NETWORK TRUSTWORTHINESS EVALUATION IN P2P NETWORKS

Ming Xiang

A THESIS SUBMITTED TO AUCKLAND UNIVERSITY OF TECHNOLOGY
IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY (PHD)

Januray 2018

School of Engineering, Computer and Mathematical Sciences

Abstract

Trust and reputation management emerges as a significant research trend, in term of soft security to tackle the security issues in computer networks. It is different from the traditional security mechanisms such as cryptography that is described as hard security. The basic idea is that every entity in the network, as an individual, can rate each other based on previous experiences. This rating on trust can assist other machines in deciding whether to collaborate with that machine in the future. Recently there has been rapid increase in literature on trust and reputation management that mainly focuses on algorithmically modelling and evaluating the trust to effectively detect and avoid various malicious attacks. These trust algorithms can isolate the malicious entities from the local trust aspect. While the concept of trust in the computer network is derived from the sociology, and in sociology, it is defined as belief that trustees will have a positive expectation of intention and behaviours. Moreover, the trustee at different positions will behave differently, such as at the Structural Hole or the position surrounded with Simmelian Ties. Do these position-based phenomena also exist in computer networks? In other words, in computer networks, is the location of a node will affect its behaviours, especially in the emerging peer-to-peer (P2P) network architecture?

Motivated by above research questions, in this thesis, we have focused on studying how the underlying network topological connectivity can affect the overlay trust behaviours from the global network perspective. This thesis has four main contributions. Firstly, we have revealed the underlying topology impact on the overlay trust behaviours *in P2P networks. We have confirmed the correlations between the topological structures of Simmelian Ties and Structural Hole, and the node trustworthiness behaviours. Secondly, we have defined a new term of network trustworthiness to describe the trust level on a network topology. This is followed by introducing the Network Trustworthiness as a Service (NTaaS) concept, which can be adopted to accommodate the different levels of trust service demands from the users. Thirdly, we have proposed the T value and Trustworthiness Tolerance Margin (MTT) based evaluation framework to evaluate the trustworthiness of the network topologies from the global aspect. Lastly, we have proposed a mathematical approach to optimise the network topology by adding a link in the most critical position so that the underlying network structures can best resist various unwanted behaviours and network failures.*

Contents

List of Tables

List of Figures

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the qualification of any other degree or diploma of a university or other institution of higher learning.

Signature of student

Publications

Journal Papers:

Xiang, M., Liu, W., Bai, Q., & Al-Anbuky, A. (2016). The critical role of structural hole in forming trust for Securing Wireless Sensor Networks. International Journal of Information, Communication Technology and Applications, 2(1), 66–84.

Xiang, M., Liu, W., & Bai, Q. (2016). A fuzzy logic-based sustainable and trusted routing for P2P enabled smart grid. International Journal of Computational Science and Engineering, 13(2), 165–174. https://doi.org/10.1504/IJCSE.2016.078445

Xiang, M., Liu, W., Bai, Q., & Al-Anbuky, A. (2015). Avoiding the Opportunist: The Role of Simmelian Ties in Fostering the Trust in Sensor-Cloud Networks. International Journal of Distributed Sensor Networks, 11(10), 873941.

Xiang, M., Bai, Q., & Liu, W. (2014). Trust-based Adaptive Routing for Smart Grid Systems. Journal of Information Processing, 22(2), 210–218.

Book Chapter:

Xiang, M., Liu, W., Bai, Q., & Al-Anbuky, A. (2016). Dynamic Trust Elective Geo Routing to Secure Smart Grid Communication Networks. In Smart Grid as a Solution for Renewable and Efficient Energy (pp. 323–343).

Conference Papers:

Xiang, M., Liu, W., Bai, Q., Al-Anbuky, A., Wu, J., & Sathiaseelan, A. (2017). NTaaS: Network trustworthiness as a service. In 2017 27th International Telecommunication Networks and Applications Conference (ITNAC) (pp. 1–6). https://doi.org/10.1109/ATNAC.2017.8215437

Mamta, N., **Ming, X.**, William, L., Jairo, G., Luca, C., Arjuna, S., & Arvind, M. (2017). UAV-assisted Edge Infrastructure for Challenged Networks. Presented at the IEEE Infocom Workshop on Wireless Communications and Networking in Extreme Environments (WCNEE 2017), Atlanta, USA.

Xiang, M., Liu, W., Bai, Q., & Al-Anbuky, A. (2015). The double-edged sword: Revealing the critical role of structural hole in forming trust for securing Wireless sensor networks. In Telecommunication Networks and Applications Conference (ITNAC), 2015 International (pp. 286–291).

Xiang, M., Liu, W., Bai, Q., & Al-Anbuky, A. (2015). Simmelian Ties and Structural Holes: Exploring Their Topological Roles in Forming Trust for Securing Wireless Sensor Networks. In 2015 IEEE Trustcom/BigDataSE/ISPA (Vol. 1, pp. 96–103).

Xiang, M., Tauch, S., & Liu, W. (2014). Dependability and Resource Optimation Analysis for Smart Grid Communication Networks. In 2014 IEEE Fourth International Conference on Big Data and Cloud Computing (pp. 676–681).

Acknowledgements

Firstly, I would like to express my sincere gratitude to my primary supervisor, philosopher, and guide, Dr. William Liu, for the continuous support of my PhD study and related research, for his patience, motivation, and immense knowledge. This guidance helped me throughout the duration of the research and writing of this thesis. I could not have imagined having a better supervisor for my PhD study.

Besides my primary supervisor, I would like to thank the rest of my supervisory team: Dr Quan Bai and Prof. Adnan Al-Anbuky for their insightful comments and encouragement, and also for their challenging questioning which encouraged me to widen my research to look at various perspectives.

My sincere thanks also go to the School of Engineering, Computer, and Information Sciences, which supported me with the full scholarships to fund my three years of PhD research. Without it, I cannot full dedicate myself into the PhD research journey.

I am also grateful to my PhD colleagues for all the support they have given.

Last but not the least, I would like to thank my family: my parents and especially my wife for supporting me spiritually throughout the writing of this thesis and my life in general.

Chapter 1

Introduction

1.1 Background

The characteristics of Peer-to-Peer (P2P) networks are decentralised and nodes self-organised. These remove the threat of single point-failure, which is a typical issue in the traditional centralised network. It also makes them easier to scale. However, the decentralised and self-organized characteristics can cause many new security threats, which the centralised networks do not have (Washbourne, 2015), such as Leechers (selfish behaviours), Distributed Denial of Services (DDoS) attacks, node authentication, and malware (malicious behaviours). Leechers are the users in Bit-torrent transmission who only download resources but rarely share their resources. In Wireless Sensor Networks (WSN) scenarios they are the selfish nodes who deny doing the tasks, such as packet forwarding. In such cases, they can preserve their energy. There is no central node in P2P network as all the nodes are self-organised, this causes authentication between each other to become difficult. In addition, behaviours of nodes can be hard to predict. Taking the WSN as an example, Sharma and Ghose have listed the known security threats in their study (Sharma & Ghose, 2010). These security threats include Denial of Services (DoS) (e.g. black hole attack, grey hole attack, signal jamming, etc.), wormhole attack; Sybil attack, selective forwarding attack, sink-hole attack, node capturing, false or malicious node, passive information gathering, and the hello flood attack. Besides, this study also suggested some security mechanisms to defend against these attacks, such as authorization and redundancy to defend against the DoS attacks, identity certificates to defend against the Sybil attack, encryption to defend against node capture.

The traditional security mechanisms are considered as hard security to defend the system or network from known security threats (Rasmusson & Jansson, 1996). If the malicious parties find a way to bypass them (backdoor, bugs) the system or network will be vulnerable. Thus, the soft security approach, which is Trust and Reputation Management (TRM), is emerging to tackle these soft security threats. Nevertheless, the existing studies on trust modelling are only focusing on the node level so that little existing work has focused on the trustworthiness of the networks.

1.2 Research Questions

The objective of this thesis is to study the co-evaluation changes in network entities' status (behaviours) and network topologies in the P2P environment. It focuses on how the changes in network topologies can affect the overlay network entities' (nodes') behaviours. When the nodes in the network change their behaviours, they can become better or malicious. If the nodes become malicious, the trust-based routing algorithms are proposed to detect these kinds of activities and avoid them. In this thesis, 'avoid' means the isolation of the malicious nodes from the rest of the legitimate nodes. In such a case, the change of node behaviours changes the network topology as well. Recently many existing studies have already focused on these. Unfortunately, there is little existing work on the reverse situation, i.e., how different network structures or changes of network topologies can affect the overlay node behaviours.

Sociology studies have long addressed the issues on how social network structures can affect individual behaviours and performance in the network. They can either sustain or hinder a wide range of performance-related outcome, both at individual and collective levels (Latora, Nicosia & Panzarasa, 2013). Can the findings on these studies be applied to the computer network? In regard to this three research questions are raised in this thesis, as below.

Q1 What is the impact of the network topological structure on the overlay node trustworthiness and the performance of trust-based routing algorithm?

Research Question 1: The change of the nodes' behaviours in the network can result in an isolation by trust algorithms. However, how can the changes in network topologies affect the overlay nodes' behaviours? There has not been much work in this area. In other words, the current studies on trust routing focus on the local aspect (nodes) rather than the global aspect (network). There are two examples to show why network topologies are important for the trust-based routing algorithms. When the trust-based routing algorithm detects any malicious activity in the network, it isolates the malicious node to avoid the attacks. On the other hand, if there is no alternative route to avoid, isolation of this malicious node can result in a disconnection of this network. Moreover, most of the trust-based routing algorithms have a reputation system to support the trust evaluation. However, if there is no neighbour, the reputation system will not work as there is no one to consult. In sociology, there are two typical social structures which have been well-studied and debated as to their impact on individual's performances and behaviours in the network; they are Simmelian Ties (Krackhardt, 1999) and Structural Hole (Burt, 2009). In another study (Engle, 1999), Engle has summarised the relationship between these two structures in both individual and collective levels in sociology. Can Engle's (1999) summary framework apply to the computer network as well?

Q2 What topological metrics can be used to evaluate the network trustworthiness?

Research Question 2: The change of nodes' behaviours in a network happen often and this change can be good or bad. When the node becomes malicious or uncooperative, this node will be isolated in the network by the trust-based routing algorithm. This is how trust can affect the change of network topologies, and there have already been a large number of studies on it. On the other hand, there is little research on how to evaluate the trustworthiness of the network as a whole. As mentioned before, the trust-based routing algorithms in different network topologies can have different performances. If the factors have been found, trust-based routing algorithms can improve the trust establishment and convergence in the network, and the factors, which can hinder them from the first research question can be used to evaluate the network trustworthiness. In other words, what are the metrics that can be used to evaluate the network topology as a whole? In such case, we can see how trustworthy this network is in achieving its objectives, such as network availability, etc.

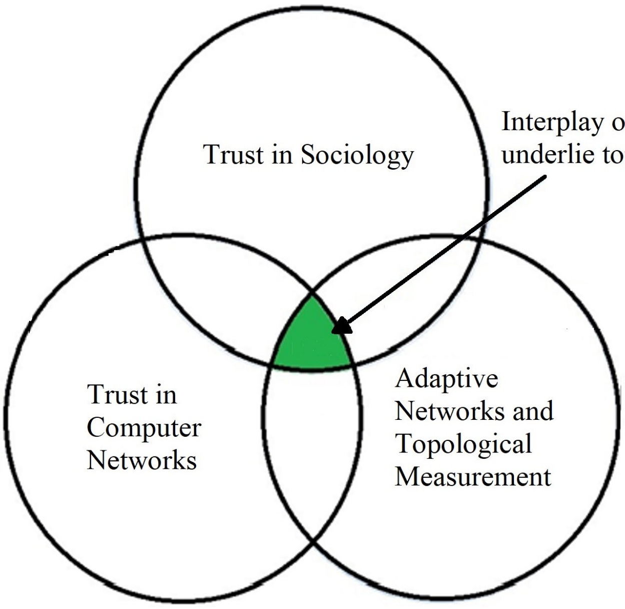
Q3 How to optimise the network topology so that the local trust can be influenced in a positive way and sustain trust-based routing algorithm performance outcome?

Research Question 3: Once the network topologies have been evaluated, if the network topologies are not trustworthy in achieving their objectives, how can we optimise or remedy the network topology by adding a link?

In the next section, the contributions of this thesis will be listed and discussed.

1.3 Contributions

This thesis studies the co-evolution changes On and Of the network. Most of the existing studies have focused on trust modelling at the network entity level (local trust), which is the change in node behaviours which can cause the change of network topologies.


Trust in Sociology
Trust in Computer Networks
Adaptive Networks and Topological Measurement

Interplay of trust behaviours vs. underlie topological connectivity

Xiang, M., et al. (2017). NTaaS: Network Trustworthiness as a Service. In Telecommunication Networks and Applications Conference (ITNAC), 2017.
Xiang, M., et al. (2016). The critical role of structural hole in forming trust for Securing Wireless Sensor Networks. *International Journal of Information, Communication Technology and Applications*, *2*(1), 66–84.
Xiang, M., et al. (2015). Avoiding the Opportunist: The Role of Simmelian Ties in Fostering the Trust in Sensor-Cloud Networks. *International Journal of Distributed Sensor Networks*, *11*(10), 873941.
Xiang, M., et al. (2015). Simmelian Ties and Structural Holes: Exploring Their Topological Roles in Forming Trust for Securing Wireless Sensor Networks. In *2015 IEEE Trustcom/BigDataSE/ISPA* (Vol. 1, pp. 96–103).
Xiang, M., et al. (2015). The double-edged sword: Revealing the critical role of structural hole in forming trust for securing Wireless sensor networks. In *Telecommunication Networks and Applications Conference (ITNAC), 2015 International* (pp. 286–291).

Figure 1.1: Research Gap

However, it seems that there has been limited work done on how the underlying network topological connectivity can affect the trust behaviours of the network nodes (global trust). Therefore, there are four main contributions as follow to cover the research gap.

Motivation: We have identified a research gap in ’trust’ across the computer network, sociology, and adaptive network areas.

We have extensively reviewed the literature on ’trust’ in the computer network, ’trust’ in sociology, and adaptive network. A research gap has been identified, which is shown in Figure 1.1. The interplay of trust behaviours versus underlying topological connectivity is the research gap we have identified. This finding has become the motivation for this thesis.

Contribution 1: We have confirmed the relationship between Simmelian Ties, Structural Holes, and Node Trustworthiness in Routing in the P2P network environment.

The Simmelian Tie and Structural Hole structures are the two typical social structures in sociology. They have played a crucial role in improving or obstructing the performance-related outcome (Latora et al., 2013). Sociologists have long debated which structures are better in improving the performance outcome. Engle (1999) has proposed a framework to explain the relationship between the Simmelian Ties, Structural Hole, and performance. Inspired by this framework, this thesis has proposed a framework to explain the relationship between the Simmelian Ties, Structural Hole, and Node Trustworthiness in Routing in the distributed P2P network environment. Adapting Engle's framework (Engle, 1999), we believe the Simmelian Ties have a positive impact on the performance of trust-based routing algorithms and the node trustworthiness. Structural Hole has a negative impact on the performance of trust-based routing algorithms and the node trustworthiness. Moreover, the Structural Hole also has a negative impact on the positive impact from the Simmelian Ties. We use the recommended topological metric, which is the Clustering Coefficient to evaluate the Simmelian Ties in the network. We ran the simulation studies with a benchmark trust-based routing algorithm, which is DTEGR (Xiang, Bai & Liu, 2012). We have validated that the network which has higher average clustering coefficient results in a lower packet loss while it was under attacks. When there is a Structural Hole in the network and the node, which is at the Structural Hole position is under attack, the packet loss results are always high. Moreover, when a network has a high average clustering coefficient and also has Structural Hole, the packet loss results are very high as well. All these findings have validated the three hypotheses on the relationship among Simmelian Ties, Structural Hole, and the node trustworthiness in routing. This has answered the first research question.

Contribution 2: We have proposed a new term, which is network trustworthiness. For this new term, we introduced a new service platform base on it, which has been named Network Trustworthiness as a Service (NTaaS).

This thesis has introduced a new term, which is Network Trustworthiness and proposed a new service platform based on it, which is Network Trustworthiness as a Service (NTaaS). It treats Network Trustworthiness as a service and provides this service to the P2P network users. The network trustworthiness models and evaluates the trust on P2P network from a global perspective.

Contribution 3: We have proposed the network trustworthiness T evaluation framework for NTaaS. In this framework, we introduced a weighted clustering coefficient for the evaluation of the Simmelian Ties for the whole network, and the Structural Hole Locator (SHL) to locate and evaluate the Structural Hole in the network.

For the network trustworthiness T evaluation framework. As T evaluation only focuses on the target attacks scenarios, thus a Trustworthiness Tolerance Margin (TTM) is proposed for the random attacks scenarios evaluation. In different scenarios, the networks are designed for different objectives. Therefore, the metrics for a network trustworthiness evaluation should be adjusted accordingly depending on its objectives. In other words, there is no universal metric to evaluate the network trustworthiness. This thesis considers a scenario, in which the objectives are ensuring the network availability and trust-based routing algorithm performance. With the findings from the first research question, the Clustering Coefficient is selected to evaluate the Simmelian Ties in the network. The effective size was used in most of the sociology studies for the evaluation of the Structural Hole in the network. Unfortunately, this metric is not able to detect the actual Structural Hole in the network. In such a case, this thesis proposed a new method to detect the physical and logical Structural Hole in the network. Moreover, with the nodes in the network at different positions, their relative importance to the network is different. For example, the nodes at the centre of the network are more important than the nodes at the edge of the network. In such cases, the topological metric Betweenness is also needed for an evaluation of the trustworthiness. There are random network topologies generated and deployed in the simulation scenarios for the validation of the T value. The simulation results have shown the T value is able to measure the trustworthiness of the network, and it is flexible for the different network objectives. This has answered the second research question.

Contribution 4: We have proposed a mathematical approach to remedy the network trustworthiness by adding a link at a critical position.

The last contribution of this thesis is the remediation of the network topology. Based on the network trustworthiness evaluation results, which is the T value, if the network is not trustworthy enough to satisfy its need to serve the objectives, obviously, this network needs to be remedied. Depending on the evaluation metrics of T value, the remediation method would be different. According to the scenario in the last chapter, the physical Structural Hole is the most critical threat to the network, then logical Structural Hole, and finally the low Clustering Coefficient. The simulation results show the remediation framework is able to achieve the largest increment on the T value by adding a link. This has answered the last research question.

In the next section, the methodology for this thesis is introduced.

1.4 Methodology

The methodology for this research is modelling and simulation. We use the simulation platform J-Sim (Sobeih et al., 2006), Omnet++ (Varga & Hornig, 2008), and ONE (Keränen, Ott & Kärkkäinen, 2009) for the network environment and trust-based routing algorithms simulation. The UCINET (Borgatti, Everett & Freeman, 2002) software can be used to calculate most existing network topological metrics. The use of this software can assist the validation of the ideas and hypotheses proposed in this thesis.

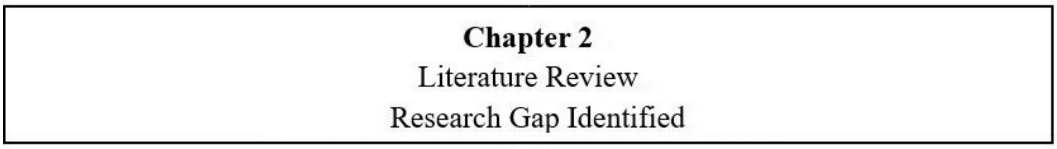
For the simulation and validation, the Simmelian Ties and Structural Hole need to be evaluated in the network, as these two structures are what we focus on in this thesis. The different network topologies and the trust-based routing algorithm performance difference can be compared and used to validate the impact of the Simmelian Tie and Structural Hole characterised network structures. The network topologies for the simulation are randomly generated by the Omnent++ to verify that the evaluation framework can be adopted in any network topologies in the P2P environment. A study in (Latora et al., 2013) has suggested the clustering coefficient can evaluate the Simmelian Tie, and the effective size and Simmelian brokerage can be used to evaluate the Structural Hole in the network in the local aspect. However, both effective size and Simmelian brokerage have been tried in different scenarios such that neither both of the metrics can detect the actual Structural Hole in some network topologies. Thus, a better approach is required to locate the Structural Hole in the networks. The trust-based routing algorithm performance can be evaluated by the number of packet loss and packet latency while the network is under attack. As the main purpose of the trust-based routing algorithm is to detect and avoid the malicious activities. The less packet loss number means the algorithm can better detect and avoid the malicious nodes. Lower packet latency means a shorter route to the destination, which normally means less sacrifice for safety and trust routing. The packet loss number is used to compare the trust-based routing algorithm performance in many studies, such as (Crosby, Pissinou & Gadze, 2006; Mahalle, Thakre, Prasad & Prasad, 2013; Mu & Yuan, 2010; Xia, Jia, Ju, Li & Zhu, 2011; Hui-hui, Ya-jun, Zhong-qiang & Hao, 2009; Michiardi & Molva, 2002; Tchepnda & Riguidel, 2006; Bao, Chen, Chang & Cho, 2012; Castelfranchi, Falcone & Pezzulo, 2003; Buchegger & Le Boudec, 2002; Liqin, Chuang & Tieguo, 2006; Peng, He & Meng, 2008; Chen, Guo, Bao & Cho, 2014).

In the next section, the thesis structure will be introduced.

1.5 Thesis Structure

This thesis starts with an introduction to relevant background knowledge, motivation, and research questions in the first chapter. This is then followed by a description and comprehensive discussion of three frameworks, which are the framework for Simmelian Ties, Structural Hole, and Node Trustworthiness relationship; the NTaaS, and the framework for network trustworthiness evaluation. Moreover, a mathematical approach to remedy network topologies is proposed. The extensive simulation studies are then presented for each framework to validate these new frameworks. The chapters of the thesis are organised as follows:

Chapter 2, as shown in Figure 1.2, it gives a rough review of the traditional security approaches to explain why the soft security is needed. It is followed by a survey on the existing trust-based routing algorithms. As ’trust’ in the computer network is derived from the sociology, a detailed overview of ’trust’ in sociology is given. From the studies on social structures, a new concept adaptive network is introduced. Inspired by this concept, a research gap is identified among these three research areas. Then a further review of the complex networks is given in seeking the current work on network topology evaluation metrics.



Chapter 2
Literature Review
Research Gap Identified

Figure 1.2: Chapter 2 Structure

Chapter 3 has given three hypotheses on the relationship between the Simmelian Ties, Structural Hole, and node trustworthiness in routing. The metrics to evaluate the Simmelian Ties and Structural Hole in the network are introduced following the hypotheses, which are used in most of the studies in sociology and complex networks. Finally, simulation studies have given to validate the hypotheses. It is shown in Figure 1.3.

Chapter 3 has confirmed the different network structures can affect the underlying nodes’ behaviours. Thus, chapter 4 defined a new term Network Trustworthiness. With this new term, we proposed the concept of Network Trustworthiness as a Service

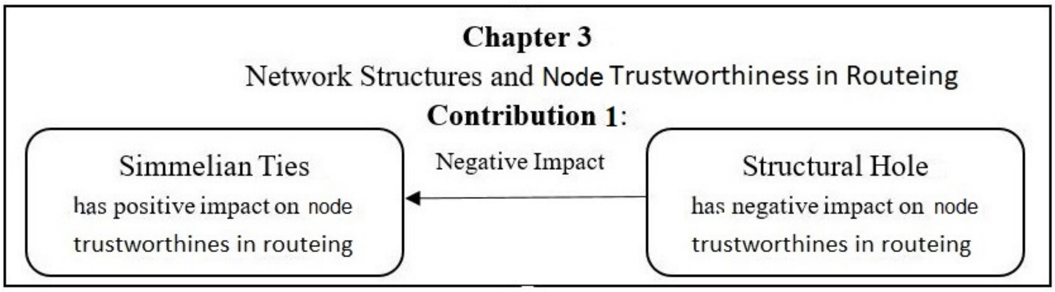

Chapter 3
Network Structures and Node Trustworthiness in Routeing
Contribution 1:
Simmelian Ties
has positive impact on node
trustworthines in routeing
Negative Impact
Structural Hole
has negative impact on node
trustworthines in routeing

Figure 1.3: Chapter 3 Structure

(NTaaS). Following this, a framework T for the evaluation of the network trustworthiness for NTaaS is proposed. A scenario is created with predefined network objectives, which are the network availability and node trustworthiness in routing. In this scenario, the findings from Chapter 3 are applicable, in which the Simmelian Ties are the preferred structure and Structural Hole structure should be avoided. This chapter also proposed an algorithm to detect the Structural Hole. In such case, the metrics to form the T are clustering coefficient, Betweenness, and the Structural Hole algorithm. At the end of this chapter, the extensive simulation studies are provided. It validated the accuracy of the T for the evaluation. As the T is mainly focused on the target attack, we also proposed the Trustworthiness Tolerance Margin (TTM) for the evaluation of random attacks' scenarios. It is shown in Figure 1.4 on the following page.

If the T for the network is not satisfied by the need to serve its objectives, which is determined by the evaluation framework proposed in chapter 4, then this network requires remediation. The remediation approaches for the network topology are discussed in Chapter 5, which is shown in Figure 1.5 on the next page. For remediation, the network topologies can be optimized by adding an additional link or node. The key factor is where the link or node should be added to achieve the best improvement in the T value. The T considers the Simmelian Ties, Betweenness, and the Structural Hole. From a threat-critical level point of the view, the physical Structural Hole is the most critical threat, then logical Structural Hole, and finally, the number of Simmelian Ties

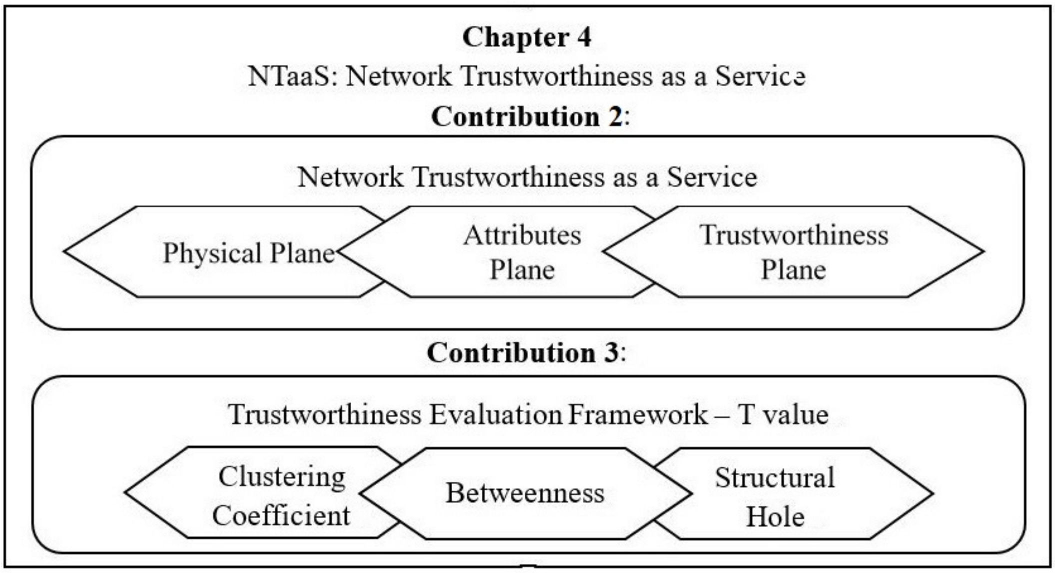


Figure 1.4: Chapter 4 Structure

in the network. In such cases, there are three scenarios that should be considered in the remediation, which are physical Structural Hole scenario, logical Structural Hole scenario, and the scenario without any Structural Hole in the network. The simulation studies also provide for the validation of the remediation approaches. Finally, possible nodes recruitment approaches for the network structure remedy are discussed.

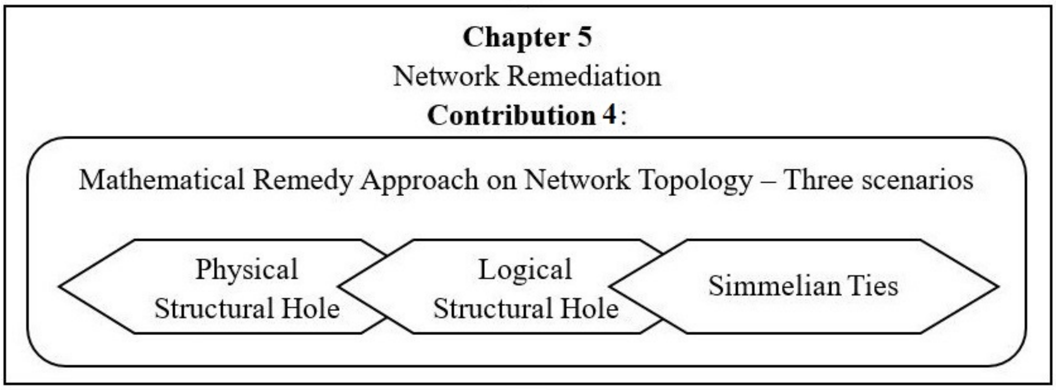


Figure 1.5: Chapter 5 Structure

Chapter 6 concludes all the contributions and the findings in this thesis. In addition, the limitation of this research and future work are discussed at the end.

Overall, the complete thesis structure is shown in Figure 1.6.

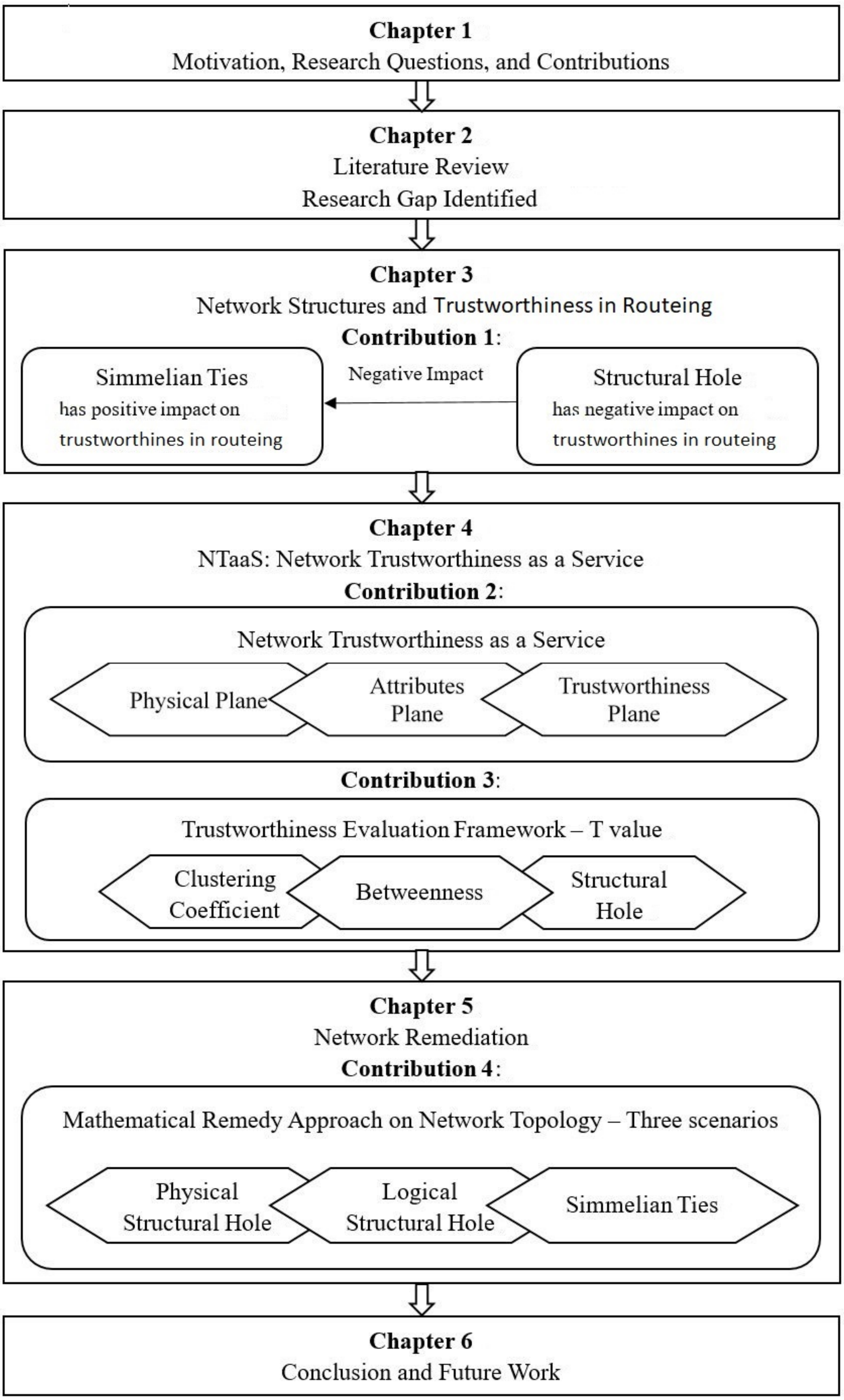


Figure 1.6: Thesis Structure

Chapter 2

Literature Review

2.1 Introduction

In this chapter, the traditional security mechanisms will be reviewed, so as to find the reasons why ’trust’ needs to be used in the P2P environment. Then the state of art in regard to trust in the computer network will be reviewed later. Meanwhile, as the concept ’trust’ comes from Sociology, the review on trust in Sociology will be considered as well. Finally, the adaptive network concept will be introduced and the relationship between trust and network structures in the computer network will be explored.

2.2 Traditional Security Approaches

2.2.1 Cryptography

Cryptography refers to ’secret writing’, which is believed to be the most powerful mechanism against many kinds of security threats (Pfleeger & Pfleeger, 2002). Encryption can protect information privacy and data integrity because it translates the data into an unreadable form. In such a case, it makes unauthorised manipulation almost impossible. There are two basic encryption methods, which are substitution and transposition (Kahate, 2013). The substitution methods replace plaintext characters with other characters, symbols, or numbers. However, it is believed that these methods are vulnerable to frequency analysis (Jakobsen, 1995; Lee, Teh & Tan, 2006). Therefore, the symmetric-key algorithms are introduced to overcome this issue.

Nowadays symmetric-key encryption is very strong such that it is almost impossible to decrypt without a key such as Advanced Encryption Standard (AES) (Heron, 2009), which has been approved and used by the US Government. However, if malicious parties find a way to obtain the key, they can easily decrypt the message without too many difficulties, no matter how complicated and robust the cyphers are. In such cases, the key distribution is crucial in symmetric encryption. To overcome this problem, asymmetric encryption (Kahate, 2013) is proposed. In asymmetric encryption, there are two kinds of key, which are private and public keys. The public key is generated by the private key and it is not reversible to the private key. A typical algorithm is RSA (R. L. Rivest, Shamir & Adleman, 1978). Unfortunately, some people have discovered that the men in the middle can cause the asymmetric encryption to become vulnerable.

From these examples, we can see that the key distribution obviously is a problem, which needs to be resolved, so as to ensure the cryptographic mechanisms work probably and securely. The following is an overview of key management techniques to resolve the ’men in the middle’ issue.

2.2.2 Public Key Infrastructure (PKI)

Cryptographic algorithms can encrypt and decrypt data with cryptographic keys while obtaining the keys by malicious parties can expose the encrypted data and allow unauthorised manipulation. Therefore, the distribution of encryption key to authorised parties is a critical issue in cryptography. Asymmetric keys are not only used to encrypt and decrypt data, but can also be used to identify the various entities. As mentioned in the encryption section, the ’man in the middle’ can lead to the asymmetric algorithm vulnerable. To resolve this problem, the third party of a trust agency is introduced, such as Key Distribution Centre (KDC) or Certification Authority (CA) is involved in certifying the ownership of the public key by its digital certificates (Sun, Trappe & Liu, 2004). Trust is the key factor in this scheme as the CA needs to be trusted by both parties in order to exchange keys, so as to allow this key management structure to run properly. The X.509 (Wazan, Laborde, Barrère & Benzekri, 2008) is a telecommunication standard for PKI which are developed by the International Telecommunication Union (ITU). However, in the P2P distributed network environment, the old PKI is no longer efficient for a centralised network as the trust agency require centralised network structures. In such a case, a distributed PKI scheme is introduced as suggested by studies (Lesueur, Me & Tong, 2009).

2.2.3 Hash Functions

The hash function is a computationally efficient function mapping binary strings of arbitrary length to binary strings of some fixed length, called hash-values, and it is one of the fundamental primitives in Cryptography (Menezes, Oorschot & Vanstone, 1996). It is also called the one-way hash function in that it generates the hash-values based on the different input texts, it should be not reversed, and any change in the original text should generate a completely different hash-value. In such a case, though it would not prevent the users to read and understand the messages like encryption does, any manipulation in the message should be detected, so as to ensured the data integrity and also support the authentication mechanisms. The hash-values also can be used to store the passwords on the computers. When users type in passwords to log in, the system will calculate the hash-values of the passwords and compare them with their store hash-values in the database, the users can log in if the hash-values match. Should anyone hack into the password database, they cannot see the passwords but only the random hash-values. On the other hands, as hash algorithms such as Message Digest 5 (MD5) (R. Rivest, n.d.), Secure Hash Algorithm 3 (SHA-3) (Aumasson, Henzen, Meier & Phan, 2008) are all known by the public, hackers can also obtain the algorithm easily. In such cases, hackers can use popular passwords and calculate their hash-values with these algorithms and compare with the password database, so as to elicit the passwords. This is why we should not use simple and/or popular passwords.

2.2.4 Authentication

Authentication is the technique used to verify the identities of each component in P2P, so as to only allow the authorised parties being granted access to authorised resources and contents while keeping unauthorised parties away from classified data. This can help protect information privacy and data integrity. Useful techniques can be passwords, biometrics such as fingerprints, security cards, and digital certificates from CA. The study by Gao (2012) has defined three factors for authentication, which is something you know (e.g. passwords), something you have (e.g. security card), or something you are (e.g. fingerprint). For these techniques, it is important to keep the keys (e.g. passwords, token, fingerprint, etc.) safe and secret, so malicious parties are not able to obtain them. Moreover, the passwords should be complicated, random, and long enough, so they can defend against different attacks, such as dictionary attacks, and brutal attacks.

2.2.5 Access Control

Access control is the policy of enforcement to ensure only authorised parties be granted access to what they have been allowed. There are three categories of the access control approach, which are Access Control List (ACL), Mandatory Access Control (MAC), and Role-Based Access Control (RBAC) (Bai & Zheng, 2011). The policies in ACL are a set of access rights for users on every protected resource. MAC is an operating system which constrains the ability of a subject or initiator to access or generally perform some sort of operation on an object or target, and the RBAC is usually used in large enterprises so that it can implement MAC. The permissions in RBAC are defined by users' roles. The principle for assigning the access right to the users is always allocating them with the minimum privilege. Access control is focused on defending data integrity and against information privacy attacks. However, even following the minimum privilege rule, if the misconfiguration on the access control occurs, it could lead to availability issues because the authorised users cannot access to their authorised resources. Moreover, as there can be millions of devices in the P2P environment, the access control can become very complex and the resource exhausted. Finally, if the legitimate users or nodes are compromised, the access control is not able to deal with this situation.

2.2.6 Hard Security vs. Soft Security

In summary, these traditional security mechanisms are becoming more and more robust to defend different malicious cyber attacks nowadays, as more and more security threats and bugs are identified. On the other hand, Rasmusson and Jansson (1996) suggested that if these mechanisms have been bypassed, the system and data will be unprotected. For example, if malicious parties are able to obtain the decryption key from somewhere else, the encryption mechanism will be vulnerable. In authentication mechanisms, if the malicious parties are able to obtain the passwords, fingerprints, etc. Then they will be identified as legitimate users as well. Rasmusson and Jansson (1996) said: *"There shall never be a key that uncritically opens up all locks on the system."* They suggested the traditional security mechanisms are hard security, which is only able to defend against known security threats, but inefficient in dealing with unexpected behaviours. Once malicious parties have found the bugs (unknown security threats) in the system, they can bypass the hard security. Thus, there will be no more protection for the system and data anymore. Moreover, hard security is not able to protect the system from the compromised or selfish nodes or users as well, as they all have the "key" to open up all locks on the system. In such circumstances, the soft security approaches are introduced, which is "Social control" or trust. Soft security considers users' actions are acceptable as long as their actions are not harmful to anyone else, but once their behaviours become malicious, they will lose their ability to act accordingly. In such case, the malicious parties may bypass the hard security, or obtain the legitimate identity, but once they act maliciously or selfishly, they shall be detected and lose their ability to act accordingly.

In the next section, the state of art on 'trust' in a computer network will be reviewed.

2.3 Trust Modelling

Trust is a very important term in our daily life because all relationships rely on it in the human society, and each interaction with other people involves trust as well. For example, when customers are shopping in stores, staffs will recommend products to them. If customers trust the staff, they will consider their suggestions and might buy the products, but if they do not trust the staff, there is no sale happening. Trust is not only important in human society, but also in the computer security area. Take the public key cryptography as an example, it requires the key only can be accessed by the authorized person. Otherwise, this security mechanism becomes compromised with which involves trust on both sides of communications.

While trust is hard to be defined as it can mean different things in different areas. The Oxford dictionary has defined the 'trust' as a synonym of six words: confidence, belief, strength, goodness, responsibility, and reliability (Hornby, 1988). In detail, the trust can give confidence; trust is a subjective matter which is a strong belief in someone or something in the goodness; trust can give strength; trust can make people rely on someone and something.

There are some other definitions regarding the more general aspects of trust. Mayer, Davis and Schoorman (1995) define trust as "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other party will perform a particular action important to the trustor, irrespective of the ability to monitor or control the party". In (Jøsang, 1996), Josang interprets trust as the belief that an agent will behave without malicious intent for passionate entities (e.g., humans), and as the belief that it will resist malicious manipulation for rational entities (e.g., systems). Denning also gives more explanation about trust in (Denning, 1993). He claims that trust cannot be treated as a property of a trusted system but rather an assessment based on the experience that is shared through networks of people.

These three definitions can be summarised as the expectation of the other parties' performance and without malicious intent, and that this expectation will become the experiences shared through the networks of people. It can be seen that this includes two parts: the first part has defined the direct trust experience, and the second part can be interpreted as the 'word of mouth' which is about reputation i.e., indirect trust.

2.3.1 Trust in Computer Network

In the field of Ad Hoc wireless network routing, the most popular definition for trust is the probability of an individual node will behave as expected, and the studies in

(Fernandez-Gago, Roman & Lopez, 2007; Guo & Wang, 2007; Marsh, 1994; Mu & Yuan, 2010; Yang, Huang, Wang, Wang & Zhang, 2009; Zahariadis et al., 2009) are all use a value from 0 to 1 or -1 to 1 to represent the level of trust. Eschenauer, Gligor and Baras (2002) define the trust in a communication network environment as "a set of relations among entities that participate in a protocol. These relations are based on the evidence generated by the previous interactions of entities within a protocol. In general, if the interactions have been faithful to the protocol, then trust will accumulate between these entities." This definition highlight two points. First, they define trust in the communication network is being built up with previous interactions, namely experiences. Secondly, probability is used in the approach to describe and model the trust level. Moreover, these definitions of trust in the wireless network relate more to the behaviours of the entities. Other studies in (Cho, Swami & Chen, 2011; Gonzalez, Anwar & Joshi, 2011) have defined five properties of trust in the wireless ad-hoc network environment, which are context-dependent, asymmetric, dynamic, subjective, and not transitive, as shown in Figure 2.1 on the following page below.

For example, the node A can be trusted in packet forwarding performance but might be 'malicious' in a data integrity aspect. Node B can be trusted in data integrity but might fail to pass the authentication. These are the context-dependent properties in trust. As mentioned by Schoorman, Mayer and Davis (2007), the trust is not reciprocal. Node B forwards the packets for A every time, so node A trust in node B. But node B does not necessarily need to trust node A in return; this is the asymmetric property of the trust. The definitions in studies (Adams & Davis, 2005; Gambetta, 1988) describe that the trust as a dynamic probability which states that trust is a dynamic value. For example, node A trusts node B previously, but node B betrays node A at a later stage, which causes node A to no longer trust node B. This is the sample of the dynamic property in trust. Gambetta (1988) has defined trust as a particular level of subjective probability in that different people might have different levels of trust of the same person or things,

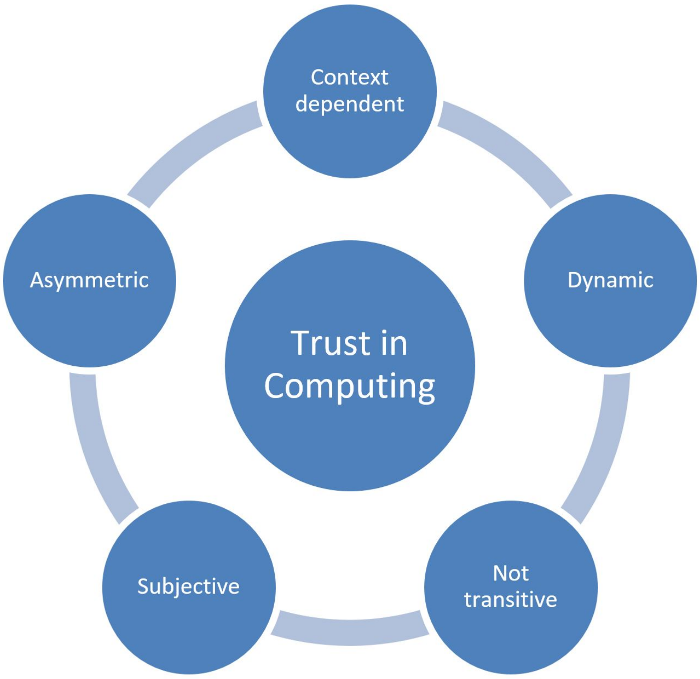


Figure 2.1: Trust Properties

and this is the subjective property in trust. The last property that trust is not transitive, for example, node A trusts in node B, and node B trusts in node C, in such circumstance, it does not indicate that node A should trust in node C. Moreover, a study in (Gonzalez et al., 2011) has defined the sixth property, which is that trust is a measure of uncertainty. It can be seen that the trust has been defined most frequently as the ’probability’ in (Fernandez-Gago et al., 2007; Guo & Wang, 2007; Marsh, 1994; Mu & Yuan, 2010; Yang et al., 2009; Zahariadis et al., 2009; Adams & Davis, 2005; Gambetta, 1988). When the probability value is not either 0 i.e., not possible at all, or 1 i.e., definitely happen, that creates the uncertainty in a particular behaviour occurs or not.

One study (Sherchan, Nepal & Paris, 2013) has classified ’trust’ in computer science into two categories which are ’user’ and ’system’. The ’user’ category is the trust amongst users on the Internet, such as the feedback system of online shops eBay or Amazon for past interactions between members (Ruohomaa, Kutvonen & Koutrouli, 2007; Resnick, Kuwabara, Zeckhauser & Friedman, 2000). The system category is the expectation of a device or system will faithfully behave in a particular manner to fulfil its intended purpose (Yao, Chen, Nepal, Levy & Zic, 2010). It should apply to both software and hardware. This thesis is focusing on trust in computer network routing which belongs in the system category.

2.3.2 Processes of Trust Evaluation

The existing distributed trust management routing algorithms normally involve three stages. The first stage is *data aggregation*, in which the 'trustor' collects evidence about 'trustee' (direct evidence) or evidence from its neighbours (reputation from 3rd parties). The second stage is *data syncretization*. The last stage is *decision making*, as shown in Figure 2.2.

Data Aggregation

From the trust definition has been defined before, we can see there are two parts of the behaviour trust which are direct trust and reputation, and this can be seen in studies (Xia et al., 2011; Hui-hui et al., 2009; Tchepnda & Riguidel, 2006; Bao et al., 2012; Liqin et al., 2006; Peng et al., 2008). 'Direct trust' is the direct experience with trustee from trustor, which is the first-hand evidence, and the reputation (indirect trust) is the second-hand information from neighbours of both trustor and trustee. For the data aggregation, the main question here is how to quantify or measure the trust.

There are three most critical security issues on a computer network, they are *network availability*, *data integrity*, and *information privacy*. In such case, most of the studies on trust measure in routing network are focused on these three issues.

The most common metric included for trust evaluation in routing networks is the packet forwarding success ratio, as this ratio can detect the common malicious behaviours in P2P, such as black hole attack, grey hole attack and selfish behaviours, so

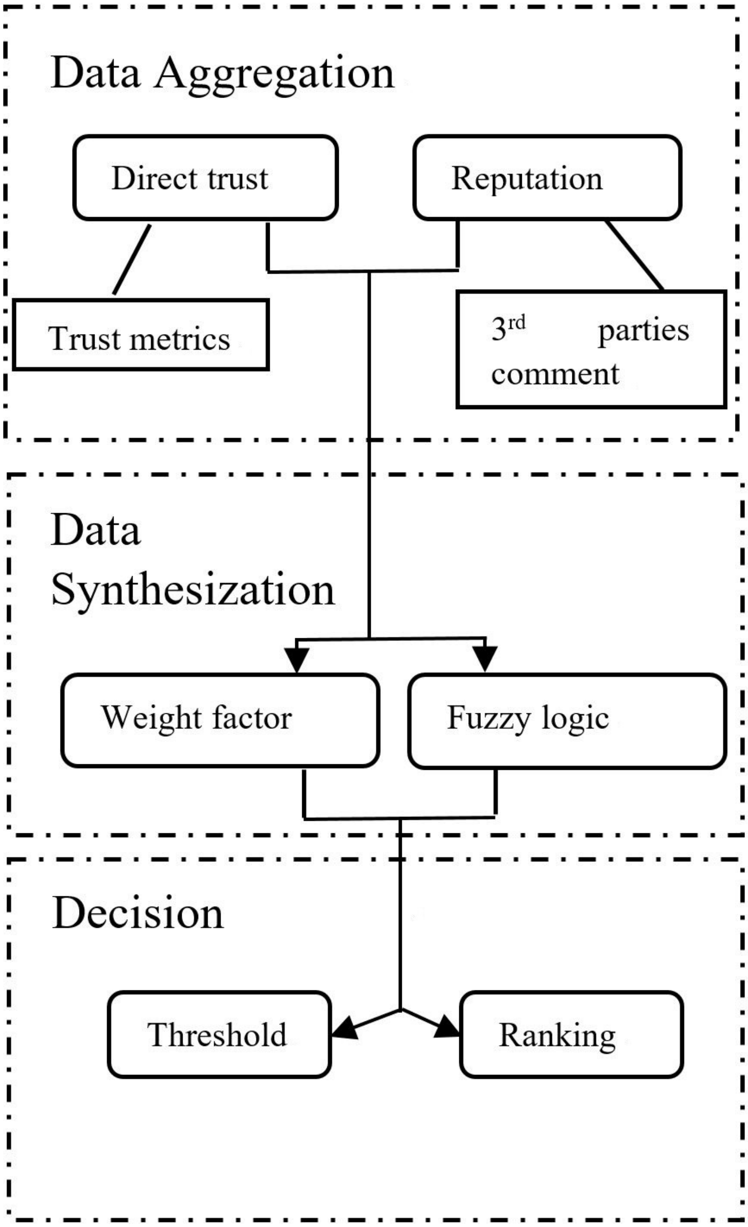


Figure 2.2: Trust Evaluation Processes

as to ensure network availability. As can be seen in Table 2.1 below, most of the existing trust algorithms have included this metric. Data integrity, cryptography, and routing protocol execution are all Boolean metrics. Data integrity is another consideration so as to ensures the data has not been manipulated by malicious parties. Cryptography states whether trustee capable for cryptography as a node capable of data encryption is more trustworthy. Routing Protocol Execution monitors a trustee to see whether it has done anything outside the protocol allowed. Battery life level is measured to ensure the target node still has sufficient energy to complete the given task. Link condition is used to check whether there is congestion or poor wireless reception to the target node. The use of Social Ties is first introduced by Bao et al. (2012) and Chen et al. (2014), is to check if the trustee has a background and reason in the real world to process such task or request. As the personal device like mobile phone, tablet, etc. become smaller and smaller, these kinds of nodes can appear more and more in the network. In such case, the social ties metric can become increasingly important as well.

Table 2.1: Typical Trust Metrics

Trust Metrics	**Literature**
Packet Forwarding Ratio	AFSTrust (Xia et al., 2011), Multi-angle trust (Hui-hui et al., 2009), Distributed Trust Infrastructure (Tchepnda & Riguidel, 2006), ATSR (Zahariadis et al., 2009), Evaluating Initial Trust Value (Mu & Yuan, 2010), SQTrust (Chen et al., 2014), FTBAC (Mahalle et al., 2013)
Data Integrity	Multi-angle trust (Hui-hui et al., 2009), Quantitative Analysis (Liqin et al., 2006), ATSR (Zahariadis et al., 2009)
Cryptography	Quantitative Analysis (Liqin et al., 2006), ATSR (Zahariadis et al., 2009)
Routing Protocol Execution	Hierarchical Dynamic Trust Management (Bao et al., 2012), SQTrust (Chen et al., 2014)
Battery Life	Multi-angle trust (Hui-hui et al., 2009), Hierarchical Dynamic Trust Management (Bao et al., 2012), ATSR (Zahariadis et al., 2009), SQTrust (Chen et al., 2014)
Link condition	Distributed Trust Infrastructure (Tchepnda & Riguidel, 2006)
Social Ties	Hierarchical Dynamic Trust Management (Bao et al., 2012), SQTrust (Chen et al., 2014)

’Reputation’ is the direct trust results from neighbours of their trustee. The 3rd parties can be malicious in that they will provide a misleading reputation of the trustee for purposes, such as ’badmouth attack’ and ’ballot stuffing attack’ (Chen, Bao &

Guo, 2016). 'Badmouthing' is providing false bad reputation against the trustee and 'ballot stuffing' is providing a false good reputation of the trustee. The algorithm CONFIDANT (Buchegger & Le Boudec, 2002) is one of the typical early trust-based routing algorithms that it does not validate the reputation from nodes, so malicious nodes can easily exclude a particular node in the network by spreading a false reputation. To guard against these attack, studies (Zahariadis et al., 2009) setup a margin threshold on the reputation from different neighbours, as they think the information from 3rd parties on the trustee should not have much difference. The reputation results from malicious nodes should be considered as low credibility as well. In such case, the reputation value from neighbours can be multiplied by their trust value (Xia et al., 2011), i.e., the more trustworthy the neighbours, the more credibility can be placed on their feedback. Moreover, as mentioned above regarding the trust properties, it states trust is dynamic. The evidence collected by a trustor should be time dependent.

The study in Xia et al. (2011) identified four metrics to evaluate 'trust' in the sensor network, including 'direct trust', 'recommendation', 'incentive function', and 'active degree'. 'Direct trust' applies an attenuation algorithm to guard against a cheating attack, which is the dynamic property of the trust. The 'recommendation' value is the trust value from 3rd parties multiplying their credibility (trust value) to reflect the usefulness of the recommendation. The 'incentive function' reflects the penalty on uncooperative entities, and 'active degree' state more active entities normally have higher trust. One paper (Hui-hui et al., 2009) categorised trust computing in a sensor network into three factors which are 'communication', 'data', and 'energy trust'. Direct and indirect trust are the communication factor, data integrity for data trust, and considering the energy is an important factor to affect an entity's performance and selfishness levels as lower battery level normally will become more selfish to save energy. Another paper (Liqin et al., 2006) categorized trust attributes on the Internet as security, dependability, and performance. Security considers, for example, whether entities authenticated, data integrity, etc. ’Dependability’ is more like behaviour trust, which concerns the no-fault-service rate, etc. ’Performance’ is concerned with service latency, throughput, etc. A third paper (Tchepnda & Riguidel, 2006), it defined a new metric other than direct and indirect trust that it is called ’basic trust’ or ’willingness to trust’. It utilises ’initialization’ and an update of trust processes. In initialization, it just like a first impression in the human society (an optimistic or pessimistic basis) to decide whether to trust this entity in the network. In the update case, there are four possible results for basic trust, optimistic on success or failure of interactions, pessimistic on success or failure of interactions. The pre-set value for the basic trust is required to be set up by the security administrators. There is no simulation test in the paper, only a framework of security procedures on how these trust should work.

Additionally, in some of the trust-based routing algorithms, nodes elect a powerful node to become a cluster head. In this case, nodes will forward the evidence they collected to the cluster head for trust evaluation, and cluster heads will forward their own direct evidence to the base station for their own trust evaluation. The cluster size is defined by the user, e.g. less than two hops away from the cluster head, These algorithms are called hierarchical trust routing algorithms, a typical one is CONFIDANT (Buchegger & Le Boudec, 2002), which mentioned before. It is uses selected nodes as trust management nodes in charge of collection and management of reputation from all other nodes in the network, so other nodes can check the reputation of others from these nodes. Another hierarchical trust-based routing algorithm is (Bao et al., 2012) that it defines trust as ’social trust’ and ’QoS trust’. ’Social trust’ consists of intimacy and honesty where intimacy indicates the closeness based on previous experiences, and honesty indicates regularity. ’QoS trust’ concerns about the energy and selfishness of the nodes where they believe a compromised node will consume more energy, and a low energy level node will most likely become selfishness.

In summary, for trust metrics, we can see that trust in different scenarios will have different concerns or objectives so as to have different trust metrics included in the measurement. In the next section, the methods to synthesise these metrics will be reviewed.

Data Synthesization

After the data aggregation from the trustee and third parties (neighbours), the next step is to synthesise the data into one final result so they can be measured and compared. There are two typical ways to synthesise the aggregated data in the current trust algorithms, which are ’weight factor’ and ’fuzzy logic’. The weight factor algorithm assigns a weight factor to each metric to state which metric is more significant or relevant for the trust measurement and which is less significant or relevant. All weight factors added up together should equal to 1, so as to sum up these metrics with weight factors into a final result; exemplified in several studies (Mahalle et al., 2013; Mu & Yuan, 2010; Xia et al., 2011; Hui-hui et al., 2009; Zahariadis et al., 2009; Tchepnda & Riguidel, 2006; Liqin et al., 2006). Fuzzy logic sets up a membership function for each metric, then use a rule table and defuzzification functions to conclude a final result, as in a number of studies (Mahalle et al., 2013; Mu & Yuan, 2010; Xia et al., 2011). Some studies (Xia et al., 2011; Tchepnda & Riguidel, 2006) appear in both syncretization mechanisms as they use both mechanisms in their algorithm. Firstly, they use fuzzy logic to synthesise all direct trust metric into one final direct trust result and then use weight factor to combine the direct trust result and reputation together.

Decision Making

At the final stage, after the trust evaluation result on a trustee is calculated, there are two typical ways to determine whether the trustee should be trusted or whether it should be the next hop to forward the packet, there are two typical ways to determine. They are threshold-based and ranking-based. The threshold-based measurement is a pre-set threshold trust value, if neighbours have trust value above the threshold then they are trusted, vice versa they might be malicious, exemplified in various studies (Tchepnda & Riguidel, 2006; Jiang & Baras, 2005). The ranking-based measurement set up a value range like between -1 to 0 is distrust, 0 is neutral, 0 to 1 is trustworthy, and normally the algorithm will select the neighbour with the highest result as next hop until reaching the destination, such as (Xia et al., 2011; Hui-hui et al., 2009). In Table 2.2 are some of the comparisons of existing trust-based routing algorithms.

Table 2.2: Comparison of Existing Trust-based Routing Algorithms

Algorithms	**Trust properties**	**Data Aggregation**	**Data synthesization**	**Decision making**
THAODV (S. Singh, Mishra & Singh, 2016)	Dynamic, context dependent [1]	Direct	Weight factors	Threshold
TBE-LEACH (Miglani, Bhatia & Goel, 2015)	Dynamic, context dependent	Direct and indirect	Weight factors	Ranking
SQTrust (Chen et al., 2014)	-[2]	Direct and indirect	Weight factors	Ranking
FTBAC (Mahalle et al., 2013)	Dynamic	Direct and indirect	Fuzzy logic	Ranking
Hierarchical Dynamic Trust Management (Bao et al., 2012)	-	Direct and indirect	Weight factors	Threshold
AFSTrust (Xia et al., 2011)	Context dependent	Direct and indirect	Fuzzy logic and Weight factors	Ranking
Evaluating Initial Trust Value (Mu & Yuan, 2010)	Dynamic	Direct and indirect	Fuzzy logic and Weight factors	Ranking
ATSR (Zahariadis et al., 2009)	Dynamic	Direct and indirect	Weight factors	Ranking

Continued over page

[1] the properties in the cell are the one missing from those five defined properties mentioned above

[2] "-" symbol means this literature considered all trust properties

Table 2.2: Comparison of… *(continued)*

Algorithms	**Trust properties**	**Data Aggregation**	**Data synthesization**	**Decision making**
Multi-angle Trust (Hui-hui et al., 2009)	Dynamic	Direct and indirect	Weight factors	Ranking
Reputation-based Trust Update (Peng et al., 2008)	Dynamic, context dependent	Direct and indirect	Weight factors	Ranking
Distributed Trust Infrastructure (Tchepnda & Riguidel, 2006)	Dynamic	Direct and indirect	n/a	Threshold
Quantitative Analysis (Liqin et al., 2006)	-	Direct only	Weight factors	Threshold
Trust-based Cluster Head Election (Crosby et al., 2006)	Dynamic	Indirect only	Weight factors	Ranking
Autonomous Trust Establishment (Jiang & Baras, 2005)	Dynamic, context dependent	Indirect only	Weight factors	Threshold
FCM (Castelfranchi et al., 2003)	Dynamic	Direct and indirect	Fuzzy logic	Ranking
Core (Michiardi & Molva, 2002)	Dynamic	Indirect only	Weight factors	Ranking
CONFIDANT (Buchegger & Le Boudec, 2002)	Dynamic	Direct and indirect	Weight factors	Ranking

2.3.3 Trust in Different P2P Environment

Zhang, Wang and Sun (2013) have designed a trust system in the Smart Grid network. It selects the trust nodes in the network to perform trust evaluation and security analysis rather than every entity in the network. The whole article investigates how to calculate the best placement of trust nodes in the network and also the coverage, so the number of trust nodes' can be minimised and at the same time they are able to cover the whole network.

One paper (Mahalle et al., 2013) proposed a fuzzy logic based trust algorithm for IoT. It is focuses on the access right such as read-only, write etc. This fuzzy logic-based trust algorithm concerns the experience (direct trust), recommendation (indirect trust), and knowledge (number of activities records) as metrics for the trust evaluation. The knowledge metric is to validate whether there is sufficient evidence for the trust evaluation of the target, so the trustor can be confident in the evaluation results. Another paper (Li, Xuan & Wen, 2011) has proposed a trusted security framework for IoT with five modules; they are 'trusted users', 'trusted perception', 'trusted terminal', 'trusted network', and 'trusted agent'. The trusted user module is pre-eminent as it is the most active factor in the security of IoT, which a good authentication mechanism is required. The basic idea of this framework is to suggest every entity in the network should be fully authenticated, encrypted, and comprehensive security policies applied. In such case, the IoT should leave no chance for any kind of malicious attacks. Some papers (Bao, Chen & Guo, 2013; Bao & Chen, 2012) have considered the community of interest (CoI) in their trust algorithm under the IoT environment. For example, one paper (Bao et al., 2013) suggest the nodes only keeps the trust information for those nodes under the same communities so as to save the storage space. Moreover, in paper (Bao & Chen, 2012), the trust algorithm considers entities' owners' social relationship with the service providers, if there is no strong relationship between then activities can be considered malicious. It also suggested checking the common friends' list and communities list to validate the trust level, which it names as cooperativeness and communities interest.

In Cloud Computing, one paper (Caton et al., 2012) has defined trust as "Trust is a positive expectation or assumption of future outcomes that results from proven contextualized personal interaction-histories corresponding to conventional relationship types and can be leveraged by formal and informal rules and conventions within a Social Cloud to facilitate as well as influence the scope of collaborative exchange.” As the users cannot see what’s behind the Cloud service providers, for trust establishment in cloud computing environment, paper (Chadwick, Lievens, Den Hartog, Pashalidis & Alhadeff, 2011) suggest a third party organisation to audit and rate the Cloud service providers is necessary. Another paper (Savas, Jin & Deng, 2013) proposed a trust-based algorithm under a cloud-integrated wireless sensor network which has three subsystems; they are ’sensors’, ’network’, and ’cloud-based data servers’. It suggested in the trust algorithm, the trust should increase slowly with good behaviours but decrease quickly with dissatisfied behaviours to ensure detect malicious activities as quickly as possible. The authors also suggest the users have more trust in the cloud service provider who can provide more security mechanisms on their servers, access control, transparent data process activities, etc. These indicate that trust builds on the knowledge of the target and the weapons the provider has to protect users’ security.

From the literature review on trust in computer networks, we can see that the focus is on how to model and quantify trust, so as to detect and predict any malicious or unexpected behaviours in the network. The concept of ’trust’ comes from Sociology. In the next section, a literature review of the trust in Sociology will be given.

2.4 Trust in Sociology

’Trust’ in the computing network is derived from the interactions of our everyday life. To better understand and model the trust in the computer network further study on ’trust’ in sociology is necessary.

2.4.1 Definition of Trust in Psychology and Sociology

In psychology, most studies focus on are how the trust is set up on someone or something. Studies (Rotter, 1967; Rousseau, Sitkin, Burt & Camerer, 1998; Tyler, 2006) describe trust as a psychological state for an individual to have a positive expectation of a trustee's intentions or behaviour. This positive expectation is not necessarily reciprocal. In another words, node A trusts node B, but node B may not necessarily to trust node A. A study in (Marsh, 1994) has defined 'trust' by giving an example. An individual is in front of an ambiguous path which can either lead to an even of benefit or harm, which depends on another person's behaviour. The other perceives what could be more harmful than beneficial on this path. If the individual decides to take the ambiguous path then this individual decides to trust the other; to do otherwise, is to distrust the other. This example implicates trust can affect the individual's decision. The definition is more related to a belief in someone or in the goodness of something.

Trust in sociology is considered as a foundation of the relationship between people. Studies in (Schoorman et al., 2007; Eschenauer et al., 2002) have both defined 'trust' as the dynamic probability of someone will behave as expected. The study in (Marsh, 1994), Luhmann suggested the concept of trust is a means of reducing complexity in society. This is saying that when people make a decision, there are always assumptions about the situation so as to make the trusting decision. These assumptions can be understood as 'trust'. One studies (Adams & Davis, 2005) defined trust as is a measurement of confidence that an entity will behave in an expected manner without any assurance. Other studies (Rotter, 1967; Williamson, 1993; Coleman & Coleman, 1994) describe trust as risk and interdependence. The risk is regarding the uncertainty of a trustee's intention. The formation of trust requires interaction with other parties which mean interdependence.

2.4.2 The Formation of Trust

The studies of Fernandez-Gago et al. (2007) and Guo and Wang (2007) both suggest that the past experiences have a great impact on building the trust. A further study in (G. G. V. D. Bunt, Duijn & Snijders, 1999) has described the formation of human friendship as ego observes his own behaviour and that of alters during the interaction and evaluates this in terms of his own values, norms, interests, etc. Ego does not only pay attention to interaction in which he plays the part himself but also to interaction among alters. The more information ego has collected about all alters, the more reliable not only ego's estimation of the suitability of alters in terms of the continuation of the relationship but also his estimation of the willingness of alters to reciprocate his personal interest. If alters have a positive match of ego's expectation, ego will put more time and effort to interact with these alters so as to hope these friendly relationships become friendship. The study of Barrera and Bunt (2009) suggested if ego has more trust on alters in the past, then he will more likely trust more in the present, and vice versa. Also, if ego hears more trust from a third party about alters, ego has more trust on alters as well. However, when ego hears distrust from a third party, it will have more impact than the advice of a good reputation from a third party. These two studies implied that the establishment of trust in human relationships is based on previous interaction and reputation from third parties.

Beatty is suggested that there are three aspects of trust which are cognitive, behavioural, and emotive (Beatty, Reay, Dick & Miller, 2011). The cognitive aspect of trust is based on rational and reasonable behaviour (Kuan & Bock, 2005; Lewis & Weigert, 1985). Emotive aspect is irrational trust that the emotional security or comfort which enables the trustor to go beyond the cognitive trust in relying on someone (Holmes, 1991). Finally, behavioural trust is referred to as to commit some actions that make the trustor vulnerable to trustee (Schlenker, Helm & Tedeschi, 1973). Emotional trust can be the formation of trust by the trustor to have positive or negative expectation of the trustee. Cognitive trust can also affect the emotional trust. Belief influences attitude, then leads to behavioural intention and finally leads to behaviour (Fishbein & Ajzen, 1975). G. G. v. d. Bunt, Wittek and Klepper (2005) have further concluded six mechanisms for the formation of the trust relationship, they fall into two motives which are the expressive motive (trust and affect) and instrumental motive (trust and control). The mechanisms of expressive motive are ’homophily’, ’balancing’, and ’gossiping effect’. The remainder of the mechanisms are ’signalling’, ’sharing group’, and ’structural holes effect’ which make up the instrumental motive. ’Homophily’ is people with similar background and/or personality will most likely attract each other. ’Balancing’ is friends of my friend will become my friends and the asymmetric relationship will either become a mutual relationship or a null relationship. The ’Gossiping effect’ states trust will more likely develop between gossip-mongers. Gossip-mongers use gossip to create social solidity and affection on specified others. ’Signalling mechanism’ state management continuously and consistently signals its ‘good’ intentions to the worker by investments into workers that are costly for the firm and imply that the firm makes itself to some degree vulnerable, because whether or not, the actions will produce a payoff is at the discretion of the worker, this trust relationship is formed at a vertical level. The ’sharing group’ is a horizontal trust relationship where workmate rely on each other to complete the given task. Finally, the ’structural holes effect’ occurs when people have been tied to at least two none related actors. A person who has many structural holes is most likely acting as a broker in the network.

2.4.3 Social Networks and Their Structures

For the Structural Holes effect, mentioned in the previous section, there is another structure which is opposite to it - Simmelian Tie. In sociology, there is an ongoing debate on these two different social structures as to whether they are sustaining or hindering the performance-related outcome in individual and collective level (M. Granovetter, 2005; M. S. Granovetter, 1973). Simmelian tie in one study (Engle, 1999; Krackhardt, 1999) is defined as the triad ties or closed structure ties that are embedded in cliques as shown in Figure 2.3. Simmelian ties were first introduced by German sociologist Georg Simmel and further developed by David Krackhardt (Krackhardt, 1999) in 1999 as an alternative to the structural hole which was introduced by Ronald Burt (Burt, 2009) in 1993. Most studies on Simmelian ties nowadays are based on Krackhardt's work (Krackhardt, 1999), including that of Latora et al. (2013) and Engle (1999). In sociology, it is believed that the Simmelian tie is stronger than other regular strong ties between two actors as it discourages misbehaviour by introducing a third party to become a "shadow of the others" and a "shadow of the future". This strongly fosters a normative environment against opportunism, and engenders mutual trust, reciprocity norms, and shared identity. It facilitates the collaborative efforts by making the actors more willing to exchange information.

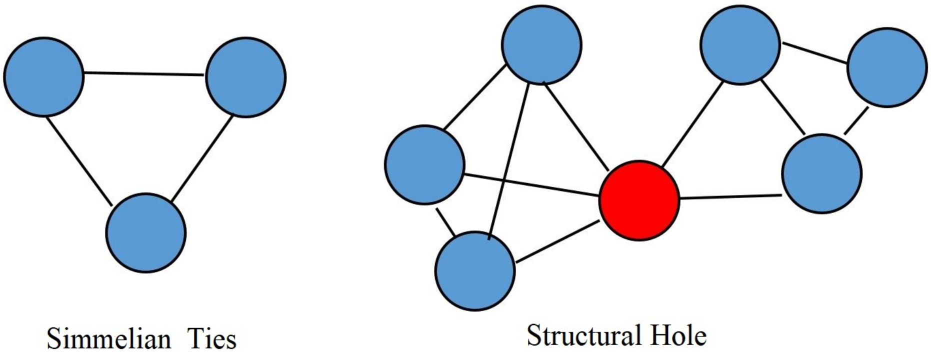


Figure 2.3: Simmelian Ties vs. Structural Hole

On the other hand, the Structural Hole (Burt, 2009) is an actor which connects between two or more actors or parties who are not related or connected. It is opposite to the closed structure of Simmelian tie, and it is an open structure tie as shown in Figure 2.3, the red node connects the left and the right-hand side of the network. In such a case, this actor normally acts as broker or gatekeeper which has the advantage of position to control the information flows among the disconnected networks. This feature makes it crucial in its position, as once it is broken, these connected networks will be disconnected again. While nodes in different networks need to communicate, these nodes can quickly forward the packets to the hole for the direction to the destination rather than asking everyone where it should go.

Structural hole theory was introduced by Burt (Burt, 2009) from three sociological theories, which consider 'weak ties in the social structure' (M. S. Granovetter, 1973), the value of exclusive exchange partner (Cook and Emerson 1978), and the benefits of Betweenness centrality (Freeman, 1978).

The weak ties in the social structure are believed as an important source of information, and it serve as the bridge between disconnected parties. For example, A and B are very good friends, which means A and B have strong ties with each other. In such case, A and B are more likely to share information with each other. In another word, if A knows something, most likely B knows that as well. However, B and C have a weak tie with each other, which means they are less likely to share all the information. In such case, if A or B want some new information, C is more likely to have it.

The value of exclusive exchange partner is the power to control. If B has the exclusive information or product, it has the power to control either A or C. B can share with it, or even withhold it.

The benefits of Betweenness centrality is it provides substantial network control for the individual in that position (Freeman, 1978). Betweenness is calculated as the number of the shortest paths between any node pair route through a node, which is used to measure the centrality of the node in the network.

Together these three theories make up the structural hole theory, where weak-ties are one of the sources for generating new information generating, exclusive exchange partner provide the power for control, and Betweenness provides the substantial network control for the individual.

Burt (2009) argues that the contacts of the individual can determine the opportunities to gain benefit for himself. The Structural Hole is believed as one of the social network structures to gain competitive advantages for an individual so as to earn benefit from it. Burt (2009) also point out the competitive advantages of Structural Hole includes information benefit and control benefit, where the information benefit is access to exclusive information in the group, earlier access time, and having information shared in the group. Moreover, there exists deep Structural Holes and small Structural Holes, which depends on the number of disconnected parties connecting on the hole (Burt, 2009). Of course, the larger the number, the deeper the hole it is, and it is believed the larger the competitive advantages. In the case of three nodes are connecting to the Structural Hole, two may has a direct connection with each other. This structural hole is connecting two disconnect parties but it is connecting three nodes, which this is considered as less efficient or a smaller structural hole. In the case of three nodes again are connecting to the Structural Hole, if two have an indirect connection with their neighbours, this is called structural equivalence. This is also considered as less efficient or a smaller hole, but compared to the previous case, it is deeper.

From the trust establishment point of view, the individual in the Structural Hole position is taking advantage to control the information flow among the disconnected networks, which this can provide the individual opportunities to act unethically toward all other parties without fear of the other person learning of his act. Being positioned in Structural Hole, the individual becomes a gatekeeper of information that might otherwise be transmitted between contacts. In addition to the opportunity to withhold critical information, the individual which spans structural hole many also has a great opportunity to distort or terminate information flows that pass through between each party. Moreover, from the network attack point of view, this individual is usually more attractive to attack because of its bottleneck position. Take the network routing as an example, it means that there is no alternative route to be selected and it cannot avoid a malicious attack if the individual is compromised or even itself has misbehaviour.

The study (Burt, 2009) suggested that Structural Hole is good in the individual task as the actor can have easier access to the exclusive information from different parties so as to better deliver the task. While it also presents the opportunities for misconduct because when an individual is spanning a gap between otherwise unrelated contacts, this individual is positioned to act unethically toward another individual or group without fear of the other person learning of the act. It is a double-edged sword. This is the same with the Simmelian ties. There is long debate on which social structure is better (Latora et al., 2013), but Burt (Burt, 2009) argued that this depends on the context. In a computer network, from the network availability point of view, more redundant routes are preferred as this ensures most of the networks stay connected while under attacks or network failures. In such cases, the Structural Hole is believed can cause greater damage to the network as it is acting as the gateway to connect different clusters in the network, once it malfunctions or becomes malicious, it will disconnect these clusters. In another case, if a node is performing flooding attacks in the network where the affected area need to be controlled in a smaller area. In such case, the Structural Hole can do a better job as the nodes in the Structural Hole positions have the power to control the data flow.

Understanding the two edge of the Structural Hole is necessary, so we know when Structural Hole can improve the trustworthiness of network and when it will be an obstacle to it. Armed with this knowledge, nodes in the network establishing the connection or designing a network topology can be done in a smarter way.

As can be seen in Figure 2.4, proposed by Engle (1999), and the study by Latora et al. (2013) agree on that the Simmelian Tie can be good in some case like an interdependent task which requires teamwork cooperation (H4b). It can also be negative as it enforces the group behaviour that limits innovation for the individual (H2). On the other hand, the

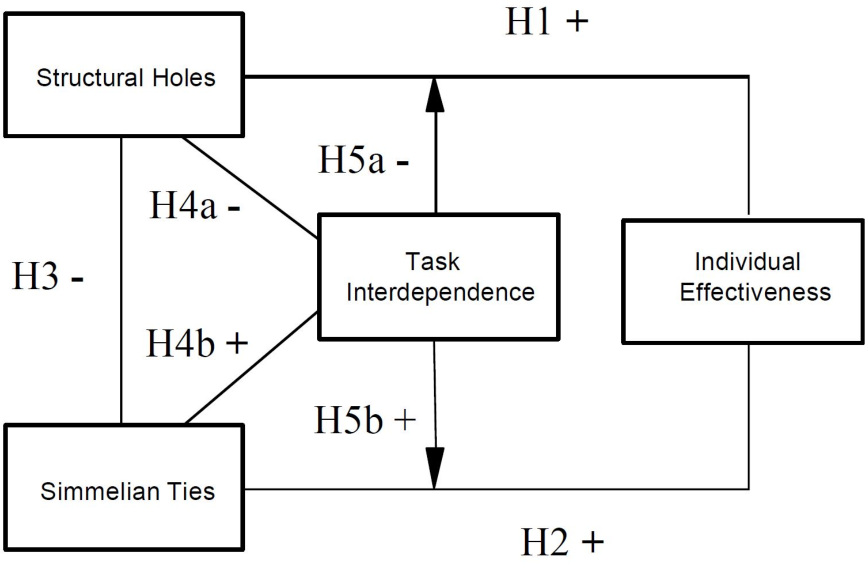


Figure 2.4: Simmelian Ties vs. Structural Hole Framework (*Source, Engle, 1999*)

Structural Hole is the same. It is good for individual tasks as the actor can easily access the exclusive information from different parties to better deliver the task (H1). While it also presents opportunities for misconduct because when an individual spanning a gap between otherwise unrelated contacts, this individual is positioned to act unethically toward another individual or group without fear of the other person learning of the act (H4a). Engle (1999) has further suggested the Structural Hole has a negative impact on the positive impact which is the result of the Simmelian Ties (H3).

Simmelian Tie and Structural Hole are different structures that affect the individual behaviours at the position, the formation of the trust is how the status of an individual or party affects the formation of the social network structure. There is a concept to describe this co-evolution which is the adaptive network and it will be introduced in the next section.

2.5 Adaptive Network

'Adaptive network' is a combination of two concepts which are dynamics *On* networks and dynamics *Of* networks (Gross & Blasius, 2008; Gross & Sayama, 2009; Sayama et al., 2013; McCabe, Watson, Prichard & Hall, 2011). Dynamics on networks is the status change on network entities, and dynamics of networks is the change of network topology. These two concept affect each other which it calls co-evolution. This concept is hardly new as it is happens in our everyday life, in addition there a significant quantity of research on the adaptive network, but such research either focuses on the network entities status transition or network topology design. Gross and Blasius (2008) pointed out that in recent years the focus of research in adaptive network employ conceptual model and found it was based on the simple local rule that networks can self-organize robustly toward phase transitions. In one paper (McCabe et al., 2011) has suggested, in the Internet web, states (behaviours) affect how topologies (structure) changes and topologies affect how states change as well. For example, the users change their behaviours in that they do more online shopping. This could cause more online shops to be established. In another way, the search results on positioning could affect user's preference in accessing content; this is topologies affecting the state. In such a case, the broken link or the website with faulty information will be removed as a result of user complaints and helpful websites will become larger because more users are visiting. In the following sections, the existing research related to adaptive networks will be reviewed.

2.5.1 Network Modelling

There is a lot of research in the adaptive network which tries to model the network evolution in the human social network such as telephone communication, co-authors network (Gross & Blasius, 2008), epidemiological network the Susceptible–Infected–Susceptible

(SIS) model (Sayama et al., 2013), etc. Moreover, the study by Sayama et al. (2013) proposed a universal model call Generative Network Automata (GNA) to describe three stages in network evolution which are ’extraction’, ’production’, and ’embedding’. ’Extraction’ is the process decide which part of the network is going to change, ’production’ is decided how this part is going to change, and finally, ’embedding’ is embedding the new part into the network. There are two well-known existing models which are Small World and Scale-Free Networks. Small World effect was introduced by Watts and Strogatz (1998). It is defined as any nodes in the network which can reach any other nodes within ’k’ hops. In the real world, sometimes you will be surprised that a person very far away from you could be your friend’s friend. In such case, he is actually very close to you; this is the small world effect. The Scale Free network was introduced by Barabási and Albert (1999) which features a power-law degree distribution. Many studies describe this scale-free network as “robust, yet fragile” (Zhao, Kumar & Yen, 2011; X. F. Wang & Chen, 2003; Onnela et al., 2007). A Free scale network is robust against random malicious attacks, but fragile while malicious parties attack its central hops. For example, random attacks need to disable 10 nodes to achieve disconnection of the network, but this same damage can be achieved by disabling one central node as well. In the next section, a literature review on how to mathematically model the network is given.

2.5.2 Link Weight and Node Strength

A directed or undirected and weight or unweighted graph $G(V, E)$ is usually used to represent a complex network such as studies (H. Wang & Van Mieghem, 2008; Sydney, Scoglio & Gruenbacher, 2013). V is the node or entity set in the network and E is the edge or link set which is connecting the Vs. A directed graph G means the Edge has direction like the link is connected from A to B but not necessarily vice versa. An undirected graph means the link in the network is reciprocated. Using the same idea, a weighted graph mean the links have weight in the network, such as the connection between A and B is the motorway, but between B and C is only an alley (M. S. Granovetter, 1973). In such a case, the link weight between A and B is larger. The strength of a node is the sum of link weights to all its neighbours, and a high strength node is normally attracts more nodes to connect, which is "rich get richer" concept (Xie, Wang & Wang, 2007). On the other hand, the removal of the link from the weak link can increase the speed for the network to become fragmented (Toivonen et al., 2007). One study (Onnela et al., 2007) has the similar finding in that in communication networks the removal of the weak ties results in a phase transition-like network collapse, although the removal of strong ties has little impact on the network's overall integrity. In the distributed trust management system, the reputation of the particular node can be considered as the node strength, and the link weight can be considered as the direct trust from the trustor to the trustee. A node with a high reputation around the neighbours, of course, has more attraction to the trustor. After reviewing the graph G, in the next section, how rust can be adopted into adaptive network concept will be discussed.

2.5.3 Trust and Topologies

The behaviours of a particular node in the network can be considered as the state of the node, once this node's behaviour has changed this also means the node's state has changed. In the trust routing algorithm, if a node becomes malicious, the state changes, the algorithm will detect the change has happened then disconnect this node and look for an alternative route to the destination. This is how the state change in network nodes can affect a change of network topology. Vice versa, a network topology change should affect the state of the node as well. For example, there are two routes between node A and B, so A can select either route to connect to B. However, if one of the routes is disconnected, that make node A has no choice but select the remaining route to stay connected with B. This can cause the remaining route to become overloaded as all traffic now travel through this route, and some of the nodes might become selfish to preserve energy, so as to stay up longer in the network. This example shows that trustworthy nodes can become untrustworthy under pressure due to a change of network topology.

Moreover, the change of network topology can affect the trust routing algorithm efficiency (Oren, Griffiths & Luck, 2013). Such studies as Mahalle et al. (2013) found out that the trust converging speed is much faster in a Small World network than in a Lattice network. This is because the longest distance between two nodes in a Small World network is not bigger than k hops. In such cases, a lattice network, it obviously has a much longer distance that makes the converging speed decrease compared to a Small World network. The trust routing algorithm can be very inefficient in some network topologies (Oren et al., 2013), such as Scale Free network and Star network. This is because such a network normally does not have an alternative route between any two nodes in the network. Moreover, the algorithm Dynamic Trust Elective Geo Routing (DTEGR) (Xiang et al., 2012) can adjust the trust threshold value depending on the number of neighbours in the trust forwarding list. When the trust neighbour list is low the threshold value will be lower, and it will adjust back when the list is increasing back to a sufficient amount until the threshold returns to the preset value again. One study (Khan, Midi, Khan & Bertino, 2015) has even gone further. When the node degree is higher, the threshold should be higher and vice versa; this is similar to DTEGR algorithm. They also consider the neighbour's degree; when it is higher, the threshold should be lower as this neighbour has become more important in the forwarding role, and vice versa. Though the algorithms in these two studies can adjust the trust threshold value based on the local topology, it did not address how the network topology can affect the node behaviour change. In the next section, a review of routing algorithm involving network topology metrics will be given.

2.5.4 Clustering

There are some studies of the routing algorithm which try to change the network topology, so as to improve the efficiency of the routing by clustering the network (S. K. Singh, Singh & Singh, 2010; Liu, Yu, Cheng & Wang, 2011; Younis & Fahmy, 2004; Heinzelman, Chandrakasan & Balakrishnan, 2000; Kour & Sharma, 2010). The most typical cluster-based routing protocol is LEACH (Heinzelman et al., 2000), which randomly select a cluster head within the network with random opportunity of every node. All the cluster members use the cluster head as the gateway to communicate inside or outside. Then all cluster heads connect directly to the base station, in such a case to simplify the routing table, reduce routing control messages, and achieve energy efficiency. One study (Younis & Fahmy, 2004) has proposed a Hybhird, Energy-Efficient Distributed clustering (HEED) algorithm based on LEACH; it includes the residue energy level as the condition to select a cluster head as the cluster head in charge for all cluster data transfer. The energy metric is to ensure the node has enough energy to act as a head cluster. Another study (Kour & Sharma, 2010) proposed H-HEED to enhance the HEED algorithm by making the cluster multi-level. In regard to the trust routing algorithm, paper (Liu et al., 2011) proposed a trust-based routing protocol but with more focus on the network topologies control. The nodes on the network are self-organized to form many 3-nodes triangle trust clusters, so the network will become a two level hierarchy in order to reduce network overhead and network delay. The simulation they ran verified it, but this algorithm requires a dense network topology so as to provide enough nodes to form the triangle clusters.

Back to the studies in sociology, the study (Engle, 1999) has defined five hypotheses regarding the relationship among Simmelian Tie, Structural Hole, Individual task, and Group task which are mentioned in Figure 2.4 on page 40. These hypotheses are made on the social network, will it apply to computing network as well?

The next question is how to identify or characterise Simmelian Ties and Structural Hole in the network. In the next section, we discuss the network robustness, and the measurement of the robustness of networks, so they can be compared and improved to become more resistant to the malicious attack, and tolerant of any nodes or links failures. More importantly, these measurement metrics can characterise the Simmelian Ties and the Structural Hole in the network. In such a case, the network topology can be improved to increase the trust algorithm performance by tuning these metrics.

2.5.5 Network Robustness

Network are ubiquitous in our world; we have social networks, the Internet network, traffic networks etc. Sometimes a network failure can cause significant damage to individuals, to companies, and to society.A large scale power outage can cause huge financial loss. Network Robustness is crucial to prevent these situations happening. However, what is Network Robustness? A study (Ellens & Kooij, 2013) defined the term as "the ability of a network to continue performing well when it is subject to failures or attacks". Study (H. Wang, Van Mieghem, TU Delft: Electrical Engineering, Mathematics and Computer Science: Telecommunications & TU Delft, Delft University of Technology, 2009) suggested as a network is more robust if the service on the network performs better, where the performance of the service is assessed when the network is either (a) in a conventional state or (b) under perturbations, e.g. failures, virus spreading etc. Robustness has a different definition in different scenarios.

2.5.6 Network Robustness Quantification (Metrics)

To define network robustness systematically, we need to quantify the robustness. One study (Ellens & Kooij, 2013) has listed four classical graph metric categories to measure robustness, they are 'connectivity', 'distance', 'Betweenness', and 'clustering'.

’Connectivity’ is calculated as the percentage of connecting pairs in the network. A fully connected network has a connectivity of 1, while a completely disconnected network has a connectivity of 0. There are another two metrics under connectivity, which are vertex connectivity and edge connectivity. The vertex connectivity is the number of nodes needed to removed so as to disconnect the network. A similar idea, edge connectivity is the number of edges needed to be disconnected so as to disconnect the network. Distance has the average hop count of all node pair connections, and the longest hop count is the diameter of the network. Betweenness is calculated as the number of the shortest path between any node pair route through node i. Clustering is using the clustering coefficient to measure the percentage of the connected triangle cluster in all connected triples (3 nodes).

Moreover, the spectral graph measure was introduced in study by Ellens and Kooij (2013). One of popular metric is algebraic connectivity. It uses the Laplacian matrix to represent the network graph, and the second smallest eigen value is the algebraic connectivity. Effective Resistance is another spectral metric to measure the network robustness. It uses Kirchhoff’s circuit laws to calculate the resistance between two vertexes, and the sum of all node pairs’ resistance is the effective resistance. The smaller the value the more robust the network is believed to be.

A study (H. Wang et al., 2009) has also categorised metrics into Distance class, Connection class, and Spectral class. It bears similarity with other studies (Ellens & Kooij, 2013) but more metrics have been listed open in each class, it also put the ’Betweenness’ is considered as part of the distance class, and ’clustering’ part of the connection class. These additional metrics are similar to the given metrics mentioned above. ’Connectivity’ is focused on the possibility of the alternative route in the network, and distance is focused on the hop-count to travel to reach the destination, as a shorter distance is believed to have less chance of encountering attack or failures. However, alternative routes could solve this problem, but distance metrics did not consider this.

The clustering is first designed to describe friends' of my friend are my friends in the social network. In addition, it can also measure the alternative routes in a network, as the more clustering that exists in a network, the more alternative routes that are likely to exist in the network. 'Betweenness' is the measure of centrality of a particular node in the network. For the algebraic connectivity, studies (Ellens & Kooij, 2013; H. Wang et al., 2009) suggested that if the second smallest eigenvalue is multiple, the algebraic connectivity will not change with additional links added, and study (Ellens & Kooij, 2013) has given an example to prove this. 'Effective resistance' does not have such problem so that it can be more a suitable measure of network robustness by the focus on alternative routes as study (Ellens & Kooij, 2013) declared.

There is no single metric here that can fully measure the robustness of the network, as robustness has different definitions in different scenarios. In the next section, the existing remedy methods for the network topology will be reviewed.

2.5.7 Discover Network Weakness and Improve Network Robustness

After the robustness of the network has been measured, if the network is not robust, redesign the network is normally too costly. Thus, how to improve the network robustness is another question here. Rewiring is the current solution to improve the network robustness. Several papers (H. Wang & Van Mieghem, 2008; Sydney et al., 2013) investigating the rewiring issue ask the questions "Where to add a link in the network will result in the most increase in algebraic connectivity" and "Where to remove a link from the network which will result in the least decrease of algebraic connectivity". One paper (Sydney et al., 2013) concluded the answer as to decreasing algebraic connectivity the least, we should remove an edge that connects two strongly connected vertices. Conversely, to increase algebraic connectivity the most, we should insert an edge between two weakly connected vertices. Moreover, it also discovered that beyond a certain rewiring threshold, which can range from 8% to 20% for the graphs presented, algebraic connectivity is constant. However, another study (H. Wang & Van Mieghem, 2008) also stated that topology dependence is such that no certain link can be added to improve algebraic connectivity for every topology.

Accordingly, it can be concluded that there is no universal way to calculate and find out where to add a link which can most improve network robustness at any moment. As mentioned before, different scenarios require different metrics to measure the robustness, and in such case require a different approach to improve network robustness. Both these studies (H. Wang & Van Mieghem, 2008; Sydney et al., 2013) use algebraic connectivity to measure robustness. One studies (H. Wang et al., 2009) suggested different topology and scenarios have different requirements and robustness metrics; in such cases, the same optimisation strategy work might work for on topology but most likely not for many others.

2.6 Summary

In this literature review chapter, we have first reviewed the traditional security approaches that act as 'hard security'. Hard security is not efficient to against the unknown or unexpected behaviours (malicious or selfish). Thus, soft security (trust) is introduced to tackle these soft security threats. The existing works on trust in the computer network is focused on trust modelling, so as to detect or even predict any malicious or unexpected behaviours in the network. The routing algorithm can avoid such behaviours after they have been identified as not trustworthy. The trust evaluation normally consists of three stages, which are 'data aggregation', 'data syncretization', 'decision making'. The data aggregation collects trust evidence such as the metric packet forwarding ratio, etc. Depending on the objectives of the task, the metrics can be different. Data syncretization is put all the metrics into one single value, so it can be compared and measured. Decision making is based on the final trust evaluation results to determine the node is trustworthy or not.

As trust is derived from Sociology, the review of trust in Sociology is given as well. From the definition of trust, we say trust is both a risk and independent. 'Risk' means 'trust' is a belief in someone or something will behave as the trustor expected without any assurance. Trust requires at least two entities to be established. From the review of formation of trust, we found that the start of trust relationship can be expressive or instrumental. The establishment of the trust requires interactions, and base on previous interactions, the trust will be established between entities. In Sociology, the social network structure is one of the important factors that affecting the individual's behaviours and performance. The two typical structures are Simmelian Ties and Structural Hole. It is believed that the Simmelian Ties have a positive impact on the dependent tasks and a negative impact on the independent tasks. Structural Hole has a negative impact on the dependent tasks, a positive impact on the independent tasks, and a negative impact on the positive impact from Simmelian Ties. Trust establishment can be considered as the change of a node in the network affecting a change of network topology. From other perspective, the change in network structures can also affect the node behaviour change; this is the concept of the adaptive network.

From the literature review on the three different research areas, a research gap has been identified, which is the interplay of trust behaviours and the underlying topological connectivity as shown in Figure 1.1 on page 5. To investigate the relationship between network topology and trust behaviours, we need to know how to measure the network topology and trust. In the next chapter, based on the assumption on the Simmelian Ties and Structural Hole in the computer network routing, we will introduce evaluation methods for the Simmelian Ties and Structural Hole, then conduct simulation studies in different network topologies for validation.

Chapter 3

Network Structures and Node Trustworthiness

3.1 Introduction

In the last chapter, the literature review on Figure 1.1 on page 5 has revealed the relationship between Simmelian Ties, Structural Hole, interdependent tasks, and independent tasks in the Sociology. The Simmelian Ties have a positive impact on the interdependent tasks and a negative impact on the independent tasks. Structural Hole has a negative impact on the interdependent tasks and a positive impact on the independent tasks. Moreover, the Structural Hole has a negative impact on the positive impact created by Simmelian Ties. These statements have well been validated in the study (Engle, 1999) in Sociology. Would these statements apply to the P2P environment as well?

The literature review has revealed the state of art of trust in P2P routing has been focussed on the trust modelling between the nodes in the network. However, there is little existing work on how the underlying topology can affect the individual trustworthiness in the network. In this chapter, we are going to explore the relationship between the Simmelian Ties, Structural Hole, and node trustworthiness in routing.

First of all, based on the P2P network routing scenarios, the assumptions of the relationship between the Simmelian Ties, Structural Hole, and node trustworthiness in routing will be made in the following section. After the assumptions have been made, the validation is carried out with simulation studies. To do that, the Simmelian Ties and Structural Hole need to be evaluated and quantified in the network, so they can be compared within different networks, and the assumptions are able to be verified. Secondly, the equations and methods are used in evaluating these two network structures in sociology and complex network studies will be introduced. Thirdly, after the measurement methods for the Simmelian Ties and Structural Hole are confirmed, the simulation study is carried out. Fourthly, after the assumptions have been validated, the dilemma of these two structures will be discussed as well. The assumptions only consider one side of the Simmelian Ties and Structural Hole, the exploration of their other sides is necessary. Finally, the summary is given at the end of this chapter.

3.2 Simmelian Ties, Structural Hole, and Trustworthiness in Routing in P2P Environment

In the distributed P2P network environment, nodes are self-organized in the network. In other words, nodes in the P2P network can decide their own action. In such a case, we can treat each node in the P2P network like a human entity in the social network. There are studies defining the relationship between Simmelian Ties, Structural Hole, interdependent task, and independent tasks in the sociology. If the nodes in the network can be considered as human entities in the social network, this relationship framework should be able to be applied to the computer network in the P2P environment as well.

Routing in the distributed networks can be considered as an interdependent task. Unlike the traditional centralised management network, all the traffic and nodes in the network are centrally managed so that the routing task is like an independent job. In the distributed networks, all the nodes are self-organized, the nodes in the network need to work with their neighbours so as to get their packets sent to the desired destination. The Simmelian Ties and Structural Hole relationship framework which is proposed by Engle (1999) has defined these two social structures in two different scenarios, which are interdependent tasks and independent task. As the routing in the distributed P2P networks is considered as the interdependent tasks, the assumptions on the relationship between Simmelian Ties, Structural Hole, and node trustworthiness in routing in distributed P2P network environment can use the interdependent task scenario in Engle's framework. The assumptions are shown in Figure 3.1.

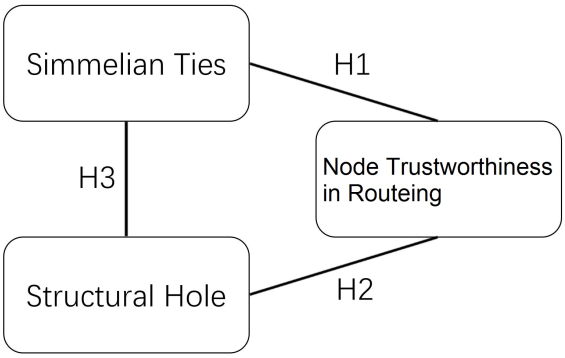


Figure 3.1: Hypotheses between Simmelian Ties, Structural Hole, and Node Trustworthiness in Routing

H1. Simmelian Ties characterised network structure has a positive impact on node trustworthiness in routing.

H2. Structural Hole characterised network structure has a negative impact on node trustworthiness in routing.

H3. Structural Hole characterised network structure has a negative impact on the positive impact from Simmelian Ties characterised network structure.

The first assumption is that the Simmelian Ties have a positive impact on the node trustworthiness in routing. As mentioned in the literature review chapter, the Simmelian Tie discourages misbehaviour in the network by introducing a third party to monitor each party in the triangle structure. This strongly fosters a normative environment against opportunist attacks, and engenders mutual trust, reciprocity norms, and a shared identity. It facilitates the collaborative efforts by making the actors more willing to exchange information. For the trust-based routing algorithm in the distributed P2P network environment, the Simmelian Ties provides a third party that enables the reputation and indirect trust system for the algorithm, as these reputations and indirect trust information all come from the third parties. In such cases, with the help of reputations and indirect trust information from third parties, the trust algorithm can detect the malicious or selfish behaviours in a more efficient manner. Efficient here means faster and more accurate. Additionally, when the nodes are monitoring each other, they are less likely to be malicious or selfish, as they will be isolated by doing so. Finally, the Simmelian Ties also provides the redundant routes once the malicious or selfish behaviours are detected by the algorithm, there is at least an alternative route for the algorithm to select, so the malicious or selfish behaviours can be avoided immediately. Thus, we believe the Simmelian Ties has a positive effect on the node trustworthiness in routing.

The second assumption is the Structural Hole has a negative impact on the node trustworthiness in routing. The person at the Structural Hole position normally acts as the agent, broker, or gateway between two or more parties. These people have the position advantage to control the information flows among the networks. It is a very critical position such that once it is broken, the whole network is disconnected. As there is no third party monitoring at this position, the people in this position have an opportunity to act unethically toward all other parties without fear of the other person learning of his act. In the distributed P2P network environment, when the node at the Structural Hole position behaves maliciously or selfishly, as there is no third party for the trust-based routing algorithm reputation or indirect trust evaluation, the detection of malicious or unexpected behaviour will be slower and less efficient. Moreover, even though the algorithm finally detected the unwanted behaviours in the network, there is no alternative route as a detour. Thus, these unwanted behaviours cannot be avoided. In such cases, this leaves the trust-based routing algorithm no choice, but continue to trust and deal with this malicious or selfish node at the Structural Hole position. This makes the trust-based routing algorithm become infeasible with such network structures. Therefore, the Structural Hole position is also more attractive to the attacker due to its gateway nature. Additionally, the person at the Structural Hole position can easily become malicious or selfish as they can act maliciously without fear of other parties learning of their act. In the distributed P2P network environment, the nodes at the Structural Hole position normally are under a heavy task load as they are acting as gateway most of the time. In such case, their energy consumption will be higher compared to the other nodes. Thus, they are more likely to act selfishly so as to preserve the energy for a longer stay in the network. Because of these reasons, we believe the Structural Hole structures have a negative impact on the node trustworthiness in routing.

The last assumption is made between Simmelian Ties and Structural Hole, in which the Structural Hole has a negative impact on the positive impact from the Simmelian Ties. Simmelian Ties benefit the node trustworthiness in routing as they provide a redundant route and third parties in their local area (e.g. one hop distance area). Structural Hole is believed to hinder the node trustworthiness in routing as they are the gateway connecting different clusters in the network. Once they are disconnected, these clusters will be disconnected as well. Moreover, as the nodes at the Structural Hole position are normally under heavy tasks load, they are more likely to become selfish. The Structural Hole hindrance is at the global level in the network, whereas the Simmelian Ties are at the local level. Once the attack or failure happens on the nodes at the Structural Hole position, the whole network will be disconnected, even though the Simmelian Ties are still benefiting the node trustworthiness in routing in their belonged clusters. Thus, we believed that the Structural Hole has a negative impact on the positive impact of the Simmelian Ties.

This section has given three assumptions on the Simmelian Ties, Structural Hole, and node trustworthiness in routing in the distributed P2P network environment. These assumptions need to be validated through a comparison of network evaluation and simulation studies. Therefore, in the next two sections, the existing evaluation methods for Simmelian Ties and Structural Hole will be introduced and discussed.

3.3 Simmelian Ties and Its Measurement Methods

Simmelian Tie is defined as triad ties or closed structure ties that are ties embedded in the cliques (Krackhardt, 1999) as shown in Figure 2.3 on page 36. For an evaluation of the Simmelian Ties, the key point is how to calculate the number of the Ties or triangles in the network. In the complex network and Sociology studies such as (Ellens & Kooij, 2013; Latora et al., 2013), the clustering coefficient is considered as the most commonly used metric for the evaluation of Simmelian Ties in a network. First of all, for any network evaluation, the network itself needs a mathematical equation to model it, which it has been mentioned in the last chapter in the complex network section. We first assume the networks are undirected and unweighted. In such a case, the network as a graph can be represented as shown in Equation 3.1:

$$G(V,E) \tag{3.1}$$

In Equation 3.2 on the next page, V is the set of vertices (nodes) and E is the set of the edges (links), which are connecting the nodes in the network G. v_i is the node i in the network, and $e_{i,j}$ is the link connecting nodes i and j. When there is a connection between node i and j, $e_{i,j}$ = 1, otherwise, $e_{i,j}$ = 0.

$$\begin{aligned} V &\in (v_1, v_2, v_3 \ldots v_i, v_j) \\ E &\in (e_{1,2}, e_{2,3}, e_{1,3} \ldots e_{i,j}) \end{aligned} \tag{3.2}$$

The idea of the clustering coefficient is to calculate the percentage of the Simmelian triangles, which are connecting to a particular node in the network out of the maximum possible number that particular node can connect with its existing neighbours. First of all, the maximum possible number of the Simmelian triangles, which are connecting to a particular node, can be calculated as shown in Equation 3.3. The variable n_i is the node degree or number of neighbours for node i. As a node want to form a triangle, it requires at least two neighbours. Equation 3.3 can discover all the possible pairs of neighbours of node i's in order to calculate the maximum possible triangle, which node i can have.

$$\frac{n_i \times (n_i - 1)}{2} \tag{3.3}$$

The next step is to calculate the actual Simmelian triangles which are connecting to the node i, which is shown in Equation 3.4.

$$\sum_{j=1, j \neq i}^{n_i} \sum_{k=j+1, k \neq i}^{n_i} e_{ij} e_{ik} e_{jk} \tag{3.4}$$

The variable e_{ij} is the link connecting node i and j, in such logic, e_{ik} is the link connecting node i and k, and e_{jk} is the link connecting node j and k. As mentioned before, if the link between two nodes exists, then $e = 1$, otherwise, $e = 0$. Therefore, when all links e_{ij}, e_{ik}, and e_{jk} exist, then this will count as 1 triangle, otherwise, it is counted as 0.

Finally, the clustering coefficient equation is the actual Simmelian triangles number divided by maximum possible Simmelian triangles number, which is shown in Equation 3.5. c_i is the clustering coefficient for node i in the network. Moreover, as mentioned previously, if node i needs to form a Simmelian triangle, it requires at least two neighbours, which means n_i needs to greater than 2. At last, as nodes j and k are defined as neighbours of node i, so the connection links e_{ij} and e_{ik} are always exist and equal to 1. In such case, the variable e_{ij} and e_{ik} can be removed from the Equation 3.4 on the preceding page. The link e_{jk} can determine if this is a Simmelian triangle or not. Otherwise, $c_i = 0$.

$$c_i = \begin{cases} \frac{2}{n_i \times (n_i - 1)} \times \sum_{j=1, j \neq i}^{n_i} \sum_{k=j+1, k \neq i}^{n_i} e_{jk} & n_i \geq 2 \\ 0 & n_i < 2 \end{cases} \tag{3.5}$$

As c_i is only the clustering coefficient for node i in the network. For the evaluation of the whole network, we can sum up every node's clustering coefficient in the network, and divide by its total number to have the average clustering coefficient, which is shown in Equation 3.6.

$$AC = \frac{\sum_{i=1}^{n} c_i}{n} \tag{3.6}$$

In the following, this section has also set up two sample network topologies for demonstration of the calculation for clustering coefficient and their comparison. The two network topologies are shown in Figures 3.2 and 3.3.

In both networks, they are composed of five nodes and seven connection links. Take the node 3 in both network as examples, node 3 in the network I has four neighbours and three neighbours in the network II. Thus, according to Equation 3.3 on the previous page, $n_3 = 4$ in the network I and $n_3 = 3$ in the network II. For the maximum possible connected Simmelian triangles on node 3 in the network I would be $(4 \times (4 - 1)) \div 2 = 6$, and in the network II would be $(3 \times (3 - 1)) \div 2 = 3$. Then the next step is to calculate the

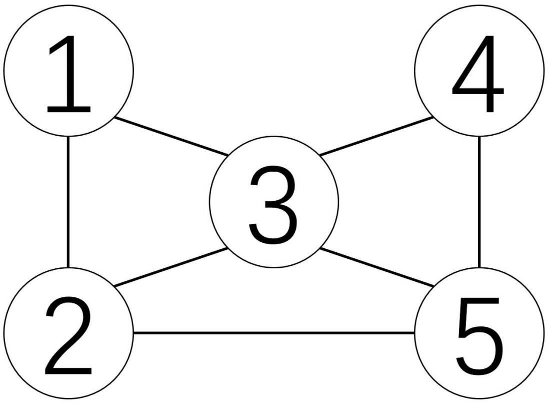


Figure 3.2: Sample Topology, Network I

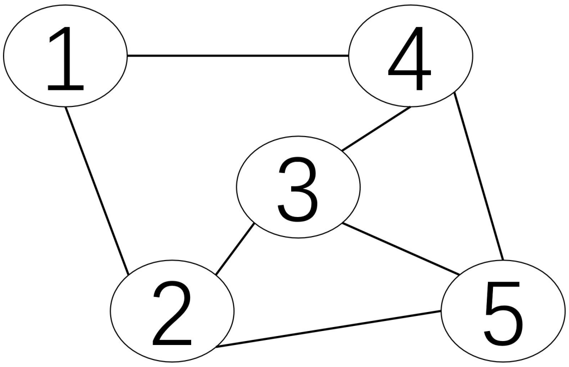


Figure 3.3: Sample Topology, Network II

actual Simmelian triangles that are connected with node 3 in both networks. In network I, nodes 1, 2, 4, and 5 are the neighbours of node 3. Among these neighbours, there are three pairs of nodes are connected, which are nodes 1 and 2, 2 and 5, and 4 and 5. Thus, the number of the actual connected Simmelian triangles on node 3 in the network I is 3. Using the same logic, the number of the actual connected Simmelian triangles on node 3 in the network II is 2. Therefore, the clustering coefficient for node 3 in the network I is $3 \div 6 = 0.5$ and in the network II is $2 \div 3 = 0.67$. This means node 3 in the network II has higher a clustering coefficient. In other words, node 3 is in a better position in network II than its position in network I. The rest of clustering coefficient results for network I and II are shown in Table 3.1. The average clustering coefficient is higher for network I, even though the node 3 in network II has a higher clustering coefficient.

This means the network I topology can be more beneficial to the node trustworthiness compared to the network II. In the next section, the evaluation methods for Structural Hole, which has the opposite structure of Simmelian Ties, will be discussed.

Table 3.1: Clustering Coefficient for Networks I and II

Nodes	**1**	**2**	**3**	**4**	**5**	**Average Clustering Coefficient**
Network I	1	0.67	0.5	1	0.67	0.768
Network II	0	0.33	0.67	0.33	0.67	0.4

3.4 Structural Hole and Its Measurement Methods

The Structural Hole (Burt, 2009) is an 'actor' connection between two or more actors or parties who are not related or connected. It is the opposite structure of the closed structure Simmelian tie in that it is an open structure tie. It is shown in Figure 2.3 on page 36. In such a case, this actor normally acts as a broker or gatekeeper, which has positioning advantage to control the information flows among the networks. It plays a very critical role in that once it is broken then the whole network is disconnected.

Since Structural Hole is the opposite structure of the Simmelian Ties, the clustering coefficient is used to measure the redundant and Simmelian triangles in the network. The metrics used to measure the Structural Hole should be opposite as well. The effective size is the metric commonly used in complex networks and sociology for the Structural Hole measurement (Ellens & Kooij, 2013; Latora et al., 2013). In a complex network, Latora et al. (2013) defined the equations as shown in Equation 3.7.

$$ES_i = n_i - \frac{1}{n} \sum_{j=1, j \neq i}^{n_i} \sum_{k=j+1, k \neq i}^{n_i} e_{ij} e_{ik} e_{jk} \tag{3.7}$$

As can be seen in the equation that part of the equation is the same as the clustering coefficient in Equation 3.4. In this case, the clustering coefficient Equation 3.5 can be substituted into the Equation 3.7 to become the new equation as shown in Equation 3.8.

$$ES_i = n_i - \frac{n_i - 1}{2} \times c_i \tag{3.8}$$

In the following, another sample topology is given to demonstrate the calculation of the effective size. The topology is shown in Figure 3.4.

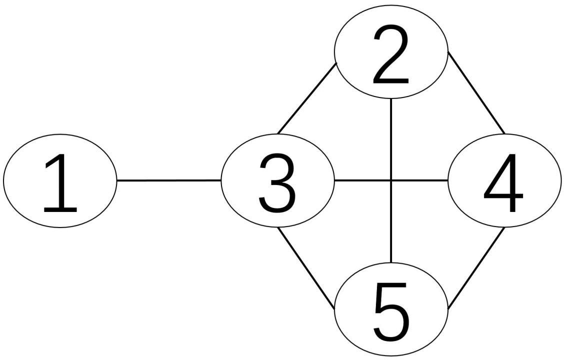


Figure 3.4: Sample Topology, Network III

In this topology, node 3 is chosen as the example to demonstrate the calculation of the effective size. Node 3 in the network III has four neighbours, which means $n_3 = 4$. The maximum number of possible connected Simmelian triangles on node 3 is 6. There are three pairs of neighbours to which node 3 is connected; they are nodes 2 and 4, 2 and 5, and 4 and 5. This means the number of actual connected Simmelian triangles on node 3 is 3. Therefore, the clustering coefficient of node 3 is $3 \div 6 = 0.5$. Finally, the effective size on node 3 is $4 - (4 - 1) \div 2 \times 0.5 = 3.25$. The effective sizes for the rest of the nodes are shown in Table 3.2. As can be seen in Table 3.2, in the network III, node 3 has the highest effective size. This can mean node 3 in network III is more likely at the Structural Hole position (and it is). However, in the network I, the node 3 also have the highest effective size, which is not at the Structural Hole position. The node 3 in both networks I and III are having four neighbours, which are the nodes with the largest number of neighbours connected. This is one of the main reason node 3 has the highest effective size in networks I and III. In other words, the effective size is not able to identify if there is any Structural Hole in the network. Instead, it is measuring the local (one hop distance) brokerage level. Brokerage level mean how many unconnected clusters are connecting to the target node in a one hop distance.

Table 3.2: Effective Size Evaluation for Networks I, II, and III

Nodes	**1**	**2**	**3**	**4**	**5**
Network I	1.5	2.33	3.25	1.5	2.33
Network II	2	2.67	2.33	2.67	2.33
Network III	1	2	3.25	2	2

Latora et al. (2013) have also proposed a metric call Simmelian Brokerage to evaluate the Structural Hole, as they think the effective size is not able to reflect the difference of the brokerage level in some scenarios. For example, for the two network topologies shown in Figure 3.5, node 3 in both networks has the same number of neighbours which is five, and the same number of the Simmelian triangles, which is three. In such a case, node 3 in both networks should have the same clustering coefficient and same effective size. However, if the node 3 is disabled on the left-hand side network, node 1 and 2 will be disconnected from node 4, 5, and 6. On the right-hand side network, only the node 4 will disconnect from 1, 2, 5, and 6. In such case, it can be seen that node 3 in the left-hand side network can cause a bigger damage. However, the effective size fails to reveal this.

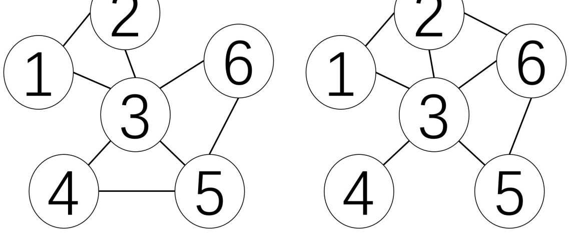


Figure 3.5: Sample Topology, Effective Size vs. Simmelian Brokerage

The Simmelian Brokerage is using a new way to calculate, which considers local efficiency instead of using the clustering coefficient. The local efficiency evaluates the reachability of the target node's neighbour after the target node is disabled. The equation is shown in Equation 3.9.

$$LE_i = \begin{cases} \frac{1}{n_i \times (n_i - 1)} \times \sum_{j=1,j \neq i}^{n_i} \sum_{k=j+1,k \neq i}^{n_i} \frac{1}{d_{jk}} & n_i \geq 2 \\ 0 & n_i < 2 \end{cases} \tag{3.9}$$

The variable d_{jk} is the distance in hop-count from node j to k. For example, the left-hand side network in Figure 3.5, for the local efficiency of node 3, the reachability of its neighbours needs to be calculated. After node 3 is disabled, from node 1 to 2 is a 1 hop distance, then $d_{1,2} = 1$. From node 1 to any of node 4, 5, and 6 are not reachable, then $1 \div d = 0$. In such logic, from node 4 to node 6 is a 2 hop distance, then $d_{4,6} = 2$, etc. Therefore, if all the neighbours are not able to reach each other anymore after node 3 is disabled, then $LE_3 = 0$. After the local efficiency is calculated, the final Simmelian Brokerage is shown in Equation 3.10.

$$SB_i = n_i - (n_i - 1) \times LE_i \tag{3.10}$$

The Equation 3.9 substituted into 3.10 will become Equation 3.11.

$$SB_i = \begin{cases} n_i - \frac{1}{n_i} \times \sum_{j=1,j \neq i}^{n_i} \sum_{k=j+1,k \neq i}^{n_i} \frac{1}{d_{jk}} & n_i \geq 2 \\ n_i & n_i < 2 \end{cases} \tag{3.11}$$

In Figure 3.5, we look back at node 3 in the two network topologies. The Simmelian Brokerage on the left-hand side network is $SB_3 = 5 - \frac{1}{5} \times (1 + 1 + 1 + \frac{1}{2}) = 2.9$, and the node 3 on the right-hand side is $SB_3 = 5 - \frac{1}{5} \times (1 + 1 + 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{3}) = 1$. From the results of the Simmelian Brokerage, it is suggested the node 3 on the right-hand side network has less brokerage level, which means the structure is better in network availability scenarios. Thus, it does prove that Simmelian Brokerage can measure the Structural

Hole more precisely than effective size does. However, the Simmelian Brokerage is same as the effective size, which is not able to locate the actual Structural Hole as well.

The effective size and Simmelian brokerage both evaluate the local one hop distance brokerage level. A node in the network might be at a Structural Hole position in one hop distance, but in two hops or further distance, those not related clusters might connect somewhere else at the further distance. Therefore, these two metrics are local evaluation metrics, and for the Structural Hole locator, a global evaluation metric is required.

A suitable metric for the Simmelian Ties evaluation in the network was introduced in the last section, which is the clustering coefficient. In this section, the effective size and Simmelian brokerage are also introduced as they are recommended to use for the evaluation of the local brokerage level, also for the Structural Hole as well. However, they are not capable to identify whether there is Structural Hole in the network or not.

In the next section, the extensive simulation studies will be provided to validate the three assumptions as shown in Figure 3.1 on page 53 with the evaluation metrics for Simmelian Ties and Structural Hole.

3.5 Simulation Studies

In this section, we have conducted extensive simulation studies to validate the three assumptions mentioned in the Figure 3.1 on page 53. They are that the Simmelian Ties characterised network structure has a positive impact on the node trustworthiness in routing; a Structural Hole characterised network structure has a negative impact on the node trustworthiness in routing; the Structural Hole characterised network structure has a negative impact on the positive impact from Simmelian Ties characterised network structure. The simulation studies are run on the J-Sim platform. J-Sim stands for JavaSim. It is a pure Java component-based, compositional simulation environment (Sobeih et al., 2006). Additionally, it is a dual-language simulation environment where the classes are coded in Java and using TCL interpreter to "glue" all the components. The benchmark trust-based routing algorithm Dynamic Trust Elective Geo Routing (DTEGR) (Xiang, Liu, Bai & Al-Anbuky, 2016) is a threshold based routing algorithm, which it will be explained in the next section.

3.5.1 Benchmark Routing Algorithm

The benchmark trust-based routing algorithm is our previously proposed algorithm Dynamic Trust Elective Geo Routing (DTEGR) (Xiang et al., 2016). We use this trust-based geographical routing model associated with trust threshold mechanism to validate our hypothesis mentioned in the Figure 3.1 on page 53. It can filter out all the neighbour nodes with trust value below the trust threshold, and select the neighbour nodes with the closest distance to the sink as next hop from the remaining qualified neighbours. The Figure 3.6 shows the decision flowchart for this algorithm.

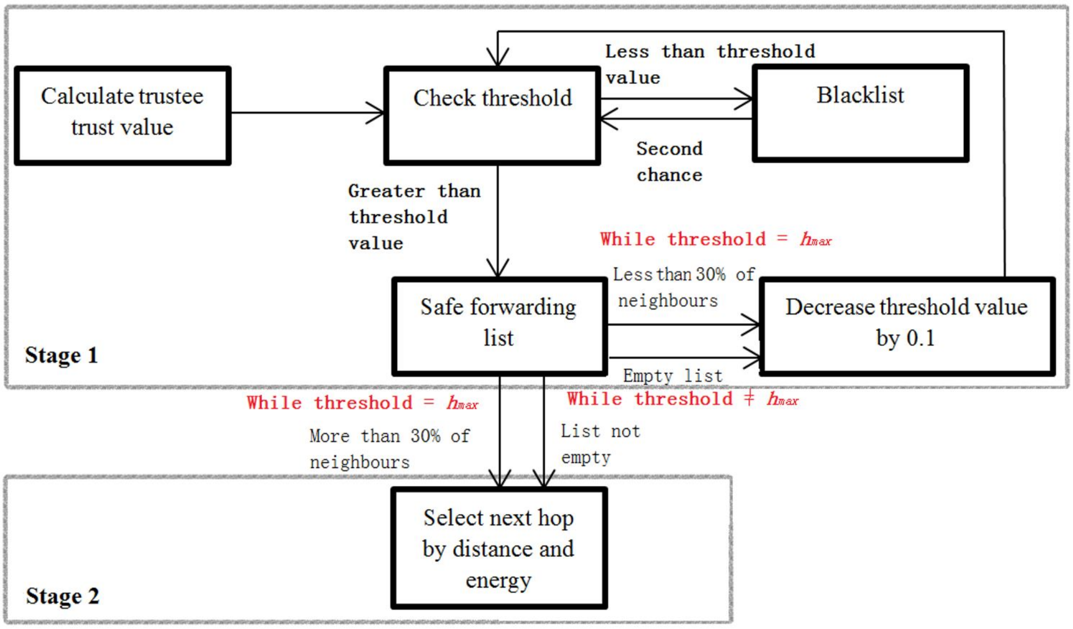

Calculate trustee trust value
Check threshold
Less than threshold value
Blacklist
Second chance
Greater than threshold value
While threshold = h_{max}
Less than 30% of neighbours
Safe forwarding list
Decrease threshold value by 0.1
Stage 1
Empty list
While threshold = h_{max}
More than 30% of neighbours
While threshold ≠ h_{max}
List not empty
Select next hop by distance and energy
Stage 2

Figure 3.6: DTEGR Algorithm Flow

It calculates trust value through direct trust and indirect trust (i.e., reputation). The direct trust is the trust evaluation performed by the trustor directly, while indirect trust is the direct trust value from other neighbours regarding the targeted node. These direct and indirect values can be combined to a final trust value of between 0 and 1 by using confidence factors (weighted factors). The direct trust metric equation is as shown in the Equation 3.12, where the s_i is the number of good behaviours for node i, and f_i is the number of unexpected behaviours.

$$t_{i,direct} = \frac{s_i}{s_i + f_i} \tag{3.12}$$

There is a chance some of the good nodes have the bad performance by accident, and the safe forwarding list size could be decreased over time. Moreover, sometimes the node can be surrounded by malicious neighbours, and can end up with an empty safe forwarding list. In such a case, the algorithm will make sure there are sufficient choices in the list and also give a second chance to the nodes, which have poor performance previously. When the threshold is not zero and the safe forwarding list size is empty, the algorithm will drop the threshold by 0.1 again until the list is not empty anymore.

$$t_{i,final} = w_{direct} \times t_{i,direct} + (1 - w_{direct}) \times t_{i,indirect} \tag{3.13}$$

The syncretization for direct trust and indirect is using the weight factor method; i.e., w_{direct} is the weight factor for the direct trust. As all the weight factors should sum up and equate to 1, so the indirect trust weight factor would be $(1 - w_{direct})$, which all have been shown in Equation 3.13. The next step for the DTEGR algorithm is to determine wehter the target node (node i) is trustworthy or not. To do so, a trust threshold is required. The algorithm will set up an initial trust threshold value which it is also the maximum trust threshold value; the equation is in Equation 3.14.

$$t_{threshold} = \frac{\sum_{i=1}^{n} t_{i,direct}}{n} - 0.1 \tag{3.14}$$

In Equation 3.14, n is the number of selected 'well-behaved' nodes and $t_{i,direct}$ is the direct trust value of node i. We minus the average direct trust value with 0.1. In such a case, this threshold can make sure all the 'well-behaved' nodes are in the safe forwarding list.

After the safe forwarding list is generated, the distance metric selects the neighbour with the shortest distance to the destination as the next hop from the list. The distance equation is shown as below:

$$d_i = \sqrt{(x_i - x_{sink})^2 + (y_i - y_{sink})^2} \tag{3.15}$$

In Equation 3.15, (x_i, y_i) and (x_{sink}, y_{sink}) are the longitude and latitude of the node i and destination. The algorithm selects the neighbour from the safe forwarding list with the shortest distance d_i as next hop to forward the packets.

In the next section, the performance metrics used for the simulation studies analysis and the simulation scenarios set up will be introduced.

3.5.2 Performance Metrics and Scenarios

In the simulation studies, we have selected two performance metrics to measure the performance of the trust-based routing algorithm DTEGR with different network topologies, and three topological metrics to evaluate the Simmelian Ties and Structural Hole in these networks, so they can be compared. They are 'packet loss ratio' and 'mean packet latency' for the DTEGR algorithm performance evaluation, and 'clustering coefficient', 'effective size', and 'Simmelian brokerage' for the topology measurement.

The packet loss ratio is the percentage of the packet loss over total packets transmitted in the transmission. This metric is effective to evaluate the sensitivity performance of detecting and avoiding malicious behaviour. The trust-based routing algorithm result of a greater packet loss indicates a worse performance in malicious node detection.

With the same trust-based routing algorithm as a benchmark, the simulation results can reveal how different network structures can affect the DTEGR algorithm performance or say node trustworthiness in routing. In such case, to make the routing algorithm simple, the direct trust evaluation metric for the DTEGR we only use a packet forwarded ratio.

The mean packet latency is the average packet delivery time from the source to the destination nodes. This metric is effective to evaluate the route finding capabilities of the algorithms in different network structures. The less packet latency indicates a shorter path is found.

The clustering coefficient, as mentioned in the previous section, is the metric used to evaluate the quantity of the Simmelian Tie in the network. The higher average clustering coefficient for the network means a higher density of the Simmelian Ties. In such cases, the networks can be compared with their average clustering coefficient, and determine if the DTEGR algorithm can have a better performance on the network with higher average clustering coefficient.

The effective size and Simmelian brokerage are both usedfor Structural Hole structure evaluation. The higher the values, the deeper for the brokerage level on that position. Brokerage level means the number of different unconnected clusters is connected to the particular node in a one hop distance. As mentioned before, the Structural Hole has a negative impact on the node trustworthiness in routing. Thus, we are expecting an attack on the position with higher effective size or Simmelian brokerage would cause more damage than the other position with the lower values.

In the following sections, we have set up two scenarios for the simulation studies. In the first scenario, there are four network topologies with the same number of nodes and connection links. One of the network topologies has Structural Hole in it. The attacks will be launched on each network, and the DTEGR algorithm will detect and avoid these attacks. With the results from all four network topologies, we can compare that with clustering coefficient, effective size, and Simmelian brokerage to find out whether they are related. In such case, all three assumptions on the Simmelian Ties, Structural Hole, and node trustworthiness in routing can be validated.

In the second scenarios, the environment set up will be the same as the first scenario, except for the network size. The number of the nodes in the network is still the same, but we increased the number of connection links in each network to identify if the topological metrics are still able to be compared with different network sizes.

3.5.3 Scenario I

In this section, we have conducted the first scenario of simulation studies on the exploration of the relationship between Simmelian Ties, Structural Hole, and node trustworthiness in routing. The main goal of this simulation is to validate the three assumptions which are made in the Figure 3.1 on page 53, through the network topologies comparison on clustering coefficient, effective size, Simmelian brokerage, packet loss, and packet latency. According to these three assumptions on the Simmelian Ties, Structural Hole, and node trustworthiness in routing, the higher of the average clustering coefficient of the network, the less packet loss the DTEGR algorithm should have achieved. When there is a Structural Hole in the network, the packet loss number should significantly increase with the same packet latency. As there is no alternative route for the detour, the routing path would not change, therefore, the packet latency should be the same. We also expect higher Simmelian brokerage and effective size for the nodes on the Structural Hole position.

We use some P2P network topologies as examples for the simulation studies. There are four network topologies designed, which all have the same number of nodes and connection links in each network. There are 16 nodes in each network with 26 connection links, which are shown in Figures 3.7, 3.8, 3.9, and 3.10. There are 900 traffic sessions from three different source nodes to three different destinations (sinks); they are nodes 1, 2, and 3 as sources nodes; and nodes 16, 14, and 13 as sinks accordingly by order. In other words, each source node will have 300 traffic sessions, and the interval is 4 seconds. Each session also forwards 1 UDP packet with 31 bytes data, and the time to live (TTL) is 128 milliseconds. The grey-hole attack with 50% packet drop ratio is selected as the malicious behaviours on the selected malicious nodes. As there are three sources nodes and three sinks in each network, so there are ten nodes left which can be selected as malicious nodes. For each network, there will be ten simulation runs. The ten nodes, which are not the sources and sinks, will be selected one by one in each simulation run to make sure the malicious attacks are attempted on all of the nodes in the network expect the sources nodes and sinks.

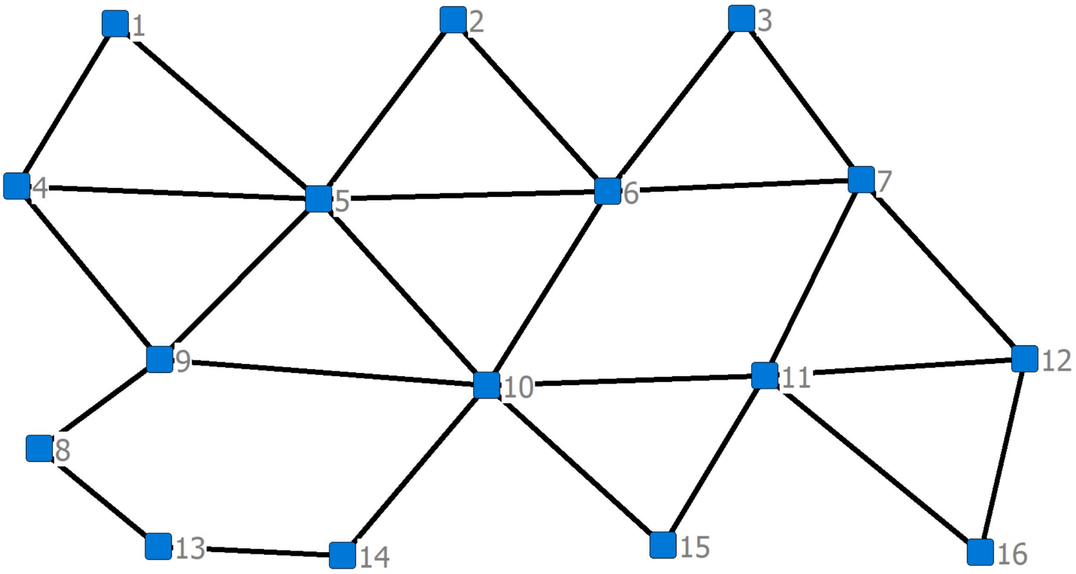


Figure 3.7: Scenario I, Network 1

As shown in the Figures 3.7, 3.8, 3.9, 3.10, the networks size are small, so the Simmelian triangles can be easily counted by human eyes. The network 1 has nine Simmelian triangles, the network 2 has seven Simmelian triangles, there are three Simmelian triangles in network 3, and network 4 has ten Simmelian triangles associated with two structural holes, which are the nodes 6 and 10. The average clustering coefficients for networks 1 to 4 are listed in Table 3.3. The values of the clustering

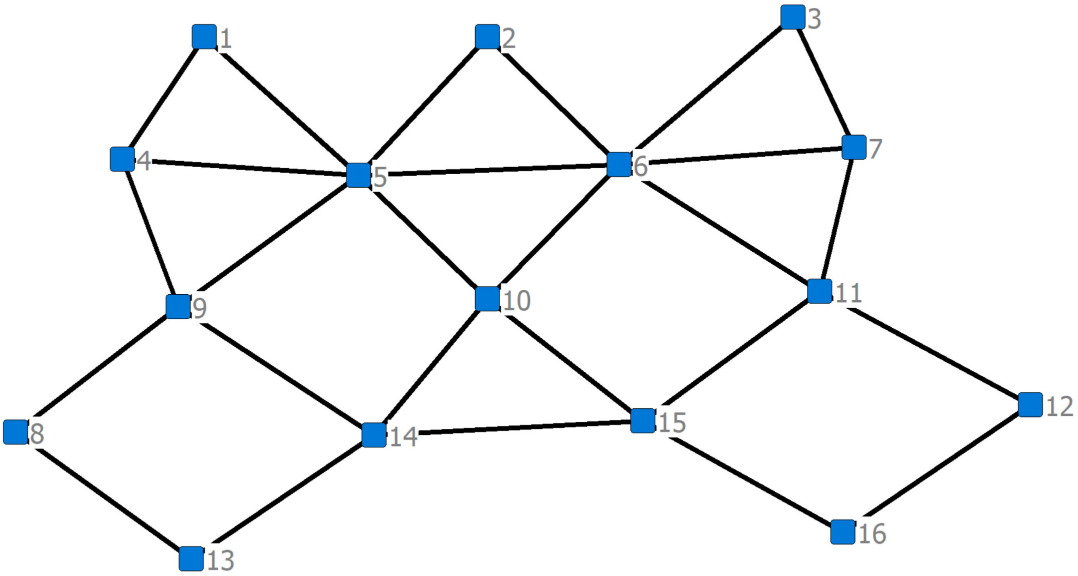


Figure 3.8: Scenario I, Network 2

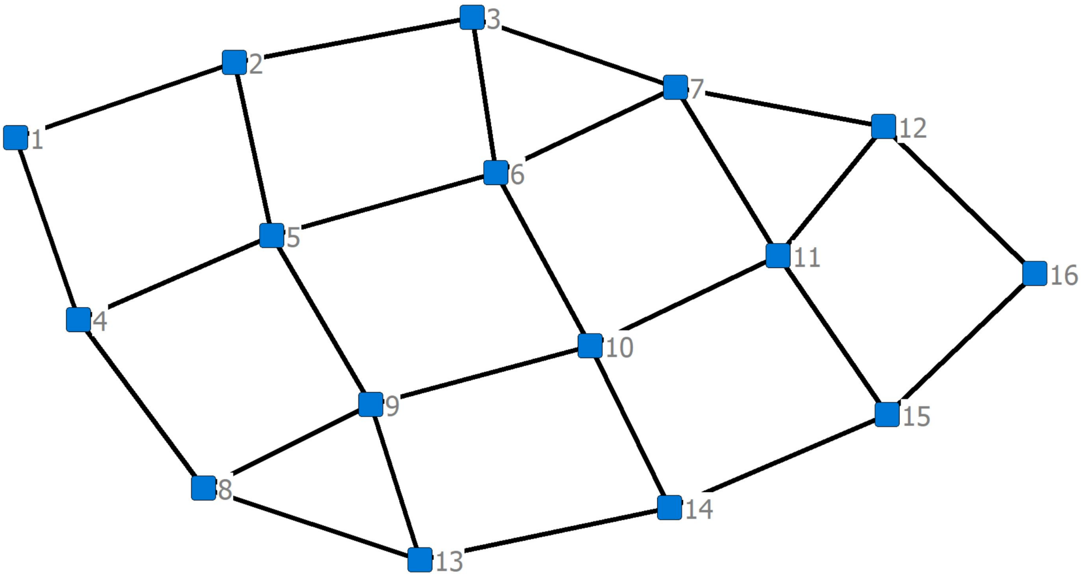


Figure 3.9: Scenario I, Network 3

coefficient reflect are reflecting the number of Simmelian triangles in the network where network 4 has highest and network 3 has the smallest coefficient. In the following, we will discuss the simulation results.

As we can see in Figure 3.11 and Table 3.4, with the same trust-based routing algorithm, DTEGR algorithm has achieved the lowest total packet loss number regarding

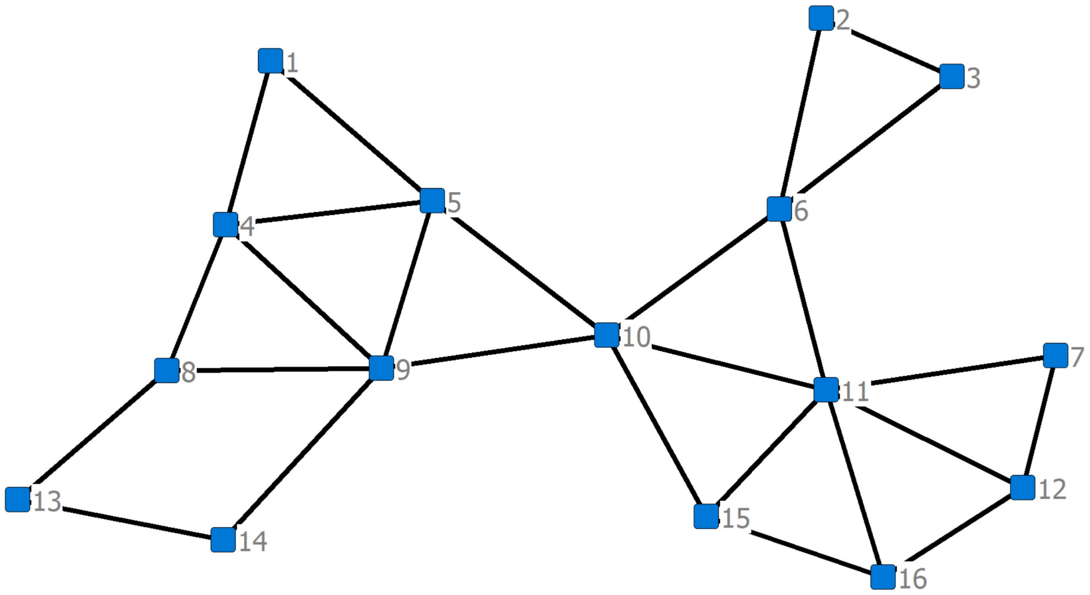


Figure 3.10: Scenario I, Network 4

Table 3.3: Average Clustering Coefficient for Networks 1 to 4

Network	**1**	**2**	**3**	**4**
Average Clustering Coefficient	0.51	0.37	0.14	0.54

all ten attack scenarios in network 1. According to the average clustering coefficient for each network in Figure 3.11 and Table 3.3, the network 1 has second highest clustering coefficient out of the four networks, then the network 2, and network 3 is smallest. These results support our first assumption that the Simmelian tie has the positive impact on the node trustworthiness in routing by achieving the lowest packet loss while under malicious attacks. Moreover, as the number of Simmelian ties increases, the average packet latency decreases which means a shorter distance to the destination node. More Simmelian ties also mean more alternative routes and the routing algorithm has more chance to find a short route to the destination. The reason for less packet loss in more Simmelian triangles network is that the Simmelian ties effectively enforce the reputation (i.e., indirect trust) shared among the nodes within the triangle, and can identify the malicious behaviours earlier and faster to avoid them, so as to have less packet loss.

For example, we have deployed a grey-hole attack on node 10 in network 3, it takes 9 packets loss to identify node 10 behaved maliciously, and then the DTEGR algorithm decided to avoid it. However, in network 1, it only cost 5 packets loss to determine that node 10 is malicious. This is because there are neighbours can to provide the negative opinion on node 10 (low indirect trust value) and so to inform or alert the other nodes.

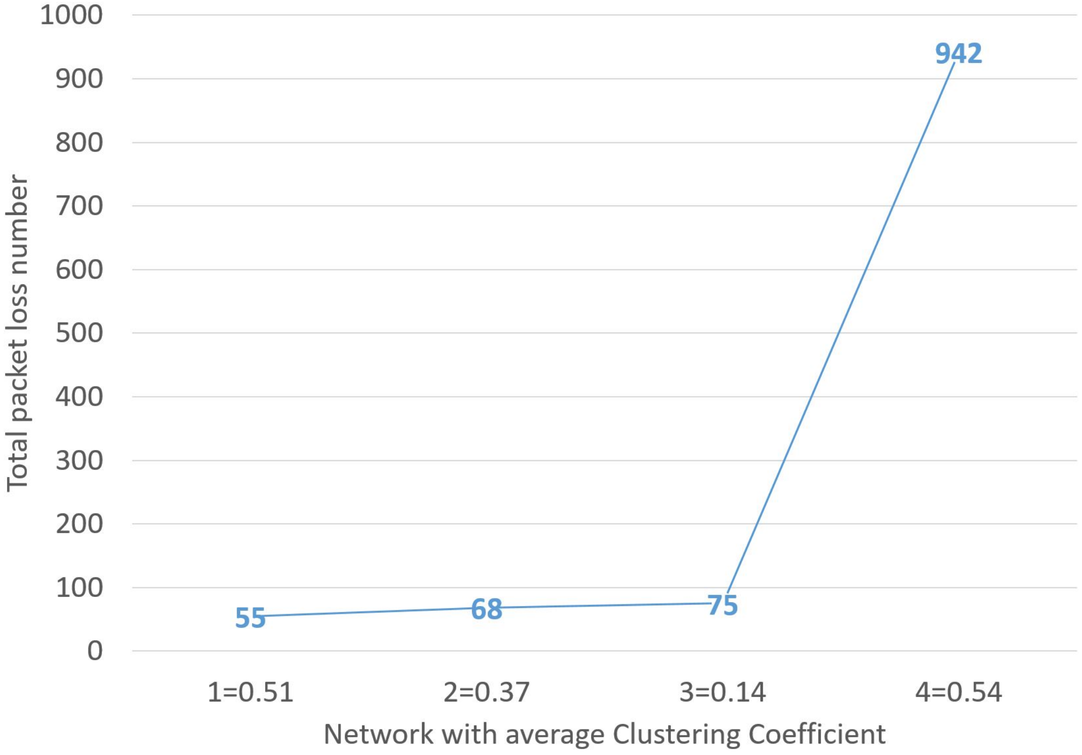


Figure 3.11: Packet Loss vs. Average Clustering Coefficient

Table 3.4: Packet Loss and Latency

Network	**1**	**2**	**3**	**4**
Total Packet Loss	55	68	75	942
Average Packet Latency(ms)	6.23	6.42	7.25	8.84

However, the network 4, which has the highest average clustering coefficient was occurred to have a huge amount of packet loss after the total of ten attack simulation runs. Although it has the most of Simmelian Ties, more significantly, it also has two Structural Holes in the network, which is the main reason causing the huge packet loss.

When we deploy the attacks on node 6 or node 10, which are at the Structural Hole positions in the network 4. Although the DTEGR algorithm has detected the malicious behaviours on these two nodes, the network did not offer an alternative route for the DTEGR algorithm to avoid the attacks. Therefore, the DTEGR algorithm has no choice but to degrades its trust threshold value to ensure network availability at a minimum level. In other words, the DTEGR algorithm will keep forwarding the packets to the malicious nodes at the Structural Hole position, and this is where the huge amount of packet loss comes from. This validates our second assumption that the Structural Hole characterized network has a heavy negative impact on the node trustworthiness in routing and degrades the effectiveness of Simmelian Ties, which this is our third assumption. Here it is worth to highlight that, the node located in the Structure Hole usually has more attractiveness to attack from outsiders, because of its significant impact on the network performance, just like the case mentioned above. In addition to this situation, the Structure Hole can also provide the node itself an unlimited opportunity to act unethically towards all other nodes without fear of the other nodes' learning about its misbehaviour. As explained above about the high packet loss and helplessness to find an alternative path to avoid this malicious node. Being positioned in the Structure Hole, the node becomes a gatekeeper of information that might otherwise be transmitted between contacts. The node spans Structural Hole has a many great opportunities to hold or distort the critical information i.e., grey-hole attack or even terminating all the information flows i.e., black-hole attack, that passes through between each party. The average packet latency was expected to be the same while encountering the attack at Structural Hole, but the simulation results have shown a difference. This is because the DTEGR algorithm tried to find an alternative route to bypass the attacks at the Structural Hole position. These attempts were the cause of the high packet latency.

Secondly, taking network 4 as an example, we have calculated the effective size for all nodes and their values are shown in the Table 3.5. It can be seen that the two

Table 3.5: Effective Size and Simmelian Brokerage in Network 4

Network	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**15**
Effective Size	2.5	2.5	3	1	2.3	3.8	3.8	4.7	1.7	1.7
Simmelian Brokerage	1.8	1.8	3	1	1.1	1.5	3.6	3.1	2	2

Structural Hole nodes are nodes 6 and 10, but the node with the highest value of effective size is node 11 rather than nodes 6 or 10. Node 10 is the second large one. It is interesting to confirm that ’effective size’ is not very effective in identifying all the nodes located in the Structural Hole, which is as reported in other literature (Latora et al., 2013). For the Simmelian brokerage value, it did pick up that node 10 is most likely at the Structural Hole position as it has the largest value. However, for node 6, which is another Structural Hole in the network, it has a smaller value than node 11 that the Simmelian brokerage did not pick it up. We can see that the Simmelian brokerage is more effective than effective size to detect the Structural Hole in network 4. Unfortunately, the Simmelian brokerage can only detect one Structural Hole in network 4. This raises a new direction for research to discover a better and effective metric for future work.

In addition, we also conclude that the Structural Hole characterized network also has a significant negative impact on the Simmelian Ties characterized network. Although network 4 has the largest number of Simmelian Ties, the existence of two Structural Holes has significantly weakened or even totally disabled the positive effect of Simmelian Ties on forming overall trustworthiness through the network; and we need to consider both characteristics of Simmelian ties and Structure Holes to evaluate the overall network trustworthiness.

We have set up the same network sizes for the comparison in this scenario. In the next scenario, the different network sizes will be compared to see whether this will make any difference.

3.5.4 Scenario II

In the second scenario for the simulation studies, we set up five network topologies with the same number of nodes (i.e., 16 nodes) and their locations in the network. Rather than having the same number of connection links as the previous case, we gradually increased the connectivity, by adding 4 links each time, to the first benchmark network 5 with 18 links in Figure 3.12. All the other derivational network topologies are shown in Figure 3.13 to 3.16. The rest of the simulation parameters are set up the same as the first scenario. There are a total of 900 packets to send from node 1 to node 16, node 2 to node 14, and node 3 to node 13 of which are 300 packets for each traffic flow. A grey-hole attack is deployed on nodes 4 to 12, and node 15 in each network case with a 50% packet drop rate.

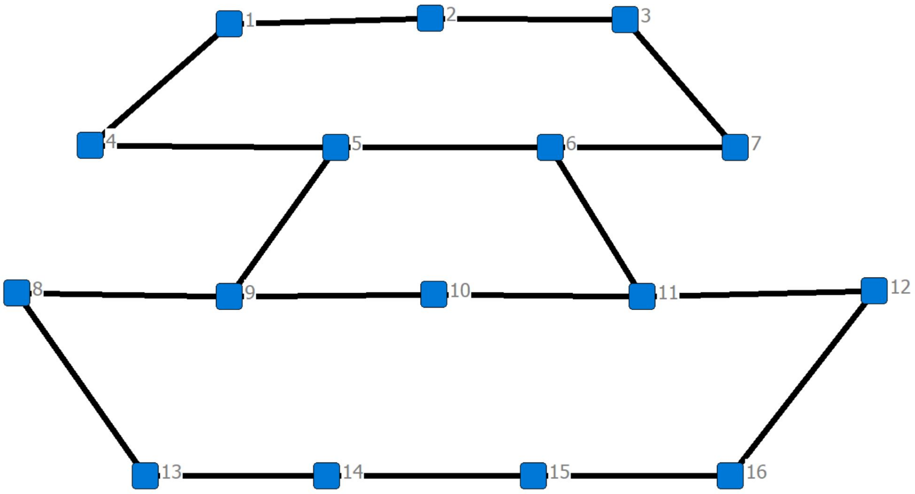


Figure 3.12: Scenario II, Network 5

It can be seen that from Figures 3.12 to 3.16, with the same number of nodes and their locations, the links are gradually being added and also the number of Simmelian triangles increases from 0 to 19 where every node is embedded in at least two Simmelian triangles. Then the average clustering coefficient and network performance are summarized in Figure 3.17 and Table 3.6. We can see that network 7 has the highest

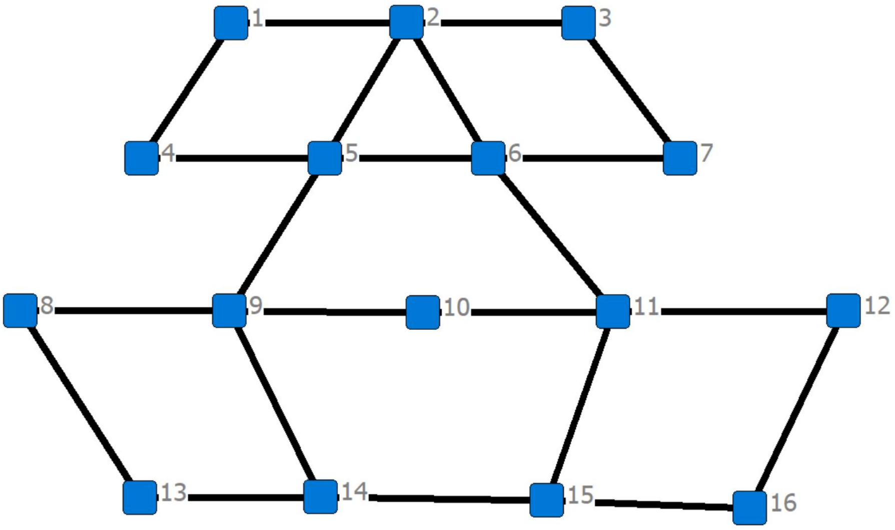


Figure 3.13: Scenario II, Network 6

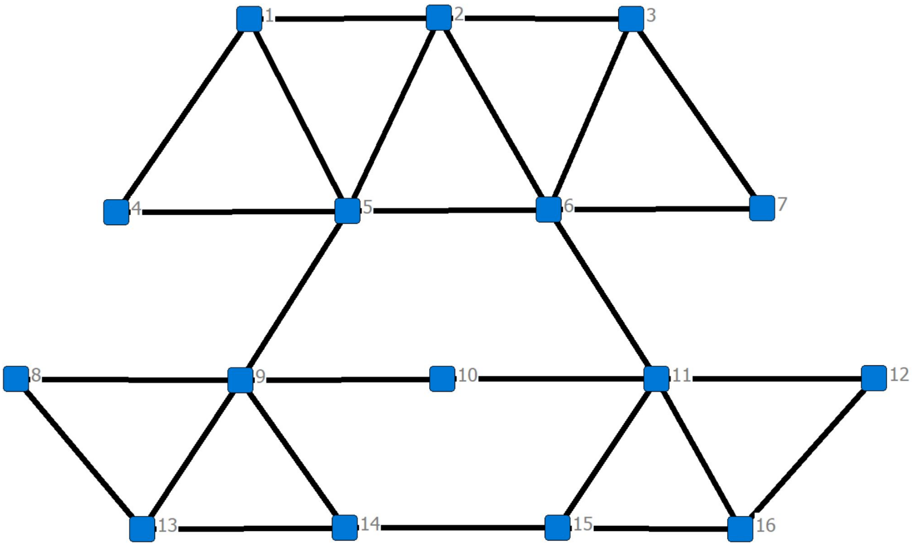


Figure 3.14: Scenario II, Network 7

average clustering coefficient of the five networks, but the best network structure, in terms of overall packet loss and latency is network 9, which with the similar lowest packet loss to network 8 and lowest average packet latency. Network 9 has the highest number of Simmelian triangles which are distributed evenly to cover the whole network to make it be the best structured. Although network 7 has many fewer links to connect all the nodes, it still enables nodes 4, 7, 8, and 12 to form the most number of essential

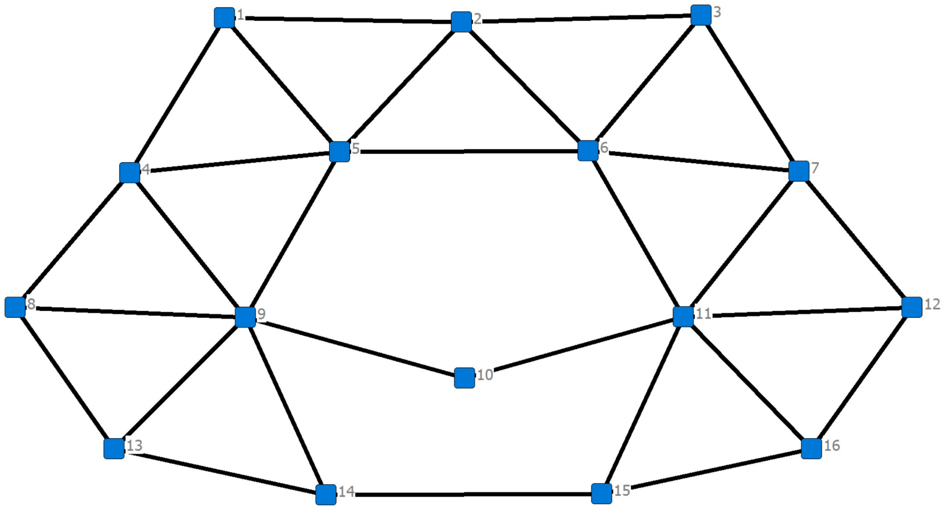


Figure 3.15: Scenario II, Network 8

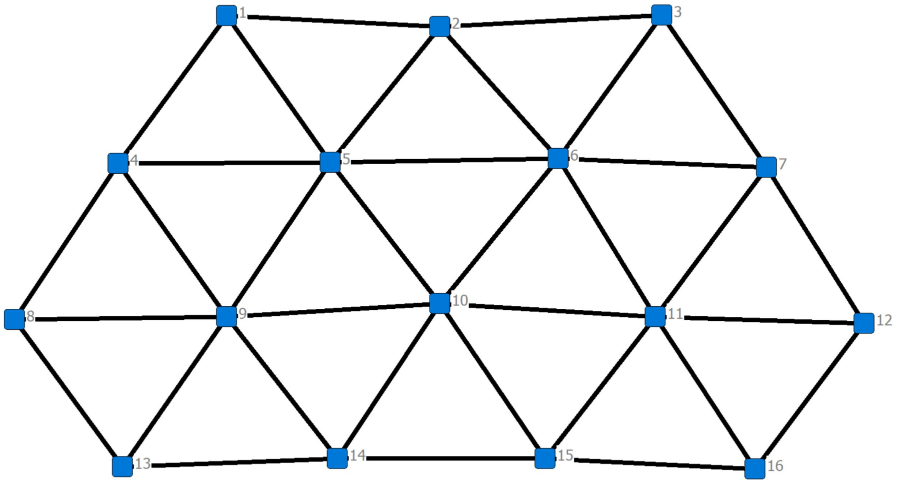


Figure 3.16: Scenario II, Network 9

Simmelian triangles within these limited links. From this fluctuation of the clustering coefficient, we can see that the average clustering coefficient is more accurate to evaluate the density of Simmelian triangles out of the maximum possible with the given amount of node degree. In other words, adding connection links in the network is not necessarily increasing the average clustering coefficient accordingly. Thus, in the first scenario simulation studies, regarding the various networks with similar numbers of nodes and links, the average clustering coefficient is not that effective to measure the various networks with a different number of links such as in this case.

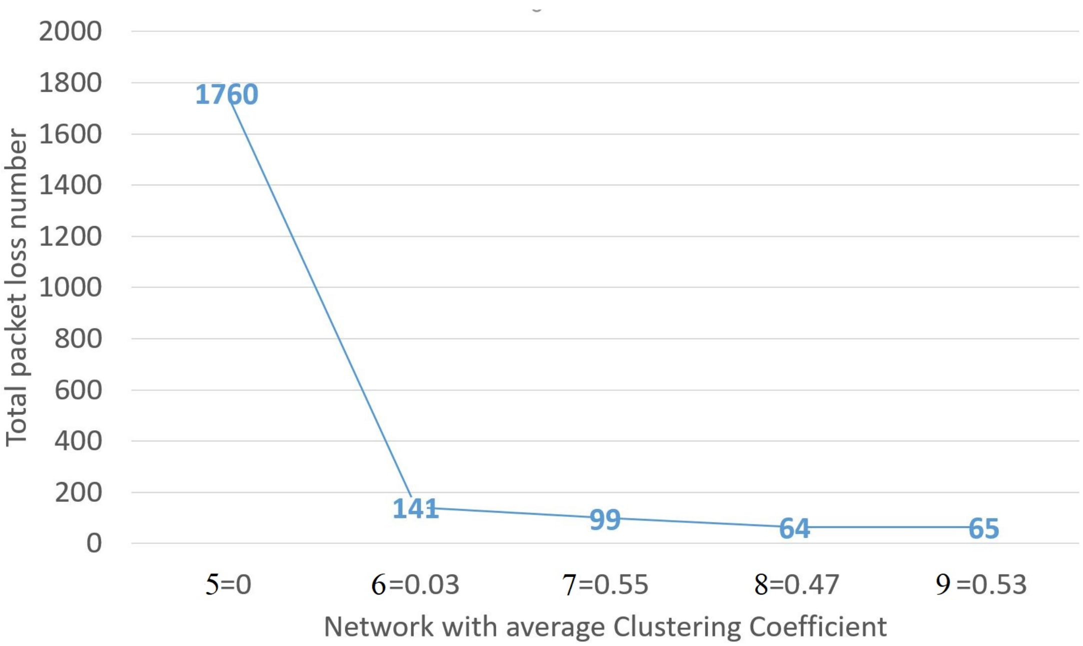


Figure 3.17: Scenario II, Packet Loss vs. Average Clustering Coefficient

Table 3.6: Scenario II, Packet Loss and Latency

Network	**5**	**6**	**7**	**8**	**9**
Clustering Coefficient	0	0.03	0.55	0.47	0.53
Total Packet Loss	1760	141	99	64	65
Average Packet Latency(ms)	36.08	8.28	6.5	6.46	5.86

Once again, it can be seen that, as the number of Simmelian triangles increases in a network, the same trust-based routing algorithm can perform better to detect and avoid any malicious attacks in terms of packet loss and packet detour latency. More Simmelian triangles embedded means more backup routes can be selected and more likely can express warning messages, that is, indirect trust and reputation value collection from other neighbours, so as to detect the malicious nodes faster. Network 5 with no Simmelian triangles is significantly constraining the trust-based routing algorithm to find a trustable end-to-end path, which causes a huge amount of packet loss and latency under malicious attack. Network 5 only has 18 links and 16 nodes so that every node only has two connections on average. Once the malicious attack occurs, most likely there would be only one route choice left to detour all the traffic flows. Another critical reason for the huge packet loss and latency is that all the detoured traffic flows will overload and congest the only leftover node or link to be routed, which contributes more packet loss and latency worsening the case caused by the primary malicious attacks. In addition, the traffic flow needs to be detoured by using a longer path to reach the destination and this contributes extra packet latency to the high average packet latency. As the links increase, the number of alternative routes increase as well. The trust-based routing algorithm has more options to select, so it can work more effectively in terms of finding better trustable end-to-end routes. This is well confirmed by the results listed in Table 3.6. The packet loss decreases as the connection links increase, and the packet latency is decreasing as the shorter route is available to be used as well.

In this section, we have the extensive simulation studies as to validation of the relationship between the Simmelian Ties, Structural Hole, and node trustworthiness in routing, which are the three assumptions have discussed in Figure 3.1 on page 53. The performance of the DTEGR algorithm is better when the network has a higher clustering coefficient. It is reflected in lower packet loss results and lower average packet latency. This has validated the first assumption in which the Simmelian Ties are believed to have a positive impact on node trustworthiness in routing. When there is a Structural Hole in the network, the attacks launched on such position have created a huge packet loss as the DTEGR algorithm cannot find an alternative way to bypass it. This has validated the second assumption, in which the Structural Hole has a negative impact on the node trustworthiness in routing. Finally, a network with higher average clustering coefficient, which it also means it has more Simmelian triangles in the network. At the same time, it has a Structural Hole in the network as well. The simulation results in this network have come back as the highest packet loss and packet latency compared to those networks with lower average clustering coefficient. This has validated the last assumption, in which the Structural Hole has a negative impact on the positive impact of Simmelian Ties.

Moreover, through the extensive simulation studies, we have found that the average clustering coefficient is not very effective in evaluating the networks with different sizes. It might require normalisation to tune the metric, so it can be used to compare the networks of different sizes. The effective size and Simmelian Brokerage is not able to detect the exact location of the Structural Hole, they are more focused on measuring the local brokerage level.

As mentioned the framework in the Figure 2.3 on page 36 in the literature review chapter, there are five assumptions regarding the Simmelian Ties, Structural Hole, independent tasks, and interdependent tasks. These indicated the two sides of the Simmelian Ties and Structural Hole. In our simulation studies in this section, only one side has been explored. In the next section, the other side of the Simmelian Ties and Structural Hole is discussed.

3.6 The Dilemma of the Simmelian Ties and Structural Hole

In the last section, the benefits of Simmelian Ties and the drawbacks of Structural Hole in P2P routing have been discussed and the assumptions validated through extensive simulation studies. As mentioned in the literature review chapter, in sociology, there is an ongoing debate on these two different social structures as to whether they are sustaining or hindering the performance-related outcome at the individual and collective level (M. Granovetter, 2005; M. S. Granovetter, 1973). Engle (1999) has proposed a framework to summary these two structures (in Figure 2.4 on page 40). In the computer network routing in the P2P environment, we have proposed an assumption on these two structures against the node trustworthiness in routing and the extensive simulation studies in the last section have already validated that assumption. However, this is only the one side of these structures. In this section, the drawback of Simmelian Ties and benefit of Structural Hole will be reviewed.

3.6.1 The Dilemma of the Simmelian Ties

As mentioned in the literature review chapter, in sociology, Simmelian ties can engender mutual trust by introducing the third parties. In the computer network, the third party can provide the feedback about its neighbour for another neighbour, in case this neighbour becomes selfish or malicious, the third party can quickly warn other neighbours, so they can be alerted. Moreover, the Simmelian ties can provide a redundant route in case of attacks or network failure.

However, everything is like a sword, which has two edges. On the other hand, in the Simmelian triangles, if both neighbours become malicious, they should do more damage to the network. For example, one neighbour keeps broadcasting the good reputation of another neighbour who is actually malicious. In such a case, the remaining node will take a longer time to discover that neighbour is malicious. This kind of attack is known as ’Ballot Stuffing’ attacks. For the other way round, which the malicious neighbours keep sending the bad feedback on the target node this is known as ’bad mouth’ attacks. As the Simmelian Ties normally form a social norm in the sociology. Thus, same ideas, the collusion attacks mentioned above can be considered as this social norm situation. Moreover, more Simmelian ties in the networks mean more redundant links in the network, which cause more energy consuming for the nodes in the network to maintain these redundant links.

We can assume the Simmelian ties can impact the trust routing in the computer network in the following ways:

1. Simmelian Ties can help node more quickly establish the trust and detect malicious or uncooperative behaviour in the network by introducing the third node.
2. Simmelian Ties provide redundant paths for the networks in the case of attacks or network failure.
3. Simmelian Ties can be a hindrance on the node trustworthiness in routing when under a ballot stuffing attack.

The first two assumptions have already been validated through the extensive simulation studies in the last section. The third assumption here regarding the Simmelian Ties needs to be validated in this section.

In this scenario, we are going to validate the assumption that the Simmelian Ties can be a hindrance on the node trustworthiness in routing when under a ballot stuffing attack. It can do more damage to the network as Simmelian Ties strongly foster a normative environment in that the node will take longer detect malicious activities when both neighbours are malicious. The setup for this scenario is the same as the scenario I in the section 3.5.3 on page 69. The only difference is the nodes are not performing grey-hole attacks will all perform ballot stuffing attacks, which they send out the indirect trust/reputation value for their neighbours with the full mark of 1. Moreover, only Networks 1, 2, and 3 are selected for the comparison, and network 4 is only for the Structural Hole evaluation. In other words, Networks 1, 2, and 3 are selected for this scenario for the simulation studies. Each network has ten simulation runs. In each run, there is one node selected to perform grey-hole attacks with 50% packet drop ratio. The rest of the nodes, not including the source and sinks, will perform ballot stuffing attacks. The weight factors between direct trust and indirect trust are 50% and 50%. Of course, those nodes without a third neighbour will only have direct trust to evaluate trust. The results are shown in Figure 3.18.

As can be seen, once every node ignores their neighbour's malicious behaviours

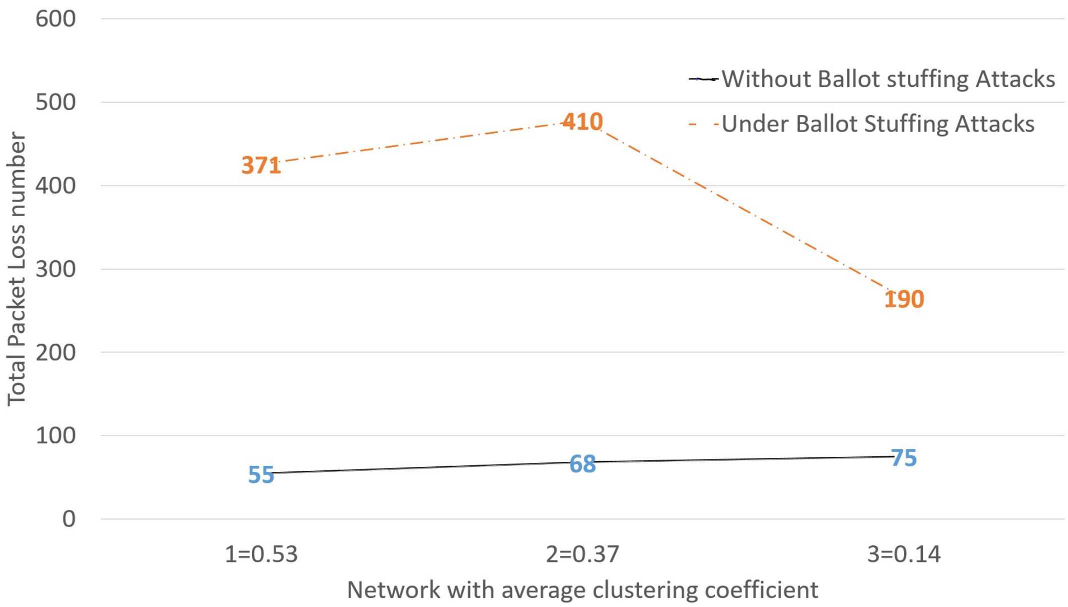


Figure 3.18: Ballot Stuffing Attacks Comparison

Table 3.7: Scenario III, Packet Loss and Latency

Network	**1**	**2**	**3**
Total Packet Loss	371	410	190
Average Packet Latency(ms)	5.95	6.29	7.36

and only sends out good reputations, the packet loss number is significantly increased compared to the results in the scenario without the ballot stuffing attacks. The packet loss results in the scenario without the ballot stuffing attacks are decreasing while the clustering coefficient is increasing. It is the opposite for the scenario under ballot stuffing attacks. As the good reputation from neighbours makes the node take longer than usual to determine whether the target nodes are malicious, the more interactions are required, which means more packet loss number in the results. Moreover, there is a trend that the lower the clustering coefficient (which means less Simmelian Ties in the network), the less affection by the ballot stuffing attacks. This is because fewer Simmelian Ties means less chance for the node to receive a false reputation value from neighbours. In Figure 3.18, network 2 has a lower clustering coefficient compared to

Network 1, but Network 2 has more packet loss at the end. Even though network 2 offers less chance for a ballot stuffing attack compare to Network 1, Network 1 has more Simmelian Ties in the network, which means more redundant routes that nodes in the network have less chance to encounter grey-hole attacks and possibly faster to reach the destination. This is reflected in the average packet latency in Table 3.7. The simulation results have validated the assumption, which the Simmelian Ties can be a hindrance on the node trustworthiness in routing when under ballot stuffing attack.

In the next section, the dilemma of Structural Hole in the distributed P2P network environment.

3.6.2 The Dilemma of the Structural Hole

As discussed in the literature review chapter, there is a long debate as to which social structure is better (Latora et al., 2013), Burt (2000) argued that this depends on the context. In the computer network, from the network availability point of view, more redundant routes are preferred as this ensures most of the networks stay connected while under attacks or network failures. In such cases, the Structural Hole is believed to cause greater damage to the network as it is acting as the gateway to connect different clusters in the network. Once it malfunctions or becomes malicious, it will disconnect these clusters. Another case is when a node is performing flooding attacks in the network where the affected area need to be controlled in a smaller area. In such a case, the Structural Hole can do a better job as the nodes in Structural Hole positions have the power to control the data flow.

The understanding of Structural Hole as having two edges is necessary, so we know when Structural Hole can improve the node trustworthiness in the network and when it will be an obstacle to it. In these cases, the network connection establishment or design of the network topology can be done in a smarter way.

The nodes at a critical position in the network acting as gateway have the power to control traffic flow, and also can act as a firewall to stop the malicious or unwanted packets being forward to another cluster. The nodes at the central position are more likely to access more information and thus, it potentially, has more power to control these information flow. The nodes at the Structural Hole should have these features.

Moreover, from a network attack point of view, the actor at the Structural Hole is usually more attractive to attack because of its bottleneck position. Take the network routing as an example, it means that there is no alternative route to be selected and cannot avoid the malicious attacks if the actor is compromised or even itself has misbehaviour. The Structural Hole is beneficial in regard to individual tasks as the actor can easily access the exclusive information from different parties to better deliver the task. It also means that for other nodes, the hole can deliver the packets to the destination more quickly in network routing. While it also presents opportunities for misconduct because when an individual is spanning a gap between otherwise unrelated contacts, this individual is positioned to act unethically toward another individuals or groups without fear of the other(s) learning of the act.

We can assume the structural hole can impact computer network routing in following way:

1. Structural Hole has the control power to connect or isolate the disconnected clusters in the network. It can be a firewall to stop malicious data or it can be a black hole to stop all the communication.

2. When attack or failure happen on Structural Hole position, the damage to the network is greater.

In the following, we have the simulation studies will conduct the two sides of Structural Hole in P2P network routing, so as to validate the assumptions made on the two sides of Structural Hole. The Structural Hole normally acts like the gateway in the computer networks, so as a gateway and a firewall, it can connect different disconnected clusters effectively and quarantine unwanted data packets within a certain area. When it malfunctions or behaves maliciously, it can also create greater damage to the network availability; the deeper the hole it is, the greater the damage can be done.

In this scenario, the algorithm is running twice in each network topology. In the first run, node 4 will perform grey-hole attacks with a 50% packet drop rate, and the second run is without any attack. There will be 300 packets sent from node 1 to node 9 with 3 seconds interval. Each packet size is 32 bytes. We set up two WSNs with different topologies where each network is composed of nine nodes and connected by twelve Wi-Fi connections. The topologies of the networks are shown in Figure 3.19. Network 10 has a Structural Hole in the network where network 11 does not have. The results are in Table 3.8.

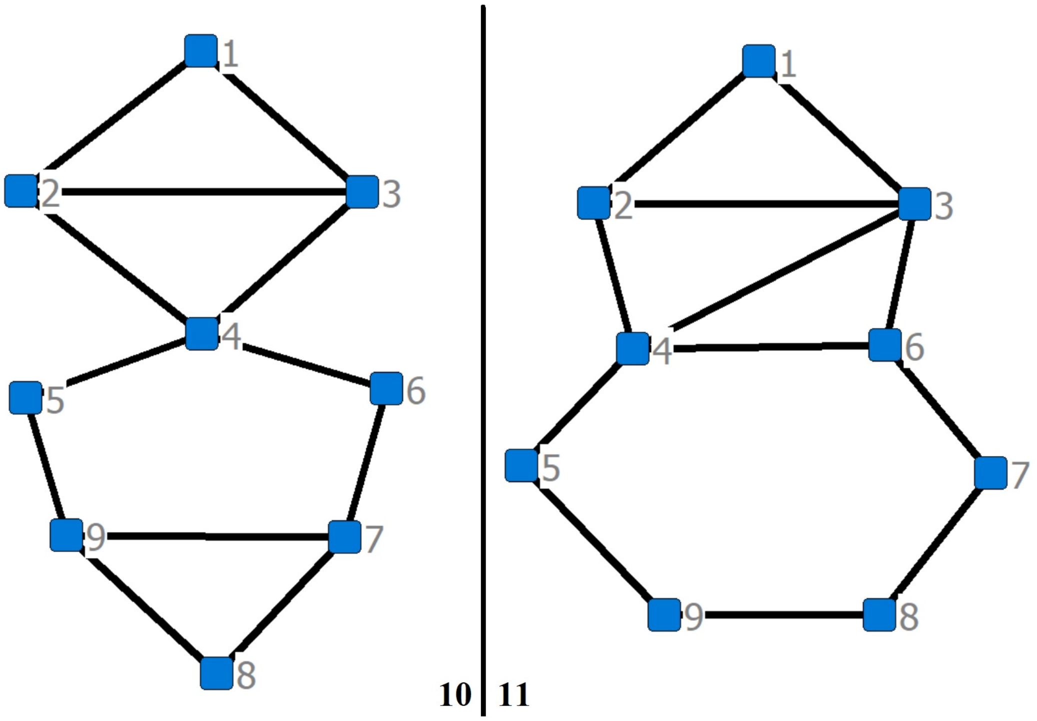


Figure 3.19: Scenario IV, Networks 10 and 11

From the results in Table 3.8, we can see that when the networks were not under attacks, both networks have the same performance which is zero packet loss and same

Table 3.8: Damage Impact vs. Effective Size & Simmelian Brokerage (Node 4 Only)

	Network 10		**Network 11**	
Under Attack	**Yes**	**No**	**Yes**	**No**
Packet Loss	154	0	10	0
Average Packet Latency(ms)	6.348	6.324	10.013	6.324
Clustering Coefficient	0.167		0.333	
Effective Size	3.5		3	
Simmelian Brokerage	3.5		2.75	

average packet latency. This is because the routing from node 1 to node 9 in both networks are exactly the same when no malicious attack was launched. However, once node 4 was performing grey-hole attacks, Network 1 has 154 packets lost out of 300 and Network 2 only has 10 packets lost. This is because node 4 in Network 1 is at the Structural Hole position that even though algorithm detected the attacks, there is no alternative route to avoid the attacks. In Network 2, there is an alternative route that allows the algorithm to avoid the attacks, though it still sacrificed 10 packets to determine node 4 is malicious. Moreover, the average packet latency in Network 2 is higher as the alternative route has further distance to the destination.

The Simmelian tie and Structural Hole appear as the opposite network structures that the clustering coefficient and effective size of node 4 also shows opposite results as well in Table 3.8. The effective size and Simmelian brokerage for node 4 in Network 1 are both have a higher value compared to Network 2. As node 4 is at the Structural Hole position in network 1, which it is the only gateway connecting the left (nodes 1, 2, and 3) and right (nodes 5, 6, 7, 8, and 9) clusters. In such a case, node 4 is considered as having higher efficiency in connecting the networks and the higher level of brokerage to act as a broker between different groups which are not connected compared to node 4 in Network 2. This result implies that the node with the higher effective size or Simmelian brokerage can more effectively connecting the networks, but at the same time, when this node behaves maliciously or is malfunctioning, it can also cause greater damage to the network.

The same case, but rather than grey-hole attack in the network, we perform flooding attacks on node 3 in both network 1 and network 2. In network 1, as long as node 4 refuses to communicate with node 3, the flooding attacks affect area can be quarantined within nodes 1, 2, and node 4. However, in network 2, even though node 4 refuse to communicate with node 3, node 3 can still spam out the unwanted packets to the rest of network by forwarding packets to node 5. In such a case, the Structural Hole is preferred so as to control the data flow.

Both Simmelian ties and Structural Hole are like a sword, which they have two edges. In different scenarios have different network objectives to define a network's trustworthiness. For example, in packet forwarding scenarios, Simmelian Ties in the network might be more preferred as it provides redundant paths to avoid malicious attacks or network failures, and the third node can provide reputation of the trustee. In a situation like a network is under flooding attacks, more Structural Hole structures in the network is preferred as the Structural Hole can act as the firewall that to stops and quarantines the unwanted packets.

3.7 Summary

In this chapter, we have studied how the underlying topologies can impact the overlay nodes' trust and reputation behaviours in the P2P environment. In particular, we have studied the positive and negative impact of the Structural Hole characteristic network structure in forming trust in the P2P network environment. Inspired by Engle's (1999) framework on Simmelian Ties and Structural Hole, we have made three assumptions with these two structures on the node trustworthiness in routing in the distributed P2P network. These have been validated through the extensive simulation studies. The Simmelian Ties has a positive impact on the node trustworthiness in routing, while

Structural Hole has a negative impact on the node trustworthiness in routing, and the Structural Hole has a negative impact on the positive impact from Simmelian Ties. Everything has two sides, as Engle's (1999) framework proposed. Thus, the Simmelian Ties and Structural Hole in distributed P2P networks should have the two sides as well. Nodes are located in the Structural Hole position in a network normally act as the gateway and they have the power to control the data flows. When they are behaving well, these gateway nodes can efficiently connect different clusters in a big network as well as serving as a firewall node to stop unwanted data flows from malicious clusters. However, when they behave maliciously, malfunction, or being attacked, these nodes can create greater damage to the network availability. The more significant the location of a Structural Hole, the greater the damage which can be caused. The extensive studies have confirmed that the Structural Hole characterized network has high risks for misbehaviour and malicious attacks. The Structural Hole can make the network fragile when it is under attack or the node has malfunctioned. Although it can efficiently connect sub-networks to the best network performance such as packet delay when it is functioning well. The situation is similar with Simmelian Ties; on one side Simmelian Ties provide a redundant route for backup and introduce a third party to engender mutual trust. On the other side, when the third party becomes malicious as well, it can create more damage to the network, such as in ballot stuffing attacks, these have been validated through the extensive simulation studies as well.

Moreover, we have introduced the Clustering Coefficient metric to calculate the percentage of Simmelian triangles out of the maximum possible on the given node degree, and the effective size and Simmelian Brokerage to evaluate the Structural Hole, which is an opposite structure of Simmelian Ties. Through the extensive simulation studies, we have identified that the Simmelian brokerage metric is capable of evaluating the Structural Hole more precisely than the effective size metric. This is because the Simmelian brokerage metric also considers the centrality of the nodes in the network, where the node in a central position can normally do more damage to the network when it behaves malicious or malfunctions. However, the Simmelian brokerage is unable to detect and locate the exact Structural Hole in the network, instead, it evaluates the local brokerage level. The locator of Structural Hole requires a global measurement metric.

In the next chapter, we will further explore the interplay of underlay network topologies and trust in the computer network with a proposed evaluation framework for the evaluation of network trustworthiness.

Publications generated from this Chapter

Xiang, M., Liu, W., Bai, Q., & Al-Anbuky, A. (2016). The critical role of structural hole in forming trust for Securing Wireless Sensor Networks. International Journal of Information, Communication Technology and Applications, 2(1), 66–84.

Xiang, M., Liu, W., Bai, Q., & Al-Anbuky, A. (2015). Avoiding the Opportunist: The Role of Simmelian Ties in Fostering the Trust in Sensor-Cloud Networks. International Journal of Distributed Sensor Networks, 11(10), 873941.

Chapter 4

NTaaS: Network Trustworthiness as a Service

4.1 Introduction

In the previous chapter, the co-evolution of the underlying distributed P2P network topologies and the node trustworthiness in routing have been explored through the extensive studies on the Simmelian Ties and the Structural Hole. These studies have confirmed that the underlying network topologies have a great impact on the efficiency of trust-based routing algorithm to resist unwanted behaviours such as malicious or selfish behaviours. From the literature review, we found the current studies on the trust network routing are focusing on the trust modelling on the local perspective or local trust, which is the trust relationship between the nodes in the network. As mentioned above, the different network structures in the network can have a different impact on these trust relationships between the nodes. They can be beneficial or impedimental to the trust routing in the P2P environment. Obviously, the more beneficial structures in the network, the trust then can be established and convergence in the more efficient way, vice versa. The current studies on trust in the P2P networks have no focus on this area yet. The studies on network topologies can improve the network design work quality, identify the potential issues in the network, so they can be optimized to the satisfaction level, improve the trust-based routing algorithm efficiency, etc. Even in a dynamic network such as mobile ad-hoc networks, with the prediction of the topology in the next moment, the topology evaluation framework can quickly identify the potential threats in the network, then with the aid of Unmanned Aerial Vehicle (UAV), to cover the potential threats temporarily.

First of all, this chapter introduces a new term, which is network trustworthiness to describe the confidence level of the networks being able to serve their objectives. This thesis focuses on the global trust for the whole network rather than focusing on the local trust level (trust on nodes), which the most existing studies focus on. Secondly, a new concept based on the new term network trustworthiness is proposed, which is' Network Trustworthiness as a Service (NTaaS)': the network trustworthiness evaluation as a service to the P2P users for trustworthy communication. This is followed by the introduction of the trustworthiness evaluation framework, which mathematically quantifies the trustworthiness of the networks for the NTaaS service paradigm. After the trustworthiness evaluation framework is introduced, a sample scenario (objectives) is provided for the validation of the trustworthiness evaluation framework. We use the findings from the last chapter on the relationship between Simmelian Ties, Structural Hole, and node trustworthiness in routing, i.e., the clustering coefficient as one of the metrics to evaluate the network trustworthiness. As the effective size and Simmelian brokerage are not able to locate the Structural Hole in the network, we then proposed the Structural Hole Locator (SHL) algorithm to identify the Structural Hole in the network. We also classified the Structural Hole into physical Structural Hole and logical Structural Hole. The detailed explanation of the differences between physical Structural Hole and logical Structural Hole comes later. In addition, the details of SHL algorithm will be explained. After this, we have used extensive simulation studies to validate the trustworthiness evaluation framework. Finally, we have introduced another concept, it is called as Trustworthiness Tolerance Margin (TTM). This concept is for the evaluation of network trustworthiness with a possible changes margin (the best case and worst case) while under the random position attacks, so as to see if this network can maintain its performance while under random attacks or network failures.

4.2 Network Trustworthiness

In the literature review chapter, this thesis has reviewed the existing work on the trust-based routing algorithm. The trust evaluation between nodes in the network is mostly based on their previous interaction experiences, capacity, social relationship, etc. The node capacity describes whether the nodes have the capability to complete the task. Such as sufficient energy level, buffer space, CPU, memory, required sensor, etc. If a node does not have enough capacity to complete the task, obviously, this node is not trustworthy for the task. In addition, if the nodes support encryption and decryption, strong authentication mechanism, high-speed processing capability, etc. They can be a bonus to increase the trust level to the trustor. Because these bonuses mean that these nodes can complete the tasks faster, more securely, or more efficiently, so as to achieve a better quality of service (QoS). This is from the local trust point of view. From the global trust point of the view, which is the capacity of the network as a whole, the network can improve the efficiency of the trust-based routing algorithm, and for example robustness while under attacks, etc. This should be the same as the local trust, and we defined this global trust as network trustworthiness.

Gabarro (1978) suggested that trustworthiness is a multifaceted construct that captures the competence and the character of the trustee. The network structure and capability can be considered as the character and competence of the trustee. In such cases, the network trustworthiness should be evaluated by the structures in the networks.

Mohammadi and Heisel (2016) defined the trustworthiness as the assurance that the system will perform as expected. Bandyszak et al. (2016) also defined the trustworthiness as fulfilling stakeholders' expectations.

Computer networks allow the computer devices inside to communicate with each other and serve their objectives. Therefore, to adapt trustworthiness definition from other areas, such as studies (Mohammadi & Heisel, 2016; Bandyszak et al., 2016) , a P2P network is 'trustworthy' when it fulfils its objectives under any expected or unexpected circumstance. This circumstance can be malicious attacks, a natural disaster, etc. An expected circumstance is a known issues occurring in the network and can be resolved with the existing solutions. The unexpected circumstance occurs when malicious parties have found back doors into the network, then bypass the security defenses and launch malicious attacks on the network.

Moreover, the reason we used the word of objectives rather than just network availability, node trustworthiness, etc. is, in different scenarios, there are different requirements for the network topologies; some are focusing on the network availability, and some are focusing on traffic control. The study by Mohammadi and Heisel (2016) also suggested that 'trustworthiness' is domain and application independent. In other words, it depends on a specific context and goals. There is no trustworthy network, but a network is 'trustworthy' in some manner. Thus, the measurement metrics for the trustworthiness can be different as well depend on the scenarios. For example, in the last chapter, we have explored the relationship between Simmelian Ties, Structural Hole, and node trustworthiness in routing. There has been a long debate on which social structure is better (Latora et al., 2013), Burt (2000) has argued that this depends on the context. In the network availability focusing scenarios, the Simmelian Ties can be beneficial to the network availability and nodes trustworthiness as it provides the redundant routes and introduces the third parties to support reputation and feedback. The Structural Hole should be avoided as it offers no alternative route and no third parties so that it will be a threats or bottleneck in the network. Therefore, when we are evaluating the trustworthiness of a networks in this scenario, the more Simmelian Ties in the network, the more trustworthy of this network. In addition, if there a Structural Hole in the network, this hole would be considered as a weakness or potential threat of the network, as once the node at this position is under attack, the network is disconnected. These have been well validated in the last chapter. There is another scenario, which is discussed in study (Van Mieghem, Omic & Kooij, 2009), which studied the epidemics spreading model such as virus spreading in a computer network. It considers the contaminating rate, which is an indicator of the epidemics spreading speed; and the recovery rate. When the contaminating ratio is higher than the recovery rate, then eventually the whole network would be contaminated, vice versa. In such scenario, we look back to the Simmelian Ties and Structural Hole again; obviously, the more redundant routes in the network, the faster the speed of epidemics spread would be. Then Simmelian Ties are not preferred in such a scenario, and the Structural Hole is more preferred as it can act as a firewall to stop the virus spreading. In such a case, the evaluation of network trustworthiness in this scenario would consider using completely different metrics to the metrics in the evaluation of the network availability scenario.

In the literature review, the Adaptive Networks concept was introduced, which claims the change in the network entity's status can affect the evolution of the network topology. At the same time, the change of network topology can affect the evolution of network entity's status. The network trustworthiness can be considered as the impact of the network entity's status evolution affected by the change of network topologies. The effect of the evolution can be good or bad. A trustworthy network should enable the network entity's evolution to become better. In next section, a novel concept is proposed based on the network trustworthiness.

4.3 Network Trustworthiness as a Service

4.3.1 TaaS: Trust as a Service

We are going to propose the NTaaS to address the research gap in this section. Before that, we consider some related work in this area. However, we can only find one study (Noor & Sheng, 2011), which proposed the trust as a service and it is more or less related to our proposed NTaaS framework. The Trust as a Service (TaaS) is focused on the credibility of the feedback system of Cloud Service Providers (CSP). The TaaS is set up between the Cloud Service Clients and CPS for the evaluation. They believe that the Cloud Specialist Clients' feedback is more credible than the amateur clients' feedback. They defined the specialist clients' as the clients who agree with most of the other feedback on the same CSP on the same service. Moreover, they also believe the older clients' feedback is more credible as they are more experienced. The TaaS provides a service to evaluate the CSP for the clients so they know which CSP is trustworthy. Our proposed NTaaS is focusing on the P2P communication routing which is introduced in the next section.

4.3.2 NTaaS: Network Trustworthiness as a Service

The NTaaS framework is shown in Figure 4.1. It uses the Device to Device (D2D) communication under a future 5G cellular network as the example for the communication environment. There are three planes in the NTaaS. The first plane at the bottom is the physical plane, where all the physical devices or sensors are located in the P2P networks; such as the mobile phone, computer, and other smart devices are connecting to the network. Moreover, the actual P2P communication will happen in this physical layer as well.

The second plane is the attributes plane. At this plane, the devices at the physical

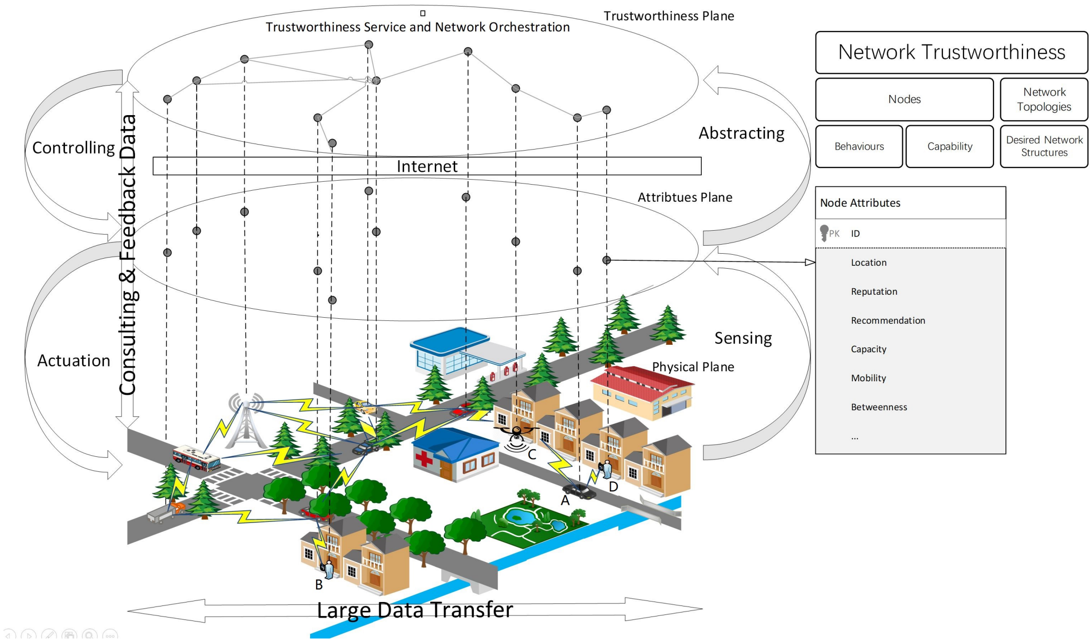


Figure 4.1: Architecture of Network Trustworthiness as a Service (NTaaS)

plane existed in the digital form that each device as a node in this plane. A node attributes table will be assigned to each node in the network, which has an ID as the primary key for the identification of the device in the physical plane. The remainder of the attributes in the table can include the 'location' of the node, 'mobility' which states whether the node is dynamic or static, the 'capacity' which used to determine whether the node is capable of the task, 'recommendation' which stores the feedback for its neighbours, or in another words, the direct trust value mentioned in the related work section. Depending on the objectives of the network, the attribute selected to store in the tables can be adjusted accordingly.

Finally, the top plane is the trustworthiness service and network orchestration plane, which is referred to as the T plane. In this plane, the nodes' attributes in the attribute plane will be abstracted and evaluated by this T plane. The network trustworthiness evaluation considers both the trustworthiness of nodes and the trustworthiness of the network topologies. The trustworthiness of the node involves the cooperativeness and the capability. For example, for capacity, if the node has enough resources to complete the tasks, such as CPU, memory, energy, etc. For cooperativeness, whether it completed the tasks faithfully in a previous interaction. After the T plane concludes the evaluation results on the target nodes, T plane sends back these results to the requesting node. The requesting node will decide the target nodes it should include in its secured neighbourhood based on its own threshold and the results from the T plane. Then the requesting node should update its secured neighbourhood to the T plane, so the T plane can have the latest topology for the network. Thus, the T plane can also evaluate the trustworthiness of the network topology, if there is any potential threat has been identified or network failure. If there is any threat, T plane will look around for nearby nodes for recruitment to remedy the network topology. We limited this evaluation scale to the single cellular tower coverage area. Otherwise, the amount of data and complexity would be overwhelming. There can be two different scenarios for this service framework, which are ’general scenarios’ in which the Internet is available and ’disaster scenarios’ in which the Internet is not available. There are two examples given in below to explain how this framework works.

4.3.3 NTaaS in General Scenarios

The flowchart in Figure 4.2 on page 102 shows the generate scenarios, in which the requesting node is the node sends out the request for the trustworthiness evaluation. The requesting node sends out the evaluation request to the T Plane. Then the T Plane starts the evaluation processes for the target node. The evaluation processes also enquire the available nodes in the area for the feedback on the target node. Once the evaluation has been done by the T Plane, the results are sent back to the requesting node. Then the requesting node based on its own trustworthiness threshold, determines whether the target node is ’trustworthy’. The requesting nodes form its local topology with only trustworthy nodes. It also updates this topological information back to the T

Plane. The T Plane forms a topology of the complete network based on this topological information. Then the topology evaluation will be performed. If there is any threat found in the network, the T Plane looks for the recruitment from the available nodes again. Take an example from Figure 4.1; there is a node A with neighbours nodes B and C in the A plane, and the node D, which is a human holding a mobile want to send some data to another person at the other end of the map. Node D want to know if node A is trustworthy for communication for its tasks. Thus, node D requests the T plane through the Internet to evaluate the trustworthiness of node A with its requirement (trustworthiness threshold). The T plane will abstract recommendations from nodes B and C on node A. Then T plane starts the evaluation process, and sends the results back to node D. Node D compares the results with its threshold and decide if node A should be included in its routing path. Then node D would update this decision back to the T plane. If node A is trustworthy, then this node would be included in T plane's orchestrated network; otherwise, it shall be excluded. Moreover, the T plane also evaluates the trustworthiness of the network topology. If there is any potential threat, network failure, or disconnected node has been identified, the T plane will look around for nearby nodes for recruitment to remedy the network topology; such as the UAV in Figure 4.1 on page 99 which has been recruited for the disconnected devices.

4.3.4 NTaaS in Disaster Scenarios

For the 'no Internet' scenario examples, the NTaaS will install a certificate on the trustworthy devices in the P2P network while the Internet is still available. Thus, even without the Internet, the P2P devices can still recognize each other through the certificate installed on the devices. Therefore, the devices in the P2P know who is trustworthy and can be confidently communicated with as shown in Figure 4.3 and in Figure 4.4 is the flowchart in disaster scenarios.

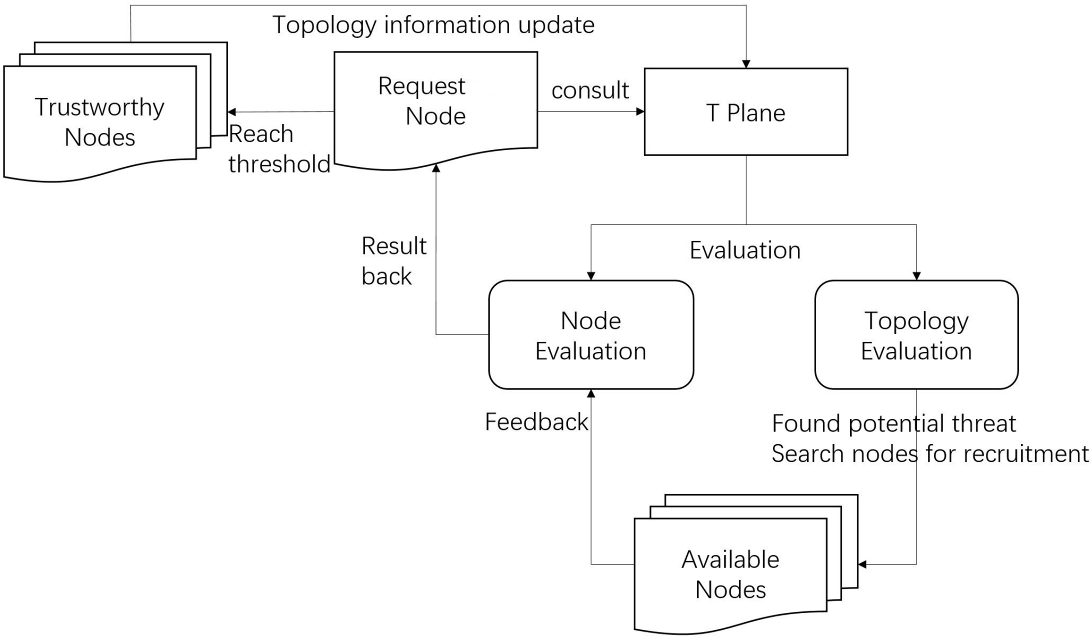


Figure 4.2: NTaaS Flow Chart in General Scenarios

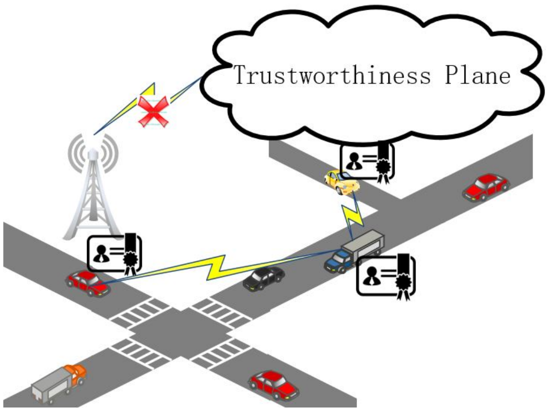


Figure 4.3: NTaaS in Disaster Scenarios

In the next section, the network trustworthiness T evaluation framework is proposed for the NTaaS.

4.4 Network Trustworthiness Measurement Framework

In the previous sections, we have defined a new term to be known as ’network trustworthiness’. It describes the confidence level of a network in being capable to complete its

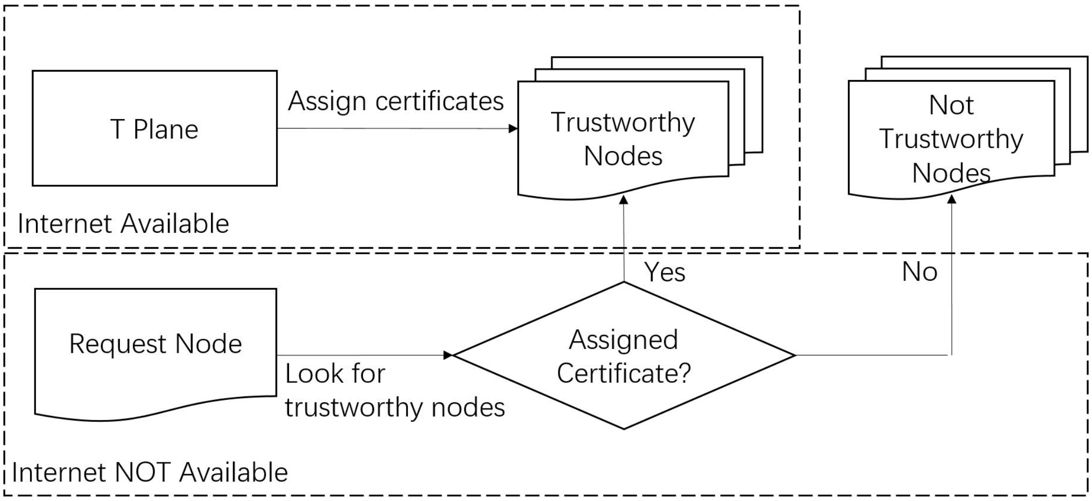


Figure 4.4: NTaaS Flow Chart in Disaster Scenarios

objectives and maintain its QoS under any expected or unexpected circumstance. This can be a guideline or benchmark for the network designer to better achieve a quality of works in different scenarios, or for the network management system to identify the potential threats or weakness in the network. Then we proposed the NTaaS as the platform for this network's trustworthiness evaluation. The NTaaS includes a node trustworthiness evaluation and a network trustworthiness evaluation. Although there are many existing works on node trustworthiness evaluation already. However, there is no existing on the network trustworthiness evaluation. Thus, this thesis focuses on the network trustworthiness evaluation. In this section, the network trustworthiness evaluation framework is introduced for the NTaaS.

First of all, we need to understand the attacks and network failures that can happen in the network. In general, there are two types of the network failures or attacks, which are 'intentional' and 'unintentional'. The intentional attacks are targeting at the maximum possible damage to the network or its objectives by malicious parties. These malicious parties analyse the network topologies and security defences, so as to identify the weakness of the topologies, and a back door to bypass the security defences. In other words, they always try to achieve the maximum damage with the minimum.

Unintentional attacks are normally referred to as network failures or internal user errors (human errors). These attacks can occur randomly in the network, and the damage level is varied as well. The framework T value is the guideline on how to evaluate the network trustworthiness in different scenarios.

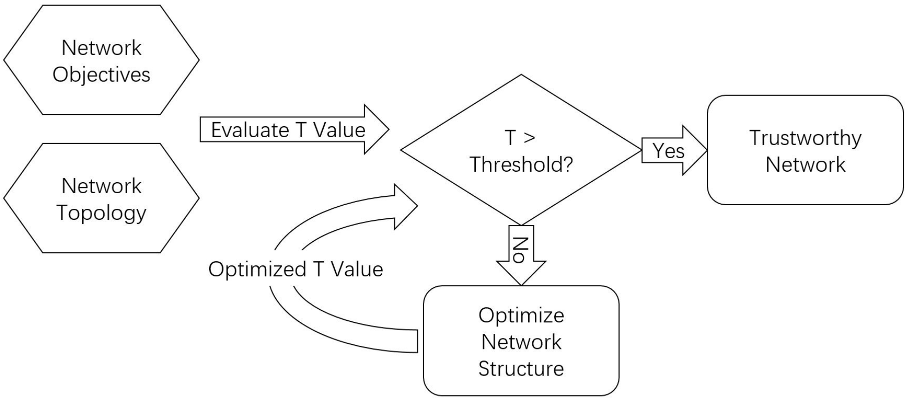


Figure 4.5: The Proposed Assessment Framework for Network Trustworthiness

Figure 4.5 is the proposed analytical framework to evaluate and quantify the trustworthiness of P2P networks for NTaaS. The trustworthiness value T can be a single metric, or combined with multi-metrics, which depend on different scenarios' network objectives. Moreover, the metrics for T can be different as well depending on the network objectives, so the trustworthiness for the different networks, which they can have different objectives, purposes, and priorities can be evaluated by this simple framework. When there is more than one metric are selected to evaluate the network trustworthiness, the trustworthiness value T_G is used in Equation 4.1, the G is standing for network G. Where m is the metric, which this measurement metric is related to particular network objectives such as network availability. In this thesis, we use the clustering coefficient metric to evaluate and quantify Simmelian Ties in the network. Regarding the traffic control, we can use effective size or Simmelian brokerage, etc. The n represents the total number of the required metrics.

$$T_G = (m_1, m_2, m_3..., m_n) \tag{4.1}$$

At the end of network trustworthiness evaluation, we only need one final value to represent the overall network trustworthiness rather than a set of different metrics' values. In such case, for the combination of all metrics, which are selected to evaluate the network trustworthiness, weight factors is one of the simplest ways to do it. It is shown in Equation 4.2. As we can see in Figure 4.5 on the preceding page, the network topology and objectives combine together to become the T value. This means the topological metrics are used to evaluate the desired and unwanted network structures in the network; these structures are related to this network's objectives. Then the weight factor for the different metrics represent the relevant level of the metrics to the particular network objective, or priority of the corresponding network objective compared to others. The w_i is the weight factor which represents the relevant level of metric i to particular network objective and the priority of this networking objective compare to others in Equation 4.2. For example, if metric A is completely irrelevant to any network objective, then $w_A = 0$. When $w_A = 1$, it means metric A can fully represent this particular network objective and this network objective is the only network objective for the network.

$$T_G = \sum_{i=1}^{n} w_i m_i, 0 \leq T_G \leq 1 \tag{4.2}$$

The T_G need to be a value between 0 and 1 for the comparison, where $T_G = 0$ represents the network G as a completely not trustworthy network and $T_G = 1$ represents the network G as a completely trustworthy network. Thus, when the metrics value is calculated, they also need to be normalised into a value between 0 and 1 to become m in Equation 4.1 and 4.2. The most common way to normalise the data set is the min-max normalisation as Equation 4.3 on the following page.

$$m' = p + \frac{(m - m_{min})(q - p)}{m_{max} - m_{min}} \tag{4.3}$$

The m' is the m value after normalisation, for the p and q they are representing the range of the data set after the normalisation. In this case, the range should be between 0 and 1 so that $p = 0$ and $q = 1$. Then we can have Equation 4.4 below.

$$m' = \frac{m - m_{min}}{m_{max} - m_{min}} \tag{4.4}$$

The reason for the normalisation on each metric m rather than on final T value is that different metrics can have a different value range. The metric with a bigger value range can compromise the metric with a smaller range. For example, some metrics such as the clustering coefficient range between 0 and 1, and some can be 0 to unlimited, such as the effective size and Simmelian brokerage which was introduced in the last chapter. In such cases, compared to the number 100, a value that no matter is 0 or 1 is not making too much difference on the final T value. Moreover, when there is more than one metric to be used to measure T and combined with weight factors, there is a chance that some metrics for the network have unacceptably low values but the final T results can be comprised by other metrics, which have a very high value. In such a case, we can set up the upper boundary and lower boundary threshold like Equation 4.5. In Equation 4.5, th_{li} is the lower bound threshold for the ith metric, and th_{ui} is the upper boundary threshold for the ith metric.

$$T = \begin{cases} \sum_{i=1}^{n} w_i m'_i, 0 \le T \le 1, th_{li} \le m'_i \le th_{ui} \\ 0, m'_i < th_{li} \\ 0, m'_i > th_{ui} \end{cases} \tag{4.5}$$

Moreover, most of the time, the actual networks in the real world are in different sizes, which means they will have a different number of nodes and connection links. Many topological metrics such as the clustering coefficient, they are not effective to compare with different network sizes. This has been well studied in the last chapter in Section 3.5.4 on page 76. Also taking the efficiency of a network (Ellens & Kooij, 2013) as an example, it uses the average hop count to measure which as shown in the following Equation 4.6.

$$E = \frac{2}{n(n-1)} \sum_{j=1}^{n} \sum_{k=j+1}^{n} \frac{1}{d_{jk}} \tag{4.6}$$

In the equation above, n is the number of the nodes in the network, and d_{jk} is the hop count for the shortest path between nodes j and k. It simply calculates the hop count of shortest paths for all possible node pairs in the network. The larger of the n, the smaller of the E will be. In other words, the larger size of the network, the lower the efficiency of the network will be. To normalise this, we can have following Equation 4.7.

$$E'_G = \frac{n_{min}}{n_G} \times E_G \tag{4.7}$$

In the equation, n_G is the number of the nodes for network G. This normalisation is only works for the efficiency metric. For other metrics, this normalisation would not be suitable, as not all the metrics count the hop count. In such case, it is impossible to have one universal normalisation equation for all the metrics; different metrics should have a different way to do it as well.

Once the network trustworthiness T value has been calculated, a threshold T value is needed, so as to have a benchmark to compare with. Therefore, it can determine whether the networks are trustworthy or not. If the network is determined as not trustworthy, then NTaaS will need to modify the network topology until its trustworthiness T value reaches the threshold. The threshold T value is different in different scenarios and circumstances. There is no universal threshold as different parties have different standards.

There are five common ways to optimize the network topology to the desired topology, ie, add the links to the network, remove the links from the network, add nodes to the network, remove nodes from the network, rewire connection link from a node to the other nodes. By changing the network topologies to the desired structures, we can increase the trustworthiness T value mathematically, as defined in the Equation 3.1 on page 56 to represent the network topology G. There are topological metrics such as the algebra connectivity which can be used in this G graph function to identify the weak point in the network, then optimize it through one of five ways mentioned above. By adding a link in the network to connect the node i and j, a value of $e_{i,j}$ needs to add into the edge set E_G for the network G, vice versa for the removal of a link. As adding a node will be more complicated, a value of v_i needs to add to the node set V_G. Moreover, if the node degree for the new node i is 2, then there should be two e add into the edge set E_G, such as $e_{i,j}$ and $e_{i,k}$, which are the two nodes j and k node i is connecting to.

For the remediation of the network topologies, depending on the different topological metrics are selected for the network trustworthiness evaluation in different scenarios, there are different ways to identify the position to add or remove the links or nodes, or rewire the links. The study by H. Wang and Van Mieghem (2008) using the algebra connectivity, the 2nd smallest eigenvalue to identify the weakness in the network. The remediation goal for this studies is to achieve the maximum increment on algebra connectivity by adding a link. The author H. Wang and Van Mieghem (2008) is try to infer the solution from the algebraic connectivity equation itself. Thus, if the metric is clustering coefficient, the remediation goal will be different, and the remedy method will be different as well.

Moreover, the network topology can be changed over time, such as a mobile ad-hoc network. It also can be caused by different events, such as node failure, move away, compromised, etc. In such case, the T value need to have a timestamp on it to state when it was evaluated, like T_{ti}, where the ti is representing the particular time period. When comparing the two networks like A and B with the T_{ti}, it should satisfy ti_A = ti_B.

In the next section, this chapter will provide a distributed P2P network scenario for the demonstration of the network trustworthiness evaluation framework, and provide simulation studies to validate the T value as well.

4.5 Evaluating the Network Trustworthiness in P2P with T Value

In the last section, the network trustworthiness evaluation framework has been proposed as a guideline to evaluate the network trustworthiness based on their network objectives. In this section, we are going to set up a distributed P2P network scenario for the demonstration of the network trustworthiness evaluation framework.

In general cases, for a distributed P2P network, the network availability is the top priority of the network objectives. As the main objectives of the network is a collection of data and then forward to the host. In distributed P2P network scenarios, the P2P networks are distributed networks so that every node has their own routing algorithm to make their own decision as to what to do in the network; in other words, self-organised. Compare to the traditional network with a central management, it is hard to control nodes' behaviours in a P2P network, and that is why the distributed trust management algorithm was introduced to let nodes in the network monitor each other; so the bad nodes can be discovered and punished, then the good nodes can be rewarded. From a global point of view, to achieve this, a good network structure is required as well, so the nodes in the network can be monitored and constrained. Thus, the NTaaS is purposed to provide an evaluation service from local (node) and global (network) aspects. As discussed in the last chapter, the Simmelian Ties can engender trust establishment, and enforce the social norm environment by introducing a third party. In such case, Simmelian Ties are believed to help improve the distributed trust management algorithm performance. Thus, it is believed the Simmelian Ties characterized structures can make the network more trustworthy in this case by increasing the resistance of unwanted behaviours, such as selfish, malicious, etc. For the Structural Hole characterized structures, as its lack of constraint and monitoring of the nodes in this position, it is believed that it makes the network less trustworthy.

In such case, the Simmelian Ties in the network are more preferred as it provides redundant routes and the third party for the reputation system. The more Simmelian Ties in the network, it is believed that the more trustworthiness of the network. However, the Structural Hole characterized network structure has a negative impact on the positive impact from Simmelian Ties characterized network structure, The Simmelian Ties can only measure the network trustworthiness accurately when there is no Structural Hole in the network. In such cases, there is a need to have Structural Hole included in the measurement metrics. The Structural Hole should be avoided because of its lack of a redundant route and limited the performance of trust-based routing algorithm.

In this section, we are using our proposed framework to evaluate the trustworthiness of network availability for the packets delivery ratio and better performance of the trust-based routing algorithm. To evaluate the Simmelian Tie and Structural Hole in the network, there are metrics recommended in sociology studies which have already introduced in the last chapter; the clustering coefficient for the Simmelian Ties measurement. However, from another perspective, the studies in 3.5.4 on page 76 have also found out that the average clustering coefficient is not efficient for the comparison of networks of different network sizes. In the following, we have proposed a solution to overcome this problem.

4.5.1 Simmelian Ties and Betweenness

In the networks, some nodes are more critical for the network availability and some are less important due to the positions they are at. Normally, the nodes at the centre of the network are believed to be more important to the network. For example, node A is located at a position that in between another two nodes; when these two nodes need to communicate, the communication needs to pass through node A. The more node pairs in the network for which node A is in between, then the more node A is believed to be more critical to its network for the network availability as once node A is under attacks or failure, the more traffic will be affected by this attack or failure. On the other side, node B is at the border of the network, which it is not in between of any node pairs. In other words, there are not many nodes need to route through node B. When it is under attack, it cannot cause much damage from this attack. In such cases, the average clustering coefficient cannot reflect the critical level for each node in the network, as the average considers all the positions in the network are equally important. In this case, the proper weight factors set up to represent the importance level of the nodes in the network will be necessary, so the clustering coefficient for each node would have the weight factors to reflect their importance level.

There are many existing topological metrics for measuring the centrality of the nodes in the networks, and the network Betweenness is one of the typical centrality measurement topological metrics. It calculates the amount of the possible traffic would need to route through the particular node in the network, so as to identify the centrality of the node in the network. In such cases, we can use the Betweenness metric as the weight factors to synthesise the clustering coefficient, rather than using the average to synthesise all nodes' clustering coefficient for the evaluation of the network trustworthiness. The weight factors in this scenario, which are also known as node Betweenness for each node to reflect the critical level of their position in the network. There are node Betweenness and link Betweenness, as in this scenarios is to measure the centrality of the node, so we are focusing on the node Betweenness. To calculate the node Betweenness value, we have the equation in Equation 4.8.

$$b_i = \sum_{j,j\neq i}^{n_i} \sum_{k,k\neq i,k\neq j}^{n_i} \frac{l_{jik}}{l_{jk}} \tag{4.8}$$

Where b_i is the Betweenness value for node i, l_{jk} is the number of preferred paths between nodes j and k, and l_{jik} is the number of preferred paths between node j and k, which route through node i. The preferred path between two nodes can be defined differently depending on different routing algorithms. An algorithm only selects the shortest route to forward as this preferred path means the shortest path. Or if the algorithm selects the nodes with the shortest distance to the sink then the preferred path will be selected by this, but this path will not necessarily be the shortest path. In such cases where there is only one preferred path between nodes j and k, the variable l_{jk} is counted as 1. If the only preferred path between nodes j and k is routing through node i, then the variable l_{jik} is counted as 1 as well. In addition, if there are more than one preferred paths between two particular nodes, we count those preferred paths which travel through node i and divide by the total preferred paths between nodes j and k. For example, there are three preferred paths between nodes j and k, but only one route through node i. Then we count this as $1/3$ toward the Betweenness value for node i. Take the network topology in Figure 4.6 as an example. For the node 2 in the network, there are node pairs 1 to 2, 1 to 4, 1 to 5, 1 to 6, and 3 to 5 with preferred paths possibly needing to route through node 2. The preferred path we define as the node with the shortest distance to the sink. For node 1 to node 5, the preferred path can be either nodes 1, 2, 3, and 5, or nodes 1, 2, 4, and 5 as the distance from the nodes 3 and 4 to the sink are the same. As both paths route through node 2, so we have the count as 1 between nodes 1 and 5. This is the same to the nodes 1 and 6, 1 and 3, and 1 and 4. In such a case, now the total count so far would be $1 + 1 + 1 + 1 = 4$. For the paths between node 3 and 4, it can be either nodes 3, 2, and 4, or nodes 3, 5, and 4, as there are two possible routes but only one path route through node 2. In this case, the count would be $1/2 = 0.5$. So the final Betweenness would be $4 + 0.5 = 4.5$. It is the same logic for the rest of the nodes Betweenness in the network.

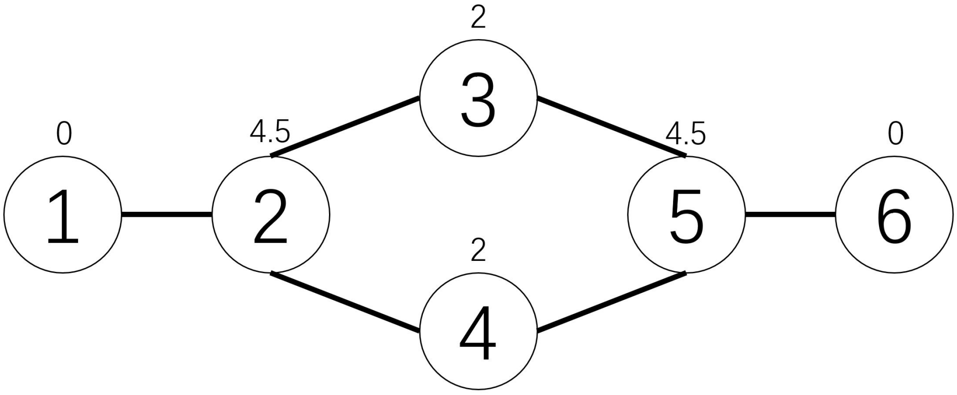


Figure 4.6: Network with Node Betweenness

Using the Betweenness calculation example, the result of Betweenness value range is shown in Equation 4.9.

$$0 \leq b_i \leq \frac{(n_G - 1) \times (n_G - 2)}{2} \tag{4.9}$$

The n_G is the number of the nodes in network G. It means the Betweenness value is between 0 and the maximum possible of different node pairs in a network with given the number of nodes n_G in the network G. However, for the weight factors, it requires that all the weight factors sum up to 1, which the w_i should satisfy as a condition as shown in Equation 4.10.

$$\sum_{n_G}^{i=1} w_i = 1 \tag{4.10}$$

Thus, the Betweenness value for each node needs to be normalised to become weight factor and the equation in Equation 4.11.

$$w_i = \frac{b_i}{\sum_{j=1}^{n_G} b_j} \tag{4.11}$$

In this case, rather than using the average function to synthesise the clustering coefficient with no reflection of the importance level of the nodes in the network, the weight factors w_i in Equation 4.11 can synthesise the clustering coefficient of all the nodes in the network in a more accurate way. This can also help the clustering coefficient overcome the accuracy issue on a different network sizes comparison. Finally, we can have the weighted clustering coefficient C_{wG} for network G as shown in Equation 4.12.

$$C_{wG} = \sum_{i=1}^{n_G} c_i \times w_i \tag{4.12}$$

In the last chapter's simulation studies, from the second scenario II section 3.5.4 on page 76, we have found the average clustering coefficient is inefficient for a comparison of network topologies in different network sizes. As we just proposed the weighted Clustering Coefficient to quantify the Simmelian Ties in the whole network, so now we have the weighted Clustering Coefficient to compare the packet loss results in networks 5 to 9. The results are shown in Figure 4.7 on the next page and Table 4.1.

Table 4.1: Average Clustering Coefficient vs. Weighted Clustering Coefficient

Network	**5**	**6**	**7**	**8**	**9**
Average Clustering Coefficient	0	0.03	0.55	0.47	0.53
Weighted Clustering Coefficient	0	0.07	0.27	0.35	0.43
Total Packet Loss	1760	141	99	64	65

As can be seen in the Figure 4.7, as the packet loss number decreases in the different networks, the average clustering coefficient is going up and down for these networks. For the new proposed weighted clustering coefficient, the trend of increase is shown very obviously. Though the network 8 has a lower weighted clustering coefficient and

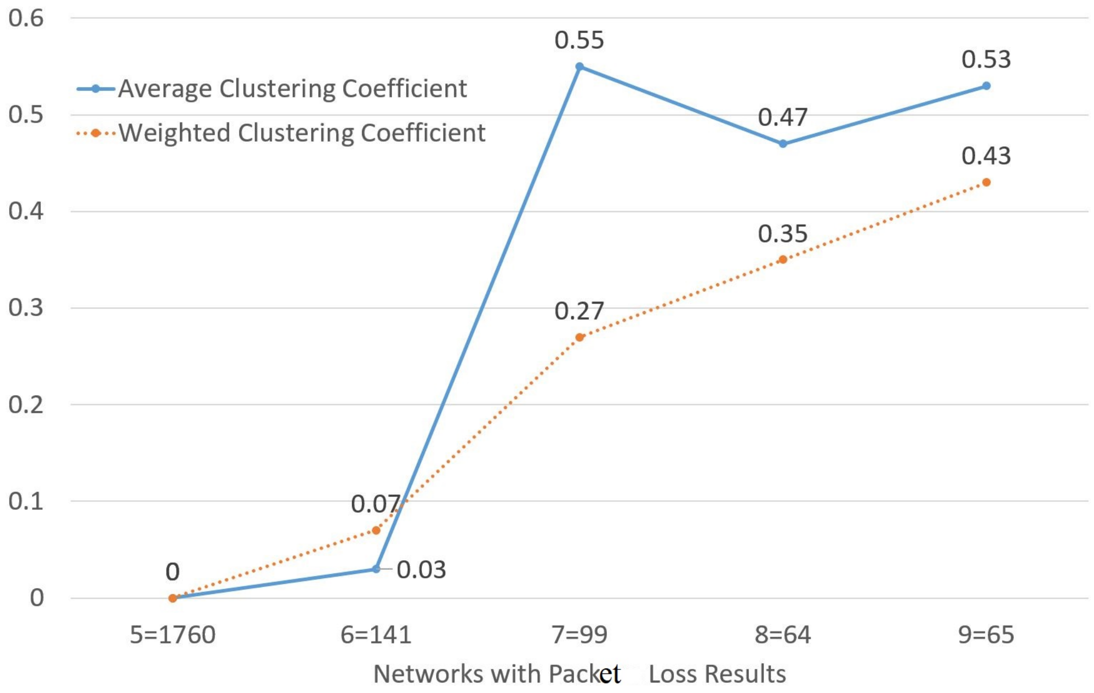


Figure 4.7: Average Clustering Coefficient vs. Weighted Clustering Coefficient in Packet Loss

low packet loss number as well compared to the network 9, the packet loss results between these two networks only have 1 packet difference, so that we can consider these two results are the same. However, network 8 has a lower weighted clustering coefficient, which means it should have a lower packet loss result. These two results are the same is because when the Simmelian Ties reach a certain number in the network, the beneficial effect will become less and less. As these certain number of the Simmelian Ties have already provided enough redundant routes and third parties to benefit trust establishment and convergence processes. The additional Simmelian Ties would not do any more good for the efficiency of the DTEGR routing algorithm. Therefore, at the end, the new proposed weighted clustering coefficient is able to measure the Simmelian Ties in the network as a whole more precisely with networks of different sizes compare to the average clustering coefficient.

Later we have proposed and validated a better way to measure the Simmelian Ties of the networks, which can be compared in networks of different sizes. We have confirmed to use this new metric, 'weighted clustering coefficient' to evaluate the Simmelian ties in the network as a whole to be part of the network trustworthiness T evaluation. However, the Structural Hole still requires to be detected, so the final T value can have accurate results. As discussed in the last chapter, the Structural Hole has the negative impact on the trust establishment and convergence of the trust-based routing algorithm. Also, it has a negative impact on the positive impact of the Simmelian Ties. Thus, the detection of a Structural Hole in the network is very crucial for the accuracy of the T value results. In the next section, we have proposed a method to detect the Structural Hole in the network.

4.5.2 Physical Structural Hole and Logical Structural Hole Locator Algorithm

In chapter 3, section 3.4 on page 60, there are two metrics were introduced to evaluate the Structural Hole in the network, which are the effective size and Simmelian brokerage. However, these two metrics have been confirmed that in the simulation studies scenario I in section 3.5.3 on page 69, they are not able to locate the actual Structural Hole in the network. They can only evaluate the Structural Hole in one hop distance, which means in one hop distance the node is Structural Hole, and in 2 hops or 3 hops or even longer, may not be a Structural Hole. The Structural Hole has a negative impact on the trust establishment and convergence as the Structural Hole is acting as the gateway in a distributed P2P network environment, once it has been attacked, the network will be disconnected. So it is crucial to locate any Structural Hole in the network as it would be a potential threat in the network to attract the attacks.

Moreover, there are two types of the Structural Hole, one is physical Structural Hole, and the second one is logical Structural Hole. The physical Structural Hole is the only position connecting different unrelated clusters or networks. The definition of the logical Structural Hole can be vary; depend on the actual scenario, the definition can be different. The general term for a logical Structural Hole is the node at a position which it is not a physical Structural Hole position, but which still controls most of the traffic flow as the alternative is either too far away or too slow, etc. In other words, if the node at the logical Structural Hole is disabled, the network might still stay connected, but the QoS will be significantly degraded. To be able to detect both a physical Structural Hole and logical Structural Hole in the network, the Structural Hole Locator (SHL) algorithm is proposed to detect any Structural Hole in the network. We use three network topologies to explain how the SHL algorithm works.

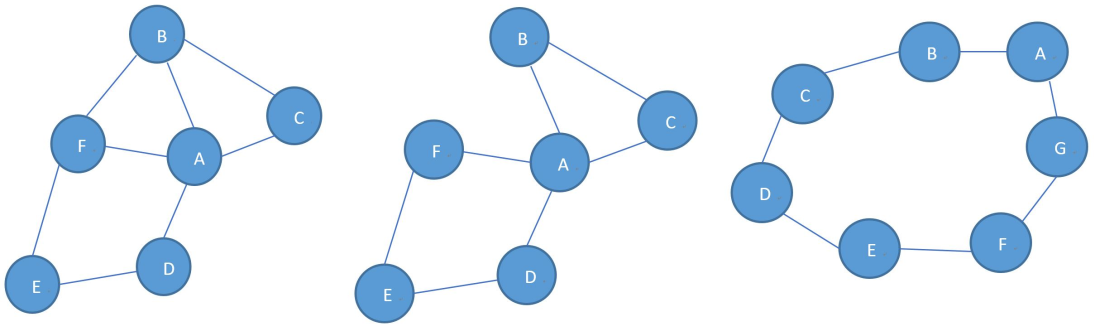


Figure 4.8: Sample Networks with Physical Structural Hole and Logical Structural Hole

As shown in Figure 4.8, there are three networks to explain the physical Structural Hole and Logical Structural Hole. Take the left network as an example. For the algorithm to determine whether the node A is at Structural Hole position, first of all, it will disable the node A in the network, then try to find the paths from node B (or any node which is a 1 hop distance neighbour of A) to the rest of node A's one-hop neighbours. In the left-hand side network, the combinations are nodes B to C, B to F, and B to D. Node B is able to reach nodes C, F, and D after node A is disabled. Thus, the node A in the left network is not at the Structural Hole position. In the centre network, after disabling the node A, node B can only reach node C and disconnect from nodes F and D. In such a case, node A in the middle network is at the physical

Structural Hole position. Moreover, in the right-hand side network, once node A is disabled, node B is still able to reach node G, but rather than a 2 hops distance through node A, now it cost 5 hops distance to reach it. It such a scenario, we can also say node A is at a logical Structural Hole position as a failure on node A can cost a lot more in communication distance between node B and G. In order to detect the logical Structural Hole in the network as well, the SHL algorithm can set up a threshold such as the hop count number, traffic load, etc. Depend on the different scenarios, the threshold metric would be different accordingly. In the right-hand side network scenarios, we set the threshold as 4 hops count, which means if node A is stopped functioning, any of its node pairs such as nodes B and G, C and F, D and G, etc. have longer than 4 hops distance to each other, then the node A in the network is at the logical structural hole position.

For the SHL algorithm to be able to look for the routing path, first of all, we need to define the network as a mathematical equation. As introduced in the last chapter, the network can be defined as $G(V, E)$. The Adjacency matrix with size of $|V|x|V|$, where the $|V|$ is the number of nodes in the network, it is used to specify the nodes and connection links in the network G. The equation for Adjacency matrix to represent the right-hand side network is shown in Equation 4.13. At the first row is node A, the first column at first row is the link between node A itself, obviously, there would not be a link exists, so it is 0. Then the first row and second column, represents the connection between node A and node B, if this connection exists, then it should be 1, according to the right-hand side network topology in Figure 4.8 on the preceding page, node A and node B are connected, so the value in the matrix would be 1. In such logic, we can have the Equation 4.13.

$$G_{right} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \tag{4.13}$$

Take the Equation 4.13 as an example, and follow the pseudo codes at Algorithm 2 on page 121, the SHL algorithm want to check if there is any node in the network is at the Structural Hole position, first of all, it will look at the first row for any 1. The columns with the value 1 means it is the neighbour of node A's. The SHL will find out all the possible neighbour node pairs to find the path. In this case, the node A in network G_{right} only have two 1 in the first row of the matrix, which means only two neighbours and one node pair. Then SHL needs to find the path between node B (second column) and G (last column) using the $FindRoute$ function in Algorithm 1 on the next page. SHL will start the loop from the second row of the matrix, which it found the first column (A) and third column (C) with value 1. As node A is excluded, so SHL can only select the third row to continue the loop, and put the node C on the excluded list, as it has been routed. Then repeats the process until it finds node G. In this case, it has repeated five times in the loop to find node G, which it means the hops count distance is 5. As we set the threshold as 4 hops count, node A is at the logical Structural Hole position.

According to the algorithm 2 on line 8, SHL need to check each node for the Structural Hole in the network. For each node, SHL needs to process the same size of

Algorithm 1 Find Route Function

```
function FindRoute(SourceNode, SinkNode, targetNode)
    $G$ ← Network topology N x N Matrix *arraylist*
    $R$ ← Routed Node *arraylist*
    $NextHop$ ← Next Hop nodes *arraylist*
    $HopCount \leftarrow 0$
    Add $SourceNode$, $targetNode$ to $R$
    $Search \leftarrow true$
    $CurrentHop$ ← Current Hop nodes *arraylist*
    Add $SourceNode$ to $CurrentHop$
    while $Search$ is $true$ do
        for $j = 0$ to $CurrentHop.size() - 1$ do
            $v \leftarrow G[CurrentHop[j]]$
            for $i = 0$ to $v.size() - 1$ do
                if $v[i]$ connection existed and $i = SinkNode$ then
                    $HopCount \leftarrow HopCount + 1$
                    $Search \leftarrow false$
                    Break the loop
                else if $v[i]$ has connection and $i \neq SinkNode$ and $i$ is not in $R$ then
                    Add $i$ to $NextHop$
                    Add $i$ to $R$
            if $Search$ is true then
                $HopCount \leftarrow HopCount + 1$
        if $NextHop$ is empty then
            $HopCount \leftarrow -1$
            Break the loop
        $CurrentHop.Clear()$
        $CurrentHop \leftarrow NextHop$
        $NextHop.Clear()$
    return $HopCount$
```

Algorithm 2 Structural Hole Locator (SHL)

```
function SEARCHSTRUCTURALHOLE
    V ← number of nodes in the network Integer
    P ← physcial Structural Hole arraylist
    L ← logical Structural Hole arraylist
    $Thrs$ = 4 ← logical Structural Hole Threshold Integer
    Lv ← logical Structural Hole level arraylist
    G ← Network topology V x V matrix data arraylist
    for $i$ = 0 to $V$ do
        $n_i$ ← neighbours of node i arraylist
        for $j$ = 0 to $V$ do
            if $G[i][j]$ has connection then
                Add $j$ to $n_i$
        if $n_i$ is empty then
            Print: node i is disconnected!
            continue to next loop
        else if $n_i.size()$ = 1 then
            Add i to P
            continue to next loop
        Initialize: $breakloop ← false, temp ← Thrs$
        for $j$ = 0 to $n_i.size() - 2$ do
            for $k = j + 1$ to $n_i.size() - 1$ do
                $SourceNode ← n_i[j]$
                $SinkNode ← n_i[k]$
                $HopCount ← FindRoute(SourceNode, SinkNode, i)$
                if $hopcount = -1$ then
                    Add $i$ to $P$
                    $breakloop ← true$
                    break the loop
                else if $HopCount > temp$ then
                    $temp ← HopCount$
                    if $L$ is not empty and last item in $L$ is not $i$ then
                        Add $i$ to $L$
            if $breakloop$ is true and $temp = Thrs$ then
                break the loop
            else if $breakloop$ is true and $temp \neq Thrs$ then
                $L.remove(L.size() - 1)$
                break the loop
        if $breakloop$ is false and $temp \neq Thrs$ then
            Add $temp$ to $Lv$
    return $P, L, Lv$
```

data to identify any neighbours for this particular node. Then SHL will have the second loop between the neighbours for the source and sink nodes pair up that SHL needs to find the route for each pair. If the route does not exist, then the physical Structural Hole is located. Therefore, for the SHL complexity analysis, the runtime of SHL algorithm can be expressed as $O(V^2 \times n_i! \times F)$. V is the number of node in the network, which denoted in the algorithm 2 line 2, and n_i is denoted as the number of neighbours for node i. F is the runtime for find a route between source and sink. According to the algorithm 1, it starts from looking for source node's neighbour, for each neighbour identified, it looks for neighbours' neighbours again until the sink is found. Depend on the density and the average shortest hop count of the network, the number of loops will be different accordingly. The denser and higher average shortest hop count, the more loops, which means longer runtime.

Once the SHL has detected the Structural Hole in the network, the Structural Hole needs to be quantified so it can be added to the T value. As mentioned before, there are physical Structural Holes and logical Structural Holes. This can be considered as two circumstances. If the physical Structural Hole is detected, this network should be considered as not satisfied. As mentioned in Equation 4.5 on page 106, there are threshold values were set up for each metric in the T; if any one of these metrics is outside the acceptable range, then T would suggest this network is not trustworthy. When there is a requirement that people want to compare the networks, where all of them have a Structural Hole in the networks, the network trustworthiness evaluation T framework can use the Betweenness value on the Structural Hole position to compare, the higher the Betweenness value is the worse network topology is. This is because, a higher Betweenness value means more traffic will route through this node, which also means more damage can be done while this node behaves maliciously or malfunctions. If there are more than two Structural Hole in the network, then the sum of the Betweenness value on these Structural Hole position can represent the damage level of the Structural

Holes for this network, which is shown in Equation 4.14. The n_{ps} is the number of physical Structural Holes in the network G, and d_{pG} is the physical Structural Hole damage level for the network G. The b_i is the Betweenness value for the node i, which it is one of the Structural Holes.

$$d_{pG} = \begin{cases} b_i & n_{ps} = 1 \\ \sum^{n_{ps}} b_i & n_{ps} > 1 \end{cases} \tag{4.14}$$

Another case is the network only have the logical Structural Hole without any physical Structural Hole. We use the Equation 4.15 to deduct the value from the final T.

$$d_i = e^{-\frac{h_{jk}}{q}}, h > 0, q > 0 \tag{4.15}$$

d_i is the deducted value for logical Structural Hole node i, h_{jk} is the hop count between nodes j and k after node i is disabled, both node j and k are the neighbours of node i's. In other words, the h_{jk} is the cost when the logical Structural Hole is under attack, and q is the factor to control the speed of attenuation. For the logical Structural Hole, we need to set up a threshold first as mentioned before, so as to determine the logical Structural Hole. For example, if the threshold is 4 hop count, a node has a hop count 5 will make $h_{jk} = 1$; if a node has hop count 6, then $h_{jk} = 2$, and so on. Once d_i is calculated, we need to multiply d_i with node i's clustering coefficient, and node Betweenness as well. That is, the bigger the cost h it is, the smaller value of d_i, and more deduction on node i's clustering coefficient.

Once the d_i is calculated, this results should between 0 and 1. Rather than use another weight factor to synthesise it with the weighted clustering coefficient metrics like the Equation 4.2 on page 105 suggested, we multiply it as shown in the Equation 4.16.

$$T_G = \begin{cases} \sum_{i=1}^{n} c_i \times w_i \times d_i & h_{jk} \neq \infty \\ 0 & h_{jk} = \infty \end{cases} \tag{4.16}$$

The node j and k are both node i's neighbours when $h_{jk} = \infty$, it means node i is at the physical Structural Hole position, then the network G should rate as not trustworthy and the $T_G = 0$. Though the proposed framework suggests the use of weight factors to synthesise different topological metrics to become the T. This framework is more like a guideline in that when there is a better way to synthesise the metrics, then the better approach is selected. The key idea of the framework for the evaluation is recommending there is no universal metric to evaluate all the networks; in different circumstances the evaluation metrics should be adjusted accordingly.

So far, we have confirmed the evaluation metrics for the evaluation of trustworthiness on network availability, which is indicated by packet delivery ratio; trust establishment and convergence on the trust-based routing algorithm. In the next section, this framework needs to be validated through the extensive simulation studies. In the next section, the extensive simulation studies are set up to validate the accuracy of the T.

4.5.3 Simmelian Ties, Betweenness, and Structural Hole

This section considers extensive simulation studies were carried out to validate the accuracy of T value on the network trustworthiness evaluation. As mentioned above, we assumed the network availability in a distributed P2P network environment is the network objective. The network availability is focusing on the packet delivery ratio while the network is under attack. Thus, the efficiency of trust establishment and convergence which is the trust-based routing algorithm performance level is the key factor to more quickly detect any malicious behaviours in the network, so the trust-based routing algorithm can quickly find an alternative to avoid them. In this simulation study, the same set up as the simulation studies in Chapter 3, Section 3.5.3 on page 69 is used for all the scenarios in this section, except the network topologies and the source nodes and sinks.

Simulation Studies on T Value - Scenario I

In this scenario, we selected some of the network topologies from Chapter 3 for the simulation studies, which are networks 1 to 4 as shown in Figures 3.7 on page 70 to 3.10 on page 72. In addition, we have set up another network topology which is network 10 shown in Figure 4.9.

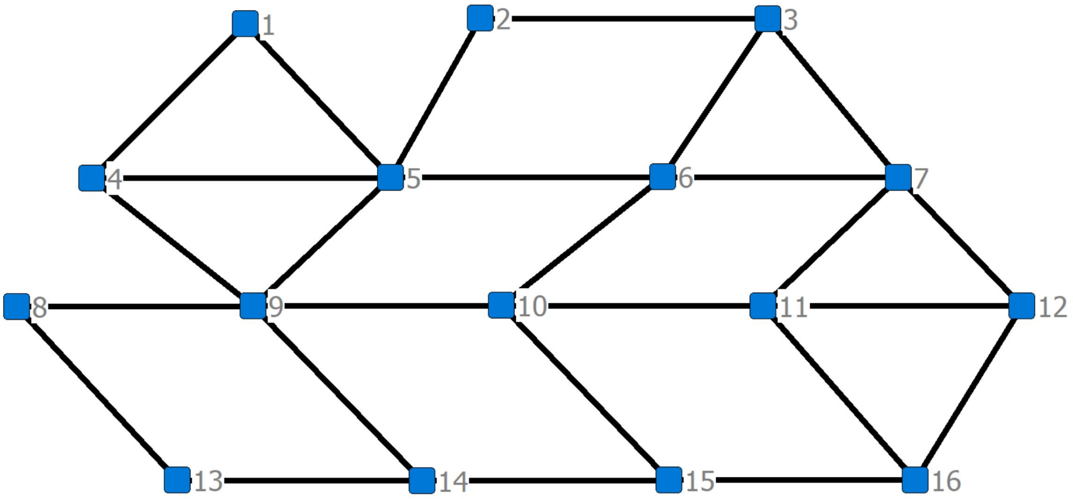


Figure 4.9: Network 10 for Scenario I

The simulation set up is same as the simulation studies in Chapter 3, Section 3.5.3 on page 69. The 900 packets are sent from nodes 1, 2, and 3 as the sources nodes; then the nodes 16, 14, and 13 as the destination accordingly by the order. The simulation results are shown in Figure 4.10 and Table 4.2. From the results, we can see that the DTEGR algorithm has achieved the highest packets loss number in Network 4 compared to the other four networks, and then Networks 10, 3, 2, then 1. The highest T value is Network 1, which means Network 1 is the most trustworthy network compared to the other four networks in network availability and improved trust-based routing algorithm performance. Then the second highest is Network 2, then 3 and 10. The Network 4 is 0 due to the existing physical Structural Holes in the network. This means our T is able to evaluate different network topologies correctly, and the packet loss numbers have reflected that. Compared to the average clustering coefficient and weighted clustering coefficient, the average clustering coefficient have shown the network 10's average clustering coefficient is higher than the Network 3's, however, this is not right as the packet loss number in Network 10 is higher than in Network 3. However, after applied the node Betweenness into the clustering coefficient to become a weighted clustering coefficient, the coefficient is significantly dropped on the Network 10, but it still slightly higher than Network 3. After considering the logical Structural Hole factor, which the T values have represented the correct measurement for the network trustworthiness. It is the same with the coefficient on the Network 4, as they did not consider the Structural Hole factors, though it has the highest coefficient, it also has the highest packet loss results as well.

Moreover, as Network 4 has two physical Structural Holes, we simply calculate the Betweenness for the two nodes on the Structural Hole positions, which are node 6 and 10. Node 6 is 26, and node 10 is 57.5. Therefore, we can state Network 4's trustworthiness is unsatisfied and with Structural Hole damage value (SHDV) at 26+57.5=83.5. when it is comparing to other networks which they also have the physical Structural Hole, through the comparison of the SHDV to reveal which network is more fragile.

The attacks on the logical Structural Hole normally cause the packet delivery latency to significantly increase, and even more packet loss as well. The increase in packet loss number is because when the nodes are looking for the alternative route, the longer distance (hop count) the alternative route it is, the more chances the nodes can encounter a malicious node or network failure again. The perfect example is in Network 10, though it has the higher weighted clustering coefficient compared to the Network 3, the DTEGR algorithm is able to achieve lower packet loss while the network is under

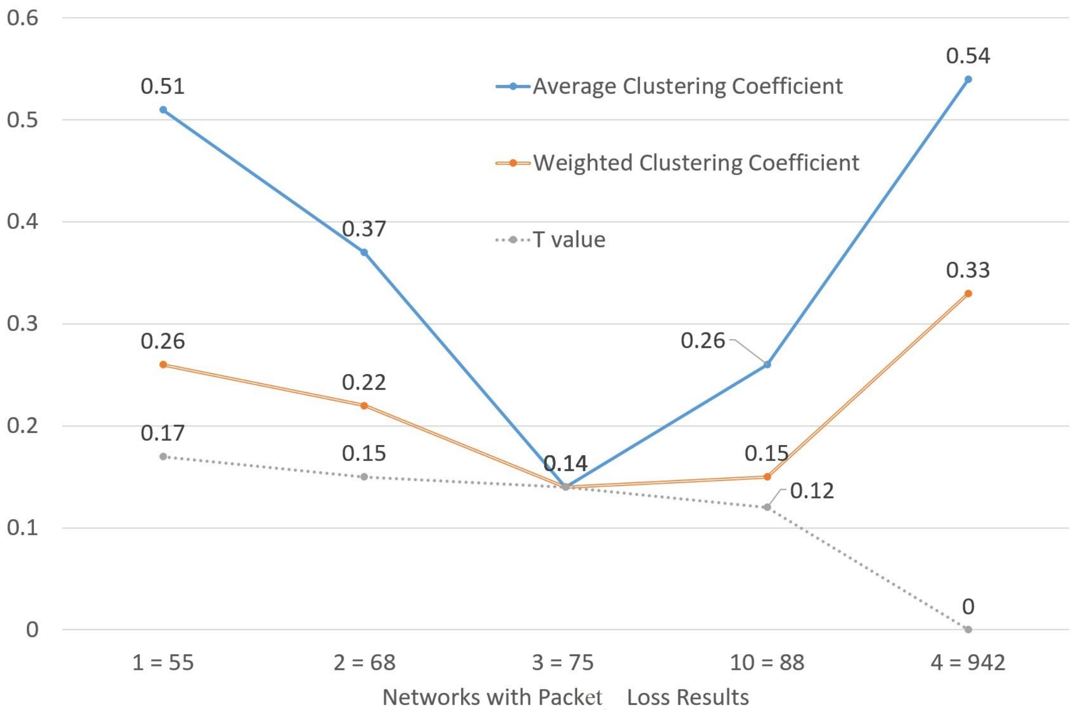


Figure 4.10: Average Clustering Coefficient, Weighted Clustering Coefficient, and T Value with Package Loss

attack. The simulation results have also validated the increase in packet delay while these logical Structural Holes were under attack; the results are shown in Figure 4.11 and Table 4.2. While the Structural Hole deduction is increasing, the trend is increasing the packet latency as well. The higher Structural Hole deduction means the more logical Structural Holes in the networks. When the deduction is 1, it means there is at least one physical Structural Hole in the network.

The new proposed SHL algorithm is able to detect any logical Structural Hole in the network with any definition, such as hop count, bandwidth, etc. In this scenario, the logical Structural Hole threshold is set up as a 4 hop counts. For example, in the Network 1, the SHL algorithm has detected the nodes 5, 6, 9, 10, and 11 are all at logical Structural Hole positions. Moreover, when node 9 in Network 4 was under attack, the packet latency was increased from 6ms to 12ms, and in Network 1 when node 11 was under attack, the latency was increased from 4ms to 10ms. For the Network 1,

Table 4.2: Average Clustering Coefficient, Weighted Clustering Coefficient, T Value with Package Loss vs. Package Latency

Networks		**1**	**2**	**3**	**10**	**4**
Average Clustering Coefficient		0.51	0.37	0.14	0.26	0.54
Weighted Clustering Coefficient		0.26	0.22	0.14	0.15	0.33
T Value		0.17	0.15	0.14	0.12	0
Total Packet Loss		55	68	75	88	942
Latency (ms) attack free	1 to 16	6.33	6.69	8.45	8.8	6.69
	2 to 14	4.48	4.5	6.38	4.45	6.34
	3 to 13	6.34	6.36	6.33	8.38	8.36
Latency (ms) under attack	1 to 16	9.4	9.25	11	8.9	28.12
	2 to 14	8.2	6.3	6.38	8.45	17.42
	3 to 13	8.2	8.89	6.36	8.46	14.64

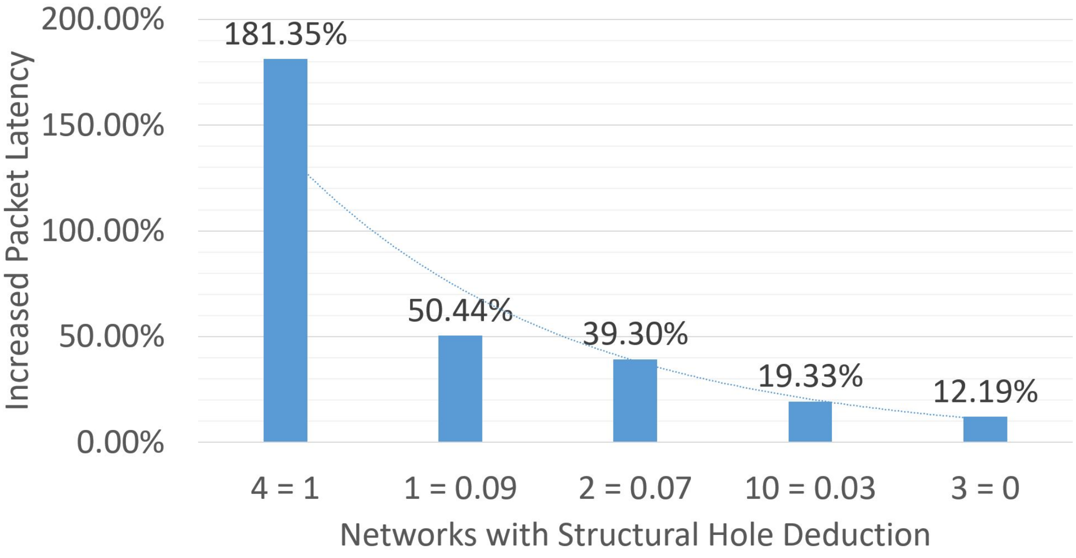


Figure 4.11: Structural Hole vs. Package Latency Increment

as mentioned in the above, there are five logical Structural Holes if we set the threshold as a 4 hop count. After we deducted the hole value from the clustering coefficient, the final clustering coefficient for the whole network reduction will be 0.17 rather than 0.26. This is the final network trustworthiness value T for the network trustworthiness as shown in Table 4.2. In the Network 3, the SHL algorithm cannot find any physical or logical Structural Hole in the network with the threshold of a 4 hop count, so there is no reduction in the clustering coefficient for the Network 3 which is 0.14. We then look at the simulation results which came back as shown in Table 4.2. In the table, from the packet loss point of view, the DTEGR algorithm can achieve better results in Network 1 than the Networks 2 and 3, this is because Network 1 has a higher clustering coefficient which means more Simmelian Ties in the network that they can provide more redundant routes. More importantly, they can also provide third parties for reputation requests, so the algorithm can detect the malicious behaviours even faster - these have been well discussed in Chapter 3. From the packet latency point of view, Network 3 can help deliver the packets faster while under the attacks compared to Network 1. We can see that in Table 4.2, when the traffic is under attacks in Network 3, only the route between nodes 1 to 16 increased less than 3 milliseconds. Compared to Network 1, there are five logical Structural Holes in the network when under the attacks, there are approx. 9 milliseconds latency in total in all three routes which increased in the latency, which is a huge increase compared to Network 3. From the Figure 4.11 we can see that, while Network 1 is under attack, there is a 50% increase in overall packet latency, compared to Network 3 where there is only a 12% increase. This validates that the logical Structural Hole does reveal the cost while the holes under attacks (in this case it is in packet latency).

Simulation Studies on T Value - Scenario II

We have designed five network topologies to validate the accuracy of the network trustworthiness T in the Scenario I. Next, we have set up five randomly generated network topologies to further valid of the network trustworthiness T. We used an OMNet++ simulation platform (Varga & Hornig, 2008) to generate five random network topologies with 50 nodes each. The connection link numbers are slightly different for the five networks, which are between 94 and 100. For the random network generation, OMNet++ randomly deployed the nodes in the certain areas. Then each node will select the three closest nodes to connect. It suggests the best size of the network is between

20 to 50 nodes; this is why we select 50 nodes as the size of the network. The network topologies are in the Figures 4.12 to 4.17.

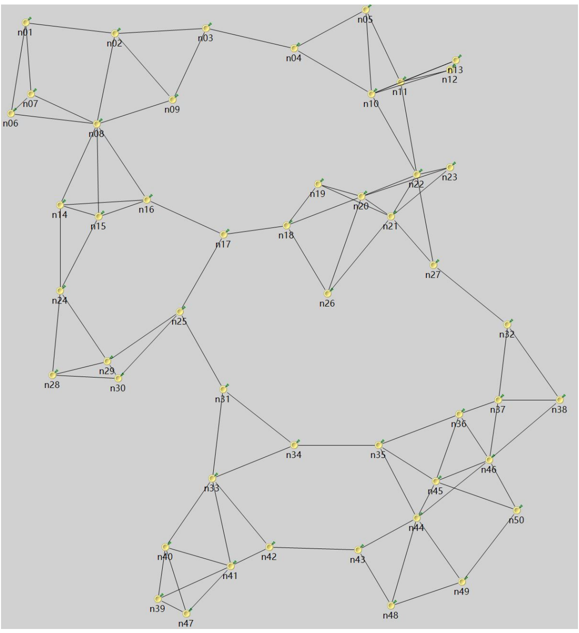


Figure 4.12: Network 11 for Scenario II

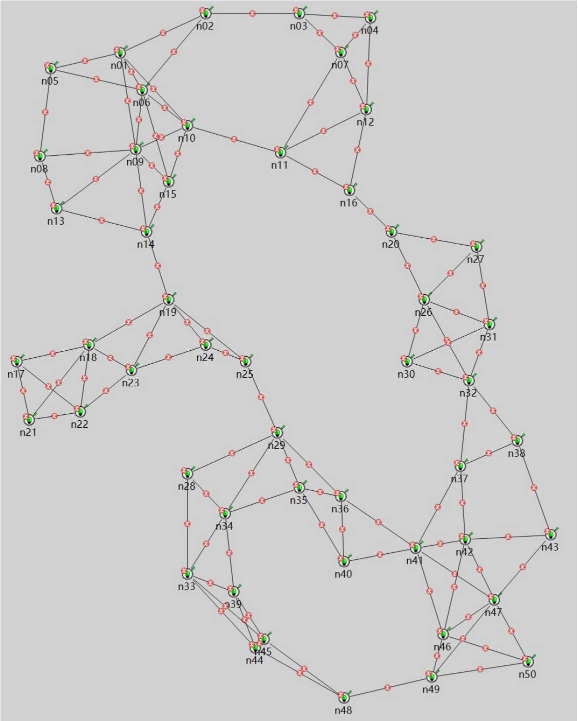


Figure 4.13: Network 12 for Scenario II

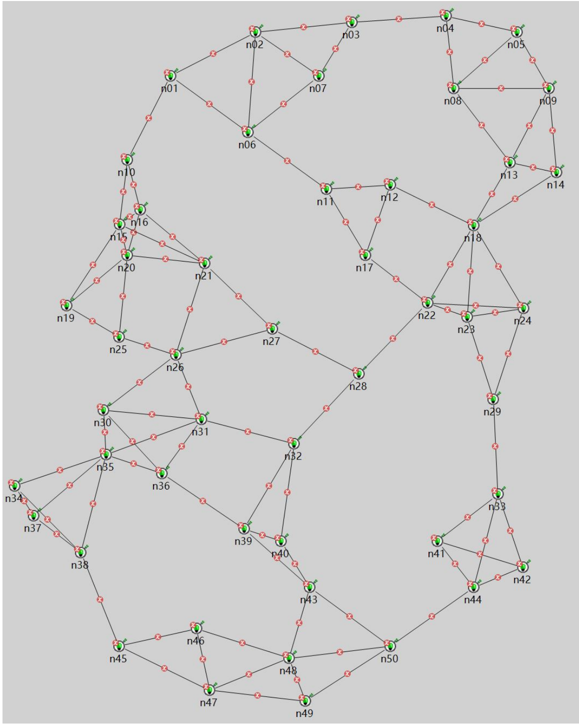


Figure 4.14: Network 13 for Scenario II

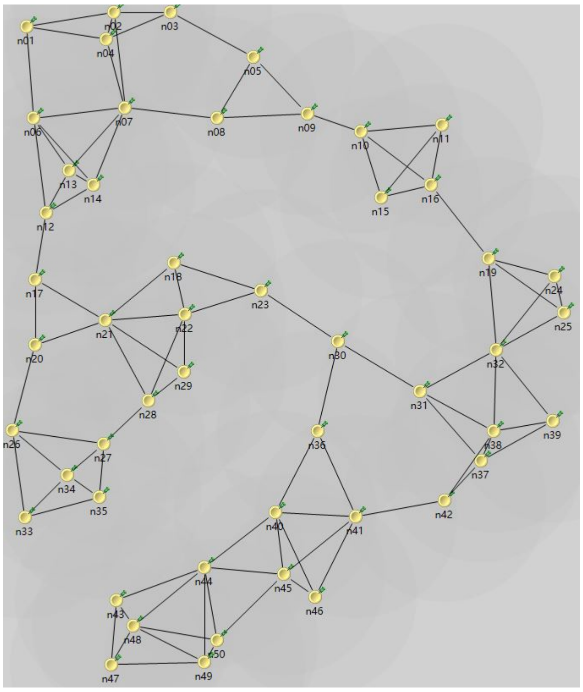


Figure 4.15: Network 14 for Scenario II

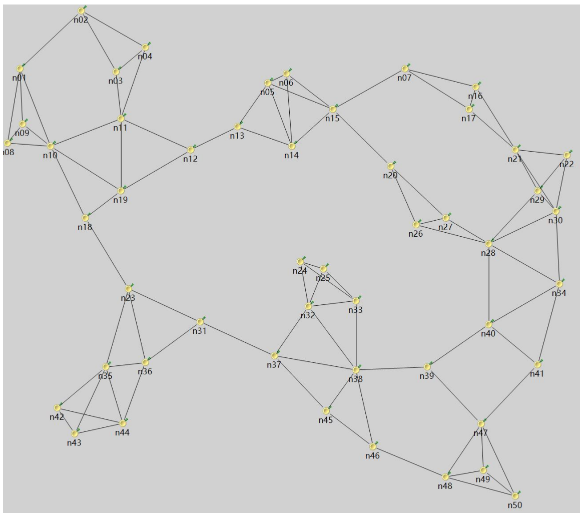


Figure 4.16: Network 15 for Scenario II

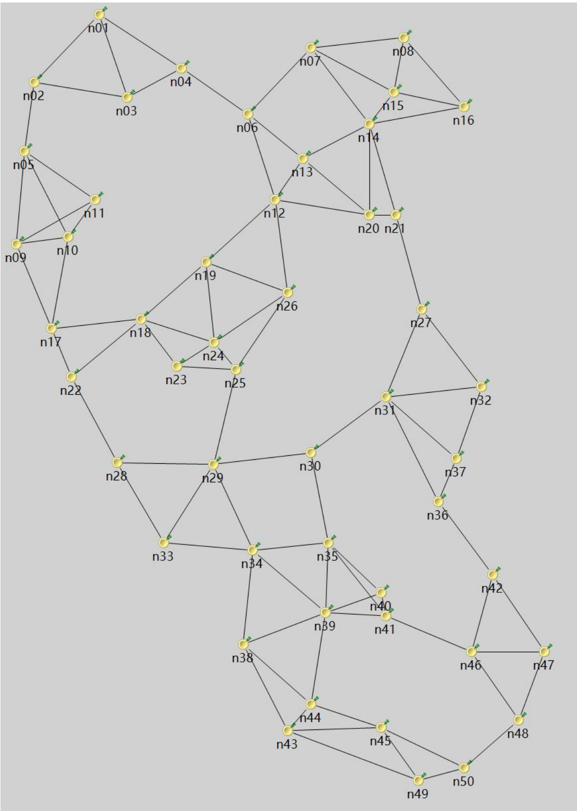


Figure 4.17: Network 16 for Scenario II

In each network from Network 11 to Network 16, four sources nodes are selected to send out the packets to different sinks, which they all located at the different positions in the network, so this traffic can cover the whole network as much as possible. In

Network 11, which is in Figure 4.12, the packets are directed from nodes 3 to 47, 8 to 49, 1 to 43, and 45 to 16. In Network 12, the directions are nodes 1 to 50, 4 to 21, 7 to 44, and 5 to 45. In Network 13, the directions are nodes 1 to 45, 2 to 44, 5 to 46, and 4 to 50. In Network 14, the directions are nodes 47 to 1, 11 to 49, 3 to 39, and 33 to 2. In Network 15, the directions are nodes 1 to 50, 22 to 43, 8 to 30, and 16 to 49. In Network 16, they are nodes 1 to 50, 5 to 47, 28 to 38, and 19 to 48. Each direction will have 300 packets send out with 128ms TTL. There is a malicious node set up in each network in each run. There are 8 nodes are selected in each network as sources and sinks, so there are 42 nodes left which can be deployed as the malicious node. The malicious nodes will perform the grey-hole attacks with a 50% of packet drop ratio. In other words, each network topology will run 42 times and sum up the total packet loss number as one of the final measurement results. Moreover, there are two different metrics are used to compare with the T value; the results are shown in Figure 4.18.

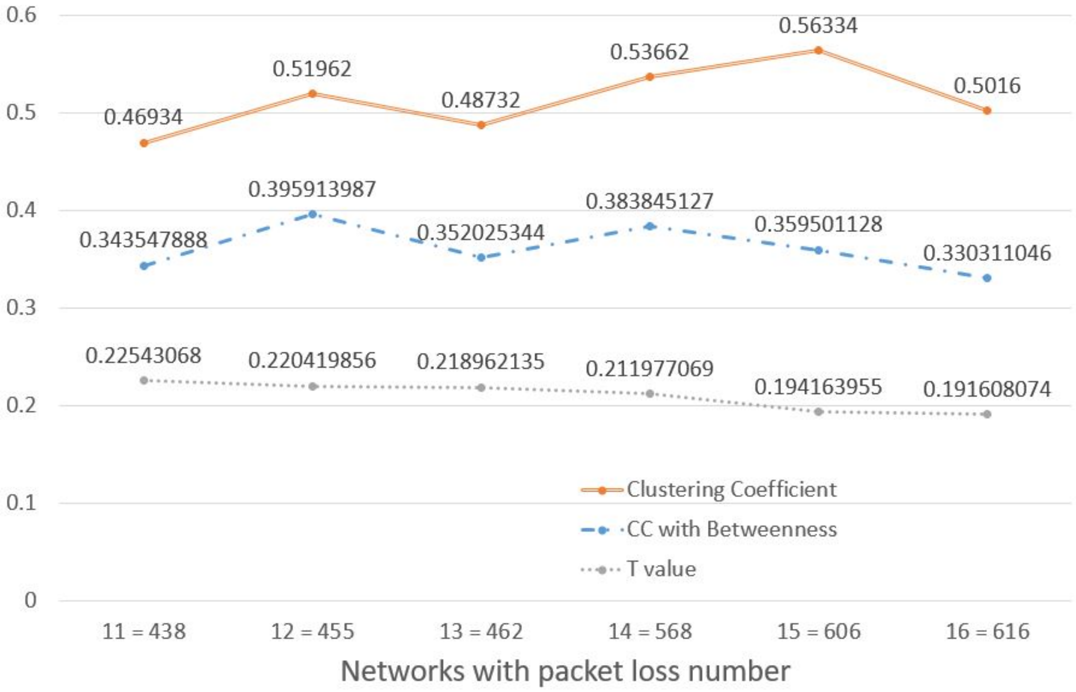


Figure 4.18: Packet Loss vs. Average Clustering Coefficient, Weighted Clustering Coefficient, and T

As can be seen in Figure 4.18 on the x-axis, from Network 11 to Network 16, the packet loss number trend increases, where the Network 11 has the smallest packet loss number, which is 438, and Network 16 has the biggest packet loss number of 616. The double line curve is the average clustering coefficient for each network, which indicates Network 15 has the highest value and Network 11 has the lowest value. However, the packet loss results for each network are telling the difference, where the higher clustering coefficient should have lower packet loss. The dashed line is the weighted clustering coefficient; it considers the Betweenness as weight factors rather than using average, so it can emphases the important nodes, which has the higher Betweenness value. Considering this Network 12 becomes the one with the highest value and network 16 the lowest value. Though the Network 16 is achieved the highest packet loss which the weighted factor showed the expected results. However, the results on Network 12 still do not reflect the trend of the actual packet loss number. As we discussed previously, the Structural Hole has a negative impact on the positive impact of the Simmelian Ties. There is no physical Structural Hole in any of these five networks, but there are logical Structural Holes in all of the five networks. The T value is the dotted line in Figure 4.18. It shows the Network 11 has the highest value, and the Network 12 is second high highest, followed by Networks 13, 14, 15, and finally, Network 16. The recommendation for the network trustworthiness in these six networks by the T value reflects a decreasing trend while the packet loss number is increasing.

Moreover, in Network 12, the weighted clustering coefficient and average clustering coefficient results are both better than that of Network 13's. Though, there are only small differences in the packet loss number. This is because there are many logical Structural Holes in the Network 12 compared to Network 13, and some of them are very deep. 'Deep' means once that node on the logical Structural Hole has been attacked, it is necessary to take an additional 10 to 15 hops as a detour and to reach the sink again. This causes further packet loss while looking for the alternative route. The packet latency also has a huge increase on these detours, which can be seen in Figure 4.19. The logical Structural Hole threshold has been set up as 4 hops for the determination of the logical Structural Hole, which means if the node takes 5 hops to detour, we consider this is $h = 1$ for the attenuation as in Equation 4.15 on page 123. The factor q which is controlling the attenuation speed is set up as $q = 3$. In Network 13, the total h value is 57, whereas the h is 84 in total in Network 12. From this, we can tell the difference in the number and depth level of the logical Structural Holes each has. The T value betweenness them has only a small difference as well, which the T did reflect the expected results precisely. Moreover, the logical Structural Hole can also significantly affected the packet latency. This can be seen in Figure 4.19.

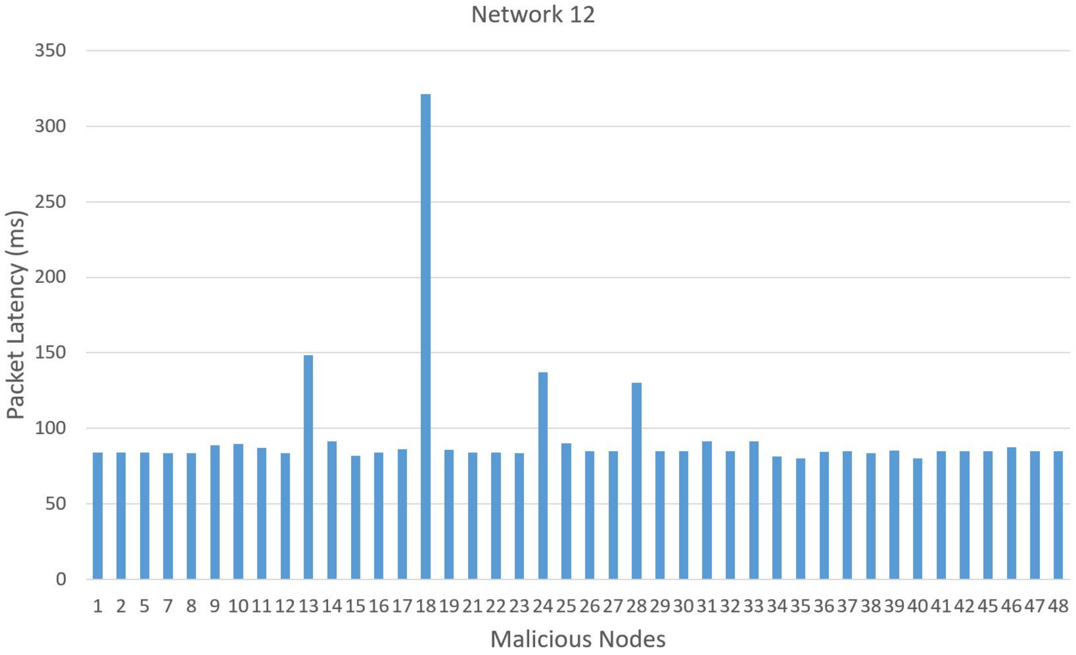


Figure 4.19: Logical Structural Hole vs. Packet Latency (ms)

In Figure 4.19,it can be seen when malicious attacks were launched on nodes 13, 18, 24, 29, and 31 in Network 12, the packet latency is significantly increased compared to the attacks were launched on the other nodes. There are 13 logical Structural Holes are identified in the Network 12, and the nodes 13, 18, 24, 29, and 31 are all logical

Structural Holes. Though there are other logical Structural Holes such as node 19, the packet latency is about average when the attacks were launched on this node. This is because the Betweenness on node 19 is much lower than that on node 18 which created a huge spike in the packet latency. The simulation traffic is not routed through node 19, so the packet latency is at the average level as well. This means the network Betweenness is also an important factor to evaluate the logical Structural Hole depth level as well.

Finally, we use the OMNET++ to randomly generate another two networks with physical Structural Holes in them. The rest of the set up is exactly the same as the set up in Networks 11 to 16. The traffic directions in Network 17 are nodes 1 to 46, 12 to 29, 2 to 50, and 11 to 38. For Network 18 the direction is from nodes 3 to 27, 1 to 50, 2 to 28, and 21 to 49. The network 17 and 18 can be seen in Figure 4.20.

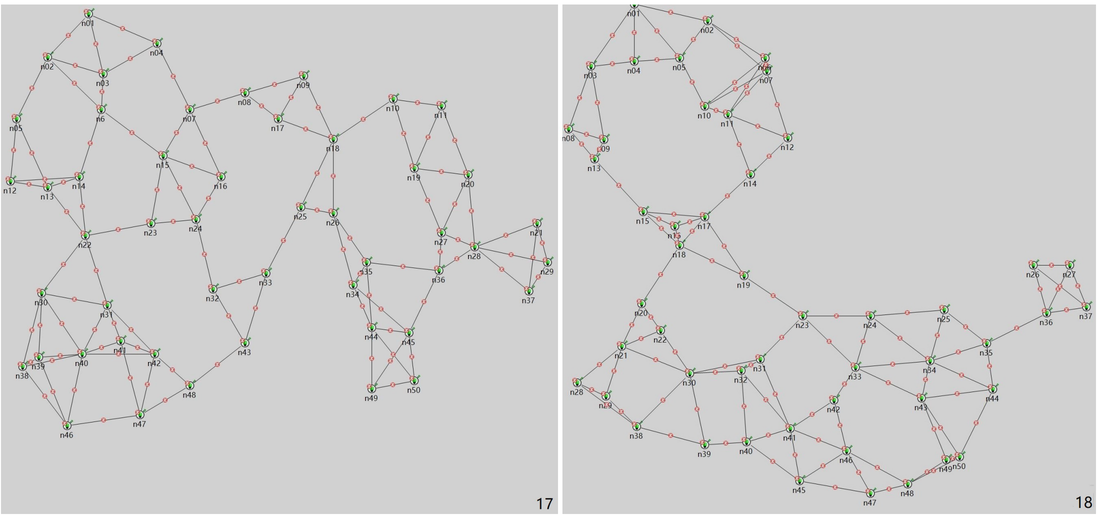


Figure 4.20: Networks 17 and 18 for Physical Structural Holes

As there are physical Structural Holes in both network, the network trustworthiness T is rated as 0. So for the comparison between these two networks, the Betweenness value on the physical Structural Holes would be used to compare. The bigger the Betweenness value, the worse it is. The Structural Hole in Network 17 is node 28 with 143.58 as Betweenness on its position, and the two physical Structural Holes in

Network 18 are nodes 35 and 36 with a 638.5 value in total. Obviously, Network 18 has a higher Betweenness value, which means Network 17 is better than Network 18. The packet loss results have reflected that as well, as the packet loss result on Network 17 is 683 and 762 for the Network 18. It means the DTEGR algorithm can achieve a much better packet loss result on Network 17, and a more effective performance compared to the performance in Network 18.

4.6 Trustworthiness Tolerance Margin

In the previous sections, this thesis has assumed the network objectives and set up the network topologies to verify the accuracy of the network trustworthiness metric T. The network objectives are the network availability and the node trustworthiness. Thus, with the findings from the Chapter 3, we know that Simmelian Ties have a positive impact on the network availability and node trustworthiness in the network by introducing the third parties, and also providing the redundant routes to better survive from malicious attacks or network failures. On the other hand, the Structural Hole has a negative impact on the node trustworthiness and network availability, as it has no third party for the trust evaluation of nodes, and more importantly, it has no redundant routes so that when it is under attacks or network failures, the network will be disconnected as well. Moreover, Structural Hole has a negative impact on the positive impact of the Simmelian Ties. All these have been verified by the extensive simulation studies in Chapter 3 and this chapter in the last section. Though, as mentioned at the beginning of this chapter, the attacks or network failures can be intentional or not intentional, which is also can be interpreted as targeted attacks or random attacks. The network trustworthiness metric T is more suitable for a targeted attacks evaluation, as the targeted attacks also focus on the minimum effort for the maximum damage; while the random attacks can be anywhere in the networks with various damage level. To be able to evaluate the network trustworthiness while under a random attack, this thesis has introduced the concept of a Trustworthiness Tolerance Margin (TTM).

Unintentional attacks can happen anywhere in the network at any time. Thus, an evaluation of the unintentional attacks will need to consider the equal probability failure on each node in the network. The worst case attack scenarios, the best case attack scenarios, and the most likely case attack scenarios, all these need to be considered in the evaluation of network trustworthiness under the random attack mode. As mentioned above, the trustworthiness T is only evaluates the worst case which is the target attacks. However, it did not consider the best case and general cases. In such case, we introduce the concept Trustworthiness Tolerance Margins (TTM) to overcome this problem. The TTM is more concerned with the stable performance (reliability) in different circumstances rather than only in the worst case scenarios. For example, there are two network topologies which are the scale-free network (Barabási & Albert, 1999) and small world network (Watts & Strogatz, 1998). The scale-free network is described by researchers as “robust, yet fragile” (Zhao et al., 2011; X. F. Wang & Chen, 2003; Onnela et al., 2007). It is robust to against the random attacks but fragile to targeted attacks. The small world network is not that robust as the scale-free network to against the random attack, but it is not fragile to the targeted attacks. Thus, the small world network is preferred compared to scale-free network in the network availability scenarios. There is a threshold margin on the upper and lower boundary of the trustworthiness of the network while under the attacks, and the more consistent network topologies can deliver the QoS while under attack. We call this threshold margin a ’tolerance margin’. Once the network topologies are outside the tolerance margin just like the free scale network topologies, these networks should be considered as less trustworthy compared to other, which have a smaller tolerance margin and similar trustworthiness T.

4.6.1 Monte Carlo Simulation

For the TTM evaluation, we selected the Monte Carlo Simulation approach. The Monte Carlo Simulation uses different computational algorithms depend on the scenarios to repeat random sampling for different purposes by means of a computer (Kroese, Brereton, Taimre & Botev, 2014). These random samples can be used to model the real world systems such as traffic networks, stock markets, etc. This modelling can greatly help in solving the deterministic problems. In this thesis's objective, the attacks on the randomly selected node in the network needs to be modelled by the Monte Carlo Simulation, so the random attack can be evaluated through the TTM. For the simulation, while the node in the network is under attack, which we assume is the node not available in the network anymore due to human error, etc. Taking Network 11 in Figure 4.12 on page 130 as an example, there are 50 nodes in the network, which means that each node in the network will have $1/50$ = 2% probability to be selected for the random attack in the network. Once the selected node is under attacks, we assume the node is no longer available on the network. In other words, the network topology is changed; the network trustworthiness T for the changed topology is what we are after for the TTM evaluation.

4.6.2 Sample Monte Carlo Simulation

In the Section 4.5 on page 109, this chapter has defined the sample scenarios for the validation of the network trustworthiness T evaluation, and the Equation 4.16 on page 124 has defined the calculation of T, which is composed of clustering coefficient, node Betweenness, and Structural Hole. Thus, once the network topology is changed, these three metrics need to be recalculated again. The input of the network graph for these three metrics is as in Equation 4.13 on page 119. In such cases, by disabling a node in the network, we simply need to remove the node data from the matrix. For example, if the node i is disabled from the network, then in the graph matrix, the data at the ith row and the ith column will be removed.

We take the Networks 11 and 12 in Figure 4.12 on page 130 and 4.13 as the examples for the Monte Carlo simulation. We run a simulation 1000 times for each network, and each run a node will be randomly selected to be disabled from the network. The changed network trustworthiness T results for the 1000 simulations run is shown in Figure 4.21 and Table 4.3.

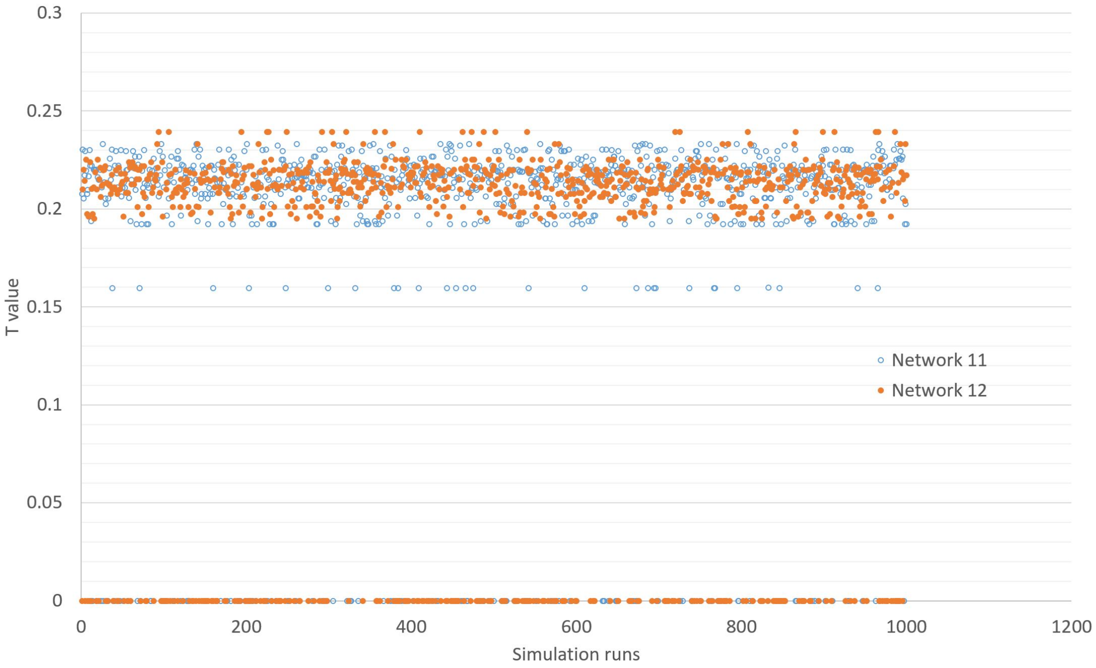


Figure 4.21: T Values Distribution for Networks 11 and 12

Table 4.3: Average and Standard Deviation T for 1000 Simulations

Network	**Average**	**Standard Deviation**	**Max**	**Min**
11	0.184876751	0.073675525	0.233180945	0
12	0.166884318	0.088819739	0.239295043	0

In Figure 4.21, Network 12's T values are mostly distributed between 0.2 to 0.232, whereas Network 11's T results are more distributed. Both networks have some 0 results are due to the physical Structural Holes in the networks. Thus, from these results, Network 12 should have a higher average T value and a low standard deviation. However, the results in Table 4.3 tell a different story. Network 11's average T results and the standard deviation are lower than that for Network 12, even though, the maximum T value in Network 12 is slightly higher than Network 11. This is because the random attacks in Network 12 are more likely to create physical Structural Hole compare to Network 11. The number of 0 results are shown in Figures 4.22 and 4.23 in below.

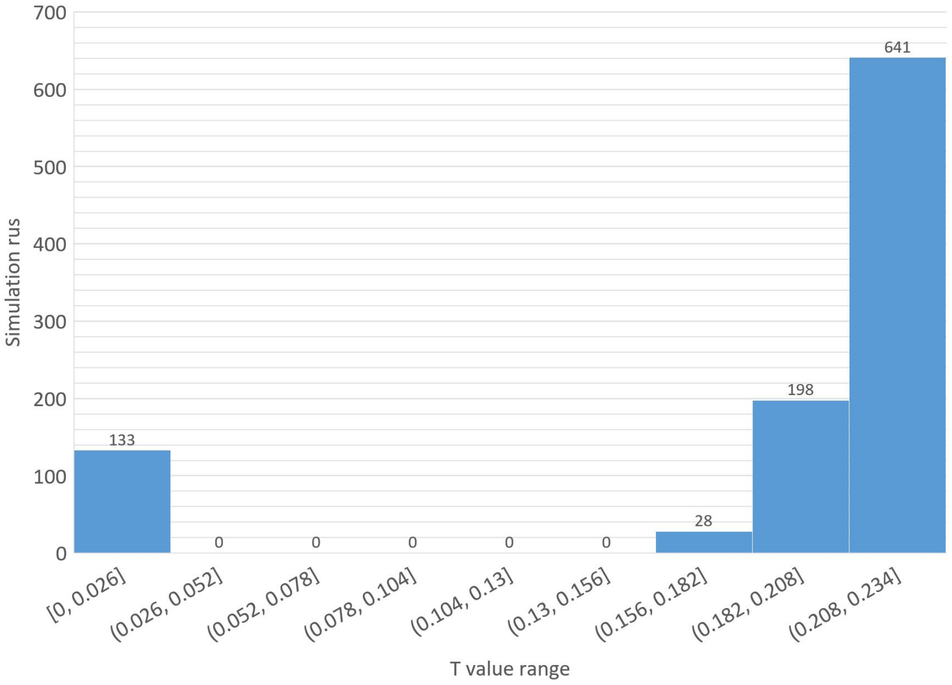


Figure 4.22: T Values Distribution for Network 11

The standard deviation is the margin of the trustworthiness tolerance. We compared the two networks with the average plus and minus the standard deviation as the upper and lower boundaries, which for the network 11, the trustworthiness tolerance margin is $(0.1112, 0.2586)$, and network 12 is $(0.0781, 0.2557)$. We can see that both network 11's upper and lower boundaries are higher than those of network 12. Thus, we consider network 11 is more trustworthy than network 12. There might be another case that the lower boundary is higher, but the upper boundary is lower. In such a case, we consider the network with lower 'lower boundary' value is more trustworthy, as the standard

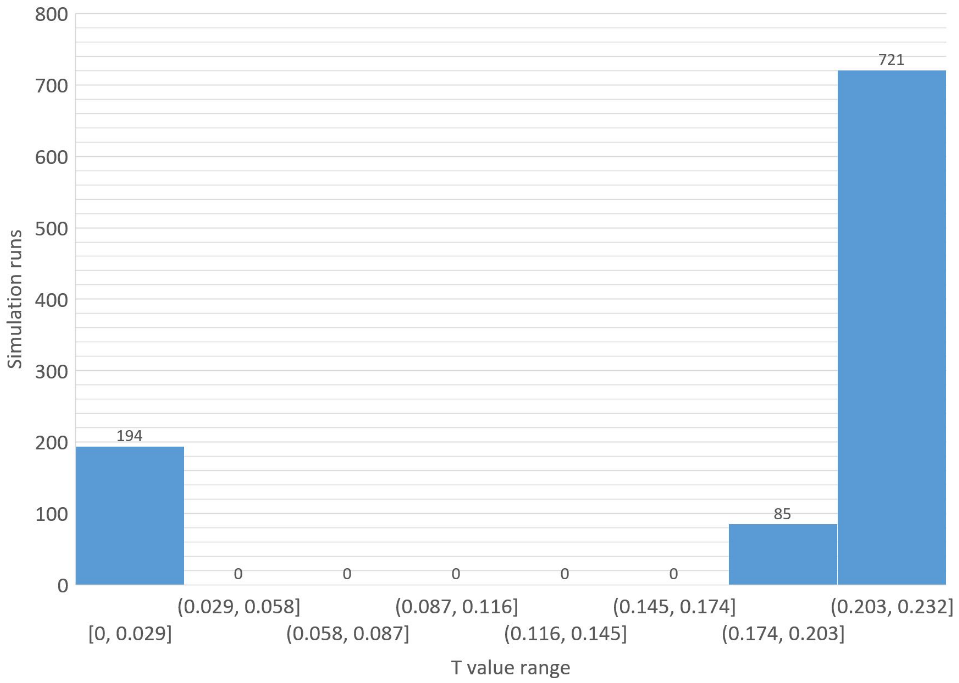


Figure 4.23: T Values Distribution for Network 12

deviation is lower, which means the trustworthiness results are more consistent.

4.7 Summary

In this chapter, first of all, we have introduced the concept network trustworthiness, where the network is considered as trustworthy when it fulfils its objectives under any expected or unexpected circumstance. Based this new defined term, we created a service platform called NTaaS. To properly provide a network trustworthiness evaluation service, we have proposed a mathematical evaluation framework for the network trustworthiness evaluation. As different scenarios have different network objectives, thus, the metrics to evaluate the network trustworthiness should be adjusted accordingly; which means there should not be a universal metric to evaluate network trustworthiness in all the scenarios. After the mathematical evaluation framework is proposed, a validation of this framework is provided.

As mentioned above, with different network objectives, the evaluation metrics should be adjusted accordingly. Thus, the first thing we do for the framework validation is to provide a network scenario that the evaluation metrics for which network trustworthiness can be decided. The network availability and node trustworthiness are selected as the network objectives. As discussed and validated in Chapter 3, the Simmelian Ties have a positive effect on the node trustworthiness as it provides the third parties for reputation analysis. More importantly, it also provides the redundant routes for the network availability. The Structural Hole has a negative impact on the node trustworthiness, as the node at this position can act maliciously without the fear known by another party. Also it does not have alternative routes that leave other nodes with no choice but to keep trusting this node. Finally, the Structural Hole has a negative impact on the positive impact of the Simmelian Ties. According to these finding, we selected the clustering coefficient to evaluate the Simmelian Ties in the network, the more Simmelian Ties in the network, the more trustworthy of the network. As the clustering coefficient is only evaluating a node in the network rather than the whole network, thus most of the existing studies have used average clustering coefficient for the evaluation of the whole network. However, some of the nodes in the network can be more important for the network availability, and some are less important. The average clustering coefficient cannot reflect this critical level as it treats every node in the network as the same. Thus, we used the node Betweenness to represent the critical level of the nodes in the network. Obviously, the more centrally of the nodes are located, the more critical of the nodes in the network. Node Betweenness is the metric to measure this. We then normalized the node Betweenness and used it as a weight factor to synthesis the node clustering coefficient as the weighted clustering coefficient for the evaluation of the whole network. For the Structural Hole, as the existing metrics ’effective size’ and ’Simmelian Brokerage’ cannot locate the exact Structural Hole in the network, therefore, we proposed the Structural Hole Locator (SHL) algorithm to locate any logical Structural Hole and physical Structural Hole in the network. We defined the logical Structural Hole in this thesis as the additional hop count for a detour if the neighbour is behaving maliciously. Through the extensive simulation studies, we have validated the accuracy of the network trustworthiness T to represent the trustworthiness of the network according to the packet loss number while under the attacks.

The network trustworthiness T is evaluating the targeted attacks in the network; such attacks are normally aim at the maximum damage with the minimum effort. There are other types of attacks which are unintentional. Normally, these attacks are human errors, disaster, network failures, etc. For evaluation of these random attacks, we introduced the concept of Trustworthiness Tolerance Margin. We used the Monte Carlo Simulation approach, to run the simulation of random attacks 1000 times, then we used the average T and standard deviation to show the upper bound and lower bound to reveal the consistency of the trustworthiness T after the attack. The more consistent one with a higher T is believed more trustworthy while under the random attacks.

After the evaluation of the network topologies, for the networks with unsatisfactory structures, NTaaS will look for the options for network optimization by recruiting the nearby available nodes. In the next chapter, we discuss the approaches for the remediation of the network topologies by adding a link and possible recruitment nodes.

Publication generated from this Chapter

Xiang, M., Liu, W., Bai, Q., Al-Anbuky, A., Wu, J., & Sathiaseelan, A. (2017). NTaaS: Network Trustworthiness as a Service. In Telecommunication Networks and Applications Conference (ITNAC), 2017.

Xiang, M., Liu, W., Bai, Q., & Al-Anbuky, A. (2015). The double-edged sword: Revealing the critical role of structural hole in forming trust for securing Wireless sensor networks. In Telecommunication Networks and Applications Conference (ITNAC), 2015 International (pp. 286–291).

Xiang, M., Liu, W., Bai, Q., & Al-Anbuky, A. (2015). Simmelian Ties and Structural Holes: Exploring Their Topological Roles in Forming Trust for Securing Wireless Sensor Networks. In 2015 IEEE Trustcom/BigDataSE/ISPA (Vol. 1, pp. 96–103).

Chapter 5

The Remediation of Network Topology

5.1 Introduction

In the last chapter, a new term 'network trustworthiness' was introduced to describe the tolerance level of the network topology can accommodate its objectives under any expected and unexpected circumstance. A new service framework of NTaaS is purposed to provide a network trustworthiness evaluation service platform to the user or devices in the P2P communication network. Thus, we have a network trustworthiness evaluation framework introduced for the NTaaS. As different scenarios have different network objectives, the evaluation metrics are different accordingly. If the network is evaluated as not trustworthy, obviously, this network topology is not capable of fulfilling its network objectives. Thus, the network topology needs to be optimized by the NTaaS, and this is what we focus in this chapter.

The evaluation metrics are different according to different network objectives. Thus, the network remediation approach should be different as well according to these different evaluation metrics. There is no universal metric to evaluate the network trustworthiness in all the scenarios. Therefore, there is no universal approach to remedy the network topologies as well. The remedial approach is determined by the evaluation metrics used in the network trustworthiness evaluation. In the last chapter, it has provided a scenario and assumed the network objectives for the network trustworthiness evaluation. The objectives are network availability and node trustworthiness in routing. The network availability is reflected in the packet delivery ratio, where the network topology should improve the efficiency of the packet routing under any expected and unexpected circumstance. The node trustworthiness in routing is that the network topology should resist any unwanted behaviours such as malicious, selfishness, etc. In other words, the network topologies should improve the efficiency of the trust-based routing algorithm.

According to the network objectives, the metrics clustering coefficient, node Betweenness, and Structural Hole Locator have been selected for the network trustworthiness evaluation. The clustering coefficient is the measurement of Simmelian Ties in the network, node Betweenness is the measurement of the node centrality in the network, and Structural Hole Locator is to locate any physical and logical Structural Holes in the network, and measure their depth level. We assumed that the resources are limited, so that only a link can be added to the network for the topology remedy. Thus, the potential threats in the network need to be prioritised where the most critical threat should be selected to optimize. The most critical threat to the network availability and node trustworthiness scenario is the physical Structural Hole structures in the network, then logical Structural Hole, and finally the insufficient Simmelian Ties. Thus, there are three sections for each of these potential threats scenarios, which as shown in Figure 5.1. First of all, the next Section 5.2 will discuss the remedy methods when there is no physical and logical Structural Hole in the network. Secondly, the section 5.3 will explain the remedy method when there is no physical Structural Hole in the network, but there is a logical Structural Hole. Thirdly, the Section 5.4 on page 158 discusses the remedy method when there is a physical Structural Hole in the network. After the remedy approaches have been explained, the possible feasible ways to execute remediation in the real world will be discussed. Finally, the conclusion will be given for this chapter.

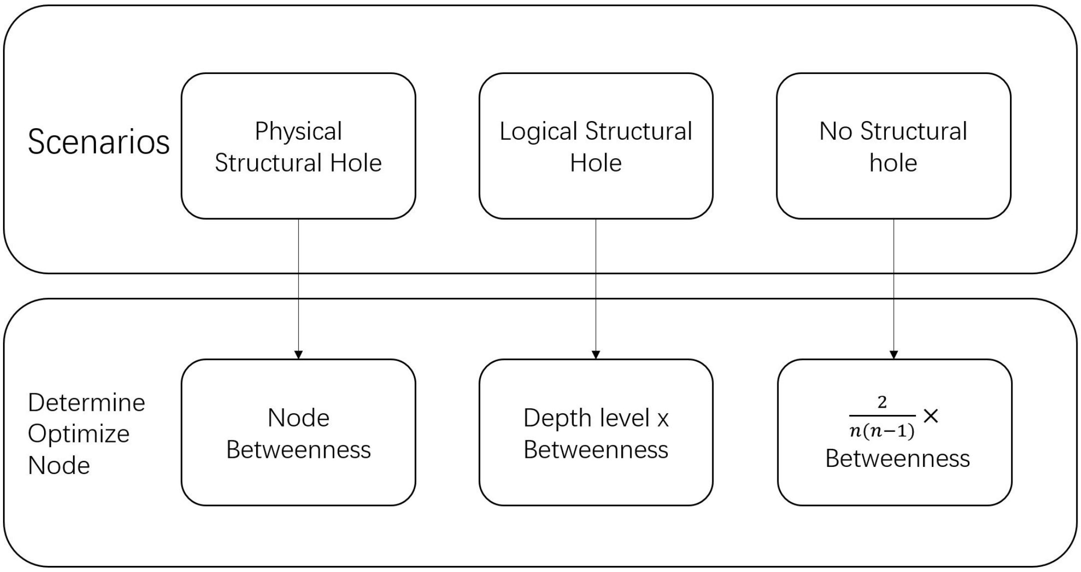


Figure 5.1: Remediation Scenarios

Before we discuss the remediation approaches for the three scenarios, the method to select the candidate neighbours for the reconnection needs to be explained (for Equation 5.1).

$$r < dis_{jk} < 2r, j \neq k \tag{5.1}$$

The dis_{jk} is the actual distance between nodes j and k. The radius of radio range for the nodes in the network is r. So the equation basically means all the neighbours of a select remediation node are the candidate neighbours for the link addition.

5.2 Non-Structural Hole Scenario

In the scenarios that there is no Structural Hole in the network, neither a physical Structural Hole nor a logical Structural Hole as well. The network trustworthiness evaluation metrics, take the Structural Hole and Simmelian Ties as the factors for the evaluation. Thus, for a network without any Structural Hole, the Simmelian Ties in the network would be the only factor needing to be considered now. The new proposed weighted clustering coefficient is selected to evaluate the Simmelian Ties in the network. This new proposed weighted clustering coefficient is the sum of each node's clustering coefficient with the node Betweenness as weight factors. In such a case, to achieve the largest increment on the network trustworthiness T, we need to achieve the largest possible increment on the weighted clustering coefficient for the network. The clustering coefficient equation is introduced in Equation 3.5 on page 58, if a link is added to increase node i's Simmelian Ties, rather than add a link to node i, the link should be added between node i's neighbours, so the Simmelian triangles connecting to node i can be increased. In this case, the increment of the clustering coefficient on node i which is $c_{i,incr}$ is shown in Equation 5.2.

$$\begin{aligned} c_{i,incr} &= \frac{2}{n_i \times (n_i - 1)} \times \Big(\sum_{j=1,j\neq i}^{n_i} \sum_{k=j+1,k\neq i}^{n_i} e_{jk} + 1 \Big) - \frac{2}{n_i \times (n_i - 1)} \times \sum_{j=1,j\neq i}^{n_i} \sum_{k=j+1,k\neq i}^{n_i} e_{jk} \\ &= \frac{2}{n_i \times (n_i - 1)} \end{aligned} \tag{5.2}$$

If the new link is added to node i's neighbours, as mentioned above, this means a new Simmelian triangle is connected to node i. Thus, we plus 1 to the $\sum_{j=1,j\neq i}^{n_i} \sum_{k=j+1,k\neq i}^{n_i} e_{jk}$ in Equation 5.2, which is the count of Simmelian triangles are connecting to the node i before the new link is added. As can be seen from the equation, the larger of the n_i, which is the node degree of node i, the smaller of the increment on the clustering coefficient by adding a link to connect its neighbours. Thus, we can probably to look for the node with the smallest node degree to improve its clustering coefficient. However, if the selected node is at the edge of the network, increasing the clustering coefficient on such node would not increase too much in the final T value due to its low

Betweenness value. In such case, as we considered the weighted clustering coefficient as the evaluation metric. The normalized node Betweenness value also needs to be considered. Thus, the new equation is in Equation 5.3.

$$c_{i,incr} = \frac{2}{n_i \times (n_i - 1)} \times w_i \tag{5.3}$$

Now we have the T increment equation by adding a link between node i's neighbours. However, when actually adding a link on node i's neighbours, the two neighbours which are selected for the new connection link will also have a clustering coefficient change for them, which is shown in Figure 5.2. In the left-hand side network, if it is assumed that we need to increase the clustering coefficient on node 1, then the only choice here is to add a link between nodes 2 and 3. By doing that, as can be seen by the dashed line, the connected nodes clustering coefficient also increased as a result of the link addition. In the left-hand side case, for each connected node such as nodes 2 and 3, each has a two Simmelian triangle increment after the link addition, as they also both connected to node 4 as well. In the right-hand side case, the increment for nodes 2 and 3 is only one Simmelian Triangle. Following this logic, for every additional node, they both connected to, the increment on the Simmelian triangle for them will increase by one accordingly. Thus, the increment on them would be shown in Equation 5.4.

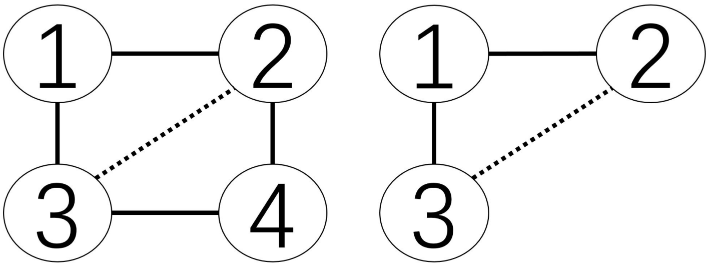


Figure 5.2: Remediation Sample Networks

$$\begin{aligned} c_{j,incr} =& \frac{2}{(n_j+1)\times(n_j-1+1)} \times (\sum_{k=1,k\neq j}^{n_j} \sum_{i=k+1,i\neq j}^{n_j} e_{ik} + 1 + n_{jk}) \\ & - \frac{2}{n_j \times (n_j-1)} \times \sum_{k=1,k\neq j}^{n_j} \sum_{i=k+1,i\neq j}^{n_j} e_{ik} \end{aligned} \tag{5.4}$$

The variable n_{jk} is the number of nodes, which are sharing the same neighbours nodes j and k, excluding node i. As node i is the node selected for the remediation, and nodes j and k are the neighbours of node i which are selected to add a link. To simplify the Equation 5.4, we first make ti_j as number of the Simmelian Triangles connected to node j, i.e., $ti_j = \sum_{k=1,k\neq j}^{n_j} \sum_{i=i+1,i\neq j}^{n_j} e_{ik}$. Then we can have the Equation 5.5 with node normalized node Betweenness.

$$\begin{aligned} c_{j,incr} &= (\frac{2}{(n_j+1)\times(n_j-1+1)} \times (ti_j + n_{jk} + 1) - \frac{2}{n_j\times(n_j-1)} \times ti_j) \times w_j \\ &= ((\frac{ti_j + n_{jk} + 1}{n_j+1} - \frac{ti_j}{n_j-1}) \times \frac{2}{n_j}) \times w_j \\ &= \frac{n_j \times n_{jk} + n_j - n_{jk} - 2ti_j - 1}{n_j^2 - 1} \times \frac{2}{n_j} \times w_j \\ &= \frac{2n_j \times n_{jk} + 2n_j - 2n_{jk} - 4ti_j - 2}{n_j^3 - n_j} \times w_j \end{aligned} \tag{5.5}$$

After we have confirmed the clustering coefficient increment on the node i's neighbour j, we have the total clustering coefficient increment on weight clustering coefficient as in Equation 5.6. In this equation, we assume the $n_{jk} = 0$, which means there are no other common neighbours of nodes j and k other than node i.

$$C_{incr} = \frac{2}{n_i \times (n_i-1)} \times w_i + (\frac{2n_j - 4ti_j - 2}{n_j^3 - n_j} \times w_j) + (\frac{2n_k - 4ti_k - 2}{n_k^3 - n_k} \times w_k) \tag{5.6}$$

There is another case can occur with the increment of weighted clustering coefficient after a link is added; Equation 5.6 is represents one of those cases. The other case is the selected candidate neighbour pair is also connected to the other nodes, in other words, $n_{jk} \geq 1$. In such a case, we need to consider four or more nodes' clustering coefficient change, depending on how many other nodes are also connected to this candidate neighbour pair. We assume there is only one more node is connecting with both candidate neighbours, which means $n_{jk} = 1$. The equation for it appears in Equation 5.7.

$$C_{incr}c = \frac{2}{n_i \times (n_i - 1)} \times w_i + (\frac{4n_j - 4ti_j - 4}{n_j^3 - n_j} \times w_j) + (\frac{4n_k - 4ti_k - 4}{n_k^3 - n_k} \times w_k) + \frac{2}{n_l \times (n_l - 1)} \times w_l \tag{5.7}$$

The node l is another node that the candidate neighbour pair is connected to. If the $n_{jk} = 2$, then we add another Equation 5.3 into Equation 5.7, just like the increment for node l did. So in summary, first of all, we use the Equation 5.3 to determine which node in the network is selected to optimize. Then we use either Equation 5.6 or 5.7 depending on the case to finally confirm the new link position is connecting with which two nodes.

Algorithm 3 on the next page is the pseudo-code for the none Structural Hole scenarios. Line 18 the $Eq53$ is using the Equation 5.3 to calculate the estimated increment on weighted clustering coefficient. Therefore, the $N[i].size()$ is the number of neighbours of node $i's$ and $w[i]$ is the weight factor for node i. If the case is the right network in Figure 5.2, the Equation 5.6 is used to calculate, which is at line 50. If the cases are the left-hand side network, or there is more than one shared neighbours existed, lines 57 to 65 are used to calculate the estimate weighted clustering coefficient increment for the additional link in place.In this algorithm, as the network size V is changing, the remedy target node selection will have roughly $O(V^2+V)$ runtime change reflect on it (lines 10 to 21). Once the target node is determined, the network density

Algorithm 3 Remediation for None Structural Hole (SH) Scenarios

```
function ClusteringCoefficientRemediation
    $V$ ← number of nodes in the network *integer*
    $C$ ← Clustering coefficient for each node *arraylist*
    $w$ ← Betweenness weight factor *arraylist*
    $G$ ← Network topology V x V matrix data *arraylist*
    $v$ ← node for remedy *integer*
    *link* ← link position *arraylist*
    $N$ ← number of neighbours for each node *arraylist*
    $n$ ← neighbhour list *arraylist*
    for $i = 0$ to $G.size() - 1$ do
        for $j = 0$ to $G[i].size() - 1$ do
            if $G[i][j]$ has connection then
                add $j$ to $n$
        add $n$ to $N$
        empty $n$
    $Incr, max \leftarrow 0$
    for $i = 0$ to $N.size() - 1$ do
        $Incr \leftarrow Eq53(N[i].size(), w[i])$
        if $max < Incr$ and $C[i] < 1$ then
            $max \leftarrow Incr$
            $v \leftarrow i$
    $pair$ ← candidate link positions around node v *arraylist*
    for $i = 0$ to $N[v].size() - 1$ do
        $temp$ ← *arraylist*
        if $G[N[v][i]][N[v][j]]$ has no connection then
            add $i$ and $j$ to $temp$
            add $temp$ to $pair$
    if $pair.size() = 1$ then
        $link[0] \leftarrow pair[0][0]$
        $link[1] \leftarrow pair[0][1]$
    else
        for $i = 0$ to $pair.size() - 1$ do
            $n1$ ← share neighbours list *arraylist*
            $l_1 \leftarrow pair[i][0]$
            $l_2 \leftarrow pair[i][1]$
            $ti_1, ti_2 \leftarrow 0$, Simmelian triangle number *integer*
            $cc$ ← increment clustering coefficient
            $max \leftarrow 0$
            for $j = 0$ to $N[l_1].size() - 2$ do
                for $k = j + 1$ to $N[l_1].size() - 1$ do
                    $ti_1 \leftarrow ti_1 + 1$
```

Algorithm 3 Remediation for None Structural Hole (SH) Scenarios (continued)

```
        for $j = 0$ to $N[l_2].size() - 2$ do
            for $k = j + 1$ to $N[l_2].size() - 1$ do
                $ti_2 \leftarrow ti_2 + 1$
        for $j = 0$ to $N[l_1].size() - 1$ do
            for $k = 0$ to $N[l_2].size() - 1$ do
                if $N[l_1][j] = N[l_2][k]$ and $N[l_2][k] \neq v$ then
                    add $N[l_2][k]$ to $n1$
        if $n1$ is empty then
            $cc \leftarrow Eq56(N[v].size, w[v], N[l_1].size(), w[l_1], ti_1,$
            $N[l_2].size(), w[l_2], ti_2)$
            if $max < cc$ then
                $max \leftarrow cc$
                $link[0] \leftarrow l_1$
                $link[1] \leftarrow l_2$
        else
            $cc \leftarrow Eq53(N[v].size(), w[v]$
            for $j = 0$ to $n1.size() - 1$ do
                $cc \leftarrow cc + Eq53(N[n1[j]], w[n1[j]])$
            $cc \leftarrow cc + Eq55(N[l_1].size(), w[l_1], ti_1, n1.size())$
            $cc \leftarrow cc + Eq55(N[l_2].size(), w[l_2], ti_2, n1.size())$
            if $max < cc$ then
                $max \leftarrow cc$
                $link[0] \leftarrow l_1$
                $link[1] \leftarrow l_2$
    return $link$
```

will be the factor to affect the runtime for the rest of the algorithm, which roughly $O(3n^3+n)$. Therefore, roughly, the complexity of this algorithm is $O(V^2+V+3n^3+n)$.

Finally, we will use Network 3 in Figure 3.9 on page 71 as an example to demonstrate how to determine where to add a link in the network to achieve the maximum increment on the weighted clustering coefficient. Network 3 has no physical nor logical Structural Hole in the network. Therefore, we calculate the increment of the clustering coefficient by adding a link between the target node's neighbours, and the results come back showing node 14 has the highest increment. There are two candidate neighbour pairs for node 14, which are nodes 10 and 13, and nodes 10 and 15. We calculate the increment of clustering coefficient using the Equation 5.7. If the link is added between node 10 and 13, the node 9 in the network would also be affected by this link addition. As can be seen in Table 5.1. The estimated increment on the clustering coefficient is 0.0671. For the nodes 10 and 15, the estimated increment is 0.0723. From this estimated number, we should add the link between node 10 and 15. We calculate the new network topologies to validate it, the original network topology's T value is 0.14493. If we put a new link between node 10 and 13, the new topology T value now becomes 0.2008, and if we put a new link between node 10 and 15 as the estimate calculation results suggested, the new T value now becomes 0.21535. This has validated the estimated calculation results.

Table 5.1: Link Addition Results Comparison in Non-Structural Hole Scenario

Link addition position	**10 & 13**	**10 & 15**
Estimate Increment	0.0671	0.0723
T **value after link addition**	0.2008	0.21535
T **value before link addition**	0.14493	

5.3 Logical Structural Hole Scenario

The second critical potential threat out of three is the logical Structural Hole in the network. The logical Structural Hole can be understood as the damage caused by the node failure is higher than the pre-defined threshold. Depending on the pre-defined threshold, the logical Structural Hole can be defined differently. The threshold can be detour hop count, traffic load limited, etc. In this case, we assume the threshold is a detour hop count of 4. As explained in Chapter 4, if the node i is disabled, and node i's neighbours node j and k need to take 5 hops to detour, then node i is at the logical Structural Hole position at a depth level 1. If this depth level is very high, once this logical Structural Hole is under attacks or suffers hardware failure, the traffic would be expected to be significantly delayed and have more of a chance to encounter another network failure or attacks. This is the reason we put the threat level of logical Structural Hole higher than insufficient Simmelian Ties in the network.

For the remediation of the logical Structural Hole, the goal is to reduce the depth level of the logical Structural Hole. In most of the cases, there is more than one logical Structural Holes in the network. With the limited resources, only one logical Structural Hole can be remedied. The deeper of the logical Structural Hole, the more damage can be done in the network. Moreover, a logical Structural Hole at the different positions in the network can have different damage levels as well. Thus, the node Betweenness is required here to reflect the critical level of the position in the network. The result is Equation 5.8 to select the most critical logical Structural Hole for remediation.

$$D_i = h_{jk} \times w_i \tag{5.8}$$

The variable h_{jk} is the detour hop count less the threshold count which is the depth level in Equation 4.15 on page 123. The D_i is the damage factor for node i, and the nodes j and k are the neighbours of node i. In such a case, the node in the network with the highest D_i shall be selected for structure remediation. After identifying the particular logical Structural Hole for remediation, we will disable the node at this hole position, and calculate the hop count that the candidate neighbour pairs require to reach each other. The pair with the highest hop count is normally to be selected to add a link between. However, the actual case is different. For example, the topology network 11 in Figure 4.12 on page 130. Node 25 is selected as the most critical logical Structural Hole in the network which it requires the remediation. The candidate neighbour pairs are node 17 and 29, node 17 and 30, node 17 and 31, node 29 and 31, and node 30 and 31. First of all, we disable the node 25 and try to calculate the hop count between nodes 17, 29, 30, and 31. We can found that node 17 is in a one-hop cluster by itself, which we name it A, nodes 29 and 30 are in another one hop cluster B, and node 31 is at the remaining one hop cluster C. Node 29 and node 30 to node 17 are at a 4 and 5 hop count distance, and to node 31 is a 14 and 15 hop count distance. From the distance, we can tell node 29 and 30 have the closer distance to node 17. Then we group node cluster A and B as one cluster. As node 29 has the closer distance to node 17, so node 29 should at the middle position of this cluster. Then we should add a link between node 29 and 31. After we have done this, now the depth level for node 25 is changed from 11 to 1, which is shown in Table 5.2. If we select the candidate neighbour pair to connect, which is node 30 and 31, the depth level for node 25 is changed from 11 to 2, which is not the best result. If we can put the link in the middle of the other cluster, that obviously can achieve the best result.

Table 5.2: Link Addition Results Comparison in Logical Structural Hole Scenario

Link addition position	**None**	**30 & 31**	**29 & 31**
Depth level	11	2	1

Algorithm 4 on page 159 is the pseudo-code for the logical Structural Hole scenarios. From line 13 to 25 is identifying the remedy target node and the neighbouring pair with largest hop count when the remedy target node is disabled. Code line 27 to 48 is to group the node into the clusters in one hop distance. If there is only one cluster, then neighbouring pair identified before will be the position for the remedy link. If the cluster number is more than two, then combine the additional clusters with the two clusters where the neighbouring pair resides in. The rule is to combine with the cluster with less average hop count distance. At last, from the remaining two clusters, find out the node with the least average hop count distance to other nodes in the clusters without travelling through the remedy target node. As the node with least average hop count distance is at the centre of the cluster, and therefore should able to reduce the most hop count distance for other nodes when the target node is disabled. Thus, the two centre nodes will be the position for the remedy link. The calculation or say runtime of this algorithm is not affected by the network size, as if focus on the particular remedy target node in the network. However, the density of the network does matter, as the more neighbour of the target node, more algorithm runtime is expected, which is roughly $O(2V + n^4 + 3n^2)$. V is the size of the network and n is the number of neighbours for the target node.

5.4 Physical Structural Hole Scenario

In this section, the remedial method for the physical Structural Hole scenario is discussed. The physical Structural Hole structure leaves no redundant route to the two sides of the network, which means once the node in this position is disabled, the two sides of the network are disconnected. In such a case, the node on the physical Structural Hole position is very attractive to the malicious parties to attack. Moreover, the node in this position normally acts as the gateway, which also means there is a heavy traffic load on this node. In this case, the energy and computing resources are drained up faster than the nodes in other positions. Thus, the nodes in such positions are more likely

Algorithm 4 Remediation for Logical Structural Hole (SH) Scenarios

```
function LOGICALSTRUCTURALHOLEREMEDIATION
    V ← number of nodes in the network integer
    L ← logical Structural Hole arraylist
    Lv ← logical Structural Hole level arraylist
    B ← Betweenness list for each node arraylist
    G ← Network topology V x V matrix data arraylist
    v ← node for remedy integer
    link ← link position arraylist
    D ← $0.0, logicalSHdamagelevel$ double
    C ← cluster group arraylist
    n ← neighbhour list arraylist
    vl ← most critical logical SH depth level integer
    for $i = 0$ to $V - 1$ do
        if $D < Lv[i] \times B[i]$ then
            $D \leftarrow Lv[i] \times B[i]$
            $vl \leftarrow Lv[i]$
            $v \leftarrow L[i]$
    for $i = 0$ to $V - 1$ do
        if $G[v][i] = 1$ then add $i$ to $n$
    $c$ ← cluster members arraylist
    for $i = 0$ to $n.size() - 2$ do
        for $j = i + 1$ to $n.size() - 1$ do
            if $FindRoute(n[i], n[j], v) = vl$ then
                $link[0] \leftarrow n[i]$
                $link[1] \leftarrow n[j]$
    $g1, g2$ ← cluster group of $link[0]$ and $link[1]$
    while $n$ is not empty do
        $con1, con2 \leftarrow false$
        add $n[n.size() - 1]$ to $c$
        $repeat \leftarrow true$
        remove $n[n.size() - 1]$
        while $repeat$ is true do
            for $j = n.size() - 1$ to 0 do
                $repeat \leftarrow false$
                for $k = 0$ to $c.size() - 1$ do
                    if $G[c[k]][n[j]]$ has connection then
                        if $c[k]$ or $n[j]$ is $link[0]$ then
                            $con1 \leftarrow true$
                        else if $c[k]$ or $n[j]$ is $link[1]$ then
                            $con2 \leftarrow true$
                        $repeat \leftarrow true$
                        add $n[j]$ to $c$
                        remove $n[j]$
        add $c$ to $C$
```

Algorithm 4 Remediation for Logical Structural Hole (SH) Scenarios (continued)

```
        if $con1$ is true then
            $g1 \leftarrow c.size() - 1$
        else if $con2$ is true then
            $g2 \leftarrow c.size() - 1$
    if $C.size() = 1$ then
        return $link$
    else if $C.size() > 2$ then
        $d1, d2 \leftarrow 0$
        for $i = 0$ to $C.size() - 1$ do
            if $i = g1$ or $i = g2$ then
                continue to next loop
            for $j = 0$ to $C[i].size() - 1$ do
                $d1 \leftarrow FindRoute(C[i][j], link[0], v) + d1$
                $d2 \leftarrow FindRoute(C[i][j], link[1], v) + d2$
            $d1 \leftarrow d1 \div (C[i].size() - 1)$
            $d2 \leftarrow d2 \div (C[i].size() - 1)$
            if $d1 > d2$ then
                add $C[i]$ to $C[g2]$
            else if $D1 < d2$ then
                add $C[i]$ to $C[g1]$
    $h0 \leftarrow$ smallest average hop count for all neighbours *double*
    $E1 \leftarrow$ selected remedy node *integer*
    if $C[g0].size() > 1$ then
        for $i = 0$ to $C[g0].size() - 2$ do
            $h1 \leftarrow 0$
            for $j = i + 1$ to $C[g0].size() - 1$ do
                $h1 \leftarrow h1 + FindRoute(C[g0][i], C[g0][j], v)$
            if $h0$ is null or $h0 > h1 \div (C[g0].size() - 1)$ then
                $h0 \leftarrow h1 \div (C[g0].size() - 1)$
                $E1 \leftarrow C[g0][i]$
        $link[0] \leftarrow E1$
    $h2 \leftarrow$ smallest average hop count for all neighbours *double*
    $E2 \leftarrow$ selected remedy node *integer*
    if $C[g1].size() > 1$ then
        for $i = 0$ to $C[g1].size() - 2$ do
            $h1 \leftarrow 0$
            for $j = i + 1$ to $C[g1].size() - 1$ do
                $h1 \leftarrow h1 + FindRoute(C[g1][i], C[g1][j], v)$
            if $h2$ is null or $h2 > h1 \div (C[g1].size() - 1)$ then
                $h2 \leftarrow h1 \div (C[g1].size() - 1)$
                $E1 \leftarrow C[g1][i]$
        $link[1] \leftarrow E2$
    return $link$
```

to become selfish, so as to preserve the energy and computing resources. This is the reason why we believed the physical Structural Hole is the most critical potential threat to a network.

For the physical Structural Hole scenarios, the ultimate goal of the remediation is to make the physical Structural Hole no longer a hole anymore. To achieve this, a link can be added as an alternative connection to those unconnected networks or clusters, which are connected by the physical Structural Hole. There are three different circumstances for the potential link candidate positions.

1. The physical Structural Hole is connecting with three or more not related clusters.
2. The candidate neighbour is also a Structural Hole (both physical or logical).
3. The candidate neighbours are not the Structural Hole, and the hole is only connected with two disconnected clusters.

The first case is an existing physical Structural Hole in the network which connects three or more disconnected clusters in the network. Such as the nodes 1, 2, 3, 5, 6, and 7 in the network 19 (shown in Figure 5.3) are in the one cluster as cluster A. The nodes 4, 8, 9, 10, and 11 are in the second cluster as cluster B, and the rest of the nodes except node 12 are in the third cluster C. In this case, one additional link would not be sufficient to back up all the disconnected clusters. We need to select the more important or say bigger disconnected clustering to back up. That is, the remediation needs to achieve the maximum possible outcome with the limited resources. The node Betweenness value is a very good metric to consider. The bigger the node Betweenness value is on the candidate nodes, which means the bigger cluster these nodes belong to. The first step is to group the candidate neighbours, which are in the same cluster. Then simply disable the hole, and see if the candidate neighbours are still able to reach each other, for those still able to reach, they are in the same cluster. The second step is select the neighbours, which are in the same cluster, then sum up their node Betweenness value as their cluster's Betweenness value. The new connection link should be added to connect the two clusters with higher Betweenness value as a redundant connection between these two clusters. In this case, the sum of node Betweenness for cluster A is nodes $5 + 6 + 7 = 43.834$. For cluster B, the sum is nodes $8 + 9 + 10 + 11 = 17$, and the cluster C is nodes $14 + 15 = 66$. As can be seen in Table 5.3, the cluster C is largest, then is cluster A. This means we need to connect clusters A and C together and leave the cluster B alone. Then the final step is to select the two nodes from these two selected clusters - one from each cluster. Then we have the candidate neighbour pairs node 5 to 14, 5 to 15, 6 to 14, 6 to 15, 7 to 14, and 7 to 15. We need to calculate the increment of clustering coefficient for each pair according to the approach in Section 5.2 on page 148 with Equation 5.7 or 5.8 to confirm the actual position the new link should be added to. We have the results indicate the candidate pair nodes 6 to 14 and 6 to 15 have the increment -0.001142. Then we select the pair with the closest distance which is nodes 6 to 14. In this case, we should add a link between nodes 6 and 14.

Table 5.3: Cluster Betweenness Comparison

Cluster	**A**	**B**	**C**
Cluster Betweenness	43.834	17	66

If the candidate neighbours are also the physical Structural Hole or logical Structural Hole as well. In such case, the neighbours on the physical Structural Hole should be first excluded if there is another choice which is not a physical Structural Hole. For a logical Structural Hole, the deeper hole should be excluded if there is another choice. Thus, we take the network 4 in Figure 3.10 on page 72 as an example again, there are seven candidate neighbour pairs. They are node 5 to 6, node 5 to 11, node 5 to 15, node 6 to 15, node 9 to 6, node 9 to 11, and node 9 to 15. After disabling the node 10, we can find that node 5 and 9 are in one cluster, and node 6, 11, and 15 are on the other

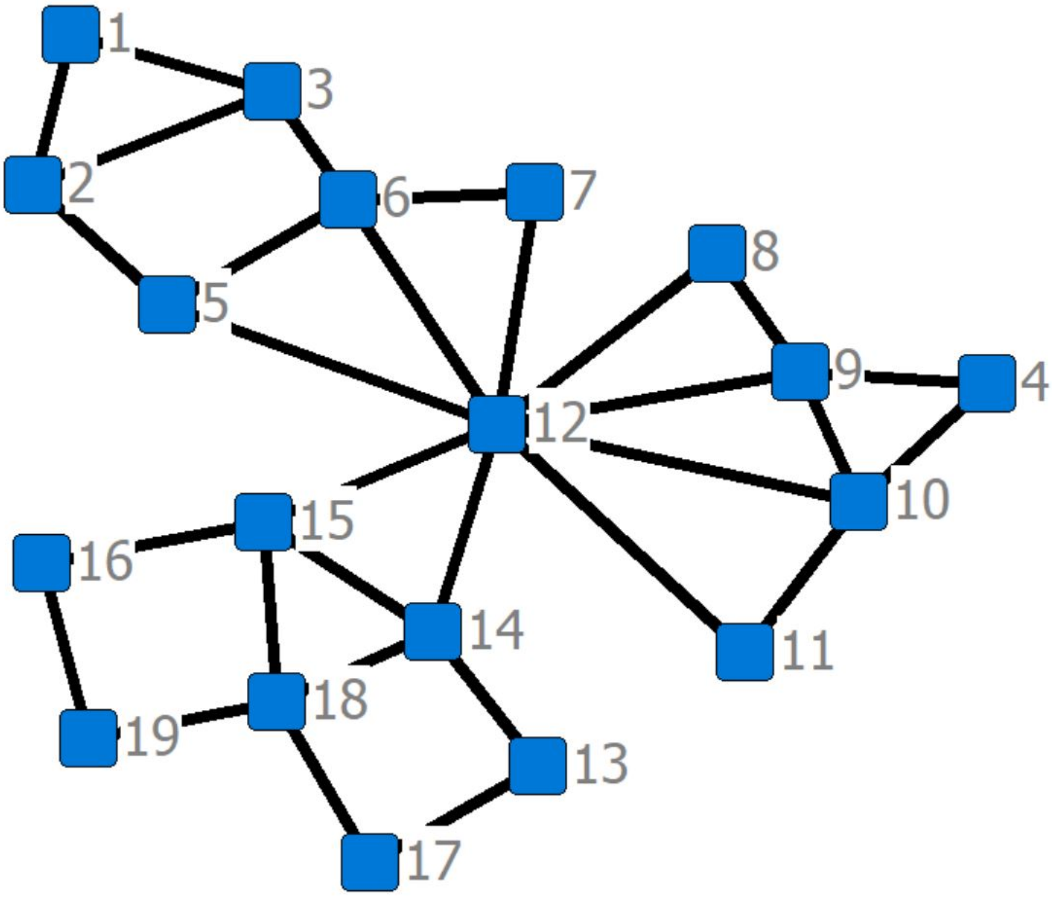


Figure 5.3: Network 19 for the First Scenario

side of the hole. Therefore, we need to exclude the node 6 to 15 as these two nodes are in the same cluster. Moreover, as node 6 is another physical Structural Hole, then we exclude the node 6 as well as there are alternatives. Therefore, node 5 to 11, node 5 to 15, node 9 to 11, and node 9 to 15 are remaining. Next step is to calculate the increment of the clustering coefficient, using the Equation 5.7 on page 152 or 5.8 and calculate that on node 5 to 11 it is -0.01586, on node 5 to 15 is -0.011199, on node 9 to 11 is -0.014567, and finally, node 9 to 15 is -0.009906. From these results, the link should be added between node 9 and 15.

Algorithm 5 Remediation for physical Structural Hole (SH) Scenarios

```
function PHYSCIALSTRUCTURALHOLEREMEDIATION
    V ← number of nodes in the network Integer
    P ← physcial Structural Hole arraylist
    L ← logical Structural Hole arraylist
    Lv ← logical Structural Hole level arraylist
    G ← Network topology V x V matrix data arraylist
    v ← node for remedy Integer
    link ← link position arraylist
    $B \leftarrow 0$
    C ← cluster group arraylist
    LC ← cluster group with logical SH arraylist
    PC ← cluster group with physical SH arraylist
    n ← neighbhour list arraylist
    for $i = 0$ to $P.size() - 1$ do
        $b \leftarrow Betweenness(P[i], G)$
        if $b > B$ then
            $B \leftarrow b$
            $v \leftarrow P[i]$
    for $i = 0$ to $V - 1$ do
        if $G[v][i] = 1$ then
            add $i$ to $n$
    while $n$ is not empty do
        $c$ ← cluster members arraylist
        $lc$ ← logical SH cluster members arraylist
        if $n[0] \in L$ then
            add $n[0]$ to $lc$
        else if $n[0] \in P$ then
            add $n[0]$ to $PC$
            remove $n[0]$
            continue to next loop
        else
            add $n[0]$ to $c$
        for $i = n.size() - 1$ to 1 do
            $r \leftarrow FindRoute(n[0], n[i], v)$
            if $r \neq -1$ then
                if $n[i] \in L$ then
                    add $n[i]$ to $lc$
                    remove $n[i]$
                else
                    add $n[i]$ to $c$
                    remove $n[i]$
        remove $n[0]$
        add $c$ to $C$
        add $lc$ to $LC$
        add $-1$ to $PC$
```

Algorithm 5 Remediation for physical Structural Hole (SH) Scenarios (continued1)

```
if $C.size() > 2$ then
    $BB \leftarrow$ Betweenness for clusters *arraylist*
    for $i = 0$ to $C.size() - 1$ do
        if $C[i]$ is not empty then
            for $j = 0$ to $C[i].size() - 1$ do
                $BB[i] \leftarrow BB[i] + Betweenness(C[i][j], G)$
    for $i = 0$ to $LC.size() - 1$ do
        if $LC[i]$ is not empty then
            for $j = 0$ to $LC[i].size() - 1$ do
                $BB[i] \leftarrow BB[i] + Betweenness(LC[i][j], G)$
    for $i = 0$ to $PC.size() - 1$ do
        if $PC[i] \neq -1$ then
            $BB[i] \leftarrow BB[i] + Betweenness(PC[i], G)$
    $B1, B2 \leftarrow$ the first and second largest cluster Betweenness *double*
    $C0, C1 \leftarrow$ the first and second largest cluster ID *integer*
    for $i = 0$ to $BB.sized() - 1$ do
        if $B1 < BB[i]$ then
            $B1 \leftarrow BB[i]$
            $C0 \leftarrow i$
        if $B2 < BB[i]$ and $BB[i] \neq B1$ then
            $B2 \leftarrow BB[i]$
            $C1 \leftarrow i$
if $C[0]$ and $C[1]$ are not empty then
    $CC \leftarrow 0.0$
    for $i = 0$ to $C[0].size() - 1$ do
        for $j = 0$ to $C[1].size() - 1$ do
            $cc = ClusteringCoefficientIncrement(C[0][i], C[1][j], v)$
            if $CC < cc$ then
                $link[0] \leftarrow C[0][i]$
                $link[1] \leftarrow C[1][j]$
else if $C[0]$ is empty and $C[1]$ is not then
    $CC \leftarrow 0.0$
    for $i = 0$ to $C[1].size() - 1$ do
        $cc \leftarrow ClusteringCoefficientIncrement(PC[0][0], C[1][i], v)$
        if $CC < cc$ then
            $CC \leftarrow cc$
            $link[0] \leftarrow C[1][i]$
    if $LC[0]$ is empty then
        $link[1] \leftarrow PC[0][0]$
    else
        $link[1] \leftarrow FindLowestSHlevelNode(LC[0])$
```

Algorithm 5 Remediation for physical Structural Hole (SH) Scenarios (continued2)

```
87:     else if $C[1]$ is empty and $C[0]$ is not then
88:         $CC \leftarrow 0.0$
89:         for $i = 0$ to $C[0].size() - 1$ do
90:             $cc \leftarrow ClusteringCoefficientIncrement(PC[1][0], C[0][i], v)$
91:             if $CC < cc$ then
92:                 $CC \leftarrow cc$
93:                 $link[0] \leftarrow C[0][i]$
94:         if $LC[1]$ is empty then
95:             $link[1] \leftarrow PC[1][0]$
96:         else
97:             $link[1] \leftarrow FindLowestSHlevelNode(LC[1])$
98:     else
99:         if $LC[0]$ and $LC[1]$ are empty then
100:            $link[0] \leftarrow PC[0][0]$
101:            $link[1] \leftarrow PC[1][0]$
102:        else if $LC[0]$ is empty $LC[1]$ is not then
103:            $link[0] \leftarrow PC[0][0]$
104:            $link[1] \leftarrow FindLowestSHlevelNode(LC[1])$
105:        else if $LC[1]$ is empty $LC[0]$ is not then
106:            $link[0] \leftarrow PC[1][0]$
107:            $link[1] \leftarrow FindLowestSHlevelNode(LC[0])$
108:        else
109:            $link[0] \leftarrow FindLowestSHlevelNode(LC[0])$
110:            $link[1] \leftarrow FindLowestSHlevelNode(LC[1])$
111:    return $link$
```

Algorithm 5 on page 164 is the pseudo-code for physical Structural Hole remediation scenarios. From line 14 to 21 are the codes to identify which physical Structural Hole is selected for remediation. Line 22 to 45 are the codes to find out the number of clusters is connected to the selected physical Structural Hole, and the neighbours belong to which cluster. Line 46 to 67 are the codes for the scenarios when there are more than two clusters connected to the Structural Hole, the remedy algorithm will select the two clusters with higher Betweenness. Line 68 to 75 is for the scenarios when there is no logical and physical Structural Hole in the two clusters. In these scenarios, the algorithm would select a node from each cluster that to connect, so the increment of clustering coefficient can be maximized. The rest of the codes are for the scenarios when there is logical or/and physical Structural Hole existed in the two clusters. When there are only logical and physical Structural Holes in the cluster, select the one with lower logical Structural Hole level. The algorithm for physical Structural Hole remediation is based on the selected remedy Structural Hole node, the change in the size of the network would not affect the runtime of the algorithm. However, the density of the network does affect the algorithm runtimes, as more neighbours for the nodes. For example, at line 33 to 45, the change in neighbour and cluster numbers can cause the algorithm runtime change at $O(n \times c)$. The overall change on runtimes for the algorithm is roughly at $O(2V + 3n^2 + 2n)$, where the V is the size of the network and n is the number of neighbours or density of the network.

We have discussed the remediation approach based on the network objectives assumed in chapter 4. The approaches are based on the network trustworthiness evaluation metrics, which are the clustering coefficient, node Betweenness, and Structural Hole locator. From this discussion, we can find out that there is no universal approach for the remediation of all networks, different scenarios will have different approaches. In the next section, we will discuss the possible feasible ways to make the remediation happen in the real world after the confirmation of the link addition position in the network, how exactly the link can be added to the network.

5.5 Feasible Remedy Methods

In the previous section, the selection of the remediation position is introduced while under the limited resources scenarios. That is, only one connection link is available to be added to the network for the network topology's remediation. In this section, the next step is to explore the possible devices which can be used for the actual network topology's remediation in the physical plane.

In a static network, the nodes in the network are static so that when the Structural Hole is detected, it will be hard or inefficient to add another node to the network to ease the problem. A possible more effective way to resolve the Structural Hole problem in a static network is making use of the existing mobile devices nearby to help forward the data when the Structural Hole is under attacks or suffering network failure. That's it, the T plane in the NTaaS.

Consider the cellular network in Auckland as an example, which is shown in Figure 5.4. In the cellular network, there are cellular towers to have the cellular signal covers the whole Auckland area. Assume there is a tower A is at the Structural Hole position, once tower A is down, the cellular network in Auckland will be separated into two. The current solution as a temporary cellular tower backup is using the cellular on wheel (CoW), which is a vehicle equipped with the cellular tower so it can to move anywhere which it is required. However, this solution can take time to get the CoW driven to the outage affected area as a backup unless the outage it is planned.

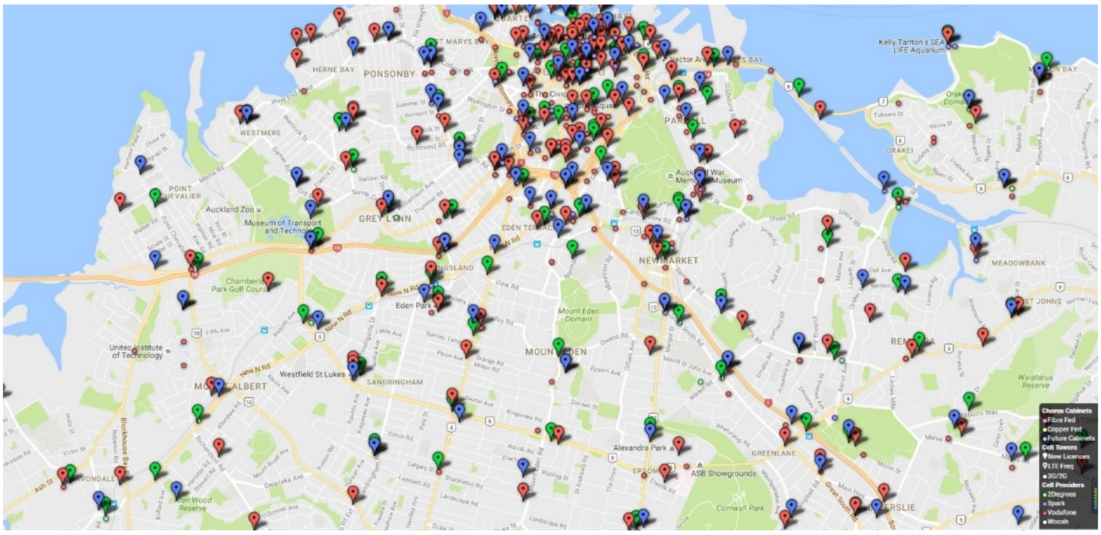


Figure 5.4: Auckland Cellular Towers Map

Another solution could be a public transport network such as the bus network. Taking the Auckland central bus network as an example, which is shown in Figure

5.5, each bus line has a fixed route and travels time so that the bus can be expected to be in a particular place at a particular time. In such case, the regular buses travelling can act as a CoW to forward the data between these separated networks during the network outages such as malicious attacks or network failures until tower A is recovered again. As mentioned above, the buses will pass by regularly, so this solution would be more efficient than the actual CoW located far away and taking a long time to arrive in the outage affect areas. However, there are requirements of bus line density and bus frequency as well. Normally, only a big urban city can satisfy such requirements. Just like the bus map is shown in Figure 5.5 on the following page. The density of the bus lines is very high in the CBD area, but further away from the CBD, the lower the density of the bus lines. In such logic, this approach is not very feasible in a small town or the countryside, as most likely the density and frequency of the buses will be low. Moreover, the cost of network equipment on the buses will be expensive as well, as the number of buses is large for a dense and high-frequency bus network.

The Unnamed Aerial Vehicle (UAV) such as a quadcopters and gliders assisted network is another approach. While UAVs have gained attention in terms of public safety, in communications sector as UAV can act as a flying base station that can be deployed rapidly to the disaster scenes such as a fire scene at the higher level of a skyscraper. A UAV base station can quickly reconnect the victims in a fire scenes to increase their chances of survival (Merwaday, Tuncer, Kumbhar & Guvenc, 2016). We investigate the UAV signal coverage in the disaster as well on study (Mamta et al., 2017). Some researchers have investigated the use of an UAV as an assistant to improve the coverage of the network (Nam, Huang, Li & Xu, 2016). In such cases, the NTaaS can recruit the UAV as a temporary base station for disconnected nodes or Structural Hole structure remediation.

We have set up the simulation studies on the UAV-assisted network with the ONE simulator (Keränen et al., 2009), which is designed for the Delay Tolerance Network

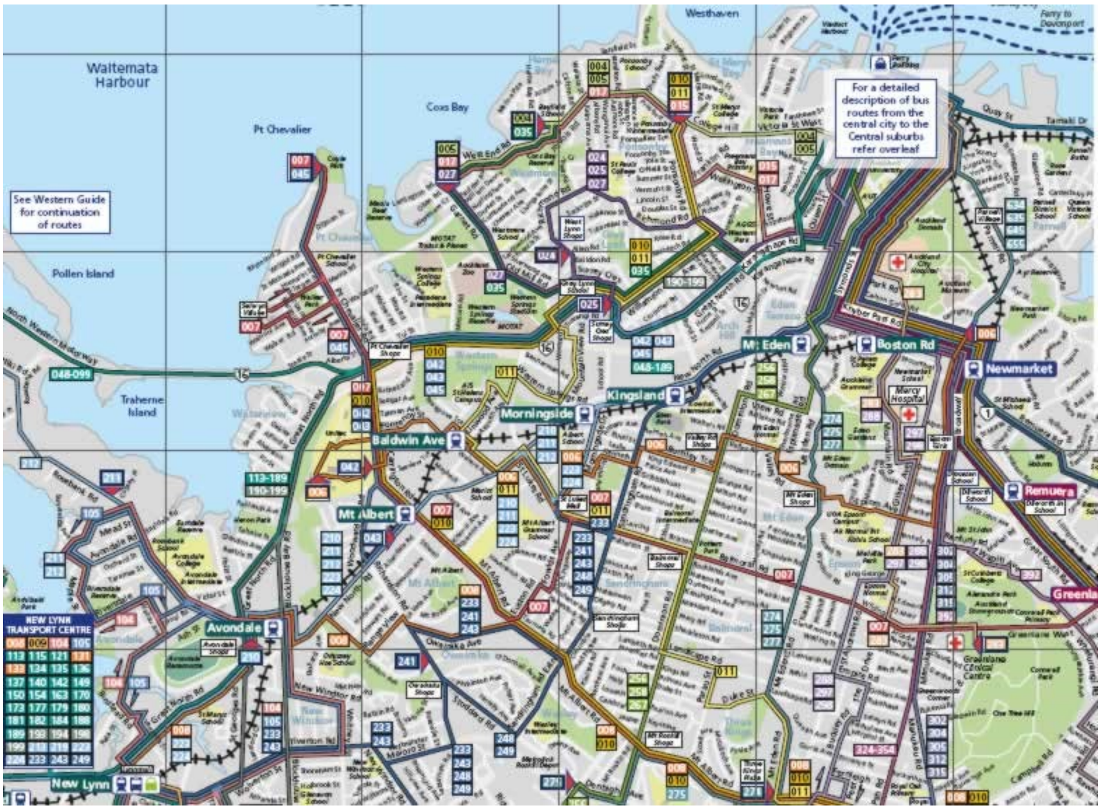


Figure 5.5: Auckland Central Bus Map

(DTN). As the J-Sim simulator is for the static network, it is not very efficient for a dynamic network simulation. We used Network 4 in Figure 3.10 on page 72 as the network topology since it has two physical Structural Holes; node 6 and node 10. We have run the simulation in two scenarios, the first one is using the DTEGR algorithm, and when nodes detect malicious attacks, they will call the UAV to fly to the location of the malicious node and act as a relay for temporary recovery. The UAV will be called as soon as any malicious node is detected. The second scenario uses the DTEGR algorithm but without the assistance of UAV. The packet will be sent from node 1 to node 16, and the attacks will launch on node 10; grey-hole attacks with 50% of packets drops. There were 1416 packets sent in each simulation run, and all nodes are using a Wi-Fi connection. The results are shown in Figure 5.6. For the network which is UAV-assisted, a resulting 95% of packets arrived at the destination sink, while without the assistance of the UAV, there are only 48% of packets arrived at the sink. When there is no UAV assistance with the DTEGR algorithm, as the attacked were launched on the Structural Hole, which there is no alternative route to avoid the attacks, the DTEGR algorithm is able to drop the trust threshold to ensure the network availability at some level. When the network is UAV-assisted, the DTEGR algorithm takes about 10 packets to determine node 10 is malicious, and the additonal 5% of packets loss occurs while waiting for the UAV to come. The faster the UAV come, the less packet loss DTEGR algorithm can achieve. So the speed and the location of the UAV are very important. There are also limitations to the UAV, such as the battery limits the fly time, and weather can have a great impact on the UAV performance and fly time since the windy weather can cost more energy than during a normal flight, or strong winds and heavy rain can even cause loss of UAV lost control. These problems still need to be tackled for the feasible deployment of this technology.

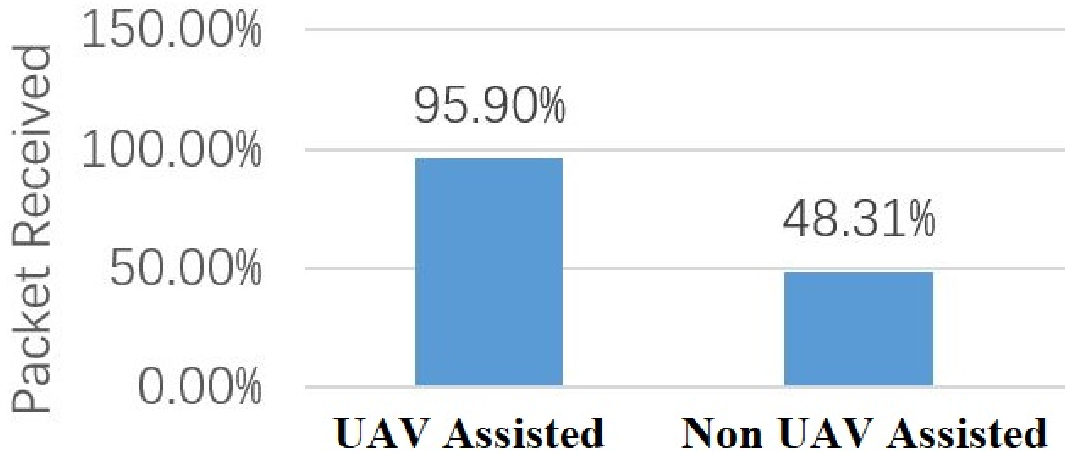


Figure 5.6: UAV-Assisted vs. Non UAV-Assisted

5.6 Summary

In this chapter, the example of network remediation calculation provided is based on the network objectives defined in Chapter 4. We assume the network objectives are the network availability and node trustworthiness in routing. With these two network objectives, we believed the Simmelian Ties structure is preferred in the network as it provides the redundant routes and provides a third party for the reputation system, and Structural Hole structure is not preferred as it does not provide the redundant route. In such case, we identified that the most critical threat to the network topology is a physical Structural Hole, then a logical Structural Hole, and finally the insufficient Simmelian Ties. Thus, we first introduce how to find the position in the network to which to add a link. This link addition can achieve the greatest increment on the weighted clustering coefficient. Then the most critical logical Structural Hole selection method is introduced as well, which the node Betweenness multiplies with the logical Structural Hole depth level. As the node Betweenness can measure the centrality of the node in the network, the more central the position of the node is, the more importance of the node in the network. Moreover, the deeper of the logical Structural Hole, it is believed the more critical it is in the network as well. Then the actual position for the link additional to remedy the logical Structural Hole structures is discussed. Finally, the selection of the physical Structural Hole is determined by the node Betweenness, the higher of the Betweenness value, the more damage of this physical Structural Hole can do while under the malicious attacks or network failure. Moreover, when the physical Structural Hole is connected with more than two disconnected clusters, as only one link can be added, this means only two disconnected clusters on this Structural Hole can be remedied. The node Betweenness of the Structural Hole's neighbours is used to determine which two clusters are selected to remedy. Obviously, the higher of the Betweenness of the cluster, the more critical of this cluster to the network.

There is no universal metric or metrics to remedy the network topologies to the desired topology. Based on the different cases, the remediation method should be added accordingly.

The feasible methods for the remediation of the network topologies in the physical plane are explored in the last section. The NTaaS is used as an upper layer with a global view of the network and recruits nearby devices for the network remediation. The use of the existing transport network can be feasible, but it requires a dense and high-frequency transport network. The UAV has caught many researchers' eyes. It can be quickly deployed and has a fast reaction time. It can also de deploy in a small town or countryside areas. However, it is also been limited by the battery resource and is highly dependent on the weather as well.

Publication related to this Chapter

Mamta, N., **Ming, X.**, William, L., Jairo, G., Luca, C., Arjuna, S., & Arvind, M. (2017). UAV-assisted Edge Infrastructure for Challenged Networks. Presented at the IEEE Infocom Workshop on Wireless Communications and Networking in Extreme Environments (WCNEE 2017), Atlanta, USA.

Ming, X., William, L., Quan, B., & Adnan, A. (2016). The critical role of structural hole in forming trust for Securing Wireless Sensor Networks. International Journal of Information, Communication Technology and Applications, 2(1), 66–84.

Chapter 6

Conclusion

The purpose of this thesis is to evaluate the network trustworthiness in P2P communication network. We purpose the novel service model NTaaS to cover the research gap as shown in the Figure 1.1 on page 5. To support this service model, we have provided an in-depth understanding of the relationship between the underlying network structures and the node trustworthiness in P2P communication network environments. We are striving to contribute new knowledge and solutions on the safe and secure routing in P2P communication networks by providing the trustworthy network structures and trustworthy nodes to effectively mitigate and avoid various malicious attacks and network failures.

6.1 Summary of Contributions

The P2P communication paradigm is becoming increasingly popular nowadays due to the rapid growth of the ICT. As a consequence, many new network paradigms have been introduced, such as D2D communication, the vehicle-to-vehicle communication, an UAV-assisted network, etc. In chapter 2, we point out that the traditional security mechanisms are considered as the hard security in that they are not efficient to tackle the soft security issues. Thus, trust has been introduced from the Sociology into the computer network. Unfortunately, most of the studies on trust in computer network routing today focus on the local aspect, which is the node trustworthiness. In sociology, there is a long debate on whether Simmelian Ties and Structural Hole can be hinder or sustain the outcome performance at the individual or collective level. The adaptive network concept also suggests the co-evolution of the dynamic On and Of network. Thus, from the literature review in these areas in Chapter 2, we have confirmed the motivation of this thesis, which is the identification of a research gap amongst the studies in the computer network, sociology, and adaptive network areas on 'trust' and the interplay of trust behaviours (node trustworthiness) versus the underlying topological connectivity.

First of all, Engle (1999) has summarised the relationship among the Simmelian Ties, Structural Hole, interdependent tasks, and independent tasks. According to his summary, the Simmelian Ties has a positive impact on the performance outcomes of interdependent tasks, but it has a negative impact on the performance outcomes of the independent tasks. For the Structural Hole, it has a positive impact on the performance outcomes of the independent tasks, and a negative impact on the performance outcomes of the interdependent tasks. Finally, the Structural Hole has a negative impact on the positive impact from the Simmelian Ties. The routing in the computer network can be considered as an interdependent tasks. Thus, in chapter 3, to adopt these summaries from the sociology into the computer network routing, we believed that:

H1. A Simmelian Ties characterised network structure has a positive impact on the node trustworthiness in routing.

H2. A Structural Hole characterised network structure has a negative impact on node trustworthiness in routing.

H3. A Structural Hole characterised network structure has a negative impact on the positive impact from a Simmelian Ties characterised network structure.

We have completed extensive simulation studies to validate our assumptions in Chapter 3. We have set up different network topologies and used the same routing algorithm DTEGR as a benchmark. We used the 'packet loss' number as the performance evaluation metric, 'average clustering coefficient' for the Simmelian Ties evaluation, and 'effective size' to measure the Structural Hole. The results show that in the network with a higher average clustering coefficient, the DTEGR algorithm can achieve lower packet loss numbers while the network is under the malicious attacks. While there is Structural Hole in the network, the packet loss number is significantly increased due to the lack of a redundant route to avoid these attacks. These results have validated our three assumptions listed above. Moreover, everything has its pros and cons; when the network is under the 'ballot stuffing' attacks, the networks with a higher clustering coefficient have made the DTEGR algorithm result in a higher packet loss. This is due to the social norm effect from the Simmelian Ties. The Structural Hole can also act as a firewall to stop unwanted data packet from the other sides, such as the virus, etc. All these findings are the first contribution of this thesis, which is the exploration of underlies topological connectivity versus the node trustworthiness in routing.

Secondly, as we have proven the underlies topological connectivities can affect the node trustworthiness in the network. This thesis has introduced a new term called 'Network Trustworthiness' in Chapter 4, which suggests a P2P network is trustworthy when it fulfils its objectives under any expected or unexpected circumstance. That is, the network trustworthiness is objective-dependent. There is no trustworthy network, but each network is trustworthy in some aspect(s). This is the second contribution of this thesis.

Thirdly, according to the literature review on trust in P2P communication routing, there is lack of trust modelling and evaluation from the global aspect in the current work.

We identified this as a current research gap. To cover this research gap, this thesis has introduced the term Network Trustworthiness as mentioned in our third contribution. The third contribution of the thesis is the proposal of a service platform called 'Network Trustworthiness as a Service (NTaaS)', which treats Network Trustworthiness as a service provided to the P2P network users. The network trustworthiness models and evaluates the 'trust' on P2P networks from the global aspect. The nodes on the physical plane want to start the P2P communication with the target node. The distance can be one hop or even ten hops away. The requesting node sends out the evaluation request to the T Plane. Then the T Plane starts the evaluation processes for the target node. The evaluation processes also enquire of the available nodes in the area for the feedback on the target node. Once the evaluation has done by the T Plane, the results are sent back to the requesting node. Then the requesting node based on its own trustworthiness threshold, determines whether the target node is 'trustworthy'. The requesting nodes form its location topology with only trustworthy nodes. It also updates this topological information back to the T Plane. The T Plane uses this topological information to form the topology of the complete network. Then the topology evaluation will be performed. If there is any threat found in the network, it looks for the recruitment from the available nodes again for the network structure remediation. In the disaster scenarios where the Internet is not available, the trustworthy nodes, which are the members of the NTaaS as well, will be assigned a certificate for proof of trustworthy node while the Internet is still available, then the subscriber of the NTaaS knows which node is trustworthy.

Fourthly, after the NTaaS framework is proposed, the core of the NTaaS, which is the evaluation of network trustworthiness needs to be addressed. Thus, the fifth contribution of this thesis is the network trustworthiness T evaluation framework as has been proposed. As mentioned in the previous paragraph, there is no trustworthy network, but a network is trustworthy in some aspect(s). Thus, the evaluation metrics for the T should be different according to the network objectives. Therefore, we used the network availability and node trustworthiness as the objectives in chapter 4 for the validation of the network trustworthiness T evaluation framework. As we have already validated the impact of Simmelian Ties and Structural Hole on node trustworthiness in routing in the P2P environment. Thus, we keep using these two structures for the validation. The extensive simulation studies have validated the accuracy of the T evaluation according to the packet loss results.

Fifthly, the clustering coefficient for the Simmelian Ties evaluation is only at the node level rather than at the network as a whole. The average clustering coefficient is a common way to synthesise the node clustering coefficient to evaluate the whole network. However, the nodes at different positions in the network can mean different things for the network availability. For example, the node at a central position can be more critical. Thus, the average clustering coefficient cannot reflect these difference. This thesis proposed the weighted clustering coefficient to represent these difference in their positions. The node Betweenness has been selected as the weight factors to synthesise the node clustering coefficient to evaluate the Simmelian Ties for the network as a whole. The simulation studies have validated that the weighted clustering coefficient is much more accurate than the average clustering coefficient.

Sixthly, the effective size and the Simmelian Brokerage are the most commonly used metrics for the evaluation of the Structural Hole in the network. They cannot confirm if there is any Structural Hole in the network, as these require calculation globally rather than only on the particular node. Thus, this thesis has proposed the Structural Hole Locator (SHL) algorithm to detect any physical and logical Structural Hole in the network. The identification of the physical or logical Structural Hole can be used in the T evaluation as the penalty structures to reduce the T accordingly. The provided simulation studies have validated that with this penalty, the network trustworthiness T is evaluated more accurately, which is reflected by packet loss number difference.

Seventhly, the network trustworthiness T is target attacks focus evaluation. Random attacks scenarios however, require a better approach for the evaluation. Thus, this thesis has introduced the Trustworthiness Tolerance Margin (TTM), which uses the Monte Carlo Simulation approach. The Monte Carlo Simulation uses a large number of random attack simulations and put the results of T on average with standard deviation. In such case, we can use this average and standard deviation of T to represent the network trustworthiness with upper and lower boundary values.

Finally, after the network trustworthiness has been evaluated, this thesis has discussed the remediation approaches, which is based on the T evaluation metrics. For the network availability and node trustworthiness in routing scenario, the most critical threat is a physical Structural Hole, then a logical Structural Hole, and finally the insufficient Simmelian Ties, so the remediation order is from the physical Structural Hole to the logical Structural Hole, and finally the clustering coefficient. As the last of chapter 5, this thesis also suggests some possible candidate nodes or methods for the actual network structure remediation in the real world for NTaaS. Such as the existing public transport system, UAV, etc.

6.2 Future Work

Considering the work covered in this thesis occurred within the constraint of a limited time period and the anticipated development of the future network, it would be useful to highlight some future areas to be further investigated.

First of all, the Simmelian Ties in the network have proven to have a positive impact on the node trustworthiness in routing and prove redundant. However, the networks without the Simmelian Ties do not mean there is no redundant routes and many Structural Holes, such as in grid network topologies. There are no Simmelian Ties in the network but it has many redundant routes. Even though it does not have a third party for the reputation system, it is also not so bad as to be rated as 0. Therefore, other than the clustering coefficient for the evaluation of the network trustworthiness T, there should be another metric involved to tackle these scenarios. For the T evaluation framework, it is flexible to have an additional metric for the evaluation; the key problem is which metric?

Secondly, Omnet++ is the software to be used for generating the random wireless networks. Due to the limitation of Omnent++ for random network generations, the network size is limited to 50 and the number of suitable random network topologies for the wireless P2P network is limited to eight networks. The number of random network topologies might not be sufficient to verify the effectiveness of T evaluation framework thoroughly. The other random network generation models such as Erdos-Renyi, Barabási–Albert, and Watts–Strogatz have been attempted by this study. However, the networks are generated by these models are not suitable for wireless P2P network simulation. Therefore, the random generation of the wireless P2P network is put in the future work here.

Thirdly, as the topological metrics can have different value ranges and different scales (e.g. node level, network level), the normalization of the metrics would be important for the accuracy of the T. This can be a future work for further exploration. Moreover, the comparison of networks with different network sizes would be an issue, which falls into this category as well.

Fourthly, for the logical Structural Hole defined with the SHL algorithm, this thesis only has one simple sample definition of logical Structural Hole. The actual logical Structural Hole can be more complicated. The SHL algorithm for the detection of logical Structural Holes would need to be adjusted accordingly in the actual scenarios.

Finally, the network trustworthiness T assumes only one node is under attack at a time; in other words, the multiple attack scenarios are not considered in this study. This can be further explored in the future work.

References

Adams, W. & Davis, I., N.J. (2005, June). Toward a decentralized trust-based access control system for dynamic collaboration. In *Information Assurance Workshop, 2005. IAW '05. Proceedings from the Sixth Annual IEEE SMC* (pp. 317–324). doi: 10.1109/IAW.2005.1495969

Aumasson, J.-P., Henzen, L., Meier, W. & Phan, R. (2008). Sha-3 proposal blake. *Submission to NIST*.

Bai, Q. h. & Zheng, Y. (2011, July). Study on the access control model. In *Proceedings of 2011 Cross Strait Quad-Regional Radio Science and Wireless Technology Conference* (Vol. 1, pp. 830–834). doi: 10.1109/CSQRWC.2011.6037079

Bandyszak, T., Moffie, M., Goldsteen, A., Melas, P., Nasser, B. I., Kalogiros, C., ... Weyer, T. (2016, July). Supporting Coordinated Maintenance of System Trustworthiness and User Trust at Runtime. In *Trust Management X* (pp. 96–112). Springer, Cham. Retrieved 2017-08-13, from `https://link-springer-com.ezproxy.aut.ac.nz/chapter/10.1007/978-3-319-41354-9_7` doi: 10.1007/978-3-319-41354-9_7

Bao, F. & Chen, I.-R. (2012, June). Trust management for the internet of things and its application to service composition. In *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2012 IEEE International Symposium on a* (pp. 1–6). doi: 10.1109/WoWMoM.2012.6263792

Bao, F., Chen, I.-R., Chang, M. & Cho, J.-H. (2012, June). Hierarchical Trust Management for Wireless Sensor Networks and its Applications to Trust-Based Routing and Intrusion Detection. *IEEE Transactions on Network and Service Management*, *9*(2), 169–183. doi: 10.1109/TCOMM.2012.031912.110179

Bao, F., Chen, I.-R. & Guo, J. (2013, March). Scalable, adaptive and survivable trust management for community of interest based Internet of Things systems. In *2013 IEEE Eleventh International Symposium on Autonomous Decentralized Systems (ISADS)* (pp. 1–7). doi: 10.1109/ISADS.2013.6513398

Barabási, A.-L. & Albert, R. (1999, October). Emergence of Scaling in Random Networks. *Science*, *286*(5439), 509–512. Retrieved 2015-02-08, from `http://www.sciencemag.org/content/286/5439/509` doi: 10.1126/science.286.5439.509

Barrera, D. & Bunt, G. G. v. d. (2009, December). Learning to Trust: Networks Effects Through Time. *European Sociological Review*, *25*(6), 709–721. Retrieved

2015-02-01, from http://esr.oxfordjournals.org.ezproxy.aut.ac.nz/content/25/6/709 doi: 10.1093/esr/jcn078

Beatty, P., Reay, I., Dick, S. & Miller, J. (2011, April). Consumer Trust in e-Commerce Web Sites: A Meta-study. *ACM Comput. Surv.*, *43*(3), 14:1–14:46. Retrieved 2015-05-27, from http://doi.acm.org/10.1145/1922649.1922651 doi: 10.1145/1922649.1922651

Borgatti, S., Everett, M. & Freeman, L. (2002). *Ucinet for Windows: Software for Social Network Analysis*. Analytic Technologies.

Buchegger, S. & Le Boudec, J.-Y. (2002). Performance Analysis of the CONFIDANT Protocol. In *Proceedings of the 3rd ACM International Symposium on Mobile Ad Hoc Networking & Computing* (pp. 226–236). New York, NY, USA: ACM. Retrieved 2016-10-27, from http://doi.acm.org/10.1145/513800.513828 doi: 10.1145/513800.513828

Bunt, G. G. V. D., Duijn, M. A. J. V. & Snijders, T. A. B. (1999, July). Friendship Networks Through Time: An Actor-Oriented Dynamic Statistical Network Model. *Computational & Mathematical Organization Theory*, *5*(2), 167–192. Retrieved 2015-02-01, from http://link.springer.com/article/10.1023/A%3A1009683123448 doi: 10.1023/A:1009683123448

Bunt, G. G. v. d., Wittek, R. P. M. & Klepper, M. C. d. (2005, September). The Evolution of Intra-Organizational Trust Networks The Case of a German Paper Factory: An Empirical Test of Six Trust Mechanisms. *International Sociology*, *20*(3), 339–369. Retrieved 2015-02-01, from http://iss.sagepub.com.ezproxy.aut.ac.nz/content/20/3/339 doi: 10.1177/0268580905055480

Burt, R. S. (2000). The network structure of social capital. *Research in Organizational Behavior, Vol 22, 2000*, *22*, 345–423. doi: 10.1016/S0191-3085(00)22009-1

Burt, R. S. (2009). *Structural holes: The social structure of competition*. Harvard university press.

Castelfranchi, C., Falcone, R. & Pezzulo, G. (2003). Integrating Trustfulness and Decision Using Fuzzy Cognitive Maps. In P. Nixon & S. Terzis (Eds.), *Trust Management* (pp. 195–210). Springer Berlin Heidelberg. Retrieved 2015-05-31, from http://link.springer.com.ezproxy.aut.ac.nz/chapter/10.1007/3-540-44875-6_14

Caton, S., Dukat, C., Grenz, T., Haas, C., Pfadenhauer, M. & Weinhardt, C. (2012, November). Foundations of Trust: Contextualising Trust in Social Clouds. In *2012 Second International Conference on Cloud and Green Computing (CGC)* (pp. 424–429). doi: 10.1109/CGC.2012.89

Chadwick, D., Lievens, S., Den Hartog, J., Pashalidis, A. & Alhadeff, J. (2011, July). My Private Cloud Overview: A Trust, Privacy and Security Infrastructure for the Cloud. In *2011 IEEE International Conference on Cloud Computing (CLOUD)* (pp. 752–753). doi: 10.1109/CLOUD.2011.113

Chen, I. R., Bao, F. & Guo, J. (2016, November). Trust-Based Service Management for Social Internet of Things Systems. *IEEE Transactions on Dependable and Secure Computing*, *13*(6), 684–696. doi: 10.1109/TDSC.2015.2420552

Chen, I.-R., Guo, J., Bao, F. & Cho, J.-H. (2014, August). Trust management in mobile ad hoc networks for bias minimization and application performance maximization. *Ad Hoc Networks*, *19*, 59–74. Retrieved 2014-06-14, from `http://www.sciencedirect.com/science/article/pii/S1570870514000419` doi: 10.1016/j.adhoc.2014.02.005

Cho, J.-H., Swami, A. & Chen, I.-R. (2011). A Survey on Trust Management for Mobile Ad Hoc Networks. *IEEE Communications Surveys Tutorials*, *13*(4), 562–583. doi: 10.1109/SURV.2011.092110.00088

Coleman, J. S. & Coleman, J. S. (1994). *Foundations of Social Theory*. Harvard University Press.

Crosby, G., Pissinou, N. & Gadze, J. (2006, April). A framework for trust-based cluster head election in wireless sensor networks. In *Second IEEE Workshop on Dependability and Security in Sensor Networks and Systems, 2006. DSSNS 2006* (pp. 10 pp.–22). doi: 10.1109/DSSNS.2006.1

Denning, D. E. (1993). A New Paradigm for Trusted Systems. In *Proceedings on the 1992-1993 Workshop on New Security Paradigms* (pp. 36–41). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/283751.283772` doi: 10.1145/283751.283772

Ellens, W. & Kooij, R. E. (2013, November). Graph measures and network robustness. *arXiv:1311.5064 [physics]*. Retrieved 2015-01-01, from `http://arxiv.org/abs/1311.5064`

Engle, S. L. (1999). Structural holes and Simmelian ties: Exploring social capital, task interdependence, and individual effectiveness. *Unt Theses & Dissertations*.

Eschenauer, L., Gligor, V. D. & Baras, J. (2002, April). On Trust Establishment in Mobile Ad-Hoc Networks. In B. Christianson, B. Crispo, J. A. Malcolm & M. Roe (Eds.), *Security Protocols* (pp. 47–66). Springer Berlin Heidelberg. Retrieved 2015-09-16, from `http://link.springer.com.ezproxy.aut.ac.nz/chapter/10.1007/978-3-540-39871-4_6`

Fernandez-Gago, M. C., Roman, R. & Lopez, J. (2007, July). A Survey on the Applicability of Trust Management Systems for Wireless Sensor Networks. In *Third International Workshop on Security, Privacy and Trust in Pervasive and Ubiquitous Computing (SecPerU 2007)* (pp. 25–30). doi: 10.1109/SECPERU.2007.3

Fishbein, M. & Ajzen, I. (1975). *Belief, attitude, intention and behavior: an introduction to theory and research*. Retrieved 2015-05-27, from `http://trid.trb.org/view.aspx?id=1150648`

Freeman, L. C. (1978, January). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239. Retrieved 2015-12-17, from `http://linkinghub.elsevier.com/retrieve/pii/0378873378900217` doi: 10.1016/0378-8733(78)90021-7

Gabarro, J. J. (1978). The development of trust, influence, and expectations. *Interpersonal behavior: Communication and understanding in relationships*, *290*, 303. Retrieved 2016-10-29, from `http://scholar.google.com/scholar?cluster=15864978835832694581&hl=en&oi=scholarr`

Gambetta, D. (1988). *Trust: Making and Breaking Cooperative Relations*. Blackwell.

Gao, Q. (2012, March). Biometric authentication in Smart Grid. In *2012 International Energy and Sustainability Conference (IESC)* (pp. 1–5). doi: 10.1109/IESC.2012.6217197

Gonzalez, J., Anwar, M. & Joshi, J. (2011, November). Trust-Based Approaches to Solve Routing Issues in Ad-Hoc Wireless Networks: A Survey. In *2011 IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)* (pp. 556–563). doi: 10.1109/TrustCom.2011.72

Granovetter, M. (2005). The Impact of Social Structure on Economic Outcome...: Full Text Finder Results. *Journal of Economic Perspectives*, *19*(1), 33–50.

Granovetter, M. S. (1973, May). The Strength of Weak Ties. *American Journal of Sociology*, *78*(6), 1360–1380. Retrieved 2015-02-01, from http://www.jstor.org/stable/2776392

Gross, T. & Blasius, B. (2008, March). Adaptive coevolutionary networks: a review. *Journal of The Royal Society Interface*, *5*(20), 259–271. Retrieved 2015-06-07, from http://rsif.royalsocietypublishing.org.ezproxy.aut.ac.nz/content/5/20/259 doi: 10.1098/rsif.2007.1229

Gross, T. & Sayama, H. (2009, January). Adaptive Networks. In T. Gross & H. Sayama (Eds.), *Adaptive Networks* (pp. 1–8). Springer Berlin Heidelberg. Retrieved 2014-06-03, from http://link.springer.com/chapter/10.1007/978-3-642-01284-6_1

Guo, Y. & Wang, Y. (2007, September). Establishing Trust Relationship in Mobile Ad-Hoc Network. In *2007 International Conference on Wireless Communications, Networking and Mobile Computing* (pp. 1562–1564). doi: 10.1109/WICOM.2007.393

Heinzelman, W., Chandrakasan, A. & Balakrishnan, H. (2000, January). Energy-efficient communication protocol for wireless microsensor networks. In *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, 2000* (pp. 10 pp. vol.2–). doi: 10.1109/HICSS.2000.926982

Heron, S. (2009, December). Advanced Encryption Standard (AES). *Network Security*, *2009*(12), 8–12. Retrieved from http://www.sciencedirect.com/science/article/pii/S1353485810700064 doi: 10.1016/S1353-4858(10)70006-4

Holmes, J. G. (1991). Trust and the appraisal process in close relationships. In W. H. Jones & D. Perlman (Eds.), *Advances in personal relationships: A research annual, Vol. 2* (pp. 57–104). Oxford, England: Jessica Kingsley Publishers.

Hornby, A. (1988). *Oxford Advanced Learner's Dictionary of Current English*. Oxford, UK: Oxford University Press.

Hui-hui, D., Ya-jun, G., Zhong-qiang, Y. & Hao, C. (2009, May). A Wireless Sensor Networks Based on Multi-angle Trust of Node. In *International Forum on Information Technology and Applications, 2009. IFITA '09* (Vol. 1, pp. 28–31). doi: 10.1109/IFITA.2009.71

Jakobsen, T. (1995, July). A Fast Method for Cryptanalysis of Substitution Ciphers. *Cryptologia*, *19*(3), 265–274. Retrieved from http://dx.doi.org/

10.1080/0161-119591883944 doi: 10.1080/0161-119591883944

Jiang, T. & Baras, J. S. (2005). Autonomous Trust Establishment. In *Proceedings INOC* (Vol. 2005).

Jøsang, A. (1996). The Right Type of Trust for Distributed Systems. In *Proceedings of the 1996 Workshop on New Security Paradigms* (pp. 119–131). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/304851.304877 doi: 10.1145/304851.304877

Kahate, A. (2013). *Cryptography and Network Security*. Tata McGraw-Hill Education. (Google-Books-ID: xCDZAgAAQBAJ)

Keränen, A., Ott, J. & Kärkkäinen, T. (2009). The ONE Simulator for DTN Protocol Evaluation. In *Proceedings of the 2nd International Conference on Simulation Tools and Techniques* (pp. 55:1–55:10). ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering). Retrieved from http://dx.doi.org/10.4108/ICST.SIMUTOOLS2009.5674 doi: 10.4108/ICST.SIMUTOOLS2009.5674

Khan, M. S., Midi, D., Khan, M. I. & Bertino, E. (2015, August). Adaptive Trust Threshold Strategy for Misbehaving Node Detection and Isolation. In *2015 IEEE Trustcom/BigDataSE/ISPA* (Vol. 1, pp. 718–725). doi: 10.1109/Trustcom.2015.439

Kour, H. & Sharma, A. K. (2010, July). Hybrid Energy Efficient Distributed Protocol for Heterogeneous Wireless Sensor Network. *International journal of computer applications*, *4*(6), 1–5.

Krackhardt, D. (1999). The ties that torture: Simmelian tie analysis in organizations. *Research in the Sociology of Organizations*, *16*(1), 183–210.

Kroese, D. P., Brereton, T., Taimre, T. & Botev, Z. I. (2014, November). Why the Monte Carlo method is so important today. *Wiley Interdisciplinary Reviews: Computational Statistics*, *6*(6), 386–392. Retrieved from http://onlinelibrary.wiley.com.ezproxy.aut.ac.nz/doi/10.1002/wics.1314/abstract doi: 10.1002/wics.1314

Kuan, H.-H. & Bock, G.-W. (2005, January). The Collective Reality of Trust: An Investigation of Social Relations and Networks on Trust in Multi-Channel Retailers. *ECIS 2005 Proceedings*. Retrieved from http://aisel.aisnet.org/ecis2005/13

Latora, V., Nicosia, V. & Panzarasa, P. (2013, March). Social Cohesion, Structural Holes, and a Tale of Two Measures. *Journal of Statistical Physics*, *151*(3-4), 745–764. Retrieved 2015-03-29, from http://link.springer.com.ezproxy.aut.ac.nz/article/10.1007/s10955-013-0722-z doi: 10.1007/s10955-013-0722-z

Lee, K., Teh, C. & Tan, Y. (2006). Decrypting english text using enhanced frequency analysis. In *National Seminar on Science, Technology and Social Sciences* (pp. 1–7).

Lesueur, F., Me, L. & Tong, V. V. T. (2009, September). An efficient distributed PKI for structured P2p networks. In *2009 IEEE Ninth International Conference on Peer-to-Peer Computing* (pp. 1–10). doi: 10.1109/P2P.2009.5284491

Lewis, J. D. & Weigert, A. (1985, June). Trust as a Social Reality. *Social Forces*, *63*(4), 967–985. Retrieved 2015-05-27, from http://www.jstor.org/stable/2578601 doi: 10.2307/2578601

Li, X., Xuan, Z. & Wen, L. (2011, March). Research on the Architecture of Trusted Security System Based on the Internet of Things. In *2011 International Conference on Intelligent Computation Technology and Automation (ICICTA)* (Vol. 2, pp. 1172–1175). doi: 10.1109/ICICTA.2011.578

Liqin, T., Chuang, L. & Tieguo, J. (2006, November). Quantitative Analysis of Trust Evidence in Internet. In *International Conference on Communication Technology, 2006. ICCT '06* (pp. 1–5). doi: 10.1109/ICCT.2006.342023

Liu, Z., Yu, L., Cheng, W. & Wang, K. (2011, September). An Independent Trust Routing Framework Based on Trust Topology Control. In *2011 7th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM)* (pp. 1–4). doi: 10.1109/wicom.2011.6040149

Mahalle, P., Thakre, P., Prasad, N. & Prasad, R. (2013, June). A fuzzy approach to trust based access control in internet of things. In *2013 3rd International Conference on Wireless Communications, Vehicular Technology, Information Theory and Aerospace Electronic Systems (VITAE)* (pp. 1–5). doi: 10.1109/VITAE.2013.6617083

Mamta, N., Ming, X., William, L., Jairo, G., Luca, C., Arjuna, S. & Arvind, M. (2017, May). UAV-assisted Edge Infrastructure for Challenged Networks. Atlanta, USA. Retrieved 2017-07-18, from https://www.researchgate.net/publication/314439784_UAV-assisted_Edge_Infrastructure_for_Challenged_Networks

Marsh, S. P. (1994). Formalising trust as a computational concept. Retrieved 2017-06-27, from http://dspace.stir.ac.uk/handle/1893/2010

Mayer, R. C., Davis, J. H. & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of management review*, *20*(3), 709–734.

McCabe, C., Watson, R. A., Prichard, J. & Hall, W. (2011). The Web As an Adaptive Network: Coevolution of Web Behavior and Web Structure. In *Proceedings of the 3rd International Web Science Conference* (pp. 22:1–22:7). New York, NY, USA: ACM. Retrieved 2014-05-13, from http://doi.acm.org.ezproxy.aut.ac.nz/10.1145/2527031.2527044 doi: 10.1145/2527031.2527044

Menezes, A. J., Oorschot, P. C. v. & Vanstone, S. A. (1996). *Handbook of Applied Cryptography*. CRC Press. (Google-Books-ID: MhvcBQAAQBAJ)

Merwaday, A., Tuncer, A., Kumbhar, A. & Guvenc, I. (2016, December). Improved Throughput Coverage in Natural Disasters: Unmanned Aerial Base Stations for Public-Safety Communications. *IEEE Vehicular Technology Magazine*, *11*(4), 53–60. doi: 10.1109/MVT.2016.2589970

Michiardi, P. & Molva, R. (2002). Core: A Collaborative Reputation Mechanism to Enforce Node Cooperation in Mobile Ad Hoc Networks. In B. Jerman-Blažič & T. Klobučar (Eds.), *Advanced Communications and Multimedia Security* (pp. 107–121). Springer US. Retrieved 2015-05-31, from http://link.springer.com.ezproxy.aut

.ac.nz/chapter/10.1007/978-0-387-35612-9_9

Miglani, A., Bhatia, T. & Goel, S. (2015, April). TRUST based energy efficient routing in LEACH for wireless sensor network. In *2015 Global Conference on Communication Technologies (GCCT)* (pp. 361–365). doi: 10.1109/GCCT.2015.7342684

Mohammadi, N. G. & Heisel, M. (2016, July). Enhancing Business Process Models with Trustworthiness Requirements. In *Trust Management X* (pp. 33–51). Springer, Cham. Retrieved 2017-08-13, from https://link-springer-com.ezproxy.aut.ac.nz/chapter/10.1007/978-3-319-41354-9_3 doi: 10.1007/978-3-319-41354-9_3

Mu, B. & Yuan, S. (2010, June). A method for evaluating initial trust value of direct trust and recommender trust. In *2010 International Conference on Computer Design and Applications (ICCDA)* (Vol. 2, pp. V2–185–V2–190). doi: 10.1109/ICCDA.2010.5541203

Nam, L. H., Huang, L., Li, X. J. & Xu, J. F. (2016, April). An approach for coverage path planning for UAVs. In *2016 IEEE 14th International Workshop on Advanced Motion Control (AMC)* (pp. 411–416). doi: 10.1109/AMC.2016.7496385

Noor, T. & Sheng, Q. (2011). Trust as a service: A framework for trust management in cloud environments. *Web Information System Engineering–WISE 2011*, 314–321.

Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., … Barabási, A.-L. (2007, May). Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, *104*(18), 7332–7336. Retrieved 2015-02-01, from http://www.pnas.org/content/104/18/7332 doi: 10.1073/pnas.0610245104

Oren, N., Griffiths, N. & Luck, M. (2013). The Effects of Trust on Convention Emergence. *Proceedings of 16th International Workshop on Trust in Agent Societies (TRUST 2013), 51-62.*

Peng, S., He, J. & Meng, Y. (2008, August). Reputation-based Trust Update in Network Environment. In *2008 International Symposium on Electronic Commerce and Security* (pp. 118–123). doi: 10.1109/ISECS.2008.211

Pfleeger, C. P. & Pfleeger, S. L. (2002). *Security in Computing* (3rd ed.). Prentice Hall Professional Technical Reference.

Rasmusson, L. & Jansson, S. (1996). Simulated Social Control for Secure Internet Commerce. In *Proceedings of the 1996 Workshop on New Security Paradigms* (pp. 18–25). New York, NY, USA: ACM. Retrieved 2017-04-30, from http://doi.acm.org/10.1145/304851.304857 doi: 10.1145/304851.304857

Resnick, P., Kuwabara, K., Zeckhauser, R. & Friedman, E. (2000, December). Reputation Systems. *Commun. ACM*, *43*(12), 45–48. Retrieved 2015-05-31, from http://doi.acm.org/10.1145/355112.355122 doi: 10.1145/355112.355122

Rivest, R. (n.d.). *The MD5 Message-Digest Algorithm.* Retrieved 2017-07-17, from https://tools.ietf.org/html/rfc1321

Rivest, R. L., Shamir, A. & Adleman, L. (1978, February). A Method for Obtaining

Digital Signatures and Public-key Cryptosystems. *Commun. ACM*, *21*(2), 120–126. Retrieved from `http://doi.acm.org/10.1145/359340.359342` doi: 10.1145/359340.359342

Rotter, J. B. (1967, December). A new scale for the measurement of interpersonal trust. *Journal of Personality*, *35*(4), 651–665. Retrieved 2015-05-27, from `http://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=8935118&site=ehost-live&scope=site` doi: 10.1111/j.1467-6494.1967.tb01454.x

Rousseau, D. M., Sitkin, S. B., Burt, R. S. & Camerer, C. (1998, July). Introduction to Special Topic Forum: Not so Different after All: A Cross-Discipline View of Trust. *The Academy of Management Review*, *23*(3), 393–404. Retrieved 2015-05-27, from `http://www.jstor.org/stable/259285`

Ruohomaa, S., Kutvonen, L. & Koutrouli, E. (2007, April). Reputation Management Survey. In *The Second International Conference on Availability, Reliability and Security, 2007. ARES 2007* (pp. 103–111). doi: 10.1109/ARES.2007.123

Savas, O., Jin, G. & Deng, J. (2013, May). Trust management in cloud-integrated Wireless Sensor Networks. In *2013 International Conference on Collaboration Technologies and Systems (CTS)* (pp. 334–341). doi: 10.1109/CTS.2013.6567251

Sayama, H., Pestov, I., Schmidt, J., Bush, B. J., Wong, C., Yamanoi, J. & Gross, T. (2013, May). Modeling complex systems with adaptive networks. *Computers & Mathematics with Applications*, *65*(10), 1645–1664. Retrieved 2014-06-03, from `http://www.sciencedirect.com/science/article/pii/S0898122112007018` doi: 10.1016/j.camwa.2012.12.005

Schlenker, B. R., Helm, B. & Tedeschi, J. T. (1973, March). The effects of personality and situational variables on behavioral trust. *Journal of Personality*, *25*(3), 419–427.

Schoorman, F. D., Mayer, R. C. & Davis, J. H. (2007, April). An Integrative Model of Organizational Trust: Past, Present, and Future. *Academy of Management Review*, *32*(2), 344–354. Retrieved 2017-06-27, from `http://amr.aom.org/content/32/2/344` doi: 10.5465/AMR.2007.24348410

Sharma, K. & Ghose, M. K. (2010). Wireless Sensor Networks: An Overview on its Security Threats. *IJCA Special Issue on MANETs*, 42–45. Retrieved from `https://www.researchgate.net/publication/46122438_Wireless_Sensor_Networks_An_Overview_on_its_Security_Threats`

Sherchan, W., Nepal, S. & Paris, C. (2013, August). A Survey of Trust in Social Networks. *ACM Comput. Surv.*, *45*(4), 47:1–47:33. Retrieved 2015-05-30, from `http://doi.acm.org/10.1145/2501654.2501661` doi: 10.1145/2501654.2501661

Singh, S., Mishra, A. & Singh, U. (2016, March). Detecting and avoiding of collaborative black hole attack on MANET using trusted AODV routing algorithm. In *2016 Symposium on Colossal Data Analysis and Networking (CDAN)* (pp. 1–6). doi: 10.1109/CDAN.2016.7570906

Singh, S. K., Singh, M. P. & Singh, D. K. (2010, September). A Survey of Energy-Efficient Hierarchical Cluster-Based Routing in Wireless Sensor Networks. *International Journal of Advanced Networking and Application (IJANA)*, *02*(02), 570–580.

Sobeih, A., Hou, J. C., Kung, L.-C., Li, N., Zhang, H., Chen, W.-P., ... Lim, H. (2006, August). J-Sim: a simulation and emulation environment for wireless sensor networks. *IEEE Wireless Communications*, *13*(4), 104–119. doi: 10.1109/MWC.2006.1678171

Sun, Y., Trappe, W. & Liu, K. J. R. (2004, August). A Scalable Multicast Key Management Scheme for Heterogeneous Wireless Networks. *IEEE/ACM Trans. Netw.*, *12*(4), 653–666. Retrieved from `http://dx.doi.org/10.1109/TNET.2004.833129` doi: 10.1109/TNET.2004.833129

Sydney, A., Scoglio, C. & Gruenbacher, D. (2013, January). Optimizing algebraic connectivity by edge rewiring. *Applied Mathematics and Computation*, *219*(10), 5465–5479. Retrieved 2014-05-16, from `http://www.sciencedirect.com/science/article/pii/S0096300312011551` doi: 10.1016/j.amc.2012.11.002

Tchepnda, C. & Riguidel, M. (2006, April). Distributed Trust Infrastructure and Trust-Security Articulation: Application to Heterogeneous Networks. In *20th International Conference on Advanced Information Networking and Applications, 2006. AINA 2006* (Vol. 2, pp. 33–38). doi: 10.1109/AINA.2006.150

Toivonen, R., Kumpula, J. M., Saramäki, J., Onnela, J.-P., Kertész, J. & Kaski, K. (2007). The role of edge weights in social networks: modelling structure and dynamics. In (Vol. 6601, pp. 66010B–66010B–8). Retrieved 2015-02-01, from `http://dx.doi.org/10.1117/12.725557` doi: 10.1117/12.725557

Tyler, T. R. (2006). *Why People Obey the Law*. Princeton University Press. (Google-Books-ID: 77G9sCO_MKIC)

Van Mieghem, P., Omic, J. & Kooij, R. (2009, February). Virus Spread in Networks. *IEEE/ACM Trans. Netw.*, *17*(1), 1–14. Retrieved from `http://dx.doi.org/10.1109/TNET.2008.925623` doi: 10.1109/TNET.2008.925623

Varga, A. & Hornig, R. (2008). An Overview of the OMNeT++ Simulation Environment. In *Proceedings of the 1st International Conference on Simulation Tools and Techniques for Communications, Networks and Systems & Workshops* (pp. 60:1–60:10). ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering). Retrieved from `http://dl.acm.org/citation.cfm?id=1416222.1416290`

Wang, H. & Van Mieghem, P. (2008). Algebraic connectivity optimization via link addition. In *Proceedings of the 3rd International Conference on Bio-Inspired Models of Network, Information and Computing Sytems* (p. 22). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

Wang, H., Van Mieghem, P., TU Delft: Electrical Engineering, Mathematics and Computer Science: Telecommunications & TU Delft, Delft University of Technology. (2009, September). *Robustness of networks*. Retrieved 2015-02-01, from `http://resolver.tudelft.nl/uuid:dc5b1158-be54`

-42d6-a4d3-b0a19462f507

Wang, X. F. & Chen, G. (2003). Complex networks: small-world, scale-free and beyond. *IEEE Circuits and Systems Magazine*, *3*(1), 6–20. doi: 10.1109/MCAS.2003.1228503

Washbourne, L. (2015, April). A Survey of P2p Network Security. *arXiv:1504.01358 [cs]*. Retrieved from http://arxiv.org/abs/1504.01358 (arXiv: 1504.01358)

Watts, D. J. & Strogatz, S. H. (1998, June). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442. Retrieved 2015-01-31, from http://www.nature.com.ezproxy.aut.ac.nz/nature/journal/v393/n6684/full/393440a0.html doi: 10.1038/30918

Wazan, A. S., Laborde, R., Barrère, F. & Benzekri, A. (2008, November). Validating X.509 Certificates Based on their Quality. In *2008 The 9th International Conference for Young Computer Scientists* (pp. 2055–2060). doi: 10.1109/ICYCS.2008.75

Williamson, O. E. (1993, April). Calculativeness, Trust, and Economic Organization. *The Journal of Law and Economics*, *36*(1, Part 2), 453–486. Retrieved from http://www.journals.uchicago.edu.ezproxy.aut.ac.nz/doi/abs/10.1086/467284 doi: 10.1086/467284

Xia, H., Jia, Z., Ju, L., Li, X. & Zhu, Y. (2011, August). A Subjective Trust Management Model with Multiple Decision Factors for MANET Based on AHP and Fuzzy Logic Rules. In *2011 IEEE/ACM International Conference on Green Computing and Communications (GreenCom)* (pp. 124–130). doi: 10.1109/GreenCom.2011.30

Xiang, M., Bai, Q. & Liu, W. (2012, December). Self-Adjustable Trust-Based Energy Efficient Routing for Smart Grid Systems. In *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology (WI-IAT)* (Vol. 3, pp. 378–382). doi: 10.1109/WI-IAT.2012.89

Xiang, M., Liu, W., Bai, Q. & Al-Anbuky, A. (2016, April). Dynamic Trust Elective Geo Routing to Secure Smart Grid Communication Networks. In *Smart Grid as a Solution for Renewable and Efficient Energy* (pp. 323–343). IGI Global. (Google-Books-ID: c6EoDAAAQBAJ)

Xie, Y.-B., Wang, W.-X. & Wang, B.-H. (2007, February). Modeling the coevolution of topology and traffic on weighted technological networks. *Physical Review E*, *75*(2), 026111. Retrieved 2015-02-01, from http://link.aps.org/doi/10.1103/PhysRevE.75.026111 doi: 10.1103/PhysRevE.75.026111

Yang, W., Huang, C., Wang, B., Wang, T. & Zhang, Z. (2009, November). A General Trust Model Based on Trust Algebra. In *2009 International Conference on Multimedia Information Networking and Security* (Vol. 1, pp. 125–129). doi: 10.1109/MINES.2009.226

Yao, J., Chen, S., Nepal, S., Levy, D. & Zic, J. (2010). TrustStore: Making Amazon S3 Trustworthy with Services Composition. In *Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing* (pp.

600–605). Washington, DC, USA: IEEE Computer Society. Retrieved 2015-05-31, from http://dx.doi.org/10.1109/CCGRID.2010.17 doi: 10.1109/CCGRID.2010.17

Younis, O. & Fahmy, S. (2004, March). Distributed clustering in ad-hoc sensor networks: a hybrid, energy-efficient approach. In *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies* (Vol. 1, pp. –640). doi: 10.1109/INFCOM.2004.1354534

Zahariadis, T., Leligou, H. C., Voliotis, S., Maniatis, S., Trakadas, P. & Karkazis, P. (2009). An Energy and Trust-aware Routing Protocol for Large Wireless Sensor Networks. In *Proceedings of the 9th WSEAS international conference on Applied informatics and communications.*

Zhang, Y., Wang, L. & Sun, W. (2013, March). Trust System Design Optimization in Smart Grid Network Infrastructure. *IEEE Transactions on Smart Grid*, *4*(1), 184–195. doi: 10.1109/TSG.2012.2224390

Zhao, K., Kumar, A. & Yen, J. (2011, May). Achieving High Robustness in Supply Distribution Networks by Rewiring. *IEEE Transactions on Engineering Management*, *58*(2), 347–362. doi: 10.1109/TEM.2010.2095503

Appendix A

Glossary

AES Advanced Encryption Standard

CBC Cipher Block Chaining

CBF Cipher Feedback

CoI Community of Interest

D2D Device to Device

DES Data Encryption Standard

DDoS Distributed Denial of Services

DoS Denial of Services

DTEGR Dynamic Trust Elective Geo Routing

HEED Hybird Energy-Efficient Distributed clustering

ICT Information and Communication Technologies

IoT Internet of Things

MAC Message Authentication Code

MD5 Message Digest 5

MDC Manipulation Detection Code

NTaaS Network Trustworthiness as a Service

PKI Public Key Infrastructure

P2P Peer to Peer

SHA-3 Secure Hash Algorithm 3

SHDV Structural Hole Damage Value

SHL Structural Hole Locator

TTM Trustworthiness Tolerance Margin

UAV Unmanned Aerial Vehicle

WSN Wireless Sensor Network