.271
CONTROL 463	0	0.519	0.500	0.080	0.525	0.015	0	0.097	0.271
CONTROL 514	0	0.515	0.500	0.090	0.510	0.015	0	0.097	0.271
CONTROL 555	0	0.519	0.500	0.082	0.519	0.015	0	0.097	0.271
CONTROL 592	0	0.519	0.500	0.082	0.519	0.015	0	0.097	0.271
CONTROL 74	0	0.199	1.000	0.005	0.827	0.015	0	0.097	0.270
CONTROL 80	0	0.199	1.000	0.005	0.827	0.015	0	0.097	0.270
CONTROL 104	0	0.199	1.000	0.005	0.827	0.015	0	0.097	0.270
CONTROL 202	0	0.511	0.500	0.090	0.513	0.015	0	0.097	0.270
CONTROL 435	0	0.511	0.500	0.088	0.515	0.015	0	0.097	0.270
CONTROL 591	0	0.346	1.000	0.037	0.531	0.015	0	0.097	0.270
CONTROL 605	0	0.519	0.500	0.080	0.514	0.015	0	0.097	0.270
CONTROL 68	0	0.346	1.000	0.023	0.551	0.015	0	0.097	0.269
CONTROL 245	0	0.515	0.500	0.080	0.515	0.015	0	0.097	0.269
CONTROL 246	0	0.515	0.500	0.080	0.515	0.015	0	0.097	0.269
CONTROL 247	0	0.515	0.500	0.080	0.515	0.015	0	0.097	0.269
CONTROL 248	0	0.515	0.500	0.080	0.515	0.015	0	0.097	0.269
CONTROL 259	0	0.511	0.500	0.082	0.517	0.015	0	0.097	0.269
CONTROL 293	0	0.515	0.500	0.080	0.513	0.015	0	0.097	0.269
CONTROL 294	0	0.515	0.500	0.080	0.513	0.015	0	0.097	0.269

Node-Title	BC	CC	ECC	EiC	SHC	TDC	BS	Criticality of Node	Node Risk Value
CONTROL 397	0	0.506	0.500	0.096	0.509	0.015	0	0.097	0.269
CONTROL 398	0	0.506	0.500	0.096	0.509	0.015	0	0.097	0.269
CONTROL 399	0	0.506	0.500	0.096	0.509	0.015	0	0.097	0.269
CONTROL 400	0	0.506	0.500	0.096	0.509	0.015	0	0.097	0.269
CONTROL 401	0	0.506	0.500	0.096	0.509	0.015	0	0.097	0.269
CONTROL 402	0	0.506	0.500	0.096	0.509	0.015	0	0.097	0.269
CONTROL 24	0	0.511	0.500	0.085	0.507	0.015	0	0.097	0.268
CONTROL 25	0	0.511	0.500	0.085	0.507	0.015	0	0.097	0.268
CONTROL 26	0	0.511	0.500	0.085	0.507	0.015	0	0.097	0.268
CONTROL 27	0	0.511	0.500	0.085	0.507	0.015	0	0.097	0.268
CONTROL 83	0	0.524	0.500	0.066	0.515	0.015	0	0.097	0.268
CONTROL 84	0	0.524	0.500	0.066	0.515	0.015	0	0.097	0.268
CONTROL 521	0	0.502	0.500	0.082	0.529	0.015	0	0.097	0.268
CONTROL 238	0	0.498	0.500	0.088	0.515	0.015	0	0.097	0.267
CONTROL 241	0	0.502	0.500	0.085	0.512	0.015	0	0.097	0.267
CONTROL 437	0	0.502	0.500	0.085	0.512	0.015	0	0.097	0.267
CONTROL 386	0	0.498	0.500	0.074	0.529	0.015	0	0.097	0.266
CONTROL 35	0	0.511	0.500	0.066	0.512	0.015	0	0.097	0.265
CONTROL 36	0	0.511	0.500	0.066	0.512	0.015	0	0.097	0.265
CONTROL 37	0	0.511	0.500	0.066	0.512	0.015	0	0.097	0.265
CONTROL 38	0	0.511	0.500	0.066	0.512	0.015	0	0.097	0.265
CONTROL 39	0	0.511	0.500	0.066	0.512	0.015	0	0.097	0.265
CONTROL 40	0	0.511	0.500	0.066	0.512	0.015	0	0.097	0.265
CONTROL 145	0	0.506	0.500	0.072	0.512	0.015	0	0.097	0.265
RISK 14	0.148	0.455	0.500	0.080	0.118	0.146	1.000	0.097	0.265
CONTROL 586	0	0.511	0.500	0.066	0.512	0.015	0	0.097	0.265
CONTROL 607	0	0.506	0.500	0.072	0.512	0.015	0	0.097	0.265
CONTROL 81	0	0.506	0.500	0.072	0.510	0.015	0	0.097	0.264
CONTROL 215	0	0.494	0.500	0.082	0.512	0.015	0	0.097	0.264
CONTROL 324	0	0.494	0.500	0.077	0.518	0.015	0	0.097	0.264
CONTROL 133	0	0.494	0.500	0.074	0.514	0.015	0	0.097	0.263
CONTROL 207	0	0.489	0.500	0.082	0.515	0.015	0	0.097	0.263
CONTROL 411	0	0.489	0.500	0.072	0.528	0.015	0	0.097	0.263
CONTROL 489	0	0.481	0.500	0.082	0.529	0.015	0	0.097	0.263
CONTROL 587	0	0.498	0.500	0.063	0.522	0.015	0	0.097	0.262
CONTROL 588	0	0.498	0.500	0.063	0.522	0.015	0	0.097	0.262
CONTROL 318	0	0.481	0.500	0.082	0.512	0.015	0	0.097	0.261
CONTROL 319	0	0.481	0.500	0.082	0.512	0.015	0	0.097	0.261
CONTROL 320	0	0.481	0.500	0.082	0.512	0.015	0	0.097	0.261
CONTROL 148	0	0.489	0.500	0.063	0.525	0.015	0	0.097	0.260
CONTROL 590	0	0.489	0.500	0.063	0.525	0.015	0	0.097	0.260
CONTROL 608	0	0.489	0.500	0.063	0.525	0.015	0	0.097	0.260
CONTROL 260	0	0.489	0.500	0.061	0.520	0.015	0	0.097	0.259
CONTROL 486	0	0.476	0.500	0.066	0.533	0.015	0	0.097	0.259
CONTROL 353	0	0.472	0.500	0.080	0.512	0.015	0	0.097	0.258
CONTROL 354	0	0.472	0.500	0.080	0.512	0.015	0	0.097	0.258
CONTROL 355	0	0.472	0.500	0.080	0.512	0.015	0	0.097	0.258
CONTROL 595	0	0.481	0.500	0.061	0.526	0.015	0	0.097	0.258
CONTROL 85	0	0.489	0.500	0.053	0.515	0.015	0	0.097	0.257
CONTROL 86	0	0.489	0.500	0.053	0.515	0.015	0	0.097	0.257
CONTROL 87	0	0.489	0.500	0.053	0.515	0.015	0	0.097	0.257
CONTROL 134	0	0.472	0.500	0.066	0.517	0.015	0	0.097	0.256
CONTROL 135	0	0.472	0.500	0.066	0.517	0.015	0	0.097	0.256
CONTROL 136	0	0.472	0.500	0.066	0.517	0.015	0	0.097	0.256
CONTROL 137	0	0.472	0.500	0.066	0.517	0.015	0	0.097	0.256
CONTROL 138	0	0.472	0.500	0.066	0.517	0.015	0	0.097	0.256
CONTROL 139	0	0.472	0.500	0.066	0.517	0.015	0	0.097	0.256
CONTROL 140	0	0.472	0.500	0.066	0.517	0.015	0	0.097	0.256
CONTROL 436	0	0.463	0.500	0.069	0.528	0.015	0	0.097	0.256
CONTROL 474	0	0.472	0.500	0.058	0.531	0.015	0	0.097	0.256
CONTROL 475	0	0.472	0.500	0.058	0.531	0.015	0	0.097	0.256
CONTROL 482	0	0.459	0.500	0.074	0.527	0.015	0	0.097	0.256
CONTROL 483	0	0.459	0.500	0.074	0.527	0.015	0	0.097	0.256
CONTROL 484	0	0.459	0.500	0.074	0.527	0.015	0	0.097	0.256
CONTROL 125	0	0.463	0.500	0.074	0.511	0.015	0	0.097	0.255
CONTROL 126	0	0.463	0.500	0.074	0.511	0.015	0	0.097	0.255
CONTROL 127	0	0.463	0.500	0.074	0.511	0.015	0	0.097	0.255
CONTROL 128	0	0.463	0.500	0.074	0.511	0.015	0	0.097	0.255
CONTROL 129	0	0.463	0.500	0.074	0.511	0.015	0	0.097	0.255
CONTROL 130	0	0.463	0.500	0.074	0.511	0.015	0	0.097	0.255
CONTROL 131	0	0.463	0.500	0.074	0.511	0.015	0	0.097	0.255
CONTROL 132	0	0.463	0.500	0.074	0.511	0.015	0	0.097	0.255

Node-Title	BC	CC	ECC	EiC	SHC	TDC	BS	Criticality of Node	Node Risk Value
CONTROL 240	0	0.450	0.500	0.069	0.528	0.015	0	0.097	0.253
CONTROL 332	0	0.463	0.500	0.061	0.518	0.015	0	0.097	0.253
CONTROL 333	0	0.463	0.500	0.061	0.518	0.015	0	0.097	0.253
CONTROL 334	0	0.463	0.500	0.061	0.518	0.015	0	0.097	0.253
CONTROL 611	0	0.463	0.500	0.061	0.517	0.015	0	0.097	0.253
CONTROL 147	0	0.407	0.500	0.039	0.642	0.015	0	0.097	0.252
CONTROL 103	0	0.459	0.500	0.047	0.528	0.015	0	0.097	0.250
CONTROL 209	0	0.442	0.500	0.066	0.528	0.015	0	0.097	0.250
CONTROL 177	0	0.299	0.500	0.018	0.821	0.015	0	0.097	0.248
CONTROL 584	0	0.450	0.500	0.045	0.530	0.015	0	0.097	0.248
CONTROL 585	0	0.450	0.500	0.045	0.530	0.015	0	0.097	0.248
CONTROL 141	0	0.433	0.500	0.047	0.547	0.015	0	0.097	0.247
CONTROL 263	0	0.437	0.500	0.050	0.537	0.015	0	0.097	0.247
CONTROL 264	0	0.437	0.500	0.050	0.537	0.015	0	0.097	0.247
CONTROL 265	0	0.437	0.500	0.050	0.537	0.015	0	0.097	0.247
CONTROL 266	0	0.437	0.500	0.050	0.537	0.015	0	0.097	0.247
CONTROL 267	0	0.437	0.500	0.050	0.537	0.015	0	0.097	0.247
CONTROL 268	0	0.437	0.500	0.050	0.537	0.015	0	0.097	0.247
CONTROL 269	0	0.437	0.500	0.050	0.537	0.015	0	0.097	0.247
CONTROL 270	0	0.437	0.500	0.050	0.537	0.015	0	0.097	0.247
CONTROL 271	0	0.437	0.500	0.050	0.537	0.015	0	0.097	0.247
CONTROL 549	0	0.420	0.500	0.053	0.559	0.015	0	0.097	0.247
CONTROL 580	0	0.446	0.500	0.053	0.520	0.015	0	0.097	0.247
CONTROL 581	0	0.446	0.500	0.053	0.520	0.015	0	0.097	0.247
CONTROL 589	0	0.446	0.500	0.053	0.520	0.015	0	0.097	0.247
CONTROL 571	0	0.446	0.500	0.047	0.520	0.015	0	0.097	0.246
CONTROL 572	0	0.446	0.500	0.047	0.520	0.015	0	0.097	0.246
CONTROL 583	0	0.446	0.500	0.047	0.520	0.015	0	0.097	0.246
CONTROL 609	0	0.446	0.500	0.042	0.522	0.015	0	0.097	0.245
CONTROL 32	0	0.429	0.500	0.037	0.548	0.015	0	0.097	0.244
CONTROL 33	0	0.429	0.500	0.037	0.548	0.015	0	0.097	0.244
CONTROL 515	0	0.411	0.500	0.053	0.559	0.015	0	0.097	0.244
CONTROL 596	0	0.446	0.500	0.037	0.520	0.015	0	0.097	0.244
CONTROL 41	0	0.437	0.500	0.045	0.519	0.015	0	0.097	0.243
CONTROL 42	0	0.437	0.500	0.045	0.519	0.015	0	0.097	0.243
CONTROL 43	0	0.437	0.500	0.045	0.519	0.015	0	0.097	0.243
CONTROL 242	0	0.416	0.500	0.037	0.556	0.015	0	0.097	0.242
CONTROL 82	0	0.368	0.500	0.023	0.645	0.015	0	0.097	0.240
CONTROL 251	0	0.411	0.500	0.037	0.555	0.015	0	0.097	0.240
CONTROL 239	0	0.411	0.500	0.053	0.523	0.015	0	0.097	0.239
CONTROL 329	0	0.420	0.500	0.042	0.527	0.015	0	0.097	0.239
CONTROL 573	0	0.420	0.500	0.037	0.524	0.015	0	0.097	0.238
CONTROL 63	0	0.203	1.000	0.005	0.585	0.015	0	0.097	0.236
CONTROL 64	0	0.203	1.000	0.005	0.585	0.015	0	0.097	0.236
CONTROL 65	0	0.203	1.000	0.005	0.585	0.015	0	0.097	0.236
CONTROL 66	0	0.203	1.000	0.005	0.585	0.015	0	0.097	0.236
CONTROL 356	0	0.398	0.500	0.042	0.526	0.015	0	0.097	0.234
RISK 1	0.148	0.303	0.500	0.018	0.060	0.269	1.000	0.097	0.233
CONTROL 208	0	0.390	0.500	0.050	0.523	0.015	0	0.097	0.233
CONTROL 603	0	0.407	0.500	0.026	0.525	0.015	0	0.097	0.233
CONTROL 604	0	0.407	0.500	0.026	0.525	0.015	0	0.097	0.233
CONTROL 16	0	0.186	1.000	0.005	0.587	0.015	0	0.097	0.232
CONTROL 17	0	0.186	1.000	0.005	0.587	0.015	0	0.097	0.232
CONTROL 262	0	0.177	1.000	0.005	0.605	0.015	0	0.097	0.232
CONTROL 597	0	0.186	1.000	0.005	0.587	0.015	0	0.097	0.232
CONTROL 88	0	0.394	0.500	0.031	0.523	0.015	0	0.097	0.230
CONTROL 89	0	0.394	0.500	0.031	0.523	0.015	0	0.097	0.230
CONTROL 90	0	0.394	0.500	0.031	0.523	0.015	0	0.097	0.230
CONTROL 91	0	0.394	0.500	0.031	0.523	0.015	0	0.097	0.230
CONTROL 92	0	0.394	0.500	0.031	0.523	0.015	0	0.097	0.230
CONTROL 93	0	0.394	0.500	0.031	0.523	0.015	0	0.097	0.230
CONTROL 94	0	0.394	0.500	0.031	0.523	0.015	0	0.097	0.230
CONTROL 95	0	0.394	0.500	0.031	0.523	0.015	0	0.097	0.230
CONTROL 96	0	0.394	0.500	0.031	0.523	0.015	0	0.097	0.230
CONTROL 97	0	0.394	0.500	0.031	0.523	0.015	0	0.097	0.230
CONTROL 98	0	0.394	0.500	0.031	0.523	0.015	0	0.097	0.230
CONTROL 99	0	0.394	0.500	0.031	0.523	0.015	0	0.097	0.230
CONTROL 100	0	0.394	0.500	0.031	0.523	0.015	0	0.097	0.230
CONTROL 101	0	0.394	0.500	0.031	0.523	0.015	0	0.097	0.230
CONTROL 102	0	0.394	0.500	0.031	0.523	0.015	0	0.097	0.230
CONTROL 574	0	0.385	0.500	0.039	0.526	0.015	0	0.097	0.230
CONTROL 575	0	0.385	0.500	0.039	0.526	0.015	0	0.097	0.230

Node-Title	BC	CC	ECC	EiC	SHC	TDC	BS	Criticality of Node	Node Risk Value
CONTROL 576	0	0.385	0.500	0.039	0.526	0.015	0	0.097	0.230
CONTROL 577	0	0.385	0.500	0.039	0.526	0.015	0	0.097	0.230
CONTROL 578	0	0.385	0.500	0.039	0.526	0.015	0	0.097	0.230
CONTROL 579	0	0.385	0.500	0.039	0.526	0.015	0	0.097	0.230
CONTROL 601	0	0.394	0.500	0.031	0.523	0.015	0	0.097	0.230
CONTROL 105	0	0.203	1.000	0.005	0.538	0.015	0	0.097	0.229
CONTROL 106	0	0.203	1.000	0.005	0.538	0.015	0	0.097	0.229
CONTROL 107	0	0.203	1.000	0.005	0.538	0.015	0	0.097	0.229
CONTROL 108	0	0.203	1.000	0.005	0.538	0.015	0	0.097	0.229
CONTROL 109	0	0.203	1.000	0.005	0.538	0.015	0	0.097	0.229
CONTROL 110	0	0.203	1.000	0.005	0.538	0.015	0	0.097	0.229
CONTROL 111	0	0.203	1.000	0.005	0.538	0.015	0	0.097	0.229
CONTROL 112	0	0.203	1.000	0.005	0.538	0.015	0	0.097	0.229
CONTROL 113	0	0.203	1.000	0.005	0.538	0.015	0	0.097	0.229
CONTROL 114	0	0.203	1.000	0.005	0.538	0.015	0	0.097	0.229
CONTROL 115	0	0.203	1.000	0.005	0.538	0.015	0	0.097	0.229
CONTROL 116	0	0.203	1.000	0.005	0.538	0.015	0	0.097	0.229
CONTROL 117	0	0.203	1.000	0.005	0.538	0.015	0	0.097	0.229
CONTROL 118	0	0.203	1.000	0.005	0.538	0.015	0	0.097	0.229
CONTROL 18	0	0.152	1.000	0.002	0.621	0.015	0	0.097	0.228
CONTROL 556	0	0.368	0.500	0.026	0.550	0.015	0	0.097	0.227
CONTROL 606	0	0.368	0.500	0.026	0.550	0.015	0	0.097	0.227
CONTROL 124	0	0.359	0.500	0.031	0.552	0.015	0	0.097	0.226
CONTROL 186	0	0.355	0.500	0.023	0.561	0.015	0	0.097	0.225
CONTROL 187	0	0.355	0.500	0.023	0.561	0.015	0	0.097	0.225
CONTROL 188	0	0.355	0.500	0.023	0.561	0.015	0	0.097	0.225
CONTROL 189	0	0.355	0.500	0.023	0.561	0.015	0	0.097	0.225
CONTROL 171	0	0.338	0.500	0.018	0.580	0.015	0	0.097	0.222
CONTROL 172	0	0.338	0.500	0.018	0.580	0.015	0	0.097	0.222
CONTROL 178	0	0.532	0	0.069	0.512	0.015	0	0.097	0.222
CONTROL 179	0	0.532	0	0.069	0.512	0.015	0	0.097	0.222
CONTROL 180	0	0.532	0	0.069	0.512	0.015	0	0.097	0.222
CONTROL 554	0	0.338	0.500	0.018	0.580	0.015	0	0.097	0.222
CONTROL 58	0	0	1.000	0	0.831	0.015	0	0.097	0.221
CONTROL 175	0	0.307	0.500	0.018	0.614	0.015	0	0.097	0.219
CONTROL 176	0	0.307	0.500	0.018	0.614	0.015	0	0.097	0.219
CONTROL 181	0	0.342	0.500	0.021	0.550	0.015	0	0.097	0.219
CONTROL 182	0	0.307	0.500	0.018	0.614	0.015	0	0.097	0.219
CONTROL 183	0	0.307	0.500	0.018	0.614	0.015	0	0.097	0.219
CONTROL 184	0	0.312	0.500	0.018	0.593	0.015	0	0.097	0.218
CONTROL 167	0	0.307	0.500	0.018	0.569	0.015	0	0.097	0.213
CONTROL 168	0	0.307	0.500	0.018	0.569	0.015	0	0.097	0.213
CONTROL 169	0	0.307	0.500	0.018	0.569	0.015	0	0.097	0.213
CONTROL 170	0	0.307	0.500	0.018	0.569	0.015	0	0.097	0.213
CONTROL 1	0	0.104	1.000	0.001	0.587	0.015	0	0.097	0.211
CONTROL 46	0	0.108	1.000	0.002	0.579	0.015	0	0.097	0.211
CONTROL 173	0	0.485	0	0.053	0.531	0.015	0	0.097	0.210
CONTROL 174	0	0.485	0	0.053	0.531	0.015	0	0.097	0.210
CONTROL 185	0	0.489	0	0.061	0.512	0.015	0	0.097	0.210
CONTROL 2	0	0.108	1.000	0.001	0.540	0.015	0	0.097	0.205
CONTROL 119	0	0.299	0.500	0.010	0.530	0.015	0	0.097	0.204
CONTROL 602	0	0.303	0.500	0.010	0.529	0.015	0	0.097	0.204
CONTROL 75	0	0.277	0.500	0.010	0.555	0.015	0	0.097	0.202
CONTROL 76	0	0.277	0.500	0.010	0.555	0.015	0	0.097	0.202
CONTROL 77	0	0.277	0.500	0.010	0.555	0.015	0	0.097	0.202
CONTROL 78	0	0.277	0.500	0.010	0.555	0.015	0	0.097	0.202
CONTROL 79	0	0.277	0.500	0.010	0.555	0.015	0	0.097	0.202
CONTROL 59	0	0.061	1.000	0	0.589	0.015	0	0.097	0.201
CONTROL 60	0	0.061	1.000	0	0.589	0.015	0	0.097	0.201
CONTROL 61	0	0.061	1.000	0	0.589	0.015	0	0.097	0.201
CONTROL 62	0	0.061	1.000	0	0.589	0.015	0	0.097	0.201
CONTROL 67	0	0.234	0.500	0.007	0.619	0.015	0	0.097	0.200
CONTROL 120	0	0.229	0.500	0.005	0.599	0.015	0	0.097	0.196
CONTROL 121	0	0.229	0.500	0.005	0.599	0.015	0	0.097	0.196
CONTROL 122	0	0.229	0.500	0.005	0.599	0.015	0	0.097	0.196
CONTROL 3	0	0.234	0.500	0.007	0.532	0.015	0	0.097	0.187
CONTROL 4	0	0.234	0.500	0.007	0.532	0.015	0	0.097	0.187
CONTROL 5	0	0.234	0.500	0.007	0.532	0.015	0	0.097	0.187
CONTROL 6	0	0.234	0.500	0.007	0.532	0.015	0	0.097	0.187
CONTROL 7	0	0.234	0.500	0.007	0.532	0.015	0	0.097	0.187
CONTROL 8	0	0.234	0.500	0.007	0.532	0.015	0	0.097	0.187
CONTROL 9	0	0.234	0.500	0.007	0.532	0.015	0	0.097	0.187

Node-Title	BC	CC	ECC	EiC	SHC	TDC	BS	Criticality of Node	Node Risk Value
CONTROL 10	0	0.234	0.500	0.007	0.532	0.015	0	0.097	0.187
CONTROL 11	0	0.234	0.500	0.007	0.532	0.015	0	0.097	0.187
CONTROL 12	0	0.234	0.500	0.007	0.532	0.015	0	0.097	0.187
CONTROL 13	0	0.234	0.500	0.007	0.532	0.015	0	0.097	0.187
CONTROL 14	0	0.234	0.500	0.007	0.532	0.015	0	0.097	0.187
CONTROL 15	0	0.234	0.500	0.007	0.532	0.015	0	0.097	0.187
CONTROL 47	0	0	1.000	0	0.555	0.015	0	0.097	0.181
CONTROL 48	0	0	1.000	0	0.555	0.015	0	0.097	0.181
CONTROL 49	0	0	1.000	0	0.555	0.015	0	0.097	0.181
CONTROL 50	0	0	1.000	0	0.555	0.015	0	0.097	0.181
CONTROL 51	0	0	1.000	0	0.555	0.015	0	0.097	0.181
CONTROL 52	0	0	1.000	0	0.555	0.015	0	0.097	0.181
CONTROL 53	0	0	1.000	0	0.555	0.015	0	0.097	0.181
CONTROL 54	0	0	1.000	0	0.555	0.015	0	0.097	0.181
CONTROL 55	0	0	1.000	0	0.555	0.015	0	0.097	0.181
CONTROL 56	0	0	1.000	0	0.555	0.015	0	0.097	0.181
CONTROL 57	0	0	1.000	0	0.555	0.015	0	0.097	0.181

REFERENCES

Aagedal, J.O., Den Braber, F., Dimitrakos, T., Gran, B.A., Raptis, D. & Stolen, K. 2002. Model-based risk assessment to improve enterprise security. (*In* Bellahsene, Z., Patel, D. & Rolland, C., *eds*. Enterprise Distributed Object Computing Conference, 2002. EDOC'02. Proceedings. Sixth International. IEEE. p. 51-62).

Agrawal, V. 2017. A comparative study on information security risk analysis methods. *Journal of Computers*, 12(1):57-67.

Ahmad, A. 2012. Type of security threats and it's prevention. *International Journal of Computer Technology Applications*, 3(2):750-752.

Aksoy, S. & Haralick, R.M. 2001. Feature normalization and likelihood-based similarity measures for image retrieval. *Pattern Recognition Letters*, 22(5):563-582.

Al-Khateeb, S., Hussain, M.N. & Agarwal, N. 2019. Leveraging social network analysis and cyber forensics approaches to study cyber propaganda campaigns. Social networks and surveillance for society. Cham: Springer. p. 19-42.

Albrechtsen, E. 2007. A qualitative study of users' view on information security. *Computers & Security*, 26(4):276-289.

AlHogail, A. 2015. Design and validation of information security culture framework. *Computers in Human Behavior*, 49:567-575.

Aloul, F.A. 2012. The need for effective information security awareness. *Journal of Advances in Information Technology*, 3(3):176-183.

Angers, M.E. 2013. 'How do we know?: The question of science and psychoanalysis in the aftermath of "post-positivism". *Issues in Psychoanalytic Psychology*, 35(1):124-142.

Arachchilage, N.A.G. & Love, S. 2014. Security awareness of computer users: A phishing threat avoidance perspective. *Computers in Human Behavior*, 38:304-312.

Armstrong, H.L., Armstrong, C. & McCulloh, I. 2010. A course applying network analysis to organizational risk in information security. (*In* Clarke, N.L., Furnell, S.M. & von Solms, R., *eds*. South African Information Security Multi-Conference. Port Elizabeth: Emerald. p. 204-214).

Armstrong, H.L. & McCulloh, I. 2010. Organizational risk using network analysis. (*In* Clarke, N.L., Furnell, S.M. & von Solms, R., *eds*. South African Information Security Multi-Conference. Port Elizabeth: Emerald. p. 132-141).

Au, C.H., Fung, W.S. & Tses, A. 2016. An investigation on the relationship between control self-assessment, cloud security, and cloud-related business performance-using partial least squares. (*In* Industrial Engineering and Engineering Management (IEEM). IEEE. p. 1879-1883).

Bach, B., Riche, N.H., Hurter, C., Marriott, K. & Dwyer, T. 2017. Towards unambiguous edge bundling: Investigating confluent drawings for network visualization. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):541-550.

Bäck, T., Kok, J.N. & Rozenberg, G. 2012. Handbook of natural computing. Heidelberg: Springer.

Beckers, K., Schmidt, H., Kuster, J.-C. & Faßbender, S. 2011. Pattern-based support for context establishment and asset identification of the ISO 27000 in the field of cloud computing. (*In* Sixth International Conference on Availability, Reliability and Security (ARES). IEEE. p. 327-333).

Bentley, L.D. & Whitten, J.L. 2007. Systems analysis and design for the global enterprise. New York: McGraw-Hill/Irwin.

Bitton, R., Finkelshtein, A., Sidi, L., Puzis, R., Rokach, L. & Shabtai, A. 2018. Taxonomy of mobile users' security awareness. *Computers & Security*, 73:266-293.

Bo, Y., Zhihui, L. & Meloche, J.A. 2010. Visualization of the Chinese academic web based on social network analysis. *Journal of Information Science*, 36(2):131-143.

Boer, H. & Seydel, E.R. 1996. Protection motivation theory. Predicting health behaviour: Research and practice with social cognition models. Berkshire: Open University Press. p. 95-120.

Borgatti, S.P. 2005. Centrality and network flow. *Social Networks*. (27):55-71.

Borgatti, S.P., Mehra, A., Brass, D.J. & Labianca, G. 2009. Network analysis in the social sciences. *Science*, 323(5916):892-895.

Boulet, R., Jouve, B., Rossi, F. & Villa, N. 2008. Batch kernel SOM and related Laplacian methods for social network analysis. *Neurocomputing*, 71(7):1257-1273.

Braber, F., Dimitrakos, T., Gran, B., S Lund, M., Stølen, K. & Aagedal, J. 2003. The CORAS methodology: Model-based risk management using UML and UP. UML and the unified process. Hershey: IRM Press. p. 332-357.

Broderick, J.S. 2001. Information security risk management — when should it be managed? *Information Security Technical Report*, 6(3):12-18.

Brodowsky, G., Stewart, K. & Anderson, B. 2018. Brand and country influences on purchase intentions: A theory-of-reasoned action approach. *Journal of Promotion Management*, 24(2):251-269.

Burt, R.S. 1992. Structural holes: The structure of competition. Cambridge: Harvard University Press.

Byrne, Z.S., Dvorak, K.J., Peters, J.M., Ray, I., Howe, A. & Sanchez, D. 2016. From the user's perspective: Perceptions of risk relative to benefit associated with using the internet. *Computers in Human Behavior*, 59:456-468.

CASOS. 2019. ORA-Lite. www.casos.cs.cmu.edu/projects/ora Date of access: 24/04/2019 2019.

Cheadle, C., Vawter, M.P., Freed, W.J. & Becker, K.G. 2003. Analysis of microarray data using z score transformation. *The Journal of Molecular Diagnostics*, 5(2):73-81.

Chen, C.C., Medlin, B.D. & Shaw, R.S. 2008. A cross-cultural investigation of situational information security awareness programs. *Information Management & Computer Security*, 16(4):360-376.

Chou, H. & Chou, C. 2016. An analysis of multiple factors relating to teachers' problematic information security behavior. *Computers in Human Behavior*, 65:334-345.

Christley, R.M., Pinchbeck, G., Bowers, R., Clancy, D., French, N., Bennett, R. & Turner, J. 2005. Infection in social networks: Using network analysis to identify high-risk individuals. *American Journal of Epidemiology*, 162(10):1024-1031.

Chung, M. 2019. Why employees matter in the fight against ransomware. *Computer Fraud & Security*, 2019(8):8-11.

Clemente, F.M., Martins, F.M.L. & Mendes, R.S. 2016. Social network analysis applied to team sports analysis. Cham: Springer.

Cormen, T.H., Leiserson, C.E., Rivest, R.L. & Stein, C. 2001. Introduction to algorithms second edition. Cambridge: The MIT Press.

Cox, J. 2012. Information systems user security: A structured model of the knowing–doing gap. *Computers in Human Behavior*, 28(5):1849-1858.

Creswell, J.W., Ebersohn, L., Eloff, I., Ferreira, R., Ivankova, N.V., Jansen, J.D., Nieuwenhuis, J., Pietersen, J., Plano Clark, V.L. & van der Westhuizen, C. 2012. First steps in research (revised edition). Pretoria: Van Schaik Publishers.

Crossler, R.E., Johnston, A.C., Lowry, P.B., Hu, Q., Warkentin, M. & Baskerville, R. 2013. Future directions for behavioral information security research. *Computers & Security*, 32:90-101.

Curtis, A., Taylor, N., Guadagno, B., Farmer, C. & Miller, P. 2018. Community awareness of patron banning in Australia: A brief report. *Journal of Police and Criminal Psychology*, 33(3):283-287.

Da Veiga, A. & Eloff, J.H.P. 2010. A framework and assessment instrument for information security culture. *Computers & Security*, 29(2):196-207.

Dang-Pham, D., Pittayachawan, S. & Bruno, V. 2017a. Applications of social network analysis in behavioural information security research: Concepts and empirical analysis. *Computers & Security*, 68:1-15.

Dang-Pham, D., Pittayachawan, S. & Bruno, V. 2017b. Applying network analysis to investigate interpersonal influence of information security behaviours in the workplace. *Information & Management*, 54(5):625-637.

Dang-Pham, D., Pittayachawan, S. & Bruno, V. 2017c. Investigation into the formation of information security influence: Network analysis of an emerging organisation. *Computers & Security*, 70:111-123.

Dang-Pham, D., Pittayachawan, S. & Bruno, V. 2017d. Why employees share information security advice? Exploring the contributing factors and structural patterns of security advice sharing in the workplace. *Computers in Human Behavior*, 67:196-206.

De la Hoz, E., De la Hoz, E., Ortiz, A., Ortega, J. & Prieto, B. 2015. PCA filtering and probabilistic SOM for network intrusion detection. *Neurocomputing*, 164(Supplement C):71-81.

Den Braber, F., Hogganvik, I., Lund, M., Stølen, K. & Vraalsen, F. 2007. Model-based security analysis in seven steps—a guided tour to the CORAS method. *BT Technology Journal*, 25(1):101-117.

Du, H., He, X., Du, W. & Feldman, M.W. 2017. Optimization of the critical diameter and average path length of social networks. *Complexity*, 2017.

Eloff, J.H.P. & Eloff, M.M. 2005. Information security architecture. *Computer Fraud & Security*, 2005(11):10-16.

Fausett, L.V. 1994. Fundamentals of neural networks: Architectures, algorithms, and applications. New Jersey: Prentice-Hall.

Feilzer, Y.M. 2010. Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research*, 4(1):6-16.

Fernandez, A. & Garcia, D.F. 2016. Complex vs. Simple asset modeling approaches for information security risk assessment: Evaluation with magerit methodology. (*In* Sixth International Conference on Innovative Computing Technology (INTECH). IEEE. p. 542-549).

Flick, U. 2015. Introducing research methodology: A beginner's guide to doing a research project. Berlin: Sage.

Frankenstein, W., Binxuan, H. & Kathleen, C. 2016. Nato trident juncture on twitter: Public discussion. www.casos.cs.cms.edu/tools/datasets/internal/index.php2019-04-16. www.casos.cs.cms.edu/tools/datasets/internal/index.php Date of access: 2019-04-16.

Fu, J., Sun, D., Chai, J., Xiao, J. & Wang, S. 2015. The "six-element" analysis method for the research on the characteristics of terrorist activities. *Annals of Operational Research*, 234:17-35.

Gardoni, P. & Murphy, C. 2014. A scale of risk. *Risk Analysis*, 34(7):1208-1227.

Gerring, J. 2007. Case study research. Boston: Cambridge University Press.

Girvan, M. & Newman, M.E. 2002. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12):7821-7826.

Glen, S. 2019. Z-score: Definition, formula and calculation. https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/z-score/ Date of access: 05 June 2019.

Gordon, L.A., Loeb, M.P., Lucyshyn, W. & Zhou, L. 2018. Empirical evidence on the determinants of cybersecurity investments in private sector firms. *Journal of Information Security*, 9(2):133-153.

Graham, J. 2005. Dynamic network analysis of the network-centric organization: Towards an understanding of cognition & performance. Pittsburgh: Carnegie Mellon University. (Thesis - PhD).

Gratian, M., Bandi, S., Cukier, M., Dykstra, J. & Ginther, A. 2018. Correlating human traits and cyber security behavior intentions. *Computers & Security*, 73:345-358.

Gross, J.L., Yellen, J. & Zhang, P. 2013. Handbook of graph theory. London: Chapman and Hall/CRC.

Gu, F. & Cheung, Y.-M. 2018. Self-organizing map-based weight design for decomposition-based many-objective evolutionary algorithm. *IEEE Transactions on Evolutionary Computation*, 22(2):211-225.

Gundu, T. & Flowerday, S. 2013. Ignorance to awareness: Towards an information security awareness process. *SAIEE Africa Research Journal*, 104(2):69-79.

Gupta, S. & Hossain, L. 2011. Towards near-real-time detection of insider trading behaviour through social networks. *Computer Fraud & Security*, 2011(1):7-16.

Halabi, T. & Bellaiche, M. 2017. Towards quantification and evaluation of security of cloud service providers. *Journal of Information Security and Applications*, 33:55-65.

Hancock, B. 2015. Survey of risk assessment practices. Raleigh: ERM Initiative.

Hanneman, R.A. & Riddle, M. 2005. Introduction to social network methods. California: University of California.

Hatala, J. 2016. Social network analysis in human resource development: A new methodology. *Human Resource Development Review*, 5(1):45-71.

He, D., Wang, H., Zhang, J. & Wang, L. 2017. Insecurity of an identity-based public auditing protocol for the outsourced data in cloud storage. *Information Sciences*, 375:48-53.

Henderson, K.A. 2011. Post-positivism and the pragmatics of leisure research. *Leisure Sciences*, 33(4):341-346.

Huang, S., Han, Z., Yang, B. & Ren, N. 2019. Factor identification and computation in the assessment of information security risks for digital libraries. *Journal of Librarianship and Information Science*, 51(1):78-94.

Hunt, R. & Hill, S. 2015. Using security logs to identify and manage user behaviour to enhance information security. (*In* 14th European Conference on Cyber Warfare and Security. Academic Conferences Limited. p. 111-119).

Ifinedo, P. 2012. Understanding information systems security policy compliance: An integration of the theory of planned behavior and the protection motivation theory. *Computers & Security*, 31(1):83-95.

Imbert, E., Ladu, L., Tani, A. & Morone, P. 2019. The transition towards a bio-based economy: A comparative study based on social network analysis. *Journal of Environmental Management*, 230:255-265.

Ishizaka, A. & Lusti, M. 2004. An expert module to improve the consistency of AHP matrices. *International Transactions in Operational Research*, 11(1):97-105.

ISO. 2009. ISO 15408-1: 2009.

ISO. 2016. ISO 27000:2016.

Jonnalagadda, A. & Kuppusamy, L. 2018. Overlapping community detection in social networks using coalitional games. *Knowledge and Information Systems*, 56(3):637-661.

Karabacak, B. & Sogukpinar, I. 2005. ISRAM: Information security risk analysis method. *Computers & Security*, 24(2):147-159.

Karlsson, F., Hedström, K. & Goldkuhl, G. 2017. Practice-based discourse analysis of information security policies. *Computers & Security*, 67:267-279.

Kearney, W.D. & Kruger, H.A. 2016. Theorising on risk homeostasis in the context of information security behaviour. *Information & Computer Security*, 24(5):496-513.

Kemper, G. 2019. Improving employees' cyber security awareness. *Computer Fraud & Security*, 2019(8):11-14.

Khajouei, H., Kazemi, M. & Moosavirad, S.H. 2017. Ranking information security controls by using fuzzy analytic hierarchy process. *Information Systems and e-Business Management*, 15(1):1-19.

Khan, B., Alghathbar, K.S., Nabi, S.I. & Khan, M.K. 2011. Effectiveness of information security awareness methods based on psychological theories. *African Journal of Business Management*, 5(26):10862.

Kim, J. & Hastak, M. 2018. Social network analysis: Characteristics of online social networks after a disaster. *International Journal of Information Management*, 38(1):86-96.

Kohonen, T. 1998. The self-organizing map. *Neurocomputing*, 21(1-3):1-6.

Kokolakis, S. 2017. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, 64:122-134.

Kong, H.K., Hong, M.K. & Kim, T.-S. 2017. Security risk assessment framework for smart car using the attack tree analysis. *Journal of Ambient Intelligence and Humanized Computing*.1-21.

Kroeze, J.H. 2012. Interpretivism in IS – a postmodernist (or postpositivist?) knowledge theory. Proceedings of the 18th Americas conference on information systems. Seattle.

Kruger, H.A. & Kearney, W.D. 2006. A prototype for assessing information security awareness. *computers & security*, 25(4):289-296.

Kuo, R.J., Rizki, M., Zulvia, F.E. & Khasanah, A.U. 2018. Integration of growing self-organizing map and bee colony optimization algorithm for part clustering. *Computers & Industrial Engineering*, 120:251-265.

Laumann, E.O., Marsden, P.V. & Prensky, D. 1989. The boundary specification problem in network analysis. Research methods in social network analysis. New Jersey: Transaction Publishers. p. 61-88.

Lebek, B., Uffen, J., Breitner, M.H., Neumann, M. & Hohler, B. 2013. Employees' information security awareness and behavior: A literature review. (*In* System Sciences (HICSS), 2013 46th Hawaii International Conference on. IEEE. p. 2978-2987).

Lee, Y. 2019. Using self-organizing map and clustering to investigate problem-solving patterns in the massive open online course: An exploratory study. *Journal of Educational Computing Research*, 57(2):471-490.

Loosemore, M. 1998. Social network analysis: Using a quantitative tool within an interpretative context to explore the management of construction crises. *Engineering, Construction and Architectural Management*, 5(4):315-326.

López, A.U., Mateo, F., Navío-Marco, J., Martínez-Martínez, J.M., Gómez-Sanchís, J., Vila-Francés, J. & Serrano-López, A.J. 2019. Analysis of computer user behavior, security incidents and fraud using self-organizing maps. *Computers & Security*, 83:38-51.

Lund, M.S., Solhaug, B. & Stølen, K. 2011. A guided tour of the CORAS method. Model-driven risk analysis. Heidelberg: Springer. p. 23-43.

Ma, Q., Schmidt, M.B. & Pearson, J.M. 2009. An integrated framework for information security management. *Review of Business*, 30(1):58-69.

Mamonov, S. & Benbunan-Fich, R. 2018. The impact of information security threat awareness on privacy-protective behaviors. *Computers in Human Behavior*, 83:32-44.

Martin, V., Zhou, X., Marshall, E., Jia, B., Fusheng, G., FrancoDixon, M.A., DeHaan, N., Pfeiffer, D.U., Magalhães, R.J.S. & Gilbert, M. 2011. Risk-based surveillance for avian

influenza control along poultry market chains in South China: The value of social network analysis. *Preventive Veterinary Medicine*, 102(3):196-205.

Martins, A. & Eloff, J.H.P. 2002. Information security culture. Security in the information society. Boston: Springer. p. 203-214.

McCulloh, I. 2009. Detecting changes in a dynamic social network. Pittsburg: Carnegie Mellon University. (Thesis - PhD).

Metalidou, E., Marinagi, C., Trivellas, P., Eberhagen, N., Skourlas, C. & Giannakopoulos, G. 2014. The human factor of information security: Unintentional damage perspective. *Procedia - Social and Behavioral Sciences*, 147:424-428.

Meyerson, A. & Tagiku, B. 2009. Minimizing average shortest path distances via shortcut edge addition. Approximation, randomization, and combinatorial optimization. Algorithms and techniques. Berkely: Springer. p. 272-285.

Mitchell, R.K., Agle, B.R. & Wood, D.J. 1997. Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. *Academy of Management Review*, 22(4):853-886.

Moody, G.D., Siponen, M. & Pahnila, S. 2018. Toward a unified model of information security policy compliance. *MIS Quarterly*, 42(1):285-311.

Nakayama, H., Sakamoto, T., Okegawa, Y., Kaminoyama, K., Fujie, M., Ichihashi, Y., Kurata, T., Motohashi, K., Al-Shehbaz, I., Sinha, N. & Kimura, S. 2018. Comparative transcriptomics with self-organizing map reveals cryptic photosynthetic differences between two accessions of North American Lake cress. *Scientific Reports*, 8(1):3302.

Ng, B., Kankanhalli, A. & Xu, Y. 2009. Studying users' computer security behavior: A health belief perspective. *Decision Support Systems*, 46(4):815-825.

Oates, B.J. 2005. Researching information systems and computing. London: Sage.

Olawumi, O., Väänänen, A., Haataja, K. & Toivanen, P. 2017. Security issues in smart home and mobile health system: Threat analysis, possible countermeasures and lessons learned. *International Journal of Information Technologies and Security*, 9(1):31-52.

Oldenburg, B., Van Duijn, M. & Veenstra, R. 2018. Defending one's friends, not one's enemies: A social network analysis of children's defending, friendship, and dislike relationships using xpnet. *PloS ONE*, 13(5):1-14.

Olivier, M.S. 2002. Database privacy. *ACM SIGKDD Explorations Newsletter*, 4(2):20-27.

Olsen, C. & St George, D. 2004. Cross-sectional study design and data analysis. *College Entrance Examination Board*, 26(03):2006.

Ongkowijoyo, C.S. & Doloi, H. 2018. Understanding of impact and propagation of risk based on social network analysis. *Procedia Engineering*, 212:1123-1130.

Pal, C., Hirayama, S., Narahari, S., Jeyabharath, M., Prakash, G. & Kulothungan, V. 2018. An insight of world health organization (WHO) accident database by cluster analysis with self-organizing map (SOM). *Traffic Injury Prevention*, 19(sup1):S15-S20.

Palmer, I.C. & Potter, G.A. 1989. Computer security risk management. London: Jessica Kingsley Publishers.

Parsons, K., McCormac, A., Butavicius, M. & Ferguson, L. Command, Control, Communications and Intelligence Division of the Australian Government Department of Defence. 2010. Human factors and information security: Individual, culture and security environment. Edinburgh: Defence Science and Technology Organisation. (DSTO-TR-2484).

Parsons, K., McCormac, A., Butavicius, M., Pattinson, M. & Jerram, C. 2014. Determining employee awareness using the human aspects of information security questionnaire (HAIS-Q). *Computers & Security*, 42:165-176.

Pfleeger, C.P. & Pfleeger, S.L. 2006. Security in computing (4th edition). Boston: Pearson.

Philips, E., Nurse, J., Goldsmith, M. & Creese, S. 2015. Applying social network analysis to security. Working paper of the Sustainable Society Network. p. 11-27.

Posthumus, S. & von Solms, R. 2004. A framework for the governance of information security. *Computers & Security*, 23(8):638-646.

Prendergast, C. 1979. The problem of the unity of knowledge in Comte's philosophy of science. *Sociological Inquiry*, 49(4):25-35.

Prentice-Dunn, S., McMath, B.F. & Cramer, R.J. 2009. Protection motivation theory and stages of change in sun protective behavior. *Journal of Health Psychology*, 14(2):297-305.

Rajbhandari, L. & Snekkenes, E. 2012. Intended actions: Risk is conflicting incentives. (*In* International Conference on Information Security. Heidelberg: Springer. p. 370-386).

Rajbhandari, L. & Snekkenes, E. 2013. Using the conflicting incentives risk analysis method. (*In* IFIP International Information Security Conference. Heidelberg: Springer. p. 315-329).

Render, B., Stair Jr, R.M. & Hanna, M.E. 2012. Quantitative analysis for management. 11th Edition. New Jersey: Pearson.

Rezazad, H. 2011. Computer network optimization. *Wiley Interdisciplinary Reviews: Computational Statistics*, 3(1):34-46.

Roca, E., Julià-Verdaguer, A., Villares, M. & Rosas-Casals, M. 2018. Applying network analysis to assess coastal risk planning. *Ocean & Coastal Management*, 162:127-136.

Rosenquist, M. 2016. Navigating the digital age: The definitive cybersecurity guide for directors and officers - Australia. Jersey City: Forbes Media.

Rubinov, M. & Sporns, O. 2010. Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage*, 52(3):1059-1069.

Russell, S.J. & Norvig, P. 2016. Artificial intelligence: A modern approach. New Jersey: Pearson Education Limited.

Saunders, M., Lewis, P. & Thornhill, A. 2019. Research methods for business students. 8th edition. Harlow: Pearson Education Limited.

Schein, E.H. 2009. The corporate culture survival guide. San Francisco: John Wiley & Sons.

Schlienger, T. & Teufel, S. 2003. Information security culture-from analysis to change. *South African Computer Journal*, 2003(31):46-52.

Schneider, B., Brief, A.P. & Guzzo, R.A. 1996. Creating a climate and culture for sustainable organizational change. *Organizational dynamics*, 24(4):7-19.

Scott, J. 2016. Social network analysis. *Sociology*, 22(1):109-127.

Scott, J. & Carrington, P.J. 2011. The SAGE handbook of social network analysis. London: SAGE Publications.

Serfontein, R. 2016. A web-based decision support system for the allocation of audit resources. Potchefstroom: NWU. (Dissertation - MSc).

Serfontein, R., Drevin, L. & Kruger, H.A. 2018. The feasibility of raising information security awareness in an academic environment using SNA. (*In* IFIP World Conference on Information Security Education. Cham: Springer. p. 69-80).

Serfontein, R., Kruger, H.A. & Drevin, L. 2019. Identifying information security risks in a social network using self-organising maps. (*In* IFIP World Conference on Information Security Education. Cham: Springer. p. 114-126).

Shedden, P., Ahmad, A., Smith, W., Tscherning, H. & Scheepers, R. 2016. Asset identification in information security risk assessment: A business practice approach. *Communications of the Association for Information Systems*, 39(1):15.

Shillair, R., Cotten, S.R., Tsai, H.S., Alhabash, S., LaRose, R. & Rifon, N.J. 2015. Online safety begins with you and me: Convincing internet users to protect themselves. *Computers in Human Behavior*, 48:199-207.

Shukla, N. & Kumar, S. 2012. A comparative study on information security risk analysis practices. *IJCA Special Issue on Issues and Challenges in Networking, Intelligence and Computing Technologies*. (3):28-33.

Siponen, M. & Willison, R. 2009. Information security management standards: Problems and solutions. *Information & Management*, 46(5):267-270.

Siponen, M.T. 2000. A conceptual foundation for organizational information security awareness. *Information Management & Computer Security*, 8(1):31-41.

Smith, G., Sabbag, L. & Rohmer, A. 2018. A comparative analysis of the roles governors play in disaster recovery. *Risk, Hazards & Crisis in Public Policy*, 9(2):205-243.

Sohrabi Safa, N., von Solms, R. & Furnell, S. 2016. Information security policy compliance model in organizations. *Computers & Security*, 56:70-82.

Soomro, Z.A., Shah, M.H. & Ahmed, J. 2016. Information security management needs more holistic approach: A literature review. *International Journal of Information Management*, 36(2):215-225.

Straub Jr, D.W. 1990. Effective IS security: An empirical study. *Information Systems Research*, 1(3):255-276.

Suh, B. & Han, I. 2003. The IS risk analysis based on a business model. *Information & Management*, 41(2):149-158.

Supriadi, L.S.R. & Pheng, L.S. 2018. Business continuity management (bcm). Business continuity management in construction. Singapore: Springer. p. 41-73.

Swanepoel, J., Swanepoel, C., Van Graan, F., Allison, J., HM, W. & Santana, L. 2008. Elementêre statistiese metodes 2de uitgawe. Noordwes-Universiteit, Potchefstroom kampus, Vakgroep Statistiek.

Tadisina, S.K., Troutt, M.D. & Bhasin, V. 1991. Selecting a doctoral programme using the analytic hierarchy process—the importance of perspective. *Journal of the Operational Research Society*, 42(8):631-638.

Tankard, C. 2017. Encryption as the cornerstone of big data security. *Network Security*, 2017(3):5-7.

Taylor, B.W. 2013. Introduction to management science 11th edition. London: Pearson Education Limited.

Tchernykh, A., Schwiegelsohn, U., Talbi, E. & Babenko, M. 2016? Towards understanding uncertainty in cloud computing with risks of confidentiality, integrity, and availability. *Journal of Computational Science* (In press).

Tesson, S., Richards, I., Porter, D., Phillips, K.A., Rankin, N., Musiello, T., Marven, M. & Butow, P. 2016. Women's preferences for contralateral prophylactic mastectomy: An investigation using protection motivation theory. *Patient Education and Counseling*, 99(5):814-822.

Theoharidou, M. & Gritazalis, D. 2007. Common body of knowledge for information security. *IEEE Security & Privacy*, 5(2):64-47.

Thomson, K.L., von Solms, R. & Louw, L. 2006. Cultivating an organizational information security culture. *Computer Fraud & Security*, 2006(10):7-11.

Thomson, M.E. & von Solms, R. 1998. Information security awareness: Educating your users effectively. *Information Management & Computer Security*, 6(4):167-173.

Tondel, I.A., Jaatun, M.G. & Meland, P.H. 2008. Security requirements for the rest of us: A survey. *IEEE software*, 25(1):20-27.

Tsohou, A., Karyda, M. & Kokolakis, S. 2015. Analysing the role of cognitive and cultural biases in the internalization of information security policies: Recommendations for information security awareness programs. *Computers & Security*, 52:128-141.

Tsui, E. & Liebowitz, J. 2005. Linking social network analysis with the analytic hierarchy process for knowledge mapping in organizations. *Journal of Knowledge Management*, 9(1):76-86.

Tu, X., Jiang, G.-P., Song, Y. & Zhang, X. 2018. Novel multiplex pagerank in multilayer networks. *IEEE Access*, 6:12530-12538.

Van Niekerk, J.F. & von Solms, R. 2010. Information security culture: A management perspective. *Computers & Security*, 29(4):476-486.

von Solms, R. 1999. Information security management: Why standards are important. *Information Management & Computer Security*, 7(1):50-58.

Wangen, G. 2015. Conflicting incentives risk analysis: A case study of the normative peer review process. *Administrative Sciences*, 5(3):125-147.

Wangen, G. 2017. Information security risk assessment: A method comparison. *Computer*, 50(4):52-61.

Wasserman, S. 1994. Advances in social network analysis: Research in the social and behavioral sciences. London: Sage.

Wasserman, S. & Faust, K. 1994. Social network analysis: Methods and applications. Vol. 8. Cambridge: Cambridge University Press.

Watson, R.N., Woodruff, J., Roe, M., Moore, S.W. & Neumann, P.G. University of Cambridge, Computer Laboratory. 2018. Capability hardware enhanced risc instructions (cheri): Notes on the meltdown and spectre attacks. (UCAM-CL-TR-916).

Weishäupl, E., Yasasin, E. & Schryen, G. 2018. Information security investments: An exploratory multiple case study on decision-making, evaluation and learning. *Computers & Security*, 77:807-823.

Whitman, M.E. & Mattord, H.J. 2011. Principles of information security. Boston: Cengage Learning.

Wiant, T.L. 2005. Information security policy's impact on reporting security incidents. *Computers & Security*, 24(6):448-459.

Widjaja, A.E., Chen, J.V., Sukoco, B.M. & Ha, Q.A. 2019. Understanding users' willingness to put their personal information on the personal cloud-based storage applications: An empirical study. *Computers in Human Behavior*, 91:167-185.

Willison, R., Warkentin, M. & Johnston, A.C. 2018. Examining employee computer abuse intentions: Insights from justice, deterrence and neutralization perspectives. *Information Systems Journal*, 28(2):266-293.

Wu, E., Torous, J., Hardaway, R. & Gutheil, T. 2017. Confidentiality and privacy for smartphone applications in child and adolescent psychiatry: Unmet needs and practical solutions. *Child and Adolescent Psychiatric Clinics of North America*, 26(1):117-124.

Wu, I.C. & Wu, C. 2011. Using internal link and social network analysis to support searches in wikipedia: A model and its evaluation. *Journal of Information Science*, 37(2):189-207.

Ying, T. & Xiao, H. 2011. Knowledge linkage. *Journal of Hospitality & Tourism Research*, 36(4):450-477.

Yu, Y., Xue, L., Au, M.H., Susilo, W., Ni, J., Zhang, Y., Vasilakos, A.V. & Shen, J. 2016. Cloud data integrity checking with an identity-based auditing mechanism from RSA. *Future Generation Computer Systems*, 62:85-91.

Yue, W.T., Çakanyıldırım, M., Ryu, Y.U. & Liu, D. 2007. Network externalities, layered protection and IT security risk management. *Decision Support Systems*, 44(1):1-16.

Žalik, K.R. 2019. Evolution algorithm for community detection in social networks using node centrality. Intelligent methods and big data in industrial applications. Cham: Springer. p. 73-87.

Zhou, L., Lü, K. & Liu, W. 2016. An approach for community detection in social networks based on cooperative games theory. *Expert Systems*, 33(2):176-188.