Modeling Root Causes of Construction Site Fatalities Using Social Network Analysis

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Seifeldeen Omar Eteifa
May 2018

ACKNOWLEDGEMENTS

First and foremost, I would like to thank god the almighty and merciful for giving me the opportunity to start my thesis at University of Tennessee, giving me the strength to complete it and the support of great people surrounding me.

Secondly, I would like to thank my thesis advisor Dr. Islam El-adaway, Associate Professor of Civil and Environmental Engineering and Construction Engineering and Management Program Coordinator at The University of Tennessee, Knoxville, for his consistent support throughout this work from inception to completion, where he always had an open door whenever I needed support. He consistently supported me both on the technical and personal level and guided me towards the right direction, all while ensuring that my work comes out to represent me in the best possible ways. He has been a true leader and big brother to me both at times of ease and hardship.

I would also like to thank my committee members Dr. Shuai Li, Assistant Professor of Civil and Environmental Engineering, and Dr. Jon Hathaway, Assistant Professor of Civil and Environmental Engineering, whose input has immensely contributed to shaping this work the way it is.

Gratitude is also due to members of Construction Industry Research and Policy Center (CIRPC) for providing access to data as well as support working with it. Special thanks are due to Mr. Edward Taylor, Executive director of CIRPC for his support and to Mr. John Wagner, Research Associate at CIRPC for familiarization with the structure of data.

I must also express my profound gratitude to my friend and senior research team member Ibrahim Abotaleb for always being there for me. His presence has always made me feel safe away from home. His continuous and selfless help has always kept me going whether in personal life or in technical work. My happy moments with him have always been a relief from work stress. I would also like to thank my colleague Noor Akroush for assisting me with collection of data and sharing all her knowledge and experience with me during her studies at the University of Tennessee.

Finally, I would like to express my gratitude to my parents and sisters who have always been in my mind and a key to my motivation through hard times. Their confidence in my abilities, sacrifices for me and endless encouragement have always been the fuel that drive me forward.

ABSTRACT

In spite of the continuing decrease in accident rates in the US Construction industry, accidents are one of the key factors constraining construction industry in terms of cost and time. Fatality rates in the US are more than other developed countries. Research into accident causation included accident occurrence mechanisms, different levels and types of accident causes as and best practices for accident prevention. These efforts, however, did not focus particularly on fatal accidents and did not prioritize the relationships amongst root causes for accidents. Accordingly, this research's goal is to quantitatively analyze relationships between fatal accident root causes as well how they are tied to direct causes commonly quoted in fatality investigations using social network analysis (SNA). Social network analysis is a set of tools derived from mathematical graph theory and utilized in several knowledge areas including social sciences, natural sciences, engineering, construction management and safety. Accordingly, a three-step methodology is devised. First, 100 case files are analyzed for accident causation data. Second, an SNA model is built utilizing the relationships between root causes and broken down according to direct causes. Third, model is analyzed, and results are interpreted and validated to provide insights into fatal accident causation and contributing underlying factors. The analysis of all social networks yield that accident root causes interact very closely, and that only a few causes contribute most to fatality networks, a relationship which can be described by a power law. The model is capable of successfully identifying the most influential root causes for fatal accidents and their relationships. It identified "lack of job specific training" as the key cause for fatal accidents particularly for the direct causes "struck by" and "caught in between". Moreover, SNA showed that this cause is most influential when combined with "absence of fall arrest system" and "lack of personal protection equipment". Benefits of this research include providing a different approach to causation of fatal accidents that provides deeper more holistic insights into their root causes. Additionally, it provides a quantitative tool for prioritization of root causes which can guide implementation of Safety Management policies and practices.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1:
INTRODUCTION

1.1 Construction Industry contribution to US economy

Construction is one of the key industries supporting the growth and development of the US economy with a contribution amounting to 4.3% of Gross Domestic Product in 2015 (US Bureau of economic analysis). This contribution to the national GDP by the construction industry has been consistently rising for the past 5 years from 2011 till the first two quarters of 2015 as shown in the Figure 1.1, which shows the potential of the industry as it recovers from the 2008 recession and continues to grow.

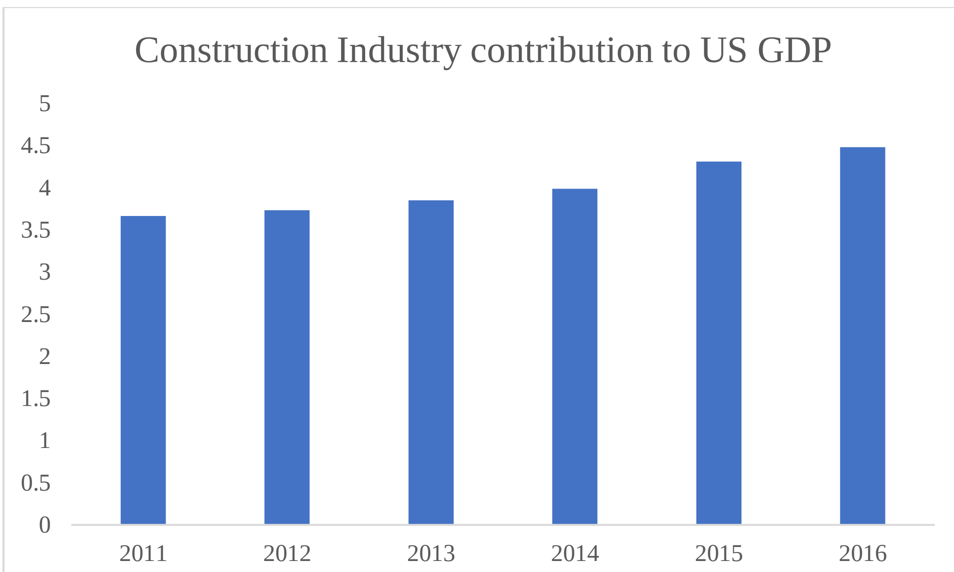

Figure 1.1: Construction industry contribution to GDP (Source: U.S. Bureau of Economic Analysis)

In addition to its contribution to the U.S. GDP, the construction industry employs 6.8 million employees as of January 2016 meaning that 4.7% of the total US workforce works in construction according to the US bureau of Labor statistics. Figure 1.2 shows the continuous growth of the Construction employment potential after the recession in 2008 which can be a key contribution to the economy.

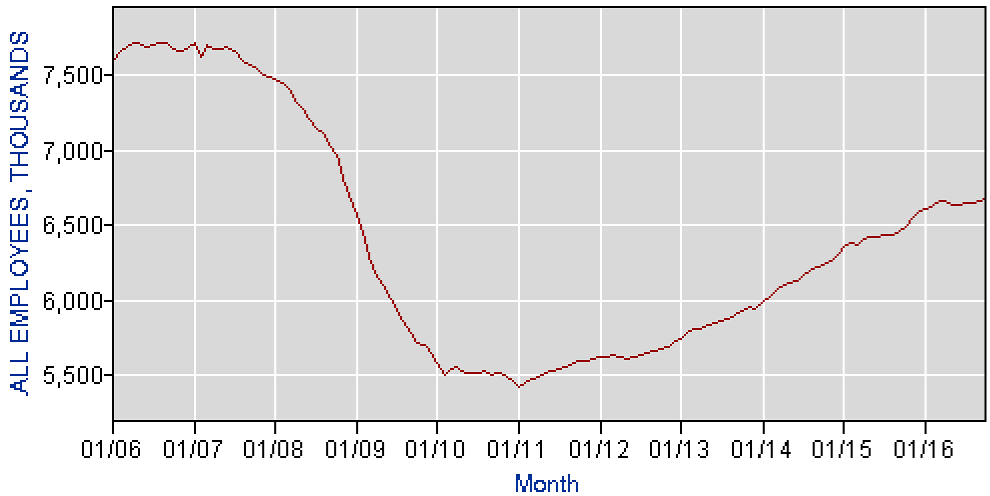

Figure 1.2: Employment in thousands contributed by construction sector (Source: Bureau of Labor Statistics)

1.2 Safety Statistics in the Construction Industry

Safety performance in construction has always been a concern for many researchers and organizations. Since its establishment in 1971, Occupational Safety and Health Administration's (OSHA's) regulations and guidelines significantly reduced fatalities and lost time injuries. This followed the issuance of Occupational Safety and Health act in 1970 which regulated workplace safety and health for the US organizations. Construction industry has experienced a significant reduction in occupational lost-time injuries which were 209.6 thousand in 1992 and were reduced to 75 thousand in 2010 as shown in Figure 1.3 (Center for Construction Research and Training 2013). This decline is even regarded as more significant as it is accompanied with an increase in the total workforce working in construction. Moreover, fatalities in construction industry have been reduced from 963 in 1992 to 802 in 2010 which is a subtler decline where fatalities increased to a peak of 1297 in 2006 before declining as shown in Figure 1.4 (Center for Construction Research and Training 2013).

Despite this improvement in safety performance, construction industry is still one of the most dangerous industries, where 17.1% of 4690 Workplace fatalities in the United States were related to Construction in 2010 as shown in Figure 1.5 (Center for Construction Research, and Training 2013). Moreover, the national fatal injury rate of the United States is 9.7 per 100,000 workers which is high compared to many other developed countries as seen in Figure 1.6.

These statistics point at that more effort should be put into accident prevention since the accident reduction achieved so far is not sufficient. The first step to prevent fatal accidents is enhancing the current understanding of their root causes and the relationships between these causes. Two levels of causation are considered in this research which are direct causes and root causes. Direct causes can be defined as the most obvious immediate causes for the accident which are usually found in accident investigations as recorded by OSHA compliance officers like "fall" or "caught in between". For root causes, on the other hand, this research adopts OSHA's definition which is "fundamental, underlying, system-related reasons why an incident occurred that identify one or more correctable system failures." This means that "defective equipment" can be the root cause of an accident where "struck by" is the direct cause. This essentially means that accidents occur due to direct causes which may result from a single root cause or multiple root causes. A key point of deviation of this study is also accounting for the interrelationships between causes which govern how they are combined together to lead to accidents. Merits of analyzing these

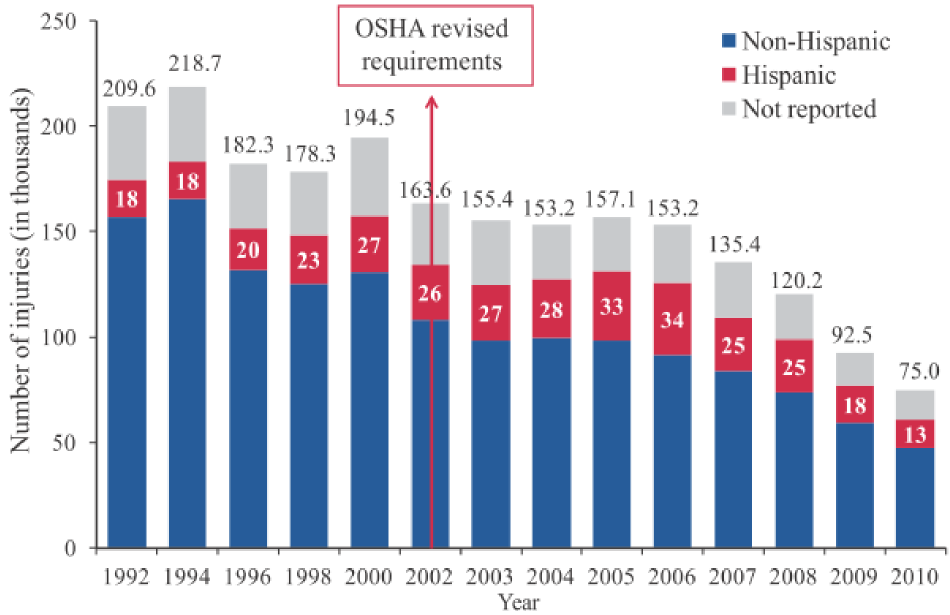

Figure 1.3: Number of lost time injuries in the United States (Source: Center for Construction Research, and Training 2013)

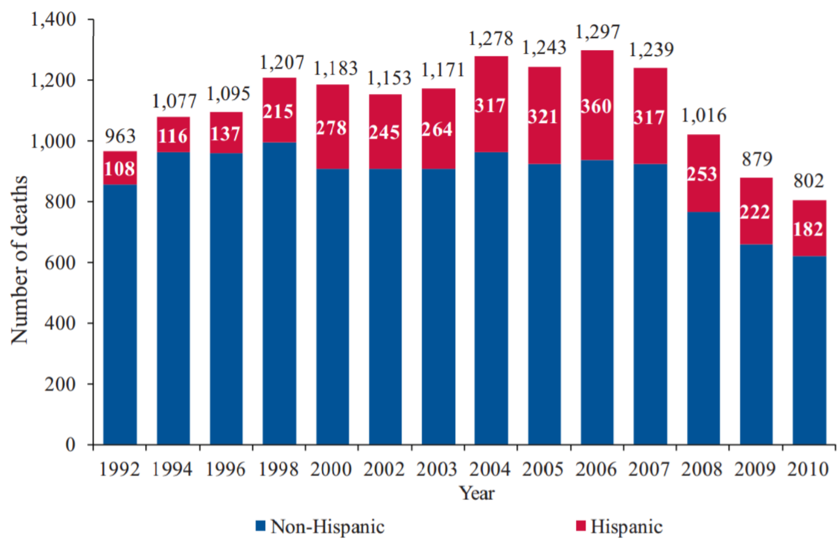

Figure 1.4: Number of fatal injuries in construction (Source: Center for Construction Research, and Training 2013)

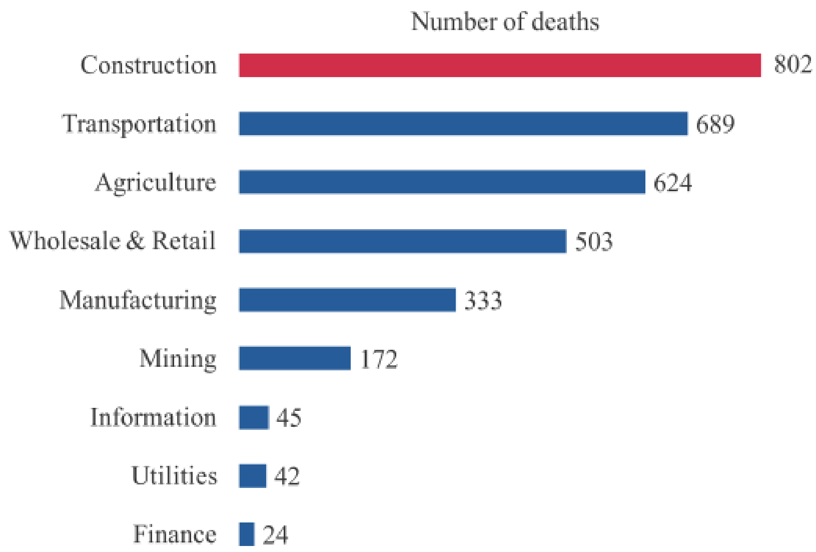

Figure 1.5: Fatalities in the United States by industry (Source: Center for Construction Research, and Training 2013)

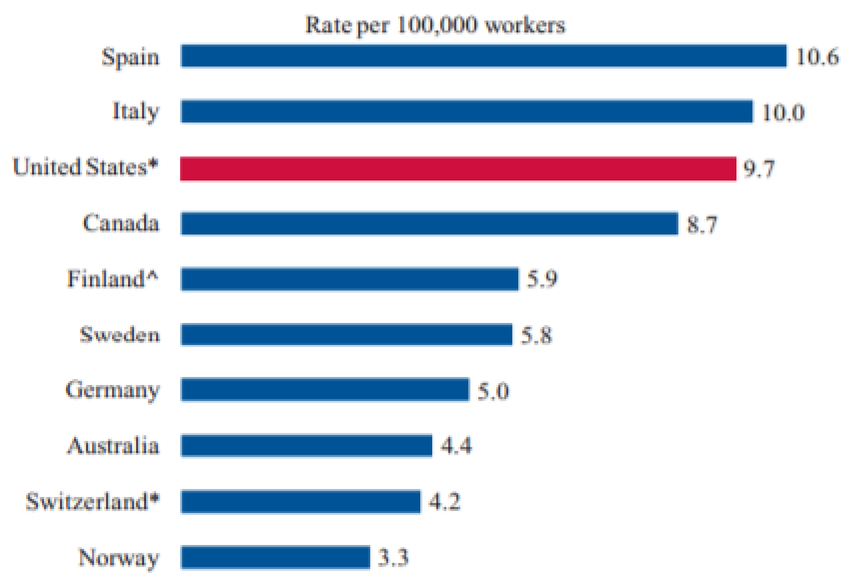

Figure 1.6: Construction fatality rate in the United States compared to other developed countries (Source: Center for Construction Research, and Training 2013)

relationships include helping to understand the strongest root cause combinations which lead to most fatalities giving an early warning the presence of such combinations on site can lead to fatal accidents. Merits also include identifying how serious an accident root cause can be through its ability to cause a variety of accidents in combination with different root causes rather than evaluating it by itself. Moreover, this approach of thinking of causes as combinations helps devise countermeasures for combinations of causes leading to improving safety management practices' efficiency and effectiveness. Finally, knowing the relationships between causes can help reveal if there is an overlap in the impact of some of the safety management practices.

1.3 Accident Causation Models

A wide variety of modeling theories was proposed to explain how accidents occur. Current literature has models with different areas of focus considering the accident causation. Various model focus on human factors such as human capacity and cognitive abilities. Others focus on organizational factors such as different influences and barriers to accident occurrence which stem from the organization. There are also sociotechnical modelling approaches that concentrate on social and technical aspects of the workers and the job being done. Finally, there are hybrid models which combine aspects from the previous models.

In addition to their area of focus, accident causation models have different perceptions of how causes of an accident interact which can be broadly classified into two groups. The first is perceiving the accident causation as a sequence of events one after the other leading all the way to the accident. The second is perception of multiple causes interacting together to cause the accident. Additionally, a large number of models proposed diagramming techniques as tools for visualizing and analyzing different levels of accident causation as well as the tie between them.

1.4 Knowledge Gap

Existing accident causation models stem from strong theoretical background and provide different approaches to accident causation. However, the models have not been developed for many years and in their current form they need updating to cope with current safety management policies and practices. Current policies and practices necessitate that objective quantitative metrics are provided to be able to assist decision making and provide priorities for safety spending. Moreover, recent research into their practical application is very limited. Finally, despite that these models help conceptualize and analyze accident root causation, they did not provide quantitative means to objectively analyze the causes and the relationships between them to understand the underlying accident dynamics.

Recent research in the field of construction safety has been concerned mostly with safety management policies and best practices intended to improve safety performance of construction firms. This research has contributed significantly to the body of knowledge pertaining to construction safety. However, this research preoccupied most researchers from research into accident causation, particularly fatal accidents which still present a major challenge to the US construction industry.

1.5 Research Goal and Objectives

1.5.1 Research Goal

The goal of this study is to investigate the relationships amongst root causes of fatalities as well as their correspondence to direct causes quoted by OSHA investigators using social network analysis (SNA). SNA is a set of tools and techniques based on graph theory. There are three main merits to the application of these techniques to fatal construction accidents. First, it enables the analysis of root causes based on their inter-relationships which cannot be otherwise attained efficiently using other techniques. Second, unlike theoretical models, it is able to determine interrelations between root causes objectively using real data from the industry. The hypothesis on which this approach is based is that the relationship between causes is stronger if they occur together in the same accident more frequently. The third merit is flexibility of SNA. SNA can facilitate the isolation of root causes based on their relation to a particular accident type, a certain policy or other groupings. This allows the analysis to be more in-depth. A demonstration of this flexibility is applied in this research where network is broken down based on direct causes for in depth analysis. This technique, however, is similarly applicable for any criterion by which accidents can be categorized.

1.5.2 Research Objectives

To be able to reach the aforementioned goal, the following set of research objectives should be achieved:

- Contemplate frequently quoted root causes and direct causes in construction industry fatalities.
- Establish most significant root causes giving rise to fatalities in construction.
- Investigate fatal accident root causes’ correspondence to their frequently quoted direct causes.

- Study fundamental organizational, human-related and technological factors leading to construction fatalities.

Accident causes investigation emphasizes the importance of the mechanics of accident occurrence referenced to the fundamental organizational, human-related, technical and physical factors contributing to the accident. The regular approach towards tackling this is breaking down accidents in the efforts to reach to their root causes in a top-down method. The approach proposed by this paper, however, aggregates all root causes which give rise to accidents together in the same network. This approach adopts a bottom up method focusing on how causes are combined together to lead to accidents rather than how to break down a single accident to its causes. The proposed analytical approach is intended to complement the current accident causation models rather than to replace them since a prerequisite to its implementation is to correctly identify root causes. This identification requires a deep understanding of current models. The proposed approach, however, enriches this understanding by adding a different dimension to it which is relationships and interactions between root causes.

CHAPTER 2:
LITERATURE REVIEW

Safety performance has been a key consideration for many businesses in the construction industry nowadays. A poor safety performance leads to accidents which incur large costs on the project. A study by Everett and Jr (1996) found that using only the well documented costs, the cost of a project grows between 7.9 and 15% due to accidents. They added that additional costs not included in their study include OSHA fines, inability to attract new high-quality workers or acquire new project due to damaged reputation and reduced morale of the employees in general. Another study by Ikpe et al. (2012) performs an elaborate cost benefit analysis taking both all the direct and indirect cost of accidents into account to assess the economic benefit from effective accident prevention. Their study concludes that the benefits are greater than the costs by a factor of 3 where for every $1 spent $3 are gained as benefit. This study extends for all small medium and large contractors where the smaller the contractor is, the greater is the accident prevention cost as a percentage of his turnover and the greater the benefit is.

In the light of these studies, a single fatality can be extremely detrimental to any organization in terms of financial costs, work stoppage, employee demotivation and loss of reputation. For this purpose, prevention of these accidents comes as a first priority. The first step to this prevention is modelling and contemplating fatal accidents to fully understand their causation mechanisms and be able to prevent them accordingly.

2.1 Accident Causation Models

At the analytical level, research has always addressed accidents investigating the causation and the process that eventually lead to accidents. To start with, Heinrich's (1969) domino theory modelled the occurrence of accidents as a series of causations in a single dimension which interact like dominoes where if one falls the next one falls. The 5-dominoes modelled are: (1) Ancestral, environmental and social background; (2) worker's mistake; (3) physical or mechanical hazard or unsafe action; (4) accident occurrence; and (5) worker injury or property damage (Heinrich 1969). Heinrich's study argues that if any of these dominoes are removed, the damage or injury can be prevented.

The one-dimensional sequence of events leading to accidents was also tackled in Occupational Accident Research Unit (OARU). OARU is a model which was established by

Kjellén and Larsson (1981). This was a two-step model based on determining the accident sequence and the determining factors leading to the accident. Unlike Domino theory, accident sequence was composed of only 3 phases: (1) Initial phase: where there's an abrupt change (deviation) from regular work processes; (2) concluding phase: where energy flow is uncontrollable as system control is lost; and (3) injury phase: where this energy leads to harm or injury to the human body. The second step involves identifying the determining factors leading to this accident in 3 main areas which are organizational, social and technical.

Petersen (1971) presented accident multiple causation theory. An accident occurrence is framed as a combination of multiple factors. These multiple factors are subdivided into behavioral and environmental factors. Behavioral factors are those factors based on human behavior such as misconduct of employees, employees not having the necessary skills or knowledge or employees being in a physical or mental state that doesn't enable them to perform well at work. Environmental factors, on the other hand, refer to poor physical conditions on the job such as equipment malfunction, unprepared work site and on site physical hazards. His theory claims that by analyzing these factors, root accident causes could be effectively identified.

Another perspective was referring to accidents as defects or flaws in organizational barriers. This was first introduced by Reason (1990; 1997) in his swiss cheese model where he classified accident causation into 4 major subgroups which are: (1) organizational impacts; (2) supervision; (3) unsafe act preconditions; and (4) unsafe acts. He makes an analogy for the layers of organizational protection from these causations as slices of swiss cheese and any shortcoming in this protection system as a hole in the swiss cheese through which hazards can pass. He made a distinction between immediate accident causes and latent causes which might lie within the organization. His claims are that if these holes align with one another, an accident is more likely to take place.

The idea of barriers protecting from accident occurrence was used to devise an accident cause investigation model known as Integrated Safety Investigation Methodology (ISIM) (Ayeko, 2002). This model investigates all information pertaining to individuals, work tasks, equipment and environment to be able to determine unsafe conditions and causal factors in the sequence of events leading to an accident. Once these factors and conditions are determined, the risk associated with each of them is quantified and the barriers defending against them are examined. These barriers might be physical, organizational or regulatory defenses and ISIM aims at identifying barriers which aren't strong enough and strengthening them. This approach tries to capture the latent conditions inside the organization that might have led to this accident. ISIM integrates elements of safety deficiency analysis and accident investigation to provide strategic recommendations to improve safety.

Jacinto and Aspinwall (2003) developed Work Accidents Investigation Technique. This technique is used in accident cause investigation and includes a questionnaire and specific guideline for collecting relevant information. It involves two phases for analyzing any occurrence. The first phase analyses the sequence of events leading to the accident and the consequences of the accident in search of what is known as active failures. Moreover, it investigates the workplace and working environment in search of influencing factors contributing to those active failures. The second phase is more in-depth analysis of human and job-related factors leading to identification of any management or organizational flaws.

Using the same barriers concept, Control Change Cause Analysis (3CA) was developed by Kingston (2007) to address the safety management system. It breaks down accident causation into unwanted changes which occur among a sequence of events. It defines in this sequence events which is named as "significant" events in terms of control reduction and giving more room for unwanted changes to occur. Based on these significant events, barriers can be identified to control these significant effects or mitigate or limit their effects. Based on the barrier identification, limitations and weaknesses of each barrier can be identified. These weaknesses are traced back to safety management system and processes that allowed the barrier to not be able to prevent particular accident occurrences. It breaks down accident causation into unwanted changes which occur among a sequence of events. It defines in this sequence events which is named as "significant" events in terms of control reduction and giving more room for unwanted changes to occur. Based on these significant events, barriers can be identified to control these significant effects or mitigate or limit their effects. Based on the barrier identification, limitations and weaknesses of each barrier can be identified. The weaknesses are then traced back to safety management system and processes that allowed the barrier to not be able to prevent particular accident occurrences.

Hale and Glendon (1987) took the approach of modelling the human factors contributing to accidents based on attribution theory. This approach is concerned with determining event causations through understanding the process of people's information processing. The method suggests that hazards are always available in any workplace and human cognition is necessary to avoid and control these hazards. It considers the different factors linking between human behavior and hazards whether human action create potential hazards dangerous to others or control these hazards reducing the magnitude of their effect.

Another classification made by Rasmussen (1987), which focuses on the human factor, particularly classifies accidents according to the level of human familiarity and control of the work environment. This approach is known as the SRK model and classifies causes according to human behavior to skill-based where the worker lacks the necessary skill to competently complete the job, rule-based where the worker violates the rule or follows a wrong rule and knowledge based where the worker lacks the necessary knowledge to address the work situation safely.

One of the accident causation models mainly based on human error is Ferrell Theory that attributes an accident to that work load is too much for human capacity (Abdelhamid and Everett 2000). This lack of compatibility could be due to a multitude of factors including carrying out an activity which is inappropriate due to not knowing how to do it or not being able to judge the risk associated with it or lacking the capacity to do it due to being overloaded.

Shappell and Wiegmann (1997; 2000) devised an approach which both has elements from the Swiss Cheese model and from Rasmussen's work known as the HFACS method. It is based on the swiss cheese model as it sub-classifies each of the 4 major subgroups mentioned to reach 17 more detailed accident causation categories. Their method has some similarity in the way it classifies human actions that lead to accidents to Rasmussen's work since it categorizes group 4 of swiss cheese model which is unsafe acts into two major subgroups which are errors and violations. Each subgroup was further broken down where errors were classified to those based on skills, decisions and perceptions. Violation, on the other hand, were divided to routine violations which happen regularly and exceptional violations which take place exceptionally at the time of the incident. Their study suggests that these 17 subgroups form a comprehensive framework for accident causation analysis.

A revolutionary approach was adopted by Rasmussen (1997) which is a sociotechnical approach. His model is based on various actors who impact the safety processes including legislators, work planners, managers and system operators and is divided by organizational levels. These different levels are then studied between different disciplines to define boundaries for the safe operation. Once these boundaries are defined, all actors shall be informed with them and given the chance to adapt to these defined boundaries.

Hollnagel (1998) developed a method which also utilized a socio-technical approach called Cognitive Reliability and Error Assessment Method (CREAM). The model correlates between control degree in socio-technical systems and reliability of performance, the less the control, the less the reliability. It can be used prospectively or retrospectively to identify failure modes in performance which are classified into strategic, tactical, opportunistic and scrambled. Links for actions preceding and succeeding this failure can be identified. When this method is used retrospectively, it differentiates between the observed actions named phenotypes and those which can only be obtained by inference named genotypes. Genotypes are subdivided into human related, technology related, and organization related.

Norwegian State Railways established a sociotechnical method called Norske Statesbaner (NSB) (Skriver et al. 2003). NSB emphasizes the interactions between three main factors contributing to accidents which are human factors, organizational factors and technical factors. The approach combines two tasks to be able to investigate accidents. The first task identifies the events contributing to the accident and their sequence as well as barriers which could have interrupted this sequence and why they were not there or broken. The second task is concerned with individuals and work tasks related with the accident and so it utilizes a questionnaire as a tool to address these and identify deficiencies.

2.1.1 Use of Diagramming Techniques in Accident Causation Models:

Various diagramming methods have lent themselves as a reliable tool in accident causation modelling. Their power lies in the ability to visualize the causation links and the process flow associated with accident causation which pushes the brainstorming, deductive power and reasoning further. This allows those investigating accidents or researching to dwell deeper to find latent causes beyond the observable causes and harness this knowledge to continually improve safety management processes and practices. Moreover, diagramming makes the models easier to understand and communicate to others which is a key element which governs the applicability of a model to real life situations since communication is one of the key elements of effective safety management. Most of the models discussed earlier would be highly enhanced by diagramming. For example, using diagrams for dominoes or swiss cheese slices or hierarchies for classifications made under many of the prementioned models could help better identify accident causation.

One of the techniques heavily based on charting is Multilinear Events Sequencing (MES) developed by Benner (1975). MES orders events on a time line chronologically. Its view of accident causation starts by a stable situation being disturbed which is then followed by a sequence of events which in turn leads to an accident. MES models actors such as individuals, equipment or materials, actions carried out by actors and events which involve actors performing actions which have an effect on safety. This model is mapped on a logic chart showing actors, actions and events in the sequence leading to the accident.

Another approach dependent on diagramming was focused mainly on organizational failures as the main problem leading to accident causation and is known as (TRIPOD) (Wagenaar et al. 1994). According to the model, an accident occurs due to failure of one or more organizational barriers. These failures are directly caused by unsafe acts; however, their occurrence is deeply rooted within mechanisms that happen within the organization known as General Failure Types including human-related, technical and organizational failures. Tripod has 11 General Failure Types and it aims to provide a bar chart that details their presence in each organization to compare different organizations and improve safety performance.

A diagram-dependent approach known as the Workgroup Occupational Risk Model (WORM) was devised by Ale (2006) and used by Aneziris et al. (2008) to make a quantitative assessment for fall from height risks. The WORM method uses a diagramming technique based on interconnected functional blocks. WORM employs a top down approach breaking down a main event into smaller and smaller simpler event until the probabilities for the simpler events can be identified to assess the risk of the main event.

Another technique based on block diagram charts is called Systematic Cause Analysis Technique (SCAT). SCAT was established by The International Loss Control Institute (ILCI) at the end of the 1980s (Katsakiori et al. 2009). SCAT has roots in Heinrich's domino theory (1969) and is presented as a block diagram with 5 blocks which are: (1) accident description; (2) common contacted categories which could have resulted into the accident, (3) immediate cause for contact with this category; (4) underlying causes for accident; and (5) preventive safety management practices. This technique is based on checklists covering human factors, job factors and different elements of the safety management. This model takes a hybrid approach towards accident prevention where it is either by removing one of the 3 intermediate blocks or by utilizing barriers to prevent the uncontrolled energy flow leading to the accident.

2.1.2 Tree-logic based Accident Causation Models:

One of the main diagramming approaches towards modelling accident causation is tree diagrams which link causes together with logical connections. Causal Tree Method (CTM) was developed by Leplat (1978) and attributes an accident to processes which deviate from the usual work processes. It classifies these variations from regular work processes into several subcategories: (1) variations related to individuals; (2) deviations in tasks performed; (3) variations related to equipment used; and (4) changes in the surrounding environment. This method also utilizes a top down approach where it starts by the unwanted incident and investigates backwards to reach to facts contributing to that incident to be able to construct the tree showing the aforementioned variations and their logical relationship with the occurrence.

Management Oversight and Risk Tree (MORT) is another accident causation model with tree logic. It was established by Johnson (1980) for use by the US Atomic Energy Commission. MORT defines accident as an undesired flow of energy due to insufficient barrier and controls to regulate energy. The aim of MORT is to find out facts to recognize hazardous energy forms on site as well as deviations from the regular process that could lead to accidents. The MORT is a tree which is composed of three major branches, namely the S factors, the R factors and the M factors. The S factors have to do with omitted overseen factors contributing to accident causation. The R factors have to do with risks which are already identified and known but to which no controls or barriers have been constructed. The M factors are those which are associated with features of management system and its particular characteristics which played a part in the accident causation. Under these three branches are elements linked to questions which should be asked by who is analyzing the accident to assess safety. This model pushes who investigates the accident to look beyond direct causes into causations which are deeply rooted in the organization itself or in the management system. It also provides a checklist and problems to look for to guide the investigator rather than just the analytical tools.

Fault tree analysis was developed by Bell Laboratories in early 1960s and further refined by Boeing Company (Ferry, 1988). The fault tree diagrams any particular accident and all possible factors contributing to it mapping the causation links and logical connections which lead to the accident. The fault tree utilizes a top down approach to models the accident occurrence using the events and conditions occurring before the event like including technical, human-related and organizational conditions and uses two type of logic gate which are the AND and OR gates to form the logical links. This results in establishing a sequence of logical relation combinations which eventually lead to the accident.

The emergence of diagramming and tree methods to analyze the causation of accidents indicates that the logical relationships and links between different levels of accident causes have always been perceived by researchers as complex and requiring closer investigation. This suggests that using networks to be able to analyze and study accident causation can unravel potential relationships and logical patterns that can help understand accident causation more. The use of networks can also provide powerful insights and different approaches towards how different causes interact together leading to accidents. Table 2.1 shows a summary of the classifications of accident causation models mentioned in this literature review.

Table 2.1: Accident Causation Models

		Focus of the Model			
		Organizational	Human-Based	Sociotechnical	Hybrid (specify)
Causation Link Type	Sequence of Events	- Swiss cheese model by Reason (1990; 1997). - WAIT by Jacinto and Aspinwall (2003) phase 1. - 3CA by Kingston (2007).	- Hale and Glendon (1987).	- MES* by Benner (1975). - CREAM by Hollnagel (1998).	- Domino Theory by Heinrich (1969) (sociotechnical and human-based). - CTM* by Leplat (1978) (human-based and physical factors). - OARU by Kjellén and Larsson (1981) (organizational and sociotechnical). - WORM* by Ale (2006) (organizational and human-based).
	Multiple Causes	- MORT* by Johnson (1980). - TRIPOD* (Wagenaar et al. 1994). - ISIM (Ayeko 2002). - WAIT by Jacinto and Aspinwall (2003) phase 2.	- SRK by Rasmussen (1987). - Ferrell Theory (Abdelhamid and Everett 2000). - HFACS by Shappell and Wiegmann (1997; 2000)	- Rasmussen (1997).	- Petersen (1971) (physical, organizational and human-based). - NSB by Skriver et al. (2003) (human-based, organizational and technical factors). - Fault tree* (Ferry 1988) (human-based, organizational and technical factors). - SCAT* (Katsakiori et al. 2009) (organizational and human based).
* indicates that the model utilizes a diagramming technique.					

In literature, the classification for accident causation models has no standardized convention. For example, models were classified by Kjellén (2000) as processes, sequences of causation, movement of energy, causation trees of logic, in addition to health, safety and environment (HSE) management models. A different classification was suggested by Hollenagel (2002) who categorized models into epidemiological, sequential and systematic models.

2.2 Construction Safety Best Practices

Apart from the theoretical analytical accident causation models that help investigate accidents and understand their causation more thoroughly, other research was directed mainly towards the practical application of practices to enhance safety performance. These are referred to as safety best practices and their application has resulted into significant improvement in safety performance in the construction industry.

There has been a growing trend of investigation of best practices to drive accident rates down to zero. In the early 90s, Construction Industry Institute (CII) obtained data by surveying members of 25 projects with superior safety performance about techniques they use to enhance safety performance (Construction Industry Institute 1993). The study concluded with identification of five techniques namely: (1) pre-task planning; (2) training and orientation; (3) written program for safety related incentives; (4) program for abuse of substances as alcohol or drugs; and (5) investigation of accidents/ incidents. A study by National Center for Construction Education and Research (NCCEER) was conducted to validate the findings of the CII 1993 study. The study concluded that the findings were valid, but that the industry-related notions of safety program implementation were changed by the late 90s (Hinze et al. 2013a).

Due to this change recognized by NCEER, CII performed another study comparing high performing construction projects by interviewing their personnel (Construction Industry Institute 2003). This study identified nine topic areas which are crucial for improvement of safety performance which are: management demonstrated commitment to safety, safety staffing, pre-task and pre-project planning, orientation and training pertaining to safety education, involvement of workforce, safety evaluation and rewarding program, management of subcontracts, investigation of accidents and incidents/near misses and testing of abuse of drugs or alcohol.

Hinze et al. (2013) performed statistical analysis quantifying the effect of 96 different safety practices implemented by world class organizations on Occupational Safety and Health

Administration (OSHA) Recordable Injury Rates (RIR). The study found a significant negative correlation between the percentage of these 96 practices implemented and the OSHA RIR. The study identified 10 practices with the highest correlation with improver RIR namely: manager site specific safety orientation, involvement of supervisors in safety policy, availability of medical facilities on site, programs of worker to worker observation, adequate safety staffing relative to workforce size, (recording first aid injuries, approval of owner of safety plan, perception surveys involving workers in safety management, 100% enforcement of steel toed boots and all contractors taking part in safety meeting.

One issue of controversy when coming up with best practices was the idea of rewards, incentives or positive reinforcement. CII's first research (1993) suggested a written program for safety related incentives. NCCEER research mentioned by Hinze et al. (2013) suggested rewards not being given on the basis of whether workers are injured but rather on whether they carry out work using the safest methods. CII (2003) recommended smaller frequent rewarding rather than larger less frequent rewards. Hinze et al. (2013) found a significant positive correlation between giving incentives for not being injured and OSHA's RIR in 27 projects meaning that safety incentives if not properly applied can be detrimental to safety performance.

2.2.1 Safety Leading Indicators

Safety leading indicators were a recent practical tool evolving from best practices that showed significant potential in accident reduction. Leading indicators can be defined as a group of measures selected to enable assessment of the safety process and safety management activities (Cipolla et al. 2009; Hinze et al. 2013b). Other researchers utilized other properties of leading indicators to define them including measurability, describing safety conditions, identifying gaps and failures in risk control system, predicting undesirable events and monitoring accident risk development (Øien 2001; HSE 2006; Grabowski et al. 2007; Kjellén 2009).

Leading indicators have been proposed for safety performance assessment as opposed to lagging indicators which were the most commonly used indicators of safety performance. The main disadvantage of lagging indicators is occurrence after the fact and relating to past performance and outcomes and being triggered mainly by accident occurrence after damage has already occurred (Construction industry institute 2012). Accordingly, safety professionals and researchers came to question that the input and insights given by lagging indicators are sufficient to be able to predict and avoid future accident occurrence (Grabowski et al. 2007; Mengolinim and

Debarberis 2008). A key distinction between lagging and leading indicator is that lagging indicators generate reactive responses that aim at trying to prevent further injuries (Hinze et al. 2013). Leading indicators, on the other hand, generate a proactive response intending to make amendments to the safety process to prevent accident occurrence. Examples of lagging indicators are OSHA's total recordable injury rate (TRIR), days away, restricted or transferred (DART) and experience modification rates (EMRs) (Construction Industry Institute 2012).

Leading indicators have been classified into active and passive indicators. Passive indicators refer to macro scale long term strategies whose implementation leads to a safer project. Therefore, their presence or absence can help predict how well the project performs and in many cases, they can't be changed during construction (Hinze et al. 2012; Construction Industry Institute 2013). Active indicators, on the other hand, are more dynamic where they can be used to measure and monitor safety performance day-to day during construction to be able to take corrective actions to improve performance whenever needed (Hinze et al. 2012; Construction Industry Institute 2013). Leading indicators were also classified into Safety Management System (SMS) Indicators and Abstract Safety Constructs (Guo and Yiu 2016). SMS Indicators relate directly to providing information to improve SMS policies and practices. Indicators of Abstract Safety Constructs, on the other hand, are tied to safety constructs, which are explanatory concepts aimed at better understanding the safety processes to better predict their outcomes.

Several strategies have been suggested for implementing leading indicators. The first step in any implementation is indicator selection which was identified by researchers as the most important step (Construction Industry Institute 2012; Guo and Yiu 2016; Rajendran 2012; Hinze et al. 2013). Key considerations identified by researchers when selecting leading indicators include current process weaknesses in safety program, organizational defects, safety culture, adaptability of current processes into indicators, the number of indicators an organization is willing to monitor, the priorities for monitoring indicators, return on investment for monitoring that particular indicator and their contribution to improvement in safety efforts (Wreathall 2009; Construction Industry Institute 2012; Hallowell et al. 2013; Guo and Yiu 2016). The implementation strategy discussed by Construction Industry Institute (2012) is an iterative process consisting of nine steps which are: indicator selection, specifying metrics related to this indicator, devising a process for indicator measurement, involving all parties responsible, implementing the indicator, analysis of information obtained, making performance public sharing it with involved parties, assessing how effective the indicator is and celebrating successful implementation.

The effectiveness of leading indicators was examined by a variety of studies. Hallowel et al. (2013) used a judgement panel of 23 safety experts to effectively evaluate 13 leading indicators with actionable metrics, measurements, required resource and response plans as useful for implementation by contractors. Guo and Yiu (2016) used a panel of 5 experts scoring to assess the effectiveness of 32 leading indicators and included that all indicators were predictable and analytically sound, 20 of them were practicable and 22 were cost effective. Construction industry Institutes (2012) analyzed 14 passive leading indicators implementation in 57 projects and found statistically significant negative correlation between the percentage of these indicators applied and the OSHA's TRIR. Rajendran (2012) was comparing between pretask planning (PTP), worker safe behavior observation (WSBO) and site safety audit scores (SSA) in a single project case study over 37 weeks. The analysis used lagging indicators namely near miss incident rate (NMR), first aid injury rate (FA), total incident rate (TI) and total recordable injury rate (TRIR) for comparison. The study found a strong significant negative correlation of PTP and WSBO with both FA and TI. SSA, on the other hand, had weak correlations with FA, NMR, TRIR and TI. Hinze et al. (2013b) assessed the correlation of two leading indicators, namely positive reinforcements and WSBO on OSHA's TRIR. The study devised a composite scoring method on 14 projects and found a significant negative correlation between the score on these two leading indicators and OSHA's TRIR.

2.2.2 Near Miss Reporting

OSHA's (2002) definition of near misses is "an incident where no property was damaged, and no personal injury sustained, but where, given a slight shift in time or position, damage and/or injury easily could have occurred." Bird and Germain (1966) identified the importance of near misses as the base of a hierarchy leading to major injuries where their hierarchy theorized that for every 600 near misses, there are 30 property damaging incidents, 10 minor injuries and finally 1 major injury. The uniqueness of near miss reporting (NMR) comes from the fact that it was classified by researchers in the field of safety as a leading indicator in some instances and as a lagging indicator in others. The logic behind defining misses as a lagging indicator was shared amongst several researchers and it is that what makes it distinct from an accident is only luck (Toellner 2001; Manuele 2009; Rajendran 2012). Other researchers viewed near misses as leading indicators

(Construction Industry Institute 2012; Hallowell et al. 2013; Hinze et al. 2013). Hinze et al. (2013) presented their argument that using NMR as leading indicators in a more proactive way increases their value. This is because lagging indicators are regarded as more negative and thus if near misses are framed that way, the safety personnel manage to obtain less valuable information from them.

Construction Industry Institute research report (2014) devised near miss reporting guidelines as a seven-step cycle namely: definition, rolling out, data collection, analysis, communication and encouragement. Marks et al. (2014) formulated the cycle as: training on NMR, Worker observation of near miss, urgency evaluation, severity evaluation, investigation of root causes, corrective action identification, communication and implementation, communicating relevant details to workforce and incorporating findings into NMR training. Key enablers for this process include management culture of leadership, blame free attitude, written program for implementation, documentation, flexibility, clear definition of near misses, commitment by owner and contractor, proper resource allocation, accessible communication, workforce involvement in reporting and investigation and timely reporting (Jones et al. 1999; Hinze and Wilson 2000; Construction Industry Institute 2014).

Several researchers evaluated the effectiveness of NMR for improving safety performance. Jones et al. (1999) examined off shore activities in two Norsk Hydro projects where it was shown that as the number of near misses reported increased for both projects, the number of lost time injuries was decreased. Hinze (2005) investigated outage work for 9 power plants and found a statistically significant negative correlation between the number of hours spent on investigating a single near miss and OSHA's RIR. CII (2014) interviewed personnel from 47 construction sites and concluded that companies which produced greater numbers of near miss reports had lower OSHA TRIR and that the near miss reporting process lead to greater employee motivation to identify site-related hazards and report them effectively.

2.2.3 Evaluation of Best Practices:

Safety best practices provide guidelines for organizations to assess their own safety performance, improve their own safety management processes and address undesirable events before they occur. However, most of the literature analyzes these practices and assess their performance by one of two ways, either expert judgement, or relationship to traditional lagging indicators such as OSHA's TRIR or FA which is, in itself, a violation of the nature of these practices. This is because these practices do not directly target reduction in accident rates, but rather focus on the underlying elements and defects in safety management system which lead to accidents thus addressing accidents in a proactive way before they occur.

This nature of these best practices makes them dependent on: (1) continuously collecting data pertaining to safety performance on site; (2) analyzing this data in a way that relates directly to the organizational safety management structure; (3) acting on the analyzed data by taking corrective action, publicizing reports, communicating results and taking these results into account during future cycles; and (4) assessing the implemented practices themselves and adding/altering them according to the nature of the safety management process and safety culture maturity within the organization. This cycle of continuous improvement continues forever.

2.3 Social Network Analysis

2.3.1 History and Applications

Social network analysis (SNA) is an analytical method for studying networks. It originates from graph theory in mathematics (Otte and Rousseau 2002). SNA emerged in the 1934 when its concepts were adopted by Moreno. Moreno and Jennings (1960) modelled the relationships between people in political and social settings as networks and sociograms. Since its introduction, social network analysis was used by researchers in a variety of applications in many fields including health care, social sciences, statistics, management sciences and computer science.

In health care, it was used to test association between drug injecting networks, sexual relationship networks with whether each human has human immunodeficiency virus (HIV) as well as the likelihood he gets infected with the virus in the future from a behavioral point of view (Friedman et al. 1997). Social networks were also useful in modelling networks with different actors in health institutions including nurses, consultants and patients in health organizations to understand their effect on health outcomes (Pow et al. 2012). SNA techniques were also useful for public health programs to model and understand the networks of partnering and collaboration between community organizations (Schoen et al. 2014).

Social network analysis was widely applied in many social sciences including political science, economics, anthropology, psychology and sociology. In political science, it was used to investigate cosponsorship networks of legislations made by congress networks in a variety of ways including how well connected the network is, how does it affect the legislative influence of congress representatives, links between sponsors and cosponsors of legislation and importance of strong and weak links for legislative success (Cranmer and Desmarais 2011; Fowler 2006; Kirkland 2011).

In the field of economics, social network analysis was used on a village in Nicaragua to investigate the effect of gender on social relationship formation and networks that provide access to more resources as well as gender-based division of labor (D'Exelle and Holvoet 2011). SNA was also used to model how social and economic networks are created or discovered by recent advancements in information technology (Sundararajan et al. 2013).This study also inquired into the social, economic and organizational effects of information going through these networks, how to make use of this information for making social and economic predictions as well as the dynamics of the network and how it evolves with the technological advancements. Another study investigates the effect of segregating a social network into groups by common properties (like age, gender or race) on the diffusion of a certain behavior within the network (fashion trend, buyer choice, new technology) (Jackson and López-Pintado 2013).

In anthropology, social network analysis was used to study the phenomenon of fission of small groups into even smaller more stable subgroup due to the bigger group members being divided by the flow of information and sentiments unequally (Zachary 1977). Applications of SNA to psychology included those by Moreno and Jennings (1934), the one who introduced SNA, where he used it to classify human relationships, which he considered the most important aspects of a human social group. Moreno and Jennings used this classification to suggest how social groups can be harmonious and capable of working together at the maximum possible efficiency. Another study was run on Facebook users as a social network to determine the influence and susceptibility of different user profiles to adopt a particular product, that is how influential is a user on other users or how susceptible he is to be influenced by other users (Aral and Walker 2012).

More than other social sciences, social network analysis added to the field of sociology. Examples of studies that contributed to sociology though SNA implementation is a study investigated the effect of different properties of social networks on the outcome of collective actions performed (Marwell et al. 1988). Another study compared between personal work networks for men and women from three different industries testing the differences between them classifying ties in the network as instrumental, expressive or overlapping (Stackman and Pinder

1999). Instrumental ties were those required to carry out job tasks, whereas expressive ones were those due to friendship while overlapping ties mixed both relationships (Stackman and Pinder 1999). This study tried to test the effect of both gender and organizational context on the formation of network ties of these three categories. Another study discussed the effect of social grouping on social structure such that each individual is connected to others with homogeneous characteristics and how all kinds of social relationships depended on common characteristics (McPherson et al. 2001). The study analyzes the effect of social grouping on dividing social structure and limiting the social world of each individual to homogeneous individuals in terms of gender, age, ethnicity, place of origin, family and other factors.

Another study discussed the social structure and its contribution to information flow in a society by mathematically modelling the tradeoff between what is known by diversity and bandwidth in a network (Aral and Van Alstyne 2011). This tradeoff is that for a high bandwidth network every individual was connected strongly to few individuals whereas in a high diversity every individual might have weaker ties but with individuals with diverse backgrounds and experiences. The study asserted that the volume of information transfer in a high bandwidth networks was large but much of this information was redundant. In diverse networks, on the other hand, though the volume of information was smaller, there was more chance of access to novel. The model attempted to optimize between those two extreme cases so as to access the most useful novel information without sacrificing the total information flow in the network.

In all the mentioned applications on social sciences, even though the studies mentioned usually target a particular field of study, it turns out to include different aspects from other social sciences. For example, a study which is focused on buyers' behavior from an economics perspective, tends to include elements from psychology, sociology, anthropology and political science. This unveils the multidisciplinary nature of social network analysis where when conducting a study about a particular population, many factors affect their behavior and choices pertaining to many areas of human knowledge.

Apart from social sciences, social networks were used in statistics in an inferential technique known as latent space network model to visually represent data of a relational nature in an interpretable spatial diagram (Raftery et al. 2012). This technique made use of Bayesian Frameworks, Monte-Carlo simulations as well as Markov chains in conjunction with social network analysis with maximum likelihood estimation (or log likelihood for larger networks) for inference (Hoff et al. 2002; Raftery et al. 2012). The aim was to analyze the effects of covariates and provide a visual representation where the relative position of nodes is indicative of the presence of a particular relationship between them thus showing both the relations as well as the statistical uncertainty associated with the spatial position (Hoff et al. 2002).

In the field of computer science, social network analysis principles were utilized to develop algorithms to approximate the optimal network for diffusion of information or viruses through networks including the web. This approach was tested by tracing and describing how news and blog articles diffuse through the web online media (Gomez Rodriguez et al. 2010). Another study visualized email networks to mathematically interpret email data using a hybrid model of Bayesian and Social networks to conduct both predictive and exploratory analyses on email communication data. This model was capable of effectively visualizing communication patterns through subdividing the main email communication networks to topic-based embedded subdivisions (Krafft et al. 2012).

In the domain of management science, social networks development had many implications on how business communication was perceived and how organizational structures were modelled. A study showed the importance of trust for knowledge transfer. It indicated that knowledge transfer through strong organizational links led to improved organizational outcomes. Moreover, knowledge transferred through weak links with the same high level of trust provided even more useful contribution to organizational outcomes (Levin and Cross 2004). Kim et al. (2011) demonstrated the usefulness of SNA metrics in the context of two main supply networks namely material flow and contractual relationships between organizations. The study identified the role implied by a high value for each of these metrics for the organization as well as the core competencies the organization should possess to fulfill this role.

2.3.2 Definitions and Metrics used in SNA

A network is essentially a graph where the main concepts on which modern social network analysis are based are deeply rooted in graph theory in mathematics. The network is comprised of nodes or actors and edges or links. Each link in the network connects two nodes together (Freeman 1978).

A node or actor is a component in the network and the different things that a node can resemble in a network is what gives social network analysis as a technique a significant level versatility and flexibility. In modern literature, nodes have been used to resemble individuals, organizations, defects, causes, computers, websites and other network components based on the application and the research context. The node is commonly represented as a dot or a circle in a network diagram. Literature include two main types of networks according to node types namely unimodal networks where all nodes are the same and bimodal networks where there are two distinct types of nodes resembling two different things for example comembership networks where there are nodes resembling members and a different type of nodes resembling a club (Borgatti and Everett 1997).

The second component of the network is the edge or link that resembles a relationship between any two given nodes. The edge also varies by application and was used in modern literature to show communication, monetary transactions, exchange of information, exchange of files, collaboration and other relationships. The edge has two main attributes which are its direction and weight. Networks where edges have a direction, for example where an actor is paying money to another, are called directed networks where edges are resembled by arrows which show the direction of the relationship (Wasserman and Faust 1994). On the other hand, networks where the edges have no direction, such as past collaborations, are called undirected networks where the edges are commonly resembled by lines connecting the nodes (Wasserman and Faust 1994). Networks where edges have a numerical weight, on the other hand, are called weighted networks where the weight resembles the strength of the link, for example the number of projects on which two firms worked together (Wasserman and Faust 1994). Whereas networks where either the link exist or not, so the presence of an edge is binary, for example whether two firms worked together before or not, are called unweighted networks and all links are assumed to have equal strength (Wasserman and Faust 1994). A structural hole within the network is said to occur when two nodes have no edges connecting between them (Borgatti et al. 2009).

Whenever two nodes have an edge linking them to one another, they are termed as adjacent. The degree of the node is defined as the number of adjacent nodes it has (Freeman 1978). The degree distribution of a network is comprised of the number of nodes in the network in each level of degree and understanding this distribution uncovers many properties of the network (Newman 2001a). For many social networks in modern literature, degree distributions follow power laws (Newman 2001b). A node which has no other nodes adjacent to it is termed an isolated node. An adjacency matrix is a different form of representation of a network from a network diagram and it utilizes matrix format to mathematically represent which nodes are adjacent to each node (Wambeke et al. 2011). The matrix element is zero if the two nodes corresponding to the row and column are not adjacent and has a value if they are adjacent (Wambeke et al. 2011). For unimodal networks, the adjacency matrix is a square n x n matrix and both the rows, and the columns of the matrix resemble the same nodes whereas for bimodal networks the adjacency matrix could be rectangular, and the rows and columns resemble the two distinct types or modes of nodes. The values of elements in the matrix could be either binary in case of unweighted networks or equal to the edge weight in case of weighted networks. In the case of unimodal networks, adjacency matrices for undirected networks are symmetric about the diagonal. Directed networks, on the other hand, are not symmetric where each side of the diagonal shows a different direction of the relationship. The diagonal elements show a relationship between the node and itself which is a rare occurrence in literature and not meaningful within this research scope.

Between any two nodes, the set of consecutive edges which connect them is known as a path (Freeman 1978). Paths which start and end at the same node are called cycles (Freeman 1978). A node is considered to be reachable from another node if there exists at least one path connecting between the two nodes (Freeman 1978). In a social network diagram, if all points are considered to be reachable from any given node in the social network, the social network is called a connected network. For networks, which are not connected, the connected subsets of the network are known as components (Wasserman and Faust 1994). The number of edges which have to be crossed along the path from one node to another is known as the distance of the path. An important concept in network theory is the geodesic which is the shortest path between any two given nodes or that consisting of the least number of links where the number of links in this case is known as the geodesic distance (Newman 2001b). The geodesic between any two given nodes might not be unique, there might be more than one shortest path connecting two given nodes. In a connected network or component, the diameter of this network is the term that describes the lengthiest geodesic which connects any of the points included in the graph (Wasserman and Faust 1994).

One tool which stems from graph theory is central to social network analysis techniques and of particular importance for this research which is an attribute of each node within the network known as centrality. Centrality, in the context of literature, has always been linked to the access of the node to the information in the network, power of the node, its importance within the network in terms of how much influence an actor has, how prestigious he is or what his status is (Sabidussi

1966; Bonacich 1987; Friedkin 1991; Katz 1953; Russo and Koesten 2005; Ibarra and Andrews 1993; Sparrowe and Liden 2005). Many factors have been considered to indicate the centrality of a particular node including the position of the node within the network, the ties each node has, how well connected the node is, whether the node has adjacent well-connected nodes, distance between the node and other nodes and number of shortest paths between nodes a given node lies on. These factors lead to development of different centrality measures, which can help identify the most powerful influential actors within the network.

2.3.2.1 Centrality measures and their relevant calculations:

- Degree Centrality

This is the simplest and most intuitive type of centrality. It calculates centrality based on the number of nodes connected to the node with edges or how many adjacent nodes are there to any given node (Freeman 1978). The formula for computing degree centrality is as follows:

$$C_D = \sum_{j=1}^{n} E_{ij}$$

Where i refers to the node being investigated, and j refers to all other n-1 nodes in the network to which node i could be connected. E_{ij} is equal to zero if there is no edge between node j and node i and is equal to one if both nodes are adjacent. To normalize this value, we divide by all the potential nodes which could have formed a link with i, which is basically all the nodes except for i itself. So, the formula for normalized degree centrality according to Freeman (1978) is:

$$C_D = \frac{\sum_{j=1}^{n} E_{ij}}{n-1}$$

This type of centrality is easiest to understand, conceptualize and explain to others. It is simple to compute and gives valuable information about each node and its number of ties which makes its computation one of the first steps in the analysis of most networks. However, in this form, it fails to recognize the different tie strengths and so treats strong ties and weak ties as equivalent. Moreover, it treats all neighbors as the same even though some might be much more connected and more important in the network than others, so it favors the number of direct ties a node has to its strategic and structural positioning within the network.

For weighted networks, the different weights of links are usually very important information that cannot be neglected when computing centrality. Therefore, another approach was devised that calculates degree centrality as the summation of all weights of links linking the node to its adjacent nodes (Barrat et al. 2004; Newman 2004):

$$C_W = \sum_{j=1}^{n} W_{ij}$$

Where i refers to the actor being studied, j refers to the n-1 actors remaining in the network and W refers to the weight of the edge between i and j which could be zero if they are not adjacent. This approach is much more suitable for weighted networks, yet it neglects the number of neighbors to which the node is connected since for example having one tie with strength 8 is equivalent to having two ties with strength 4.

One approach to make degree centrality capable of capturing edge weights while accounting for the degree of each node is using two factors multiplied together the first resembles the degree of the node and the second is the total weight of all edges connected to the node (Opsahl et al. 2010). A tuning parameter α is utilized to adjust the effect of each of the two factors, where if alpha is zero, the effect of weights is eliminated, and we get the non-weighted degree centrality and if alpha is 1, the effect of degree is eliminated, and we get the sum of weights method. The tuning parameter could be varied according to the application and whether the number of edges or the weights of edges are more important for the centrality. The formula for this is written as:

$$C_T = {C_D}^{1-\alpha} {C_W}^{\alpha}$$

- Closeness Centrality.

Closeness centrality of a node is a measure of centrality based mainly on the distance between the node and the other nodes, the closer it is to the other nodes the more central. It is commonly tied to the concept of reachability from any point within the network. According to Freeman (1978), it could be written as:

$${C_C}^{-1} = \sum_{j=1}^{n} d_{ij}$$

Where C_C^{-1} is the inverse of closeness centrality and is calculated by summing all geodesics d from node i to every other node on the graph j from 1 to n. The reason it is an inverse is that as the sum of the geodesics increases, the point is deemed to be less reachable from other points and so the centrality of the point decreases. Another form of this centrality equivalent to neutralizing other forms of centrality is written as:

$$C_C = \frac{n-1}{\sum_{j=1}^{n} d_{ij}}$$

The closeness centrality has two main key limitations, the first is that it deals with geodesic distances which means that for weighted networks, the weights are usually not taken into account. Moreover, in the cases where the graph is not connected, closeness centrality cannot be used for unconnected nodes and so the analysis based on looseness is usually limited to the largest component of any social network (Freeman 1978).

- Betweenness Centrality.

Betweenness of a node is a measure of tendency of a node to lie on the path between other nodes. Consequently, betweenness in literature has been linked to the node acting as a bridge, a broker, a mediator, a coordinator, controller of flow of information or a middleman (Barthelemy 2004; Borgatti et al. 2009; Freeman et al. 1991; Newman 2001b). All these roles stem from that a node with high betweenness centrality connects between a lot of other nodes and bridge many structural holes which makes it an essential link in the network which holds it together.

For unweighted networks, the definition for betweenness centrality of a node is simply the proportion of shortest paths within all other nodes in the network on which the node lies (Newman 2001b). This could be written as (Barthelemy 2004):

$$C_B = \sum_{i \neq j \neq k} \frac{g_{ij}(N_k)}{g_{ij}}$$

Where $g_{ij}(N_k)$ is the geodesics between node i and node j on which node k lies, whereas $g_{ij}$is the total geodesics between i and j. This is summed over all nodes i and j other than k such that i≠j. Freeman (1978) proved that the maximum value for the summation above is:

$$\frac{n^2 - 3n + 2}{2}$$

Therefore, the normalized form of betweenness centrality is:

$$\frac{2C_B}{n^2 - 3n + 2}$$

Where n refers to the number of nodes in the network. The above value for normalized centrality ranges from zero for a peripheral node or an isolated node to 1 for a node through which all geodesics between all other nodes pass. In Figure 2.1, node A has a centrality of 1 where all shortest paths between other nodes pass through it. Nodes B, C, D and E, on the other hand, are peripheral nodes where no geodesics pass through them, so they have a betweenness of zero.

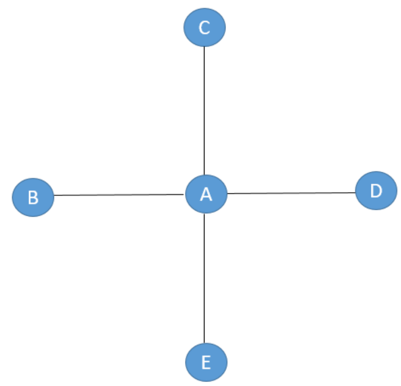

Figure 2.1: Betweenness centrality in a star network structure

Since the normal betweenness centrality depends on the proportion of geodesics between other nodes in the network on which the node lies, it doesn't take into account any weights of the ties as the length of the shortest path is simply defined as a number of edges. To address this problem, Freeman et al. (1991) suggested a different method of calculating betweenness centrality depending on the fraction of flow in the graph that MUST pass through the node under investigation. This flow can resemble a flow of information, goods, files, logic or anything based on the function of the social network. The flow usually originates from a node known as the source of the flow into another node which is defined as the sink of the flow and it moves through the edges which are defined as channels of the flow. The point of deviation between flow betweenness and regular betweenness is that weights are taken into account as the different links in the network are modelled as channels with a capacity equivalent to their weights beyond which no more flow can occur through this channel which can be expressed as:

$$f_{ij} \leq c_{ij}$$

Where f_{ij} represents the flow through the edge connecting nodes i and j and c_{ij} resembles the capacity of the channel between i and j which is the weight of the link ij. Moreover, another point of deviation is that betweenness centrality is usually concerned with only whether the node lies on the shortest path between other nodes whereas flow betweenness is concerned with all paths that connect nodes whether they are short or long. Flow is global to the network, it is not concerned with a single path or a single edge capacity, but with the whole network and the capacity of all the edges together which makes flow betweenness special as it considers the whole network for the calculation of centrality of each and every node(Freeman et al. 1991).

The flow betweenness centrality describes betweenness as the quantity of the total flow which must flow through the node k as a fraction of the total quantity of maximum flow in the network for all unordered pairs of nodes i and j assumed to be source and sink where i≠j≠k and i<j. This can be expressed as:

$$C_F = \sum_{i}^{n} \sum_{< \; j}^{n} m_{ij}(N_k) \,/ \sum_{i}^{n} \sum_{< \; j}^{n} m_{ij}$$

m_{ij} is the maximum flow in the network given that the sink is node *j* and the source is node *i*. m_{ij} (N_k) refers to the proportion of m_{ij} which must pass through node *k* (has no other alternative). To illustrate on the usage of flow betweenness, the social network in Figure 2.2 can be considered.

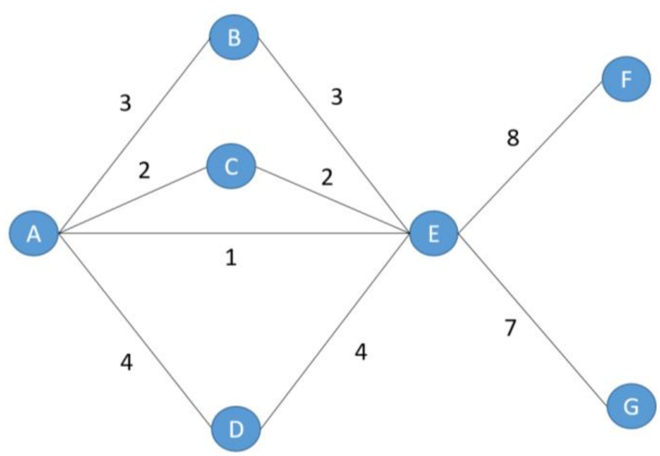

Figure 2.2: Weighted network example

The network in Figure 2.2 is an example of a simple unimodal undirected weighted network of seven nodes. Nodes B, C and D have a degree of 2. Nodes F and G have a degree of 1. Node A has a degree of 4 and Node E has a degree of 6. If the flow betweenness centrality of B is to be computed, we need to consider all other unordered pairs of nodes (A,C), (A,D), (A,E), (A,F), (A,G), (C,D), (C,E), (C,F), (C,G), (D,E), (D,F), (D,G), (E,F), (E,G) and (F,G) one of which as source and one as sink of flow. Accordingly, for every case the flow through B is calculated and the total flow is calculated. Flow through B is summed up for all 15 cases and divided by the sum of the total flow for all 15 cases to obtain flow betweenness centrality.

For the case where A is the source and F is the sink, the total flow of the network is 8 as even though A can provide up to 10 units of flow in its channels, there is only one channel leading to F whose capacity is 8 and thus the flow in the whole network will be 8. Among this 8, 4 can flow through D, 1 can flow directly from A to E and 2 can flow through C leaving only 1 unit of flow which has to pass through B. Therefore, the flow through B will be taken as 1. On the other hand, in the case where A is the source and E is the sink, the maximum flow within the network will be 10 units among which 3 will have to move through B. Moreover, in the case where A is the source and D is the sink, the maximum flow will be 8, 4 will move directly from A to D, 1 will move along the path AED, 2 will move along the path ACED and 1 will have to move along the path ABED. This means that the flow through B will again be one. A visual representation of these three cases is shown in Figure 2.3.

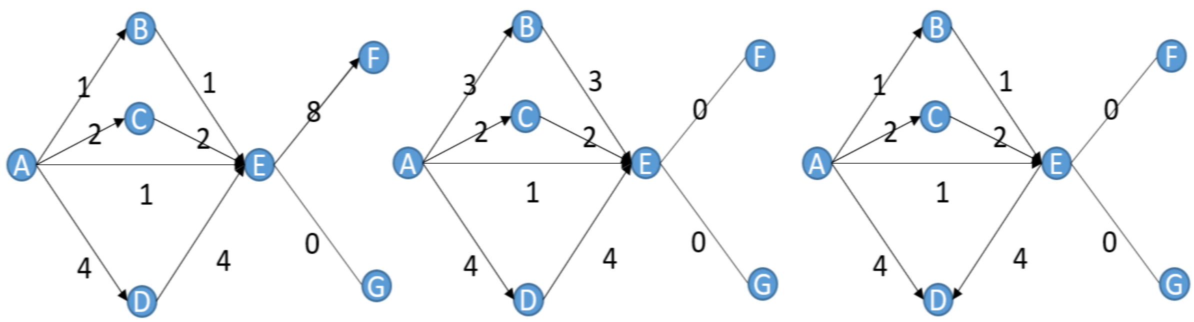

Figure 2.3: Flow Betweenness centrality calculations where source is A and (1) sink is F (2) sink is E (3) sink is D

• Eigenvector Centrality.

All the previously discussed centrality measures whether depending on degree, closeness or betweenness have a common problem. This problem is that they neglect the centralities of adjacent nodes and their effect on the node’s centrality. For example, a node connected to two highly central nodes is supposed to be treated as more central than a node connected to two peripheral nodes. This problem was tackled by Bonacich (1972) when he introduced a new centrality measure where the centrality of a given node is proportional to the centralities of its neighboring nodes. This means that centrality can be obtained by a node either by having many connections (similar to degree) or by connecting with other nodes with high centrality (Newman 2004). This centrality was obtained from the relationship written as (Bonacich 2007):

$$\lambda x_i = \sum_{j=1}^{n} a_{ij} x_j$$

Where i extends from 1 to n. This can be rewritten in matrix form as:

$$\lambda x = Ax$$

A refers to the adjacency matrix. Thus, by definition λ is matrix *A*'s eigenvalue (largest eigenvalue in this case) and *x* is the eigenvector of *A* of the dominant eigenvalue λ. This λ that gave eigenvector values corresponding to eigenvector centrality was shown by Bonacich (1972) to be the largest eigenvalue (also referred to as the dominant eigenvalue).

One of the main advantages of eigenvector centrality is that, in addition to factoring in the neighboring node centralities, it is usable for weighted networks and takes into account tie strengths as they are part of the adjacency matrix. Therefore, the neighbors with stronger ties have more effect on the node's centrality than the neighbors with weak ties.

- Bonacich Power Centrality.

Bonacich (1987) suggested a different centrality measure which has all properties of eigenvector centrality but has more flexibility. This centrality measure was developed in response to Cook et al. (1983) whose analysis of exchange networks provided that the centrality measures in exchange networks didn't necessarily describe power. For example, in some cases being connected to less central/peripheral nodes in a trading network can give more bargaining power as the node has no substitute supplier of goods which gives the node of concern the opportunity to take more advantage (Cook et al. 1983). The flexibility of Bonacich power centrality stems for the constant β, which can be chosen to be any value, typically from -1 to 1. Negative values of β mean that if neighbor centrality increases the centrality of the node decreases. Conversely, positive values mean that as neighbor centrality increases node centrality increases. In both cases, the extent to which the neighbor centrality affects the node's centrality is dependent on the magnitude of β, the greater the β, the greater the effect (Bonacich 1987). A β of value zero corresponds to beta centrality being very similar to degree centrality not giving any value to centralities of neighbors (Bonacich 1987). The equation for computation of this centrality is matrix format is:

$$c(\alpha, \beta) = \alpha(I - \beta A)^{-1} A1$$

Where A denotes the adjacency matrix, the symbol I denotes the identity matrix, 1 denotes a column vector of ones and alpha is a scaling parameter selected such that (Bonacich 1987):

$$\sum_i c_i(\alpha,\beta)^2 = n$$

The expression above means that the square of the length of the centrality vector is equivalent to the size of the network (equivalent to normalization). This means that a beta centrality of 1 for any node means that the node has a centrality which is neither high or low (Bonacich 1987). This means that centrality values much greater than one are considered to be unusually high and those much less than one are considered to be unusually low (Bonacich 1987). The values indicate the variability in centralities where if all nodes have similar centralities they will be fluctuating around one (Bonacich 1987). Beta centrality is considered to be a generalized form of eigenvector centrality where centralities approach eigenvector centralities as beta value approaches the value of the reciprocal of the largest eigenvalue of the adjacency matrix (Bonacich 1987).

2.3.2.2 Other Relevant Definitions:

Centralization

Centralization is a property of the social network which is closely related to centrality. Centralization defines the variability in centralities within the graph. For example, a highly-centralized graph can have a single central node connected to connected to peripheral node such that the peripheral nodes have only 1 edge which is linked to the central node commonly known as a star graph shown in Figure 2.4.

In the above graph the node C has 5 links whereas the 5 other nodes have only 1 link each thus resulting in a highly-centralized graph. The definition of centralization according to Freeman (1978) is the sum of the differences between the centrality of the point of maximum centrality in a graph and the centralities of other points expressed as a percentage of the maximum achievable difference in centralities for a graph of the same number of points n. For example, in Figure 2.4, the degree centralization is 1 where no more difference between maximum centrality and other node centralities can be realized for a 6-node graph. On the other hand, in a graph where centralities of all nodes are the same, the maximum centrality will be equal to all node centralities thus leading to a centralization score of zero (Freeman 1978). Therefore, when a node has the highest possible centrality in the network and the others have low centralities this leads to a high centralization

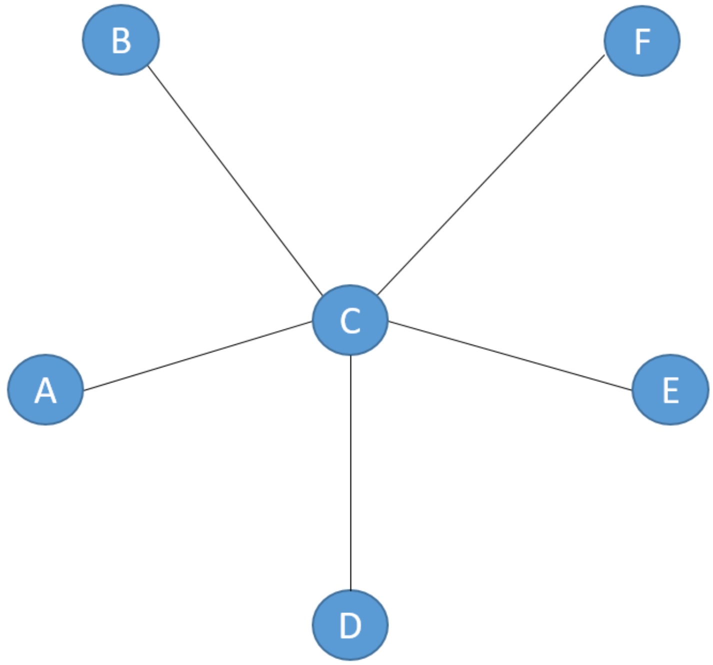

Figure 2.4: 6 node star graph

value for the network and indicates most of the centrality in the network is concentrated in one node. This could be formulated as:

$$Centralization = \frac{\sum_{i=1}^{n} C(x^*) - C(x_i)}{\max(\sum_{i=1}^{n} C(x^*) - C(x_i))}$$

Where $C(x^*)$ is the centrality of the node of largest centrality in the network, $C(x_i)$ refers to centralities of all other nodes and $\max(\sum_{i=1}^{n} C(x^*) - C(x_i))$ refers to the maximum achievable difference in centrality for any graph arrangement of n nodes (like the star arrangement for the degree centralization of the 6-node network in Figure 2.4) assuming that the network is binary.

In case of valued data, the values of the data cannot be included in the centralization analysis as $\max(\sum_{i=1}^{n} C(x^*) - C(x_i))$ is not computable. Accordingly, the centralization can only be computed by binarizing data that is converting all nonzero values into 1. The exception to this is in the case of flow betweenness centrality where the calculation does not depend on $\max(\sum_{i=1}^{n} C(x^*) - C(x_i))$ and can be expressed as (Freeman et al. 1991):

$$Centralization = \frac{\sum_{i=1}^{n} C(x^*) - C(x_i)}{n-1}$$

This centralization calculation is thus only a measure of the average difference between the node with the highest centrality and the other nodes of the network without referring to the highest possible centrality for an n-node network. This measure is, thus, not normalized to be comparable to the binary network centralization measures.

Density

Density, also referred to commonly as connectedness, refers to how well-connected nodes are within the network. Unweighted density is simply the number of links in the network as a fraction of the maximum number of possible links in the network, which is the number of links which would have existed had all the nodes in the network been connected (Hanneman and Riddle 2005). Weighted density, conversely, sums all the weights of edges in the whole network and divides that by the number of possible edges given all the nodes are linked to one another. This, when applied to undirected networks can be expressed as (Liu et al. 2009):

$$D = \frac{2\sum_{i\in N, j\in N, i\neq j} w(i,j)}{|N|.(|N|-1)}$$

Where w refers to the weight or the strength of the edge connecting nodes i and j which resemble all unordered pairs of nodes in the set of nodes in the network N. $|N|$ refers to the total number of nodes in the network. This can be applied to unweighted networks if all weights of nodes are set to 1. Moreover, it is applicable to directed network if the multiplier 2 in the numerator is removed as the total number of possible edges in the network is doubled for a directed network.

Transitivity

Transitivity is one of the unique features of social networks. In networks having high transitivity, if node 1 is connected to node 2 and node 2 is connected to node 3, this increases the probability that node 1 is connected to node 3 (Krivitsky et al. 2009).

2.3.3 Social Network Analysis in the context of construction industry

Research into the use of SNA and its application in various fields pertaining to construction has been a growing trend in the past few decades. The tools provided by SNA were found to be of particular benefit in different areas of construction management. They provided simple and powerful metrics for the analysis and description of networks in these fields to enhance the understanding of network structure. Moreover, they provide a useful assessment tool to support the decision making and allocation of resources within construction organizations.

2.3.3.1 Use of SNA in Construction Management

For the last two decades, research into Social Network Analysis application in Construction Project Management has been growing rapidly and continuously (Zheng et al. 2016). This trend was observed as the network analysis was applied to different fields pertaining to Construction Project Management. As a versatile and innovative technique, Social Network Analysis was used in a variety of ways in construction project management. Its main application in construction research to date was modelling communication, knowledge diffusion and exchange of information. This type of modelling alone had various implications in terms of the effect of Social Network Analysis on research areas in construction including communication management, knowledge management, assessment of IT Utilization, high performance team formation, site and resource management and risk management (Zheng et al. 2016).

Social Network analysis revolutionized the areas of knowledge and communication management in the construction research. It changed how communication could be perceived both at an interorganizational and an intra-organizational level.

Earlier research efforts for Social network analysis focused on intra-organizational aspects of communication (Zheng et al. 2016). The intra-organizational approach modeled a construction project as a social network with key actors in the project as nodes and the relations between them as links. This approach was based on the emerging view of any construction project as a temporary organization or a temporary network of stakeholders first discussed by Turner and Müller (2003). They suggested the project is a network utilizes complex decision making and problem solving to be able to achieve the goals of permanent organizations.

At the Intra-organizational level, Brookes et al (2006) utilized a case study to model network relationships in a network and how their density and conductivity contribute to the network' social capital. Their concept of social capital was the resources available for achieving project goals as a function of the individuals' resources and the relations between them in terms of density and conductivity. They used hypothesis testing to find a significant correlation of conductive relationships with trust, respect, relationship longevity, common background and broader social context within a network.

SNA was proposed as a performance assessment tool for communication at the intra-organizational level by several research studies. Pryke (2004) used SNA to model the construction project actors and three sets of intra-organizational relationships between them in the project namely contractual, information exchange and performance incentive relationships. His findings ascertained the importance of network centrality measures as a tool of evaluating whether a construction procurement method works as intended in terms of shaping roles and relationships among actors. Chinowsky et al. (2010) suggested a technique known as project network interdependency alignment as a measure of project effectiveness. Their technique aligns two social networks together. The first network is a theoretical network derived from interdependency of tasks of different project team members and the required level of knowledge exchange between the different project actors in order to be able to complete the task optimally. The second network is a network of actual data of communication taking place between different project actors. They proposed that by aligning these two networks and their measures together, any communication inefficiency can be identified in a proactive manner prior to causing any effect on project outcomes so that it can be addressed immediately. This approach was suggested to have evolutionary outcomes on project management processes.

Another approach on using SNA as a performance assessment tool for communication at the intra-organizational level was proposed by Pryke (2005). He demonstrated using four case studies how SNA can be used to quantify changes of roles of different project actors as well as project governance. He utilized it to test the effects of different supply change management techniques, test the effect of partnering as well as utilizing work clusters for procurement in projects. His findings demonstrate how SNA was used to comparatively analyze project governance by using SNA measures like density and centrality and present a reusable framework for analysis of other projects.

Technological implementation was another application of SNA in analyzing intra-organizational communication in construction projects. Mead (2001) was one of the first adopters of SNA in analysis of new technology implementation. He used SNA on a case study to analyze effects of using an intranet network on enhancing the communication for large construction projects. His findings showed the contribution of the system in the enhancement of communication for the project and information diffusion as well as that the system itself increased in terms of centrality as more time passed after its implementation. Thorpe and Mead (2001) used SNA measures, particularly centrality analysis, to assess the use of the project specific website. The project specific website was to be used as a central source of just in time information rather than just in case information which usually results in information overload. The findings showed the project specific website to speed information flow and overcome many communication barriers changing the way that different project teams communicate together. They also showed that this system is prone to the fact that if one of the key project actors refuses to participate, it becomes much less effective and doesn't produce the outcomes it was intended to produce.

Finally, social network analysis was used to evaluate the effect of using a lean construction system known as last planner on communication on site (Priven and Sacks 2015). The last planner system is a system functioning on short term planning of work packages filtered to ensure that their prerequisites have matured such that they can be completed with efficiency and correctness. It also functions on commitment management where it brings planning to the supervisors' level, encourages conversations and relationship buildings to acquire promises throughout the construction process. These promises should be acquired from the correct personnel at the right time to ensure a successful implementation. This aims at reduction of wastage in time and creating better coordination to ensure efficiency. The study showed an increase of strength in social network ties and communication when last planner system was implemented as well as formation of ties across different ethnic groups. The implementation also corresponded to a significant increase of centralities of crew leaders in different trades as well as increase in the number of essential communication channels which were active during construction.

The interorganizational approach, on the other hand, focused on the projects as different organizations working together rather than the single organization approach suggested by Turner and Müller (2003). These efforts are relatively recent, and their focus is on bridging the gaps between different organizations working on the same project. This is particularly true in a multinational or multicultural context in projects whose members are organizations from all around the world where many sociocultural communication barriers might affect the progress of the project.

Taylor and Levitt (2007) how the use of new innovative technology in project communication can align with preexisting project roles and the effect on overall performance. The study was carried out on project networks of totally different structures from the United States, Germany, France and Finland studying the implementation of three new innovations in three-dimensional computer aided design (3D CAD) and how they align with project roles. The findings showed that the alignment facilitated the implementation of the new innovations. The findings extended to that in the case of misalignment, the adverse effects on the new innovation diffusion is mitigated by the presence of accrued firm interests at the network level, presence of agent for change in project networks, permeability of inter-firm boundaries as well as high relational stability within the network.

Taylor and Bernstein (2009) utilized network analysis to analyze the maturity of use of Building Information Modelling (BIM) in different businesses which they termed "evolution along BIM paradigm trajectories". Information exchange was modelled as transfer of BIM electronic files through the project network either within the firm or between different firms was analyzed. This analysis was combined with other qualitative measures to assess the maturity of BIM in each of 26 cases of construction firms utilizing BIM. The paradigm trajectory was composed of 4 paradigms for BIM which are visualization, coordination, analysis and supply chain integration respectively (Taylor and Bernstein 2009). Evolving from a paradigm to another took more experience as well as lead to more information sharing through the network.

Nayak and Taylor (2009) investigated the networks of complex design outsourcing and formation of virtual teams across cultural boundaries. Their findings outlined key barriers between the clients and vendors. They found that the vendors perceived trust development and knowledge sharing as the most important elements for successful global outsourcing. Clients, on the other hand, perceived communication, schedule control and quality control as the key important factors (Nayak and Taylor 2009). Both parties perceived cultural barriers, different norms of design practices as well as redundant roles as key issues of concern (Nayak and Taylor 2009). SNA was also used to display the role of cultural boundary spanners (individuals who can bridge between multiple cultures, for example having grown in a country and worked in a different country) in facilitating and enhancing communication amongst multicultural teams in the context of global outsourcing (Di Marco et al. 2010). This research utilized a comparative case study between a team which had a cultural boundary spanner and one which didn't. Both teams on the study were working on a design schedule optimization problem. The findings showed the role of the cultural boundary spanner, whether he is nominated or emergent, as a facilitator given that the cultural boundary spanner had the highest centrality in the team. The findings extended to that the cultural boundary spanner expedited the process of link formation among other members of the team leading to a denser better-connected communication network.

Another key application of SNA in the field of knowledge and communication management is in the area of team formation. Chinowsky et al. (2008) suggested an approach for team formation in high performance projects through modeling trust, communication and knowledge sharing in the project team network. Their model was based on two main components, namely dynamics and mechanics. According to the author, dynamics are mainly what motivates an individual to increase project performance, whereas mechanics are the information exchanged to be able to complete a project. The research findings linked performance of the project teams on different project areas successfully to network attributes including density and centrality. The study concluded that if some key individuals are isolated from the rest of the network, this gap in trust and information sharing leads to unintegrated knowledge that in turn leads to suboptimal decision making. A complementary study utilized this approach on case studies for four companies using networks to resemble trust, client specific communication and knowledge exchange (Chinowsky et al. 2009). Their research was based on two main analyses which are leadership and collaboration analyses. Their goal was to identify whether leaders are playing their roles in enhancing collaboration based on centrality and whether collaboration is promoted through trust and client specific communication through network densities. The study findings indicated that when a large company focuses on each of its offices being independent and communicating through key individuals, the connections between the offices are weak. The study findings extended to that these weak links which have no redundant alternatives hinder collaboration and make knowledge transformation networks more isolated.

Construction project management research did not only benefit from SNA technique in the field of communication and knowledge management. The versatility and modelling capability of SNA extended its use in many other fields in construction management research including site management, transportation planning and risk management.

In the field of site management, SNA was utilized as a useful tool for assessing the relationship between different crews on site. Wambeke et al. (2011) conducted a study to model a network of construction project trades based on their spacial proximity. They based their findings on various centrality measures mainly degree and eigenvector centrality to identify the key trades. The study identified a core group of trades including painting, mechanical, electrical and drywall trades, as well as second tier trades which included wall finishing, ceilings, fire protection, inspections, flooring, concrete and steel fabrication. Even though their study was based on a single project, their findings present a repeatable framework that could be utilized in various areas in project management including scheduling and work breakdown structures. Moreover, their model could be utilized as a detection tool in the site management area to detect any problems or miscoordination on site and address it in a proactive manner, which is considered to be a new approach to site management.

In the field of transportation planning, a study connected between network accessibility, time and place ranks and different SNA centrality measures (Rubulotta et al. 2012). The study concluded correlations between centrality measures and different measures of accessibility. It asserted that SNA can be utilized as a tool for decision support for sustainable mobility decisions.

A study by Cheng et al. (2015) displayed the versatility of Social network analysis for use in transportation planning. SNA centrality metric was extended by the study from only studying network topology to considering the flow of commuters and delay experienced by commuters if a node was shut off. The study introduced 3 new centrality measures to account for these two factors and compared their results in terms of node centrality with conventional centrality measures.

Finally, a study by El-adaway et al. (2016) utilized social network analysis (SNA) tools and metrics to investigate transportation networks for the most critical intersections. The study showed that the performance of the whole network of roadways hinges on the performance of the most critical intersections; where criticality is determined by the different SNA centrality measures. The developed methodology was applied to 2 case studies in Mississippi where results were in line with the traffic studies conducted by the Mississippi department of transportation. The methodology was proposed as a preliminary step that is neither expensive nor time consuming to prioritize the intersections in terms of criticality. This allowed to determine which intersections should be candidates for more thorough analysis to optimize on resource usage and better evaluate transportation networks.

Within the area of risk management, Loosemore (1999) used SNA to study the effect of centrality on power, risk acceptance and responsibility within the project. His network was used to predict the power of each actor at different stages through his centrality, structural equivalences and factional data. He used several networks to monitor the different changes in behavior at different stages of a large conflict within a construction project. The study concluded that this behavior is what dictates who bears the responsibility of a risk during a conflict rather than contractual obligations. This was regarded by him as an alarming phenomenon as it usually leads to responsibilities for risks being assigned to those with lower power in the project network rather than those with sufficient expertise which leads to suboptimal project outcomes.

Within the area of risk management, a different approach was taken towards network modelling. All the networks discussed earlier in the area of construction management were modelling individuals, organizations or project actors as nodes whereas this approach modelled defect causes as nodes. Aljassmi et al. (2013) idealized defect root causes as nodes to a social network and defined their main attribute to be pathogenicity which is their capacity as to being able to result in other causes or conditions that can lead to generation of defects. They applied this approach on a case study of residential construction projects and were able to rank the causes leading to the usage of impaired material in terms of pathogenicity by analyzing their network centralities. The network they created was a visualization of the causation mechanism eventually leading to defect generation. The model was based on the swiss cheese model of causation of defects developed by Reason in 1990.

2.3.3.2 Social Network Analysis application in Construction Safety

There have been very recent research efforts utilizing the capabilities of social network analysis in the area of construction safety. The ability of social network analysis to capture important properties and patterns in communication networks has led several researchers to utilize SNA tools to analyze safety communication.

Albert and Hallowell (2014) utilized SNA to correlate worker interaction networks with their hazard awareness, identification and communication. This was carried out by including 18 crews from distinct construction trades in US projects. His findings provide a strong evidence that crews which have denser networks perform much better in terms of hazard awareness and communication than crews with poorly connected networks. This leads to the deduction that organizations sould be keen on promoting worker-worker interactions within crews to improve safety performance.

Chan et al. (2014) analysed communication patterns of 21 ethnic minority workers in Hong Kong mainly from Nepal and Pakistan using social network analysis. The study suggested the use of SNA metrics including density and centrality as well as sociograms to optimize the safety performance and communication of workers belonging to an ethnic minority. This was to be achieved by enhancing the networks of communication amongst these workers. The results suggested that frequent and open supervisor- worker communication enhanced safety performance of workers.

Alsamadani et al. (2013b) utilized social network analysis to identify how safety-related communication is influenced by language proficiency amongst multilingual crews. The sample for the study was 14 crews in the US in Denver Metropolitan Area. The study findings included that crews with a single language perform 51% better than multilingual crews, bilingual workers have a more central role than unilingual workers when the crew is multilingual and workers younger than 35 years are more central in multilingual teams. Additionally, the study found that managers play a crucial role in safety communication and knowledge sharing regardless of their language proficiency. Finally, the study mentioned the importance of the role of what is known as language boundary spanners, who are those who can communicate in several languages, as they form the core of the communication network connecting members of crews who are usually separated. On the other hand, the study mentioned the risk of having separated subnetworks in safety communication which leads to poor safety performance and weakly connected crews with high turnover rates. Additionally, this poor structure can lead to significant inability to communicate hazards leading to increased accident occurrence.

Alsamadani et al. (2013a) used SNA to identify and quantify safety communication patterns. The study was focused on crew members in construction teams in Rocky Mountain Region in the United States. The study identified some common characteristics of top performing crews in terms of safety including having formal communication with management about safety at least weekly, having safety communication that is informal every week, undergoing a formal training in safety and utilizing every proposed safety communication method at least once a month. Moreover, it was found that SNA metrics including centrality and particularly betweenness wasn't a significant measure to differentiate high performance crews from low performng ones. The only relevant metric was nework density where denser networks were found to function better for the communication mode concerned with training and management, whereas networks that are diiffuse were found to function better for informal communication mode. These findings prove that for an effective safety communication, the required network depends on the knowledge being exchange and the communication mode. Moreover, effective commication according to this study hinges on the frequency of communication as well as its mode.

CHAPTER 3:
METHODOLOGY

To be able to achieve the aim of this study which is analyzing interactions between root causes of fatal accidents as well as their relation to commonly quoted direct causes, a three-step research methodology is adopted by this study. The three steps are to (1) collect data; (2) develop a social network analysis model; and (3) utilize the model and analyze results. Utilizing this methodology will enhance the current comprehension of accident causation as well as key underlying dynamics which make accidents occur.

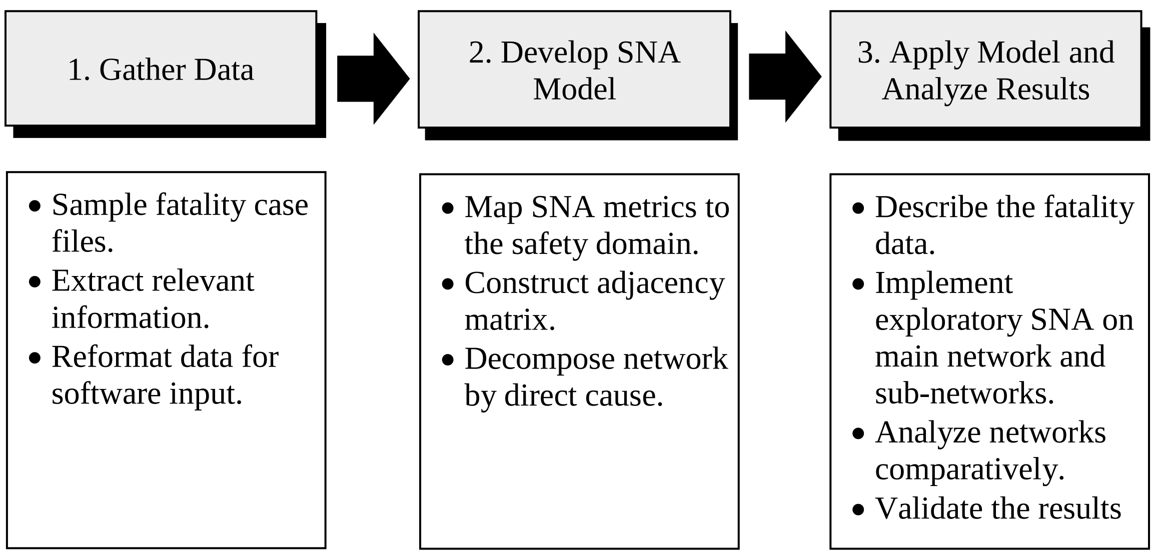

Figure 3.1: Flow chart describing research methodology

The Figure 3.1 outlines the three-step methodology utilized for this research. The following sections in the methodology go through each of these steps thoroughly and explains how it is applied to serve the research goal.

3.1 Data Gathering

3.1.1 Sampling Fatality Case Files

The first step towards analyzing fatal injuries was sampling case files from recent years from the CIRPC database at the University of Tennessee. Accident case files document fatal accident occurrences and are written by OSHA compliance officers subsequent to the occurrence of fatalities. The scope of investigation was specified to include 5 years of fatal accident occurrence.

Therefore, sampling was undertaken from the files since the year 2006 until the year 2010 which was the most recent available data that was complete. In this case the population of subjects of the study were the victims of the fatal accidents, the outcome was the fatalities and the information being collected was the attributes of the individuals, the projects, the organization and the site as well as the direct causes and the root causes of the fatal accident.

For the purpose of this study, stratified random sampling technique was utilized. To obtain a stratified random sample, the sample space is divided according to a certain property to what is known as strata (Olken 1993). Subsequently, a random sample is taken from each of the strata to represent the whole stratum (Olken 1993). The property of data utilized for division or stratification in this case is the year of the occurrence of the fatality. The sample size was equal for every year to obtain an unbiased representation of all years equally. For every year, 20 files were randomly sampled from the database of this year leading to a total of 100 case files as a sample size for the study spanning 5 years.

3.1.2 Sample Size

This study has a sample size of 100 projects from which the fatality data was drawn. There is no convention or methodology identified in literature to identify the suitable sample size for SNA. To resolve this problem, key research efforts utilizing SNA in construction and safety management and making strong contributions were analyzed and the number of samples they utilized were aggregated as a guideline for minimum sample size for this study. A summary of 17 of these studies with their purpose and the sample size they used is provided in Table 3.1. The largest number of projects from which data was collected was 82 projects used by Taylor and Levitt in 2007 when studying the alignment between roles of different project actors and the advancements in technology in the area of communication. The average of the number of projects utilized for these studies is 22 projects. Since this study draws data of fatalities from 100 different projects, the sample size is considered to be large in the context of social network analysis in construction management. This provides credibility for the results of this study and lays the ground for drawing solid conclusions from its results.

3.1.3 Extraction of relevant information

The data gathering involved a thorough investigation of each of the fatality case files to be fully able to understand the accident and extract the necessary information for the sake of the

Table 3.1: Sample sizes in reviewed construction management studies

Study	Purpose	Sample Size
Albert and Hallowell (2014)	Correlating between the interaction of workers and their situational awareness and ability to recognise and communicate hazards using SNA.	18 different construction crews from different projects and trades in the United States
Alsamadani et al. (2013b)	Analyse the effect of language proficiency on patterns of safety communication using SNA.	14 small construction crews in Denver Metropolitan region in the Unted States
Chan et al. (2014)	Analysis of communication patterns of ethnic minority workers using SNA.	21 ethnic minority workers in Hong Kong
Chinowsky et al. (2009)	Investigating the leader's roles in enhancing collaboration and whether trust and client specific communication enhance collaboration using SNA.	Four construction firms
Construction Industry Institute (1993)	Devising zero injury techniques.	25 projects with superior safety performance
Construction Industry Institutes (2012)	Analysing the application of leading indicator and its effect on OSHA's TRIR.	57 projects
Construction Industry Institutes (2014)	Analyzing the effects of NMR on OSHA's TRIR and employee motivation to identify and report site hazards.	47 projects
Di Marco et al. (2010)	Investigating the role of cultural boundary (CBS) spanners in global design outsourcing teams using SNA	2 teams one which has a CBS and one which doesn't
Hinze (2005)	Investigating the effect of number of hours investigating a near miss and OSHA's RIR.	9 power plant projects
Hinze et al. (2013a)	Quantifying the effect of safety practices implemented by world class organizations on OSHA's RIR.	27 projects
Hinze et al. (2013b)	Assessing the correlation of two leading indicators, namely positive reinforcements and WSBO on OSHA's TRIR	14 projects
Jones et al. (1999)	Investigating the effectiveness of NMR in reducing lost time injuries and in quantitative risk assessment.	Two Norsk Hydro projects
Pryke (2005)	Analysis of project governance changes as well as changes in project actor roles when using partnering, supply chain management and work clusters in procurement	4 projects followed throughout their lifecycles
Rajendran (2012)	Comparing between PTP, WSBO and SSA as leading indicators.	1 project over 37 weeks
Taylor and Bernstein (2009)	Analysis of evolution of businesses along Building Information Modelling (BIM) paradigm trajectories using SNA.	26 construction firms
Taylor and Levitt (2007)	Investigating the effect of the alignment of new innovations in communication with the underlying project roles networks using SNA.	82 construction firms
Wambeke et al. (2011)	Identifying key trades on a construction site based on spatial proximity using SNA	1 project involving 43 trades over 28 weeks.

causation analysis of the accident. Another aim of the data gathering was understanding of the attributes of the organization, project, site or individual which led to the occurrence of the fatality. Characteristic OSHA accident investigation case files include a site walkaround, an inspection summary, an inspection narrative, a safety narrative and OSHA citations as well as interviews, official forms and checklists. The data gathered had to contain the following:

1. Quoted Direct Cause of Accident.
2. Root Causes of the accident.
3. Number and codes of quoted Occupational Safety and Health Administration (OSHA) Citations linked to the accident.
4. State in which fatality occurred.
5. Work Site End Use Function.

The classification used for end use function is adapted from the end use functional classification created by U.S. Census Bureau (1997) for construction worksites for both building and non-building construction. Understanding the end use can help understand the nature of the project and thus the conditions under which the accident occurred. The data gathered for work site end-use function lied in one of the following classifications:

F1. Commercial building.
F2. Manufacturing and light industrial building.
F3. Single Family House.
F4. Apartment building.
F5. Educational building.
F6. Religious building.
F7. Health care and institutional building.
F8. Farm building.
F9. Amusement, social, and recreational building.
F10. Bridge and elevated highway.
F11. Highways, streets and related work.
F12. Power and communication transmission lines, towers, and related facilities.
F13. Power plants and cogeneration plants, except hydroelectric.
F14. Sewers, sewer lines, septic tanks and related facilities.
F15. Sewage treatment plants.

F16. Tank storage facilities other than water.

F17. Pipeline construction other than sewer or water lines.

F18. Dam and reservoir construction.

F19. Multi Family Dwelling.

F20. Heavy Construction.

F21. Other Building.

6. Whether the work the victim was executing at the time of the fatality was his regular task.
7. Victim's role in the fatal accident.

This refers to whose immediate actions lead into the fatal events. Previous research conducted onto similar CIRPC case files divided the victim's role in the fatal accident causation to five main subcategories and explained each subcategory as follows (Beavers et al. 2009):

1. "Action by victim"

   This category refers to when direct actions initiated by the victim had a substantial contribution to the fatal accident.

2. "Action by another worker"

   When the victim's own action had no substantial contribution to the accident. On the other hand, another worker's actions had a substantial contribution to the accident.

3. "Combined"

   When both the victim's actions as well as actions of another worker/workers had substantial contribution to the accident occurrence.

4. "Wrong time/wrong place"

   When neither the worker's actions nor other worker's actions had a significant contribution to accident occurrence, but the accident was due to the victim being at the wrong place at the wrong time.

5. "Unknown"

   When an accident doesn't fall in any of the pre-mentioned categories and thus classifying it in one of them is not possible.

Most of this data was readily accessible in the case files. The only implicit data was the root causes and the victim's role in the fatal accident. This data was deduced through an exhaustive examination and analysis of every accident through the provided case file. Root causes were extracted by a thorough data analysis of each file against particular defects covering different areas including training, communication, safety equipment, human-related, organizational, site-related, design-related, procedural and equipment-related factors. The collected causes were then shared with different members of the CIRPC center as well as members of the research group to assure their validity.

3.1.4 Data Reformatting for software input

This phase involved reviewing the collected data to ensure that the data gathered is organized, consistent and logical. The root causes which were written down for each fatality case were collected from all cases and arranged together to agglomerate the causes which have essentially the same meaning into one cause thus getting rid of redundant causes.

The output from this phase was the bimodal directed adjacency matrix, where one mode is the root causes (RC) and the other mode is the case files, which resemble the fatal accident (CF). The network was binary as the same root cause cannot be repeated in the same case file more than once. The edges resembled the causation links which were directed from the root causes to the fatality case files, where each fatality had directed edges coming into it from root causes that contributed into this particular fatality. A clarifying visual representation of the network of the root causes and the fatal accidents is shown below in Figure 3.3. The format in which the adjacency matrix should be in by the end of this phase is shown in Figure 3.2. The adjacency matrix formed was composed of 66 rows resembling the 66 root causes and 100 columns resembling the 100 construction fatality case files utilized to construct this network. In this matrix, a one indicates that the root cause in the row was one of the causes of the fatality in the column whereas a zero indicates that the accident could not be attributed to this cause. It should be noted that the given figures are for clarification on the format of adjacency matrix and the corresponding network. The figures have nodes resembling case files 6 to 99 as well as root causes 6 to 65 omitted to be simpler for clarification purposes.

3.2 SNA Model Development

3.2.1 Mapping SNA metrics to safety domain

The tools and metrics commonly used in social network analysis as well as their mathematical formulation and some of their variations have been thoroughly covered in the literature review. This section of the methodology deals with these tools in the context of the eight networks of fatality root causes. The objective of this section is to examine the applicability of each of the

	CF1	CF2	CF3	CF4	CF5				CF100
RC1	1	0	0	0	1	.	.	.	0
RC2	0	1	0	1	0	.	.	.	1
RC3	0	0	1	0	0	.	.	.	1
RC4	1	0	0	0	0	.	.	.	1
RC5	0	1	0	0	1	.	.	.	0
.	.	.	.	.	.	.			.
.	.	.	.	.	.		.		.
.	.	.	.	.	.			.	.
RC66	0	1	0	1	0	.	.	.	0

Figure 3.2: Adjacency matrix for bimodal directed binary network

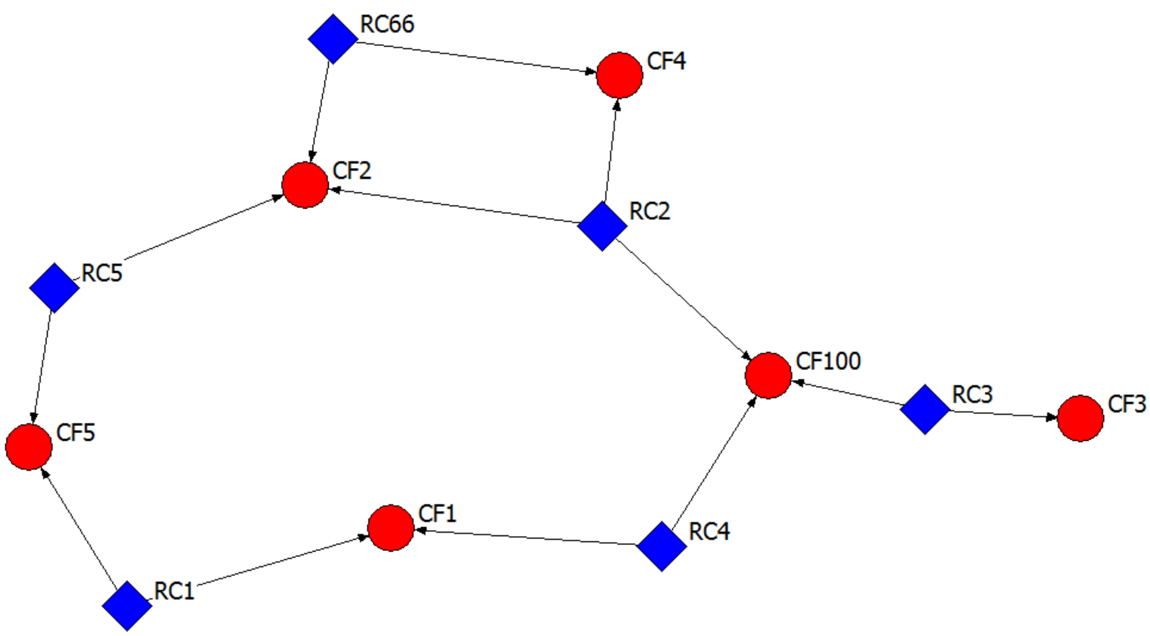

Figure 3.3: Social network diagram for bimodal directed binary network

metrics to the network and to answer questions like: which characteristic of the network that metric describes, what does that metric tell us about the network and which questions about the network can this metric be helpful in answering? Answers to these questions not only forms the most important part of the analysis of the network, but also play a key role into the expandability, flexibility and repeatability of the proposed social network model. This is because it helps understand the nature of this analysis and how it could be adapted to different types of data in the field of construction safety.

3.2.1.1 Density.

Density in general refers to how connected root causes are within the social network. The unweighted density resembles the number of the cooccurrence links between the root causes as a proportion of all the possible cooccurrence links in the network. For this purpose, a density of 1 means that every pair of root causes in the network cooccured together at least once, whereas a density of zero refers to that no root causes ever cooccured together and between those two extremes is the density which is the proportion of the root causes cooccurring together at least once.

Weighted density on the other hand is more useful in this context since if two root causes happen to cooccur in one fatality, it does not necessarily mean they are interrelated. The number of times any two given root causes cooccur, which is neglected in the former case, is the real measure of whether they are really interrelated. The weighted densities refer to the sum of the number of times every pair of root causes in the network cooccured together divided by the maximum possible number of cooccurrence links in the network. In this context, this density can be greater than one and it resembles the mean or average number of cooccurrences between any two given root causes in the network or the mean edge weight. This calculation considers both those pairs of root causes which do not cooccur as zero cooccurrence or zero edge weight and those which cooccur as the number of times they cooccur which is the value of the edge weight.

3.2.1.2 Centrality.

Several types of centrality have been mentioned in the literature review. The applicability of these to the network will be examined:

3.2.1.2.1 Degree Centrality

The unweighted variant of degree centrality shows the degree of each node which is, in this case, the number of root causes to which a particular root cause is connected. That is, how many root causes co-occurred with the root cause being investigated in at least one fatal accident.

Even though this approach is useful towards understanding how much a root cause co-occurs with other root causes, the main purpose behind our analysis is eventually to be able to reduce fatal accidents. To this end, a more interesting approach to be taken is which root cause, if removed from the network, will lead to the greatest reduction in the density or connectedness in the network.

This approach is based on the definitions of weighted density and weighted degree centrality. Weighted density shows the connectedness in the network taking edge weights into account. Weighted degree centrality, on the other hand, shows the percentage of edge weights connected to a particular node. Therefore, the weighted degree centrality can tell how much reduction in the overall density or connectedness in the network can be achieved by removing the root cause being investigated. It should be noted that even though the reduction in the network connectedness leads to a reduction in the fatalities due to removal of a certain cause, the percentage reduction in fatalities is not equivalent to reduction in network connectedness. This is due to the fact a single fatality might be due to 5 root causes and so it will be counted in each link between each two of these five causes, so it will be counted 10 (5 choose 2) times.

Another approach discussed suggests a centrality measure midway between normal and weighted degree centralities. This approach is just a middle ground between the two prementioned approaches and has no physical meaning in the context of root cause networks. Moreover, the lack of a reliable method for defining the parameter alpha which gives the relative weights given to each of the weighted and unweighted methods, makes it a subjective centrality measure.

3.2.1.2.2 Closeness Centrality

To be able to define closeness centrality in the context of the root cause networks, the geodesic has to be understood first in this context. The geodesic refers to the shortest path between two root causes which would be 1 if the two root causes cooccurred in at least one fatality. It would be 2 if one root cause co-occurred with another root cause which cooccurred with the root cause of interest. It would be 3 if one root cause co-occurred with a second root cause which cooccurred with a third root cause which cooccurred with the root cause of interest, and so on. The length of this cooccurrence path from one root cause to another root cause is thus a measure of how closely or distantly related are the root causes.

However, in the context of cooccurrence, it doesn't make a lot of difference if the geodesic is 2 or 3 as long as it is not one other than how compact or how diffuse the network is. Moreover, this centrality does not consider the weights of the edges connected to the node being investigated. This makes closeness centrality of limited usefulness when dealing with this particular type of network where the weights of the edges have a lot of value in predicting centrality. Geodesics, even though show how well connected a root cause to the others, has no way of differentiating root causes which cooccurred in many fatalities from those which happen to cooccur in a single fatality. This makes closeness centrality one of the weaker measure which, though indicative of centrality, is a weak predictor in the case of this network.

3.2.1.2.3 Betweenness Centrality

Betweenness centrality describes the proportion of the geodesics between other root causes on which a root cause of interest lies. This necessarily means that the betweenness describes how much a root cause cooccurs with other root causes which do not cooccur together. In that sense, causes with higher betweenness are those which bridge the most structural gaps within the network cooccurring with root causes that are unrelated. This essentially means that high betweenness root causes can be conceptualized as hubs which play a key role in the network connectedness. This means that addressing these causes which bridge between other causes which normally do not cooccur can potentially play the greatest role in collapsing the linkages between the nodes and making the network tend more toward isolated nodes. In the context of accident causation theories including domino theory (Heinrich 1969), multiple causation theory (Petersen 1971), swiss cheese model (Reason 1990; 1997) and other models previously discussed in the literature review, removing the cause bridging between most other causes has the potential to significantly reduce the accident occurrence risk. Accordingly, betweenness centrality proves to be an important metric which describe the tendency of a cause to bridge between unrelated causes.

However, the standard form of betweenness centrality neglects a very important element in the network which is the weights of the edges or the number of times two causes cooccurred, which limits its potential as it identifies both strong and weak links as the same. For this purpose, the flow betweenness provides even more insight into the network properties.

3.2.1.2.4 Flow Betweenness Centrality

Flow betweenness centrality is very similar to the regular betweenness as it measures the tendency of a root cause to act as a bridge between other root causes in terms of cooccurrence. However, the flow betweenness takes weights of links into accounts as capacities of these bridges and it considers all paths in the network rather than just the geodesics. It models the cooccurrence between different root causes as a flow and computes the portion of the flow that must pass through the cause of interest. This methodology provides the benefit discussed previously of identifying the cause bridging between most unrelated root causes while taking the weights of the edges into account.

3.2.1.2.5 Eigenvector Centrality

Eigenvector centrality of a given root cause is affected by how many root causes it cooccurred with (degree), the number of times it cooccurred with each root cause (tie strength) and the centrality of root causes it cooccurred with.

3.2.1.2.6 Bonacich Power Centrality

Similar to eigenvector centrality, Bonacich Power Centrality is affected by how many root causes it cooccurred with (degree), the number of times it cooccurred with each root cause (tie strength) and the centrality of root causes it cooccurred with. The difference is that Bonacich Power Centrality includes a parameter β which could be varied according to the extent to which centrality of neighboring accident root causes affects the centrality of a root cause.

Due to the exploratory nature of this study and the fact that previous data is not available on behavior of social network models of similar nature, estimating beta can be difficult. Accordingly, the value of beta used in UCINET is (0.995/largest eigenvalue) yielding a centrality which is almost identical to eigenvector centrality. The computation of Bonacich power centrality, however, has a twofold value:

1. By identifying value of beta corresponding to largest eigenvalues, the effect of being connected to a highly central root cause displayed by eigenvector centrality can be better understood. These values can also be guidelines (maximum values) for estimation of beta value in future research.
2. The method of normalization alpha makes it easier to identify the highly central causes from those with moderate to low centrality. This way, a cutoff can be identified for the number of high centrality causes which are most important to be addressed.

3.2.1.3 Centralization.

Centralization is a measure of variability in centrality of root causes. This means that a highly-centralized network of root causes refers to that there is a great variability in the centralities of root causes or that there are a few root causes which are most central in the network whereas others have much lower centralities. A centralization of zero mean that all root causes have equal centralities whereas a centralization of 1 means that the network has the maximum variability of centrality possible given its number of nodes. A high centralization is desirable in our case since it essentially means that there are a few root causes that if efficiently addressed will cause a very large impact in terms of reduction of network connectedness.

3.2.1.4 Transitivity

Transitivity answers a simple question relative to other metrics which is whether the root cause A cooccured with a root cause B and the cause B cooccured with the root cause C increases the probability that root cause A cooccured with root cause C. The answer to this question, whether yes or no, resembles an interesting behavior of the network that can help improve our understanding of root causes of fatal accidents and their relationships with each other.

3.2.2 Developing the Adjacency Matrix

Even though the bimodal network in Figures 3.2 and 3.3 contained all the information necessary to perform the social network analysis, having single fatality reports as nodes had no particular importance. This is true since identifying the root causes contributing to every single fatality had been done already in the data collection stage. Moreover, calculating social network analysis metrics for a single fatality was not of interest within the scope of this research study. This is because the research goals are directed towards analyzing the interrelationships between one cause and another rather than the relationships between one cause and one fatality.

On the other hand, having the network resembled by a bimodal or a bipartite graph added 100 nodes to the graph resembling the fatalities, which made the graph more difficult to visualize and understand. Computing social network analysis metrics for bimodal directed data is also more complex mathematically and the tools and metrics devised to deal with bipartite graphs are more limited compared to the more generic unimodal graphs (Borgatti and Everett 1997). This made having fatalities as nodes highly inconvenient, especially given that one of the goals of this Social Network Analysis model is to be highly extendable and repeatable. In this case, extending the model by adding more fatalities only added more nodes to the network which made the network even more complex and difficult to visualize and understand.

In cases where one of the modes dominates the other in terms of analytical interest, Borgatti (2009) suggested utilizing a unimodal approach when dealing with bimodal data. He recommended converting the bimodal directed binary graph into a unimodal weighted undirected graph. Considering there are two modes mode 1 and mode 2 such that the analytical interest lies in mode 1, the weight of the edge between two mode 1 nodes in the unimodal networks resembles the number of mode 2 nodes to which both nodes were connected in the bimodal network. Borgatti (2009) suggested this conversion is made using this simple equation:

$$A = BB^T$$

Where A is the unimodal adjacency matrix and B is the bimodal adjacency matrix with mode 1 as the rows of the matrix and mode 2 as the columns of the matrix. In this case, mode 1 is the root causes and mode 2 is the fatal accident case files. Accordingly, the weight of the edges between any given two root causes corresponds to the number of fatalities to which both root causes contributed together. Moreover, since the diagonal elements resemble an edge between the root cause and itself which is meaningless in the context of this analysis, the diagonal elements of the adjacency matrix were replaced by zeroes. An illustration of the matrix in Figure 3.2 and 3.3 after being converted is shown in Figures 3.4 and 3.5.

As we see from the illustrative example, only root cause 2 and root cause 66 have a higher edge weight than others where they both contributed to fatalities number 2 and 4. Each of the other two causes in this simple network contributed together to a single fatality (for example RC1 and RC5 contributed only to fatality number 5).

3.2.3 Network Decomposition according to Direct Causes

Fatal accident data collected has seven main direct causes contributing to the accidents. Four of the seven collected direct causes (fall, electrocutions, struck by, caught in between) are called the fatal four direct causes and they constitute 60.6% of construction fatalities in 2014 (Occupational safety and Health Administration 2017). Asphyxia, burn/explosion and health related were the other 3 direct causes recorded from the case files. Since one of the goals of this research is to identify the highest contributing root causes to each of fatality direct causes, the main network of root causes was decomposed or subdivided into 7 networks of root causes, such that each network

	RC1	RC2	RC3	RC4	RC5				RC66
RC1	0	0	0	1	1	.	.	.	0
RC2	0	0	1	1	1	.	.	.	2
RC3	0	1	0	1	0	.	.	.	0
RC4	1	1	1	0	0	.	.	.	0
RC5	1	1	0	0	0	.	.	.	1
.	.	.	.	.	.	.			.
.	.	.	.	.	.		.		.
.	.	.	.	.	.			.	.
RC66	0	2	0	0	1	.	.	.	0

Figure 3.4: Adjacency matrix for unimodal undirected weighted network

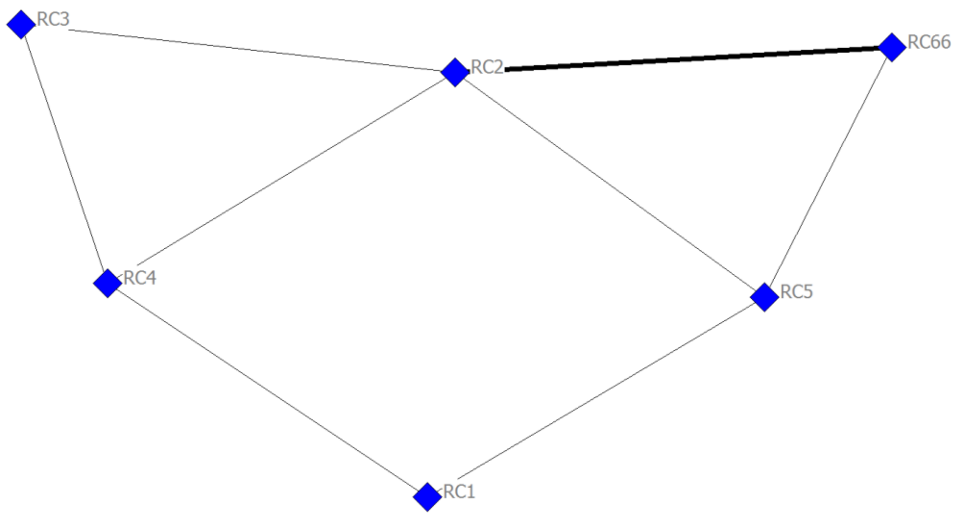

Figure 3.5: Social network diagram for unimodal undirected weighted network

resembles only the root causes which contribute to a particular direct cause and the interrelationships between them.

Isolating the subnetworks corresponding to the different root causes corresponding to each direct cause helps develop the analysis and focus on dynamics of accidents pertaining to a single particular direct cause (falls for example). This enables understanding how the root causes interact to lead to the occurrence of this particular direct cause. This is particularly important when trying to address this particular direct cause since understanding the root causes that lead to it can provide a valuable insight into how to avoid or mitigate this particular direct cause for accident occurrence.

3.3 Model application and result analysis

3.3.1 Data description

Data description is a key step of conducting a thorough analysis of the gathered fatality data. It involves two major steps, the first is categorizing the root causes obtained from the data gathering stage into subcategories. This is to be more able to recognize the accident dynamics, the interaction between these accident cause and the underlying factors which lead to these causes, which in turn lead to the accident occurrence. The second major step is providing simple statistics on the sampled fatalities including their direct cause distribution, location distribution, work site functions, whether tasks were regularly assigned and victim roles in the fatality. This provides a background on the nature of data being analyzed and an indication of the basic characteristics of the data being analyzed which affects the results of the social network analysis. This is particularly necessary such that the nature of data can be taken into account in future research when comparing results of this study with another conducted in a different area or for a different time span.

3.3.2 Implementation of exploratory SNA

The techniques utilized corresponding to social network analysis can be identified as exploratory research. Exploratory research refers to research techniques utilized in social sciences. Their purpose is to explore the social network behavior to discover emerging properties or patterns without bias or predispositions. This means that all SNA techniques and metrics will be used to test the network and then the results will be analyzed to gather as much data about the network properties as possible rather than having a fixed set of SNA metrics define the network. This approach is appropriate in this case since the nature of the network has a large degree of uncertainty given that no previous research has dealt with a similar form of networks.

The utilized social network analysis software to compute all Social Network Analysis (SNA) metrics and conduct the network analysis is known as UCINET for windows (Borgatti et al. 2002). UCINET is a social network analysis software package that includes many social network analysis tools usable to describe and characterize networks as a whole, or particular nodes within the networks (Borgatti et al. 2014). It also features some useful analytical techniques including some multivariate statistics and categorization of cohesive subgroups (Borgatti et al. 2014). Additionally, all of the graphs or network diagrams provided by this study are designed using NETDRAW which is a network diagramming software package which is associated with UCINET (Borgatti 2002).

The reason for utilizing UCINET is its ease of use, user-friendly interface, extensive documentation, wide compatibility in addition to its wide range of capabilities in terms of social network analysis tools and metrics. The data for all eight networks utilized was easily imported into UCINET as MS Excel spreadsheets and from there UCINET was utilized to compute all required social network analysis metrics and NETDRAW was utilized to generate the visual representations of the networks being investigated.

3.3.3 Analysis of results comparatively and statistically.

This section deals with the use of statistical tools and comparative techniques to compare between the metrics obtained for different categories of root causes, to compare between the results obtained from different centrality measures and analyze the differences between them, to compare between the networks contributing to different direct causes and the difference of network properties between them. Since the methods involved in comparison and the corresponding statistical techniques rely heavily on the nature of the networks, it will be covered more extensively in the results and analysis section. Statistical analysis was done using the software package (SAS 9.4®) for Windows.

3.3.4 Validation of findings

Different SNA metrics which were obtained in the results which rank the construction fatality root causes. Accordingly, the identified top ranked or key root causes are compared to the top construction fatality root causes which can be found in past studies from literature. Comparing the findings of this study with previous studies has two main benefits. The first is being able to validate the findings of this study by correlating them with past studies which utilized different methodologies. He second is identification of the potential contributions of SNA which enrich the analysis and make it deeper relative to other analytical techniques used in past studies. The validation of findings is further detailed in chapter 4.

CHAPTER 4:
RESULTS AND DISCUSSION

4.1 Descriptive Statistics

Descriptive Statistics are preliminary statistics aimed at describing and capturing basic features of the collected data. Accordingly, the following statistics have been obtained about the data analyzed:

1. Direct cause of the fatality

For the gathered data about fatal incidents, the fatal four (fall, electrocution, struck by and caught in between) had the largest contribution to fatalities with 53 fall accidents, 27 struck by accidents, 12 caught in between accidents and 9 electrocutions as shown in Figure 4.1. Burns or explosions and asphyxia were not encountered as often with 4 incidents for each. Additionally, 3 health related incidents were encountered. It should be noted that these do not add up to 100 case files because in some of the case files more than one direct cause was quoted as leading to the fatal incident.

2. Location of the fatality

Figure 4.2 shows the location distribution of the stratified random sample of 100 fatality reports collected. Through the 100 reports analyzed the states which had most fatalities was Texas with 10 cases followed by Florida and Mississippi with 9 cases each. The following was Georgia with 8 cases and Kansas with 7 cases. Colorado and Tennessee contributed by 6 cases each followed by the contribution of Ohio, Pennsylvania and Illinois each by 5 cases and Arkansas by 4 cases. The remaining states contributed by 3 or less cases as shown in Figure 4.2.

3. Work site by end use function

Figure 4.3 shows the work site end uses which had the greatest contribution to fatalities within the 100 case files analyzed coded according to the classification provided in section 3.1.2. The sites which contributed most to fatalities in our gathered data were commercial buildings and single-family houses contributing to 19 fatalities each. Highways and street related work contributed to 9 fatalities. Buildings which were not categorized under any of the given categories (other buildings) contributed to 7 fatalities. Amusement, social and recreational buildings, power and communication transmission lines, towers and other facilities and sewage treatment plants contributed by 5 fatalities for each category. Sewers, sewer lines, septic tanks and related facilities contributed by 4 fatalities and the remaining categories each contributed by 3 or less fatalities.

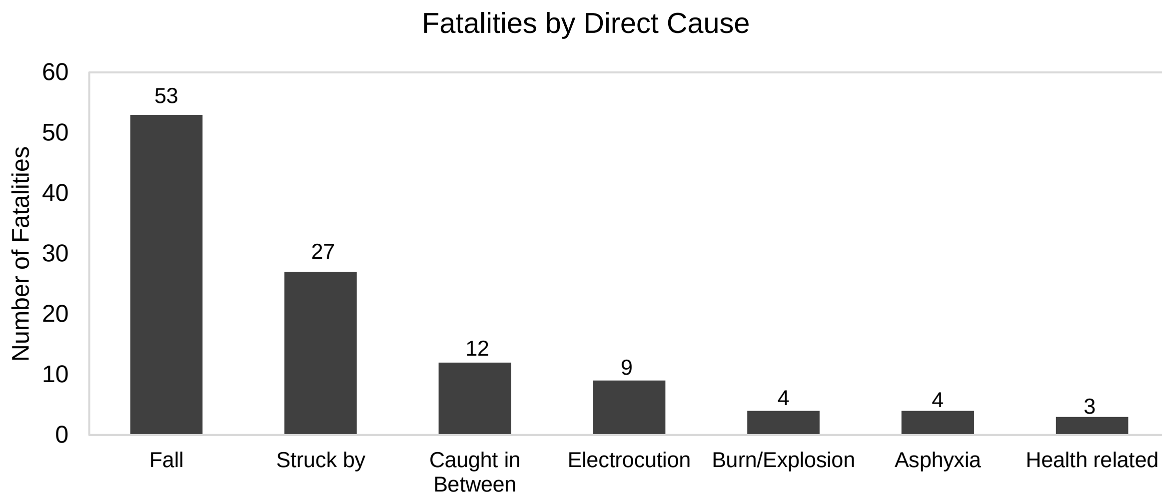

Figure 4.1: Fatality distribution by direct cause

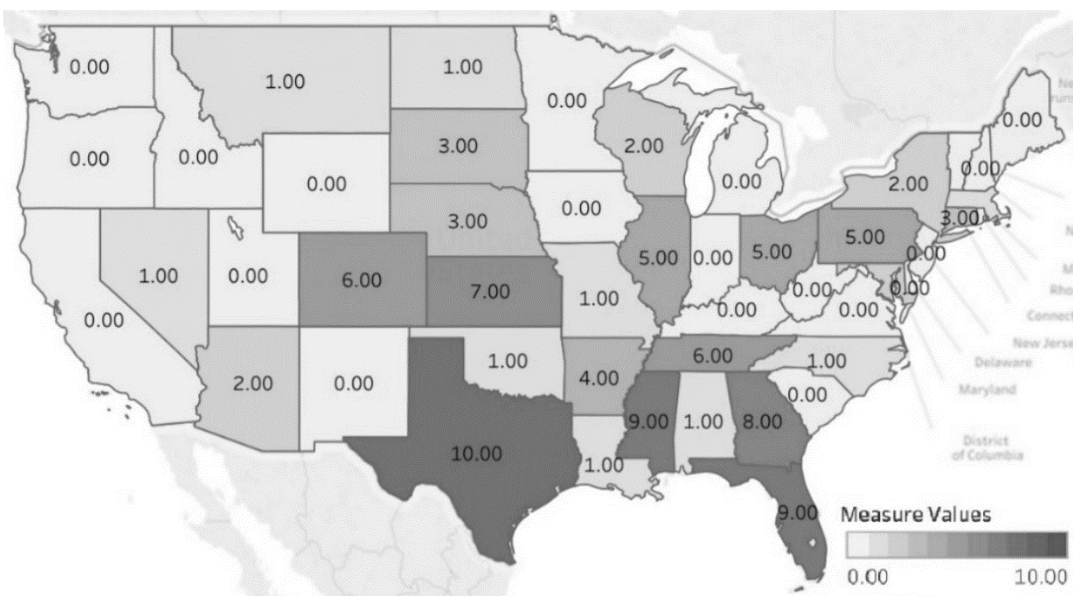

Figure 4.2: Location distribution of the fatalities in the gathered data

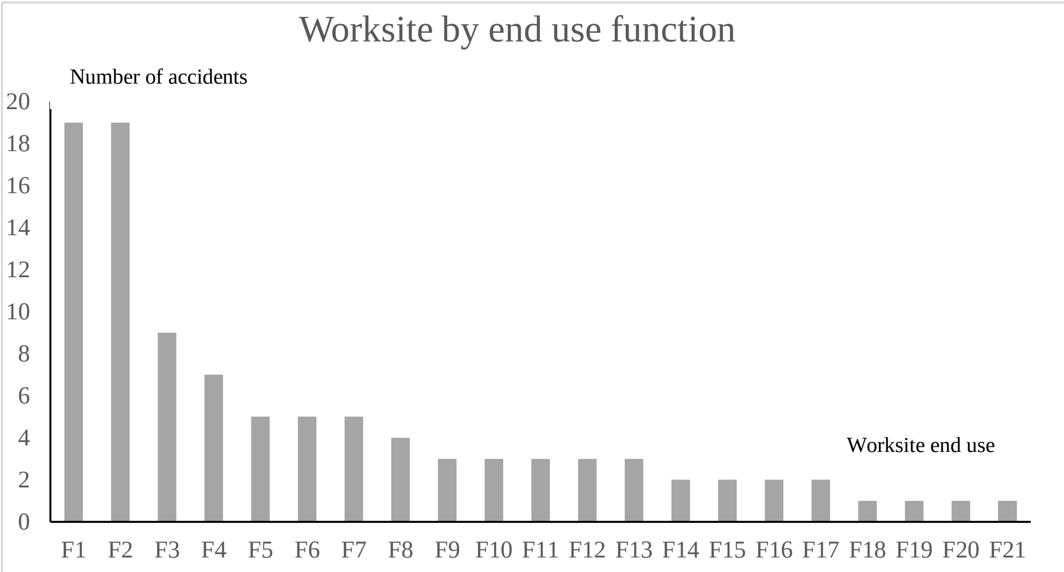

Figure 4.3: Work site by end use function for the gathered data

4. Task Regularly Assigned.

In 88 of the 100 cases analyzed, the task done by the victim at the time the fatality occurred was his regularly assigned task as shown in Figure 4.4. In only 9 instances the task being performed by the victim was different from the task he is assigned regularly and has experience with. In 3 instances whether the task was the regularly assigned task of the victim was not reported and could not be inferred from the context of the job description or the accident description in the relevant case files.

5. Victim's role in the fatal accident.

In 48 of the 100 case files analysed was the victim's own actions the main factor that initiated the fatal accidents as shown in Figure 4.5. In 22 of the cases it was his actions combined with actions of another worker or workers which lead to the accident. In 13 percent of the cases, the victim happened to be in the wrong place at the wrong time where the accident couldn't be attributed to his actions or those of other workers but was more due to being at the wrong place at the wrong time. Only 7 of the accidents were due solely to actions by workers other than the victim. 10 of the cases could not be classified under any of these categories. It should be noted that this categorization is inferential in nature and involves a lot of deduction from the description of the accident and the records of the investigations carried out which might involve some informed subjective assessment of the situation being investigated which lead to the fatal accident.

4.2 Root Cause Categories

A list of 66 root causes are found to be those leading to the 100 fatalities investigated. To be more able to recognize the accident dynamics, the interaction between these accident cause and the underlying factors which lead to these causes are to be further studied. The 66 root causes, thus, were divided into 12 main categories. The categories of root causes, the corresponding root causes and the node number corresponding to every root cause are provided in the Table 4.1.

4.3 Root Cause Pairings

Prior to conducing the network analysis and obtaining relevant SNA metrics, one important finding can be directly extracted from the adjacency matrix. This finding is which pairs of root causes have the strongest ties between them which reflects on largest edge weights. The strength of their ties means that these two root causes, when occurring together, contributed to the largest number of

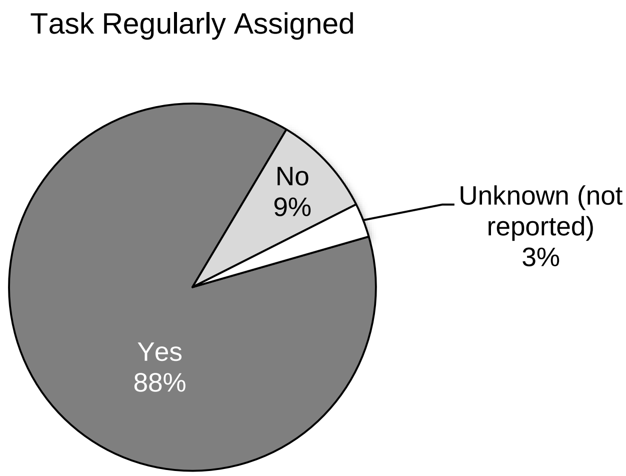

Figure 4.4: Whether task is regularly assigned in the gathered data

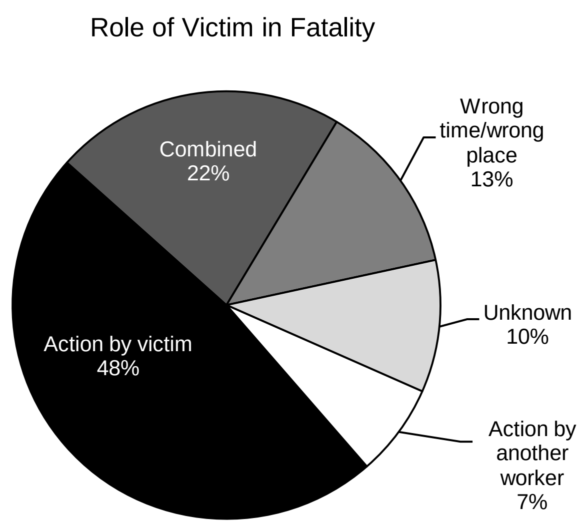

Figure 4.5: Role of the victim in the fatalities in the gathered data

Table 4.1: Root cause categories

Node	Equipment related root cause
1	Damaged/defective equipment
2	Lack of necessary Equipment
3	Poor Equipment handling
4	Use of non-suitable equipment
5	Equipment malfunction
6	Equipment/spoil on excavation edge
7	Vibration in excavation
8	Equipment tipping
Node	Safety equipment related root causes
9	Poor Labeling
10	Improper use of PPE
11	No PPE
12	No fall arrest system, guardrails or safety nets.
13	No cable insulation
14	Not wearing seatbelt
15	Safety element failure
16	No/damaged Cave in protection
17	No protection from traffic
18	Inappropriate decking
Node	Communication Related Root Causes
19	Lack of employee knowledge
20	Lack of coordination of site activities
Node	Work tools & material Related Causes
21	Poor tool handling
22	Inappropriate tools used
23	Poor Material Handling
24	Poor storage
25	High exposure to chemical
26	Wet Material
Node	Employer Related Causes
27	Employer gross negligence
28	Employer allowed employee to work in an unsafe environment
29	Lack of knowledge by employer about site conditions
30	Lack of clear employer instructions
31	Failure to properly locate Utilities
32	Lack of preventive action

Table 4.1: Continued

Node	Personnel Related Causes
33	No competent person on site
34	No first aid personnel
35	Lack of supervision
36	Vehicle observer error
37	Operation carried out by noncompetent individual
38	Absence of necessary personnel
Node	Employee related causes
39	Employee misconduct
40	Willfully exposing self to hazardous situation
41	Misjudgment of hazardous situation
42	Heart Attack
Node	Training related root causes
43	Lack of specific on the job training
44	Lack of general health and safety training
Node	Inspection related
45	Inspection related
46	Lack of Inspection for equipment and tools
Node	Structural related Causes
47	Poor Assembling of equipment/scaffold/decking/formwork
48	Error in design
49	Failure of structural element
50	Collapse of structure
51	Over excavating
Node	Site related
52	No safe access to site/scaffold/trench
53	Poor housekeeping
54	No safe exit to site
55	Working surface condition not suited to task
56	Site obstruction
57	No safe walkways
58	No site survey
59	Inappropriate lighting
Node	Procedure related
60	Not following proper work procedures
61	No hazard identification/ communication program
62	Not following proper procedure for operating equipment
63	No testing procedure for equipment
64	Noncompliance to equipment manufacturer specifications and recommendations
65	No effective emergency plan
66	Lack of safe working procedures

accidents. Accordingly, it means that if these two root causes are found together on a site, this increases the likelihood that a fatal accident occurs. Accordingly, knowing these pairings can help provide early warning of accident occurrence in case certain combinations of accident root causes are found on site. Four key pairs of root causes were identified.

1. “Lack of specific on the job training” and “No fall arrest system, guardrails or safety nets”: These two causes are connected to “fall” accidents and their presence together causes 23 of the 100 fatalities analyzed. These cases were mostly small contractors specialized in roofing who hire workmanship who aren’t adequately trained to complete their jobs properly both in terms of safety and effectiveness. This absence of training on their job increased the probability of falling while carrying out the job and the fact that there is no fall arrest, safety nets or guardrails meant that many of the falls would be fatal.

2. “No PPE” and “No fall arrest system, guardrails or safety nets”: These two root causes are also connected to “fall” direct cause, which is the most common direct cause for fatalities. This pair is connected to 20 of the 100 accidents which were analyzed. This pair simply shows the probability of having a fatal accident is amplified by absence of safety equipment for both site safety and personal protection.

3. “No PPE” and “Lack of specific on the job training”: These two causes indicate a property evident in workers who are untrained or new to the work, which is that they are mostly more prone to accidents. These two causes combined lead to 17 of the 100 investigated fatalities. It calls that enforcing the use of PPE on site very strictly by supervisors is a must particularly for untrained workers whether they are undergoing training or new workers who have been assumed to be already appropriately trained which might not be true.

4. “Poor labeling” and “Lack of specific on the job training”: This pair of causes has led to 10 of the 100 accidents analyzed. It shows that there is a correlation between training on the job and the ability of workers to identify potential risks and hazards on site. This interaction can be explained by that the ability of untrained workers to identify unsafe or hazardous conditions on site is lower than that of trained workers. This necessitates labelling these conditions very clearly for them since the probability that they identify these conditions by themselves is much lower.

4.4 Exploratory Social Network Analysis

The representation is of the full network of the 66 root causes discussed in the methodology section is shown in Figure 4.6. They are modelled as nodes and 1052 ties exist between them. The locations of these nodes are identified by an iterative graph theoretic algorithm which defines the initial positions of networks based on a technique known as Gower multidimensional scaling. This results into distances between nodes in the graph being proportional to how similar they are in many ways such as the geodesics they have to all other nodes (Hanneman and Riddle 2005). This initial condition is followed by utilizing a spring algorithm which iteratively moves nodes such that nodes with shortest path length from each other are closest together (Hanneman and Riddle 2005). The algorithm moves the nodes and assesses based on a criterion of fit badness which aims to avoid nodes being very close together and improve readability. There is an option in UCINET for node repulsion where nodes repel each other to make the network more readable (Hanneman and Riddle 2005). This graph then combines elements of multidimensional scaling and spring algorithms where distances between nodes are interpretable in terms of their cooccurrence while preserving readability. This same algorithm is utilized for plotting all subnetworks presented in the following section.

At the first look to Figure 4.6, the central area has very dense links. On the other hand, the peripheral is much less dense in terms of links. Therefore, the causes which are located in the central area (such as RC 11: No PPE) are expected to be of highest centrality and greatest structural role in the network, where they form hubs to which many other causes are connected with cooccurrence links. Nodes at the periphery (like RC 25 high exposure to chemical), on the other hand, have a much less central role in terms of cooccurrence with other root causes and addressing them (though essential in many cases) won't affect other root causes much. Note that the weights which are usually reflected in the line weights of the links are not reflected in the line weights in the network as adding line weights to the network will increase its complexity and will be detrimental to its readability.

4.5 Social Network Analysis Metrics for the Main Network

4.5.1 Network Cohesiveness Measures

- Diameter

The network has a diameter of 4 geodesics and an average of 1.87 geodesics between any two

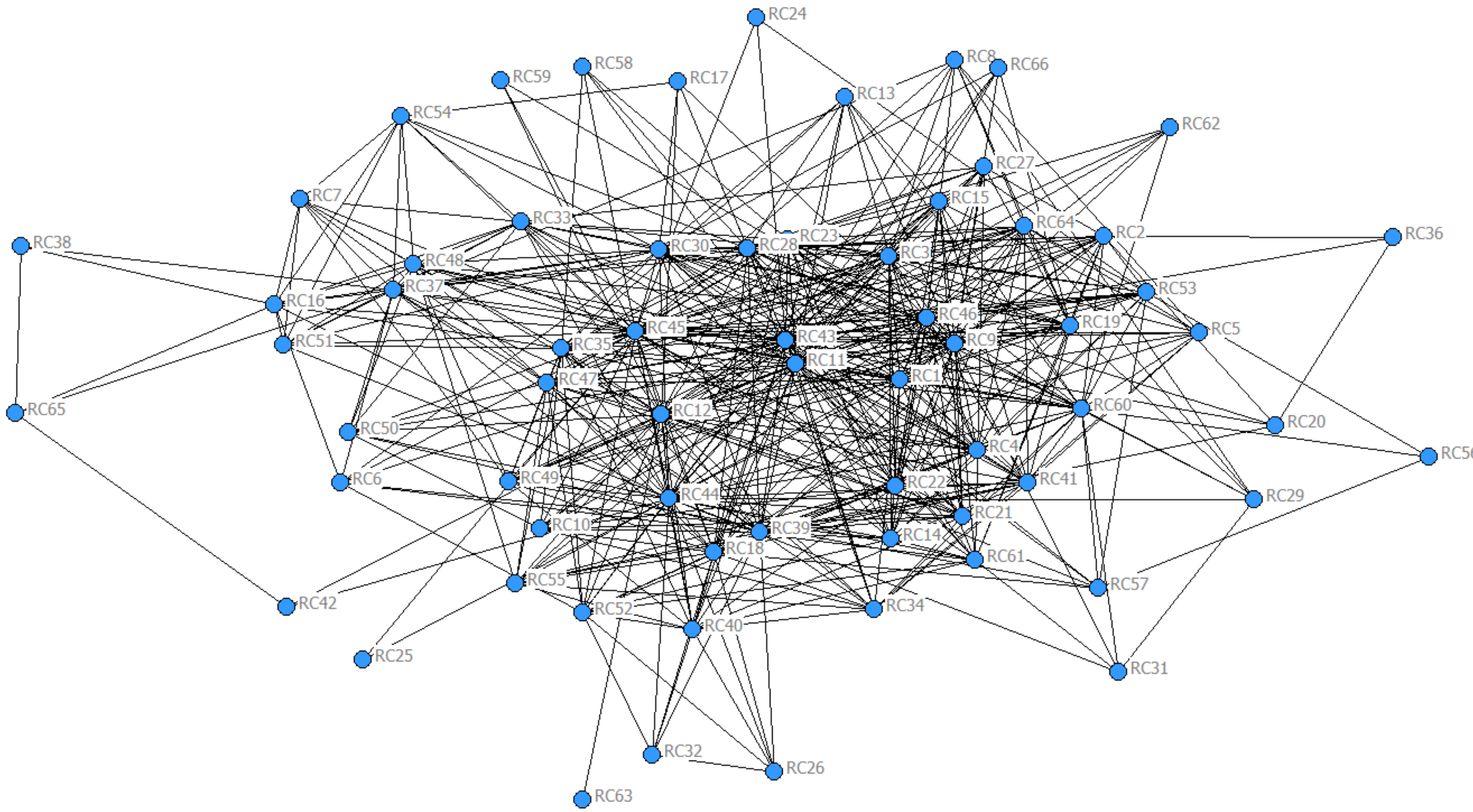

Figure 4.6: Cooccurrence network of the 66 root causes.

given nodes. This indicates that the network is very compact where the furthest root causes from each other have 4 cooccurrence links separating them (A cooccurred with B which cooccurred with C which cooccurred with D which cooccurred with E). This is a relatively short geodesic which shows that root causes for accidents are closely related and likely to occur together.

- Density

The unweighted density of this network is 0.245 which means that the number of edges in the network is approximately quarter all the possible edges in the network where the edges in the network are 1052 whereas the maximum possible number of edges is 66 * 65 which is equal to 4290. The density here resembles that among every 4 links which can occur in the network, one link is present where the presence of a link means that the two causes cooccur at least once. This value resembles a high connectedness amongst the root causes where a density of 1 means that every root cause cooccurs with all other 65 root causes at least once whereas the density of 0.245 means that each root cause on average cooccurs with quarter of the root causes which is about 16 other root causes at least once. This large percentage of cooccurrence between the root causes is shown by the dense structure of the network and the large number of edges connected to each node can be seen in Figure 4.6.

On the other hand, the weighted density of the network is equal to 0.464 which is the average value of all weights of cooccurrence ties (total cooccurrences between all pairs of causes) divided by the maximum number of possible edges so it is the average number of cooccurrences between any two given accident root causes taking the zero cooccurrences into account. Since from the unweighted density it is known that 75.5 % of the edges are of zero weight, this is what drives the average down since the sum of the weights of the weighted ties are divided by approximately four times their number, which is the total number of possible ties.

- Transitivity:

If every combination of 3 root causes in the network are examined such that root cause A is connected to root B and root cause B is connected to root cause C, transitivity examines the implication of these two connections on the probability of cooccurrence between root cause A and root cause C at least once. When full network is examined, 7705 subsets of three root causes are identified with at least two cooccurrence links between them. Among these, only 1938 three root cause subsets had the third link present. This corresponds to a transitivity value equal to 1938 divided by 7705 which is equal to 25.152 %. This value indicates the behavior of the network at a localized rather than a global level and indicates that the probability of A cooccurs with C given that A cooccurs with B and B cooccurs with C is 25.152 %. The unweighted density for this network is 0.245 so the probability of any two nodes A and C being connected regardless of the connection with B is 24.5%. This test shows that transitivity is a property which has no effect in our network and that the network data shown exhibits little to no transitivity.

4.5.2 Network Centrality Measures

The Table 4.2 shows the computed values for centralities of all 66 root causes of accidents identified from the fatality report case files. The following section is directed towards analyzing the top 15 root causes according to each centrality measure and the different patterns observed from utilizing different centrality measures. The reason why the top 15 centrality root causes were analyzed rather than any given number is derived from the Bonacich Power Centrality. According to Bonacich (1987), the scaling parameter α used to normalize the centralities is selected such that the centrality vector's squared length is equivalent to the network's size or the number of nodes. Accordingly, Bonacich (1987) asserts that a normalized beta centrality of 1 is neither high or low. As an indicator, this is used to assess how many root causes have a high centrality. Accordingly, the number of root causes with beta centrality greater than one is counted and amounts to the 15 root causes mentioned.

4.5.2.1 Degree Centrality

As discussed in the methodology section, the weighted degree centrality was the most appropriate form of degree centrality for our study. It indicates the number of cooccurrence links in the network which will no longer be present if we manage to address the root cause and thus make it no longer part of the network. In this sense and in the pre-mentioned context of accident causation theories, the reduction in network cooccurrence link connectedness corresponds to a reduction in the accident occurrence probability but not necessarily in the same proportion. The 15 root causes corresponding to the highest weighted degree centrality are shown in Table 4.3.

A closer look at the most central root causes in terms of cooccurrence with other root causes leads to a realization that the very basic elements of an appropriate safety management in terms of organization and enforcement are the leading cause of fatal accident. All of these causes resemble fatal mistakes or major failures of safety management systems which lead to the fatal incidents.

Underlying factors which can be inferred from this set of root causes include that firms do

Table 4.2: All centrality measures for all 66 root causes in the main network

	Degree	PctDegree	FlowBet	nFlowBet	Eigenvector	Beta Cent	Normalized Beta Cent
RC1	54	2.71%	66.3699	1.5954%	0.132	9836.686	1.072716594
RC2	21	1.05%	82.50172	1.9832%	0.047	3529.623	0.384914726
RC3	56	2.81%	87.55089	2.1046%	0.135	10041.71	1.095075488
RC4	33	1.66%	28.45384	0.6840%	0.085	6322.894	0.689528286
RC5	13	0.65%	17.0723	0.4104%	0.025	1899.395	0.207134008
RC6	9	0.45%	14.72819	0.3540%	0.017	1243.262	0.135580987
RC7	11	0.55%	26.24917	0.6310%	0.013	943.1992	0.102858372
RC8	13	0.65%	31.83013	0.7651%	0.025	1892.801	0.206414998
RC9	67	3.36%	143.1775	3.4418%	0.206	15349.33	1.673884869
RC10	15	0.75%	18.8364	0.4528%	0.048	3554.897	0.387670904
RC11	122	6.12%	131.2543	3.1552%	0.378	28165.79	3.071554661
RC12	144	7.23%	348.4578	8.3764%	0.431	32079.86	3.498394012
RC13	13	0.65%	5.866598	0.1410%	0.037	2757.726	0.300737292
RC14	19	0.95%	27.71196	0.6662%	0.04	2967.142	0.323574662
RC15	29	1.46%	25.46569	0.6122%	0.103	7674.072	0.836877942
RC16	26	1.31%	134.7582	3.2394%	0.03	2261.711	0.24664554
RC17	7	0.35%	6.334481	0.1523%	0.025	1841.07	0.200773522
RC18	42	2.11%	76.88148	1.8481%	0.137	10166.55	1.108689666
RC19	30	1.51%	61.34781	1.4747%	0.076	5652.821	0.616455078
RC20	5	0.25%	45.0303	1.0825%	0.012	865.9555	0.094434746
RC21	38	1.91%	31.20873	0.7502%	0.114	8491.055	0.925971985
RC22	29	1.46%	27.4324	0.6594%	0.082	6104.831	0.665747941
RC23	38	1.91%	50.36981	1.2108%	0.109	8125.384	0.886094689
RC24	3	0.15%	3.675333	0.0883%	0.01	738.3195	0.08051569
RC25	2	0.10%	0.168675	0.0041%	0.006	414.0977	0.045158446
RC26	6	0.30%	22.41305	0.5388%	0.009	696.1111	0.075912766
RC27	15	0.75%	16.81164	0.4041%	0.04	2966.069	0.323457718
RC28	55	2.76%	109.4402	2.6308%	0.148	11005	1.200124502
RC29	7	0.35%	28.09683	0.6754%	0.01	719.246	0.078435689
RC30	59	2.96%	108.2211	2.6015%	0.125	9319.025	1.016264439
RC31	6	0.30%	20.02215	0.4813%	0.011	807.8771	0.088101134
RC32	6	0.30%	22.41305	0.5388%	0.009	696.1111	0.075912766
RC33	23	1.15%	47.98152	1.1534%	0.048	3596.043	0.392158002
RC34	18	0.90%	20.23281	0.4864%	0.055	4087.559	0.445759088
RC35	46	2.31%	81.08424	1.9491%	0.124	9276.093	1.011582494
RC36	3	0.15%	25.81125	0.6205%	0.003	211.3229	0.023045326
RC37	25	1.26%	134.8171	3.2408%	0.04	2982.456	0.325244725
RC38	3	0.15%	32.50308	0.7813%	0.001	76.27709	0.008318219
RC39	56	2.81%	110.7427	2.6621%	0.157	11719.41	1.27803266
RC40	30	1.51%	57.2335	1.3758%	0.091	6776.21	0.738963664

Table 4.2: Continued

	Degree	PctDegree	FlowBet	nFlowBet	Eigenvector	Beta Cent	Normalized Beta Cent
RC41	39	1.96%	38.3119	0.9210%	0.101	7520.349	0.820114017
RC42	3	0.15%	28.61037	0.6877%	0.009	663.5998	0.072367318
RC43	208	10.44%	354.225	8.5150%	0.488	36409.97	3.970603943
RC44	83	4.17%	256.6603	6.1697%	0.213	15888.19	1.73264873
RC45	83	4.17%	308.4395	7.4144%	0.189	14089.22	1.536466837
RC46	55	2.76%	127.449	3.0637%	0.144	10758.55	1.17324841
RC47	49	2.46%	52.40611	1.2598%	0.155	11557.21	1.26034379
RC48	23	1.15%	56.51883	1.3586%	0.037	2745.143	0.299365103
RC49	23	1.15%	29.77032	0.7156%	0.07	5229.737	0.570316672
RC50	10	0.50%	13.26806	0.3189%	0.023	1747.314	0.190549225
RC51	11	0.55%	25.42696	0.6112%	0.012	896.9952	0.097819708
RC52	24	1.20%	59.85803	1.4389%	0.067	4954.597	0.540311873
RC53	19	0.95%	106.9156	2.5701%	0.04	2990.48	0.32611981
RC54	14	0.70%	44.45964	1.0687%	0.014	1087.567	0.118602052
RC55	22	1.10%	12.58915	0.3026%	0.087	6457.41	0.704197705
RC56	3	0.15%	16.18764	0.3891%	0.002	174.3487	0.019013191
RC57	8	0.40%	48.89073	1.1753%	0.023	1698.764	0.185254708
RC58	5	0.25%	2.016086	0.0485%	0.015	1089.712	0.118835948
RC59	3	0.15%	0.259625	0.0062%	0.011	848.69	0.092551894
RC60	46	2.31%	155.4149	3.7359%	0.104	7775.253	0.847912014
RC61	16	0.80%	45.33166	1.0897%	0.036	2711.112	0.295653939
RC62	7	0.35%	2.601422	0.0625%	0.017	1293.747	0.141086563
RC63	1	0.05%	0	0.0000%	0.006	442	0.04820127
RC64	28	1.41%	64.58681	1.5526%	0.06	4444.971	0.484735847
RC65	4	0.20%	86.00307	2.0674%	0.001	86.2623	0.009407132
RC66	8	0.40%	2.274534	0.0547%	0.028	2079.219	0.226744398

Table 4.3: Top fatal accident root causes according to degree centrality

Node	Root Cause	Degree	PctDegree
RC43	Lack of specific on the job training	208	12.68%
RC12	No fall arrest system, guardrails or safety nets	144	8.78%
RC11	No PPE	122	7.44%
RC44	Lack of general health and safety Training	83	5.06%
RC45	No jobsite inspection	83	5.06%
RC9	Poor Labeling	67	4.09%
RC30	Lack of clear employer instructions	59	3.60%
RC3	Poor Equipment handling	56	3.41%
RC39	Employee misconduct	56	3.41%
RC28	Employer allowed employee to work in an unsafe environment	55	3.35%
RC46	Lack of Inspection for equipment and tools	55	3.35%
RC1	Damaged/defective equipment	54	3.29%
RC47	Poor Assembling of equipment/scaffold/decking/formwork	49	2.99%
RC35	Lack of supervision	46	2.80%
RC60	Not following proper work procedures	46	2.80%

not value and invest in their employees inferred from their lack of investment in training them as well as making them work in unsafe environments. The enforcement of the safety rules seems to be an issue that needs addressing too. Not abiding by wearing the PPEs, not handling construction equipment in the adequate way, poor worker conduct, lack of competence when completing crucial jobs like decks, scaffolds or formworks, absence of adequate staff supervising workers and not abiding by the proper procedures for completing work all provide evidence of that the industry needs better communication and enforcement of safety regulations.

A key element evident to be a major contributor to fatality occurrence is reduced safety spending and cutting down on safety budget. This can be inferred from the lack of training provided whether it is safety and on the job training, not investing in fall protection equipment, not investing in maintenance and inspection for equipment and tools in Table 4.3. This can be due to an urge to reduce prices by cutting down on safety budget, compromising safety and taking some risks that can lead to fatalities.

Another area is the area of communication whether this communication is visual or verbal. The fact that no clear instructions are given to workers and that not enough labelling is done for site hazards or dangerous tools and equipment means that there is a gap in the communication links on site.

Moreover, the deficiency of basic procedural site operations seems to be one of the underlying causes of fatalities where they are evidence of a weak safety management system. Worksite inspection, equipment and tool testing and maintenance as well as availability of supervision are all very basic procedures whose absence indicates serious operational inefficiencies.

These top 15 root causes also signify the importance of safety culture to be imposed among the workers and to be strongly supported by all levels of management. The causes point at the poor culture of safety for both the employer and workers where they indicate the lack of hazard awareness and belief in the importance of safety and that it cannot be compromised. From employer responsibility perspective, poor safety culture is evident in absence of worker training, proper site safety equipment, well defined instructions and allowing workers into dangerous site conditions. This all proves the employers lack safety culture and only care about monetary savings.

Poor culture of safety is also to blame on workers as safety culture does not seem to be part of their mindset. That can be attributed to working with employers who fail to emphasize the importance of safety and nourish this culture or simply due to workers' background in terms of common knowledge, ancestry and beliefs. Again, the mindset of workers, amongst other human-related factors that lead to accidents, can be attributed to not having the knowledge simply by not being given sufficient induction or training, the safety management system not being strongly imposed in workers or serious behavioral issues by the workers in which case they might not be good enough to be selected for the job. Workers being reluctant to wear PPE or taking them off, equipment mishandling by operators who might be informed but willing to take the risk leading to accidents, not following orders, or poor conduct leading to dangerous situations, not carrying out work according to proper procedures and not giving extra care to jobs upon safety of others depend like scaffolding are all signs that workers have very poor culture and attitude towards safety.

4.5.2.2 Flow Betweenness Centrality

Flow betweenness is a form of betweenness centrality that treats the network edges as channels through which the cooccurrence flows. The analogy serves to identify the root causes which bridge across the most structural gaps in the network meaning that they cooccur with the root causes which are normally unrelated and act as a hub to which many normally unrelated root causes are connected. The 15 root causes corresponding to the highest flow betweenness centrality are shown in Table 4.4.

Comparison between degree centrality and flow betweenness centrality results:

Even though betweennness centrality measures the tendency to bridge structural gaps between unrelated root causes and weighted degree is the number of cooccurrences of the root cause with other causes the results suggest that both centralities yield results which are close to each other with twelve of the fifteen top root causes common between the two centralities even though the ordering of the root causes is significantly altered as shown in Table 4.4. The two root causes with highest centrality as well as the fourth highest and the sixth highest are the same for both centrality metrics. Three root causes increased in centrality which are lack of jobsite inspection which increased from rank 5 to rank 3, not following the proper work procedure which increased from rank 15 to rank 5 and lack of inspection for equipment and tools which increased from rank 11 to rank 10. This highlights the betweenness of basic procedural site operations and its importance in this network bridging between causes which are usually unrelated. For example, the lack of regular inspection for equipment or jobsite can contribute to fall due to slipping or tripping and can contribute to being struck by the equipment or loose material on site. In each of these two accident

Table 4.4:Top fatal accident root causes according to flow betweenness centrality

Node	Root Cause	FlowBet	nFlowBet
RC43	Lack of specific on the job training	354.225	8.5150%
RC12	No fall arrest system, guardrails or safety nets	348.4578	8.3764%
RC45	No jobsite inspection	308.4395	7.4144%
RC44	Lack of general health and safety Training	256.6603	6.1697%
RC60	Not following proper work procedures	155.4149	3.7359%
RC9	Poor Labeling	143.1775	3.4418%
RC37	Operation carried out by noncompetent individual	134.8171	3.2408%
RC16	No/damaged Cave in protection	134.7582	3.2394%
RC11	No PPE	131.2543	3.1552%
RC46	Lack of Inspection for equipment and tools	127.449	3.0637%
RC39	Employee misconduct	110.7427	2.6621%
RC28	Employer allowed employee to work in an unsafe environment	109.4402	2.6308%
RC30	Lack of clear employer instructions	108.2211	2.6015%
RC53	Poor housekeeping	106.9156	2.5701%
RC3	Poor Equipment handling	87.55089	2.1046%

scenarios, there will be totally different root causes that contribute to the accident and many of them might be unrelated but they all cooccur with lack of inspection which results in its high betweenness.

The root causes which were lower in betweenness than in degree are not using PPE which decreased from rank 3 to rank 9, misconduct of employees which decreased from rank 9 to rank 11, the employer allowing his workers to work in dangerous or unfavorable site conditions which moved to rank 12 from 10, unclear employer instructions which decreased from rank 7 to rank 13 and finally poor equipment handling which went from rank 8 to rank 15. This indicates that even though these causes play an essential role in network connectedness, their value when focusing on bridging across structural holes and cooccurring with different other root causes in different accident scenarios is reduced. The three root causes which were among the top 15 in the flow betweenness, but not weighted degree include operations on site being carried out by a non-competent individual in the 7^{th} rank, having no or damaged excavation cave in protection in the 8^{th} rank and poor housekeeping of site in the 15th rank. The reason why operation being carried out by a non-competent individual has much higher betweenness than weighted degree is that while it might have not occurred in many accidents, the results of carrying out work by non-competent individuals is not expectable and can lead to multiple catastrophic accident scenarios depending on the work being carried out. As for the cave in protection, cave in accidents have multiple causes like poor design, over excavating, vibrations, load on excavation edge and other causes which can all be avoided if proper cave in protection is put in place therefore lack cave in protection is the common cause cooccurring with all different causes of cave in fatalities and bridging the cooccurrence gap between them. Finally, poor housekeeping is expected to be of large betweenness as there are many potential hazards that can result in many unrelated accident scenarios which result only from poor housekeeping of the site.

Key trends that differentiate betweenness centrality from weighted degree include the importance of site operations being carried out in the correct safe way where noncompetent individuals, lack of on the job training and not following proper work procedures are all related to carrying out the task correctly and all have large betweenness centralities. Additionally, site condition seems to gain more importance in betweenness than in degree centrality where jobsite inspection and housekeeping gained significant increase in their centralities.

4.5.2.3 Eigenvector Centrality

Eigenvector centrality is the type of centrality that considers the cooccurrence of the root cause with the other causes, the number of times the cooccurrence was repeated as well as the centrality of the adjacent causes which cooccur with the cause being investigated. The innate assumption in eigenvector centrality is that causes which cooccur frequently with highly central root causes gain more centrality. The 15 root causes corresponding to the highest eigenvector centrality are shown in Table 4.5.

The eigenvector centrality yielded centrality results which are very similar to the weighted degree centrality even though it took the centralities of the adjacent root causes into account when calculating the centrality of the root causes as shown in Table 4.5. The resulting top 15 root causes have 14 which are the same as in the case of weighted degree centrality and one which is new. The similarity between the two centrality metric results suggests a property of the data that highly central root causes tend to cooccur with other highly central root causes. This pattern occurs such that including the centrality of adjacent root causes has a small effect on the centrality distribution in the network so there is no significant change in the most central root causes when using eigenvector centrality or weighted degree.

The first four most central root causes did not change. There were small increases in centrality of poor labelling (rank 6 to 5), employee misconduct (rank 8 to 7), employer allowing his workers to work in dangerous or unfavorable site conditions (rank 10 to 9) and inadequate inspection for tools and equipment (rank 11 to 10). Poor assembling of equipment, decking, formworks or scaffolds (rank 13 to 8) was the only cause with a significant increase in centrality rank between weighted degree and eigenvector centrality which can be interpreted as that this cause tends to cooccur with highly central root causes. This is because the only information considered by eigenvector centrality and not by weighted degree is centrality of the causes that the investigated root cause cooccurs with.

No decking was ranked 13 in this case whereas in the case of degree centrality it was ranked 16. This is not a large increase in centrality ranking but it means that the root cause tends to cooccur with higher centrality causes as well.

No jobsite inspection (rank 5 to 6), damaged or defective equipment (rank 12 to 13) and lack of supervision (rank 14 to 15) had their centrality ranking slightly reduced. Absence of clear instructions by the employer (rank 7 to 14) and poor equipment handling (rank 9 to 12) had their

Table 4.5: Top fatal accident root causes according to eigenvector centrality

Node	Root Cause	Eigenvector
RC43	Lack of specific on the job training	0.488
RC12	No fall arrest system, guardrails or safety nets	0.431
RC11	No PPE	0.378
RC44	Lack of general health and safety Training	0.213
RC9	Poor Labeling	0.206
RC45	No jobsite inspection	0.189
RC39	Employee misconduct	0.157
RC47	Poor Assembling of equipment/scaffold/decking/formwork	0.155
RC28	Employer allowed employee to work in an unsafe environment	0.148
RC46	Lack of Inspection for equipment and tools	0.144
RC18	No decking	0.137
RC3	Poor Equipment handling	0.135
RC1	Damaged/defective equipment	0.132
RC30	Lack of clear employer instructions	0.125
RC35	Lack of supervision	0.124

centrality ranking significantly reduced which means that they happen to cooccur with less central root causes.

The eigenvector centrality is useful in the case of comparison with weighted degree to assess whether a root cause frequently cooccurs with others with high centrality. Whether the centrality of the root cause that a given root cause happens to cooccur with frequently should be considered when judging the centrality of this root cause is not a property of the network which should be assumed. This is why the information given by eigenvector centrality can only be utilized in cases where there is either theoretical, numerical or empirical evidence of this network property. Therefore, the results for eigenvector centrality require validation of the innate assumption that the centrality of a root cause is directly proportional to the centrality of other causes it cooccurs with. This assumption, though might be valid, undermines the eigenvector centrality as a measure.

4.5.2.4 Bonacich Power (Beta) Centrality

Bonacich power or beta centrality is more generic than the eigenvector centrality in the sense that it takes into account the root cause's degree and tie weights as well as the centrality of the neighboring root causes. This is possible through the coefficient β which could be adjusted according to the required effect to be positive or negative or zero. In this case, UCINET computes β as (0.995/largest eigenvalue) making the results identical to eigenvector centrality.

For the main network of cooccurrence between all root causes, the computed β value is equal to 0.0137469. A zero value of beta is the equivalent of degree centrality and a beta value which is equal to the reciprocal of the largest eigenvalue of the adjacency matrix is the equivalent to eigenvector centrality. The computed beta from the data is very close to the reciprocal of the largest eigenvalue of the adjacency matrix which implies that Bonacich power centrality gives almost the same results as eigenvector centrality. However, the beta value is considered very small for both Bonacich power and eigenvector centralities which shows that even though the data implies a positive relationship between the centrality of neighboring nodes and the centrality of the node being investigated, this relation factors in with a very small factor showing that it is a weak relationship. Table 4.6 shows the computed values for Bonacich Power centrality and the top 15 root causes in terms of Bonacich power centrality.

Table 4.6 shows that the ranking for Bonacich power centrality is exactly the same as in the case of eigenvector centrality. This is expected since the β value for Bonacich power centrality is very close to the reciprocal of the largest eigenvalue which means that the effect of being connected to a highly central root cause is increasing the root causes centrality. The beta value iterated is very close to that for eigenvector centrality which indicates that the two centralities are highly correlated and almost identical. Additionally, the small value of beta due to the large dominant eigenvalue indicates that the effect of having neighboring root causes with high centrality doesn't have a large impact on the centrality of a given root cause.

The normalized values of beta centrality are such that the square of the length of the centrality vector is equivalent to the size of the network which means that a centrality of 1 is the centrality which is not too high or too low. This is what was used as the basis for classifying the top 15 root causes to be those of particular interest as these are the causes which have higher than regular centrality. This means that these are the root cause which cooccur with the most other root causes (degree) taking into account the frequency of cooccurrence (weight) and the centrality of neighboring nodes (eigenvector).

Evaluation of Centrality Values:

As previously mentioned, different centrality types examine the different properties of the network root causes. The choice of the centrality value which best represents the influence of a particular node or root cause in the network depends on what particular question is the main focus of the research. A question like how much does a root cause cooccur with unrelated root causes can be answered with flow betweenness. A question like how strongly connected a root cause is to other well-connected causes in the network can be addressed with eigenvector centrality. A question of what is the effect of having connections with other well-connected root causes in the network can be answered by computing the beta value in Bonacich Power Centrality.

All these questions are applicable in a variety of studies and they are all important to examine to be able to observe particular patterns and behaviors within the data. In fact, here is a degree of conceptual correlation between different centrality measures making them partially dependent on one another in a sense that they are neither too low in terms of correlation to be considered independent or too high in terms of correlation to be considered redundant. Which of them is most important relies mainly on the nature and the focus of the particular study. The question most important for this particular study, for example, is which of the root causes, when removed from the network, will correspond to the greatest decrease in the network connectedness. This question correlates directly with weighted degree centrality, which makes weighted degree centrality the main determinant of how influential a root cause is within the scope of the networks

Table 4.6: Top fatal accident root causes according to Bonacich power centrality

Node	Root Cause	Beta Cent	Normalized Beta Cent
RC43	Lack of specific on the job training	36410	3.9706
RC12	No fall arrest system, guardrails or safety nets	32080	3.4984
RC11	No PPE	28166	3.0716
RC44	Lack of general health and safety Training	15888	1.7326
RC9	Poor Labeling	15349	1.6739
RC45	No jobsite inspection	14089	1.5365
RC39	Employee misconduct	11719	1.2780
RC47	Poor Assembling of equipment/scaffold/decking/formwork	11557	1.2603
RC28	Employer allowed employee to work in an unsafe environment	11005	1.2001
RC46	Lack of Inspection for equipment and tools	10759	1.1732
RC18	No decking	10167	1.1087
RC3	Poor Equipment handling	10042	1.0951
RC1	Damaged/defective equipment	9837	1.0727
RC30	Lack of clear employer instructions	9319	1.0163
RC35	Lack of supervision	9276	1.0116

being investigated. To this end, weighted degree will be used as the key centrality measure for the analysis of the seven subnetworks contributing to direct causes. Other centrality values will also be examined and interpreted.

4.5.2.5 Centralization:

Centralization shows to which extent there is variability on centrality of root causes or to which extent some root causes dominate in terms of centrality. A highly-centralized network means that a few root causes have most of the centrality in the network while others have low centrality. A low centralization network, on the other hand, refers to centrality being distributed evenly amongst the root causes (when centralization is closer to 0). In cases of highly centralized networks, only a few causes need to be addressed to significantly reduce network connectedness. In cases of less centralized network structures, centrality is more evenly distributed making addressing the most central causes less effective in reducing network connectedness.

Due to the difficulty associated with defining the maximum possible difference in centralities between maximum and other nodes for valued data, centralization is only computed for binary data. This can lead to misleading results due to treating all edges are equal and not taking edge weight into account. Nevertheless, it provides a useful insight into network properties which can be utilized to observe patterns and behaviors within the root cause network. This does not apply only for flow betweenness due to the fact it has an alternative computation of average difference in centrality between maximum centrality node and other nodes. This approach avoids calculation of maximum possible centrality for a n-node network and thus is irrelevant to compare with other centralization values within the network. Table 4.7 shows the centralization values for each of the metrics utilized for centrality. In this case, unweighted degree centralization was chosen to be the main centralization index as weighted degree was the main centrality measure as discussed earlier. Moreover, degree centralization is more easily interpretable than eigenvector centralization which takes the effects of neighbors, thus distributing centrality differently from its inherent distribution.

Table 4.7: Centralizations for root causes main network

Degree	50.87 %
Betweenness centralization	7.067%

In Table 4.7, An unweighted degree centralization of 50.87 % represents that the network is of average to high centralization. It shows that the difference between the highest centrality root causes and the other root causes is equivalent to 50.87% of the highest possible centrality for a network with the same number of nodes. The reason this is considered as a high value is that the highest possible centrality for an n node network resembles that all causes are connected to 1 node and no other connections exist within the network. Practically for a network to reach that level of centralization is highly improbable which makes reaching 50 % of this assumption resemble a high centralization structure for the network.

Valued betweenness centralization index, on the other hand, is 7.067% in Table 4.7. This centralization value is calculated differently from the former since it takes the value of data into account and resembles the average difference in centralization between the highest value and all the other values of flow betweenness. This value has to be interpreted with reference to the maximum value of betweenness where the maximum value is 8.515% where the average difference between that value and other values is 7.067%. This shows that many root causes have very small centrality relative to the largest centrality cause which explains why the average difference between maximum and other node centralities is 7.067%. This means that the network structure is highly centralized where the average difference between largest centrality and other centralities is very large relative to the largest centrality.

Evaluation of Centralization:

The centralization values all indicate a large variability in centrality where few of the root causes dominate in terms of centrality and many others have little centrality. This is a favorable property of this network as it indicates that if the few central causes are addressed property, the density of the cooccurrence network decreases significantly. Decreasing network density is predicted (through different accident causation models) to accident occurrence probability.

4.5.3 Dichotomizing the network

The very small negligible values of centrality associated with some of the root causes lead to applying a process to dichotomize the network. The dichotomizing process is mainly targeted at removing some of the weak links in the network in an attempt to achieve a better visual representation of the stronger edges in the network which is less complex and more interpretable. The dichotomized network is shown in Figure 4.7. This visual representation eliminates all ties of strength 1 which indicate that the two root causes cooccurred together in a single fatality report.

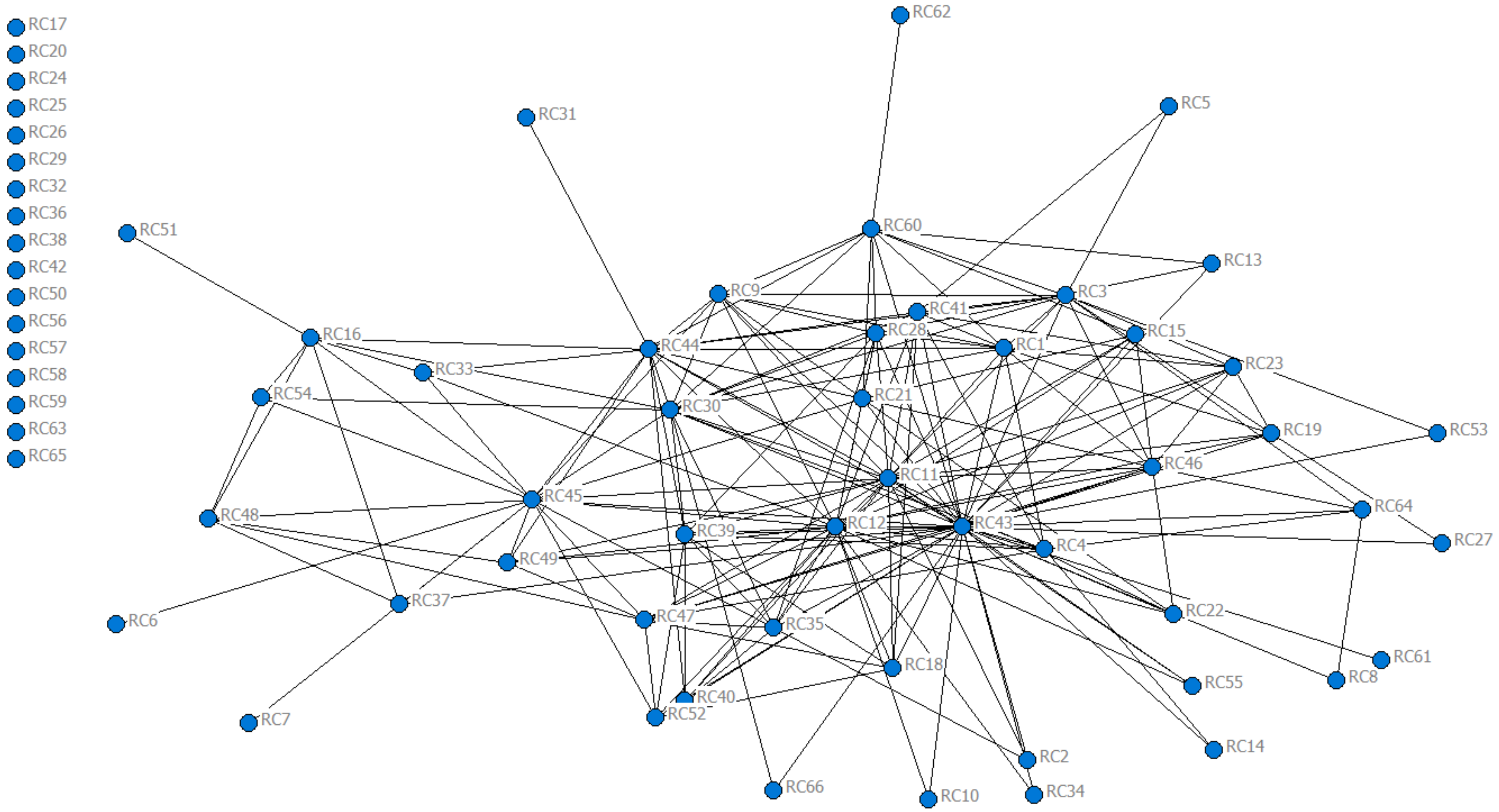

Figure 4.7: Dichotomized root cause main network

`

This representation allows to better focus on cooccurrence links which have greater impact in the network. Dichotomizing the network improves the network visual readability and interpretability while maintaining the strong links. Eliminating the links with strength 1 resulted into isolating 17 nodes which were only connected with links of weight 1 but none of these were listed as highly central nodes.

The drawback of this method, however, is that in some cases there might be a root cause which cooccurs with a large number of other root causes but only with a strength of link 1. In this case, the links of the root cause will be eliminated and even though it is important in the network, it might appear to be an isolate. For example, node 53 in this case is a high flow betweenness centrality node which means that it bridges across structural gaps and plays an important structural role in the network. In Figure 4.6, this node has many links with a wide variety of root causes from which it derives its high flow betweenness. In Figure 4.7, on the other hand, it appears to only have 2 connections where all the other connections were eliminated for having a strength of 1 which diminishes the structural role it plays in the network greatly.

In essence, dichotomizing is a useful technique to visually enhance the network and give it better readability and interpretability. However, this technique should be used carefully by the analyst since it can lead to loss of some information which can be important for the analysis.

4.6 Social Network Analysis Metrics for the Subnetworks

After identifying the most central root causes of fatalities in the main network which includes all 100 root causes, data is manipulated to decompose this main network into 7 different subnetworks corresponding to the 7 commonly quoted direct causes of accidents identified. Every subnetwork includes only root causes corresponding to a particular direct cause excluding others. For example, the fall network includes all root causes which lead into fall accidents and excludes those leading to other accident types. This approach was adopted to be able

to focus on the accident dynamics in the direct cause subnetworks and obtain a clearer view of how the interaction of root causes leads into a particular direct cause.

4.6.1 Network Cohesiveness measures:

The first step towards achieving a better understanding of the 7 subnetworks is comparison of the cohesiveness of the subnetworks. Table 4.8 shows a comparison based on four common measures of cohesiveness, namely diameter, average geodesic distance and density.

Table 4.8: Cohesiveness measures for networks of direct causes

Network	Diameter	Average Geodesic	Unweighted density	Weighted density
Main	4	1.87	0.245	0.464
Fall	3	1.808	0.279	0.504
Struck by	4	1.989	0.198	0.247
Caught in between	3	1.749	0.291	0.348
Electrocution	3	1.73	0.333	0.397
Asphyxia	3	1.549	0.495	0.571
Burn	3	1.495	0.543	0.562
Health Related	2	1.286	0.333	0.333

As in the case of the main network, all subnetworks show dense structures and close relations between root causes contributing to each direct cause as shown in Table 4.8. All subnetworks have diameters which are less than or equal to that of the main network which shows how closely related root causes contributing to a single direct cause are. All subnetworks have a diameter of 3 except for “Struck by” which has a diameter of 4 and “Health Related” which has a diameter of 2. The average geodesic in the network is in line with the diameter with “Struck by” having the maximum value and “Health related” having the minimum value. The order in between is “Fall”, “Caught in between”,” Electrocution”, “Asphyxia” and “Burn” respectively. This average geodesic is a measure of how close root causes are linked together.

In terms of unweighted density, the densest networks are Burn and Asphyxia networks where 54.3% and 49.5% of all possible links between causes are present. This is followed by “Electrocution” and “Health related” where 33.3% of all possible links are present. “Caught in between” and “Fall” follow closely with 29.1% and 27.9% of possible links present. Finally, “Struck by” network has only 19.8% of all links present. This reflects that on average in these networks, a cause cooccurs at least once with 19.8 to 54.3 of other causes which means they are generally dense structures.

When these values are adjusted to reflect link weights, the result is the weighted density which is the average link strength in the network including all the nonexistent links as zero strength links. The use of weights increases density of “Asphyxia” (0.571) to be the greatest followed by “Burn” (0.562). The inclusion of weights highly boosts the density of “Fall” (0.504) as it the most recurring direct cause and so it is reasonable for its contributing root causes to have higher edge strengths. This is followed by “Electrocution” (0.397), “Caught in between” (0.348) and “Health related” (0.333) respectively. Finally, “Struck” (0.247) has the lowest density.

Diameter, average geodesic and unweighted density are all indicators of the geometry of the network which help identify how sparse or how cohesive a network is in terms of geometry. The most cohesive networks using this definition are the “Asphyxia” and “Burn” networks where they have the highest unweighted densities and the smallest diameters and average geodesics. The reason why “Health related” is not considered to be one of the most cohesive is that it only has 5 nodes in two components and so the diameter and small geodesic are heavily restricted by the size of the network. Moreover, the isolated node makes geodesic calculation only valid for the larger component. The network with longest diameter, largest average geodesic and lowest unweighted density is the struck by network. The fact that the “Struck by” network is the sparsest indicates that “Stuck by” hazards result from a large variety of root causes. These root causes variety means that the network consists of root causes which are less closely related than the other networks.

Weighted density, on the other hand, takes weight into account rather than just geometry of the network. Despite that weights vary between the networks, there is little change to the ordering of densities. “Asphyxia” and “Burn” networks are the densest and struck by is the least dense. Weighted density only has significant effect on “Fall” and “Health related” networks order. “Fall” which gains a lot of density from weights and becomes the third most dense. “Health related”, on the other hand, gains no density from weights and becomes the second least dense.

4.6.1.1 Interpretation of network cohesiveness measures:

The cohesiveness measures attributed the greatest cohesiveness values to asphyxia and burn networks. The high density as well as low average geodesic in these networks indicates that both burn, and asphyxia happen due to a limited set of causes strongly related to one another. This means that the network for these direct causes are predictable since all the root causes for these networks are strongly linked and related to one another meaning that these accidents. In other words, the means that these two accident types occur due to mostly the same reasons which repeat themselves over and over.

This can be contrasted with the less predictable “Struck by” network, where the network is very sparse and unpredictable with low density and the longest average geodesic. This structure means that struck by accidents are much less predictable where the root causes for these accidents are not as related compared to other networks. This means that they result from a wide range of root causes and that their prevention is more difficult since there are more causes to address to achieve the same accident reduction.

4.6.2 Network Centrality measures:

Examining centralities of root causes pertaining to each direct cause helps understand the causation sequence leading to a particular accident. This is due to the fact that the root causes with higher centrality in each direct cause network are those with highest influence in the network. This influence means that, as previously discussed, removing them from the cooccurrence network significantly reduces the overall connectedness in the network. Accordingly, if a particular root cause with high centrality is efficiently addressed this significantly reduces the probability of accident occurrence which in turn reduces fatalities.

4.6.2.1 Fall Network

As falls are the leading cause of fatalities in Construction Industry in the United States, analysis of the highest centrality root causes contributing most to the fall network is essential. Figure 4.8 is a representation of the network of root causes leading to falls.

The network graph in Figure 4.8 shows that 45 out of the total 66 root causes contribute to fall accidents. The network structure looks very similar to the main network with a dense core and a less dense periphery. The dense core contains nodes cooccurring with many others and thus central to the network while the periphery contains those connected with a few others. To obtain a more elaborate quantitative view, Table 4.9 provides data about the causes with highest centrality.

Analysis of Centrality Values

The centrality values shown in Table 4.9 are obtained by including the top 9 causes for all centrality measures including degree, flow betweenness, eigenvector and Bonacich power respectively. The cutoff 9 was chosen as only 9 values had a Bonacich power centrality (above moderate). As previously discussed the degree centrality was used as the key indicator for centrality and thus the root causes were ranked by degree centrality. Eigenvector and Bonacich power centralities gave similar rankings due to the value of beta used in UCINET. Therefore, eigenvector centrality values were redundant and thus not included in the Table 4.9.

The value of beta computed for this network is 0.0200289 which explains why there is a large similarity between the top 9 highest centrality root causes for both degree and Bonacich Power measures. Having the largest value of beta centrality as 3.6 shows a large variability in centrality of data since the highest centrality root cause is 3.6 times more central than a moderate centrality cause. The minor differences between the two measures include Employee misconduct being in the seventh place in beta centrality rather than the fifth place in degree centrality. This means that employee misconduct happens to cooccur with less central root causes than the two which have higher beta centrality, namely “no decking” and “Poor Assembling of equipment/ scaffold/ decking/ formwork”. Moreover, beta centrality has “Poor Labeling” in the 9^{th} place rather than “lack of inspection for equipment and tools” which also means it is tied to more central root causes.

Conversely, flow betweenness centrality provides a different perspective on centrality though it still shows a lot of similarity with degree centrality. It ranks “lack of employee knowledge” third amongst root causes even though it is ranked very low by both degree and beta measures. This indicates that this cause tends to cooccur with a wide variety of causes which

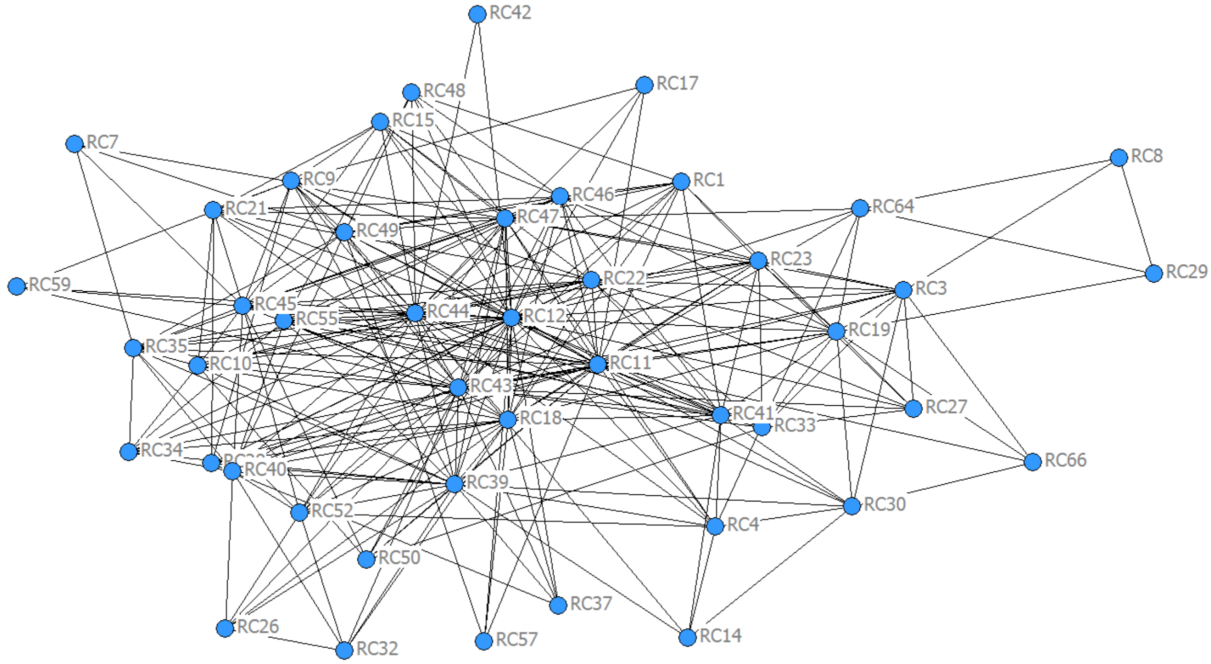

Figure 4.8: Fall root causes network

Table 4.9: Top centrality root causes for fall root cause network

Node	Root Cause	PctDegree	nFlowBet	nBeta	Rank by Degree	Rank by FlowBet	Rank by Bon/ Eig
RC12	No fall arrest system, guardrails or safety nets	13.2265%	13.1449%	3.5942	1	1	1
RC11	No PPE	8.6172%	7.3434%	2.8728	2	4	2
RC43	Lack of specific on the job training	8.2164%	7.5756%	2.3694	3	2	3
RC44	Lack of general health and safety Training	6.3126%	7.3338%	1.9092	4	5	4
RC39	Employee misconduct	4.2084%	5.1799%	1.1873	5	6	7
RC47	Poor Assembling of equipment/ scaffold/decking/formwork	4.2084%	4.0528%	1.3049	6	8	5
RC18	No decking	3.8076%	4.5290%	1.2502	7	7	6
RC45	No jobsite inspection	3.5070%	3.6835%	1.1404	8	9	8
RC46	Lack of Inspection for equipment and tools	3.1062%	1.9314%	0.8637	9	18	11
RC9	Poor Labeling	2.6052%	2.8207%	1.0604	11	12	9
RC19	Lack of employee knowledge	2.2044%	7.3441%	0.4963	15	3	21

normally don't cooccur together. In other words, there is a wide variety of accident scenarios which could occur where lack of employee knowledge can lead to a fall accident each of which has other root causes where root causes of these scenarios do not cooccur. This role of bridging across those unrelated causes is indicated by flow betweenness. "Lack of Inspection for equipment and tools" which is ranked 9th in the case of degree is ranked 18th for flow betweenness. This shows that even though this cause might cooccur frequently with other causes, it doesn't happen to bridge across normally non-cooccurring causes. Other differences between flow betweenness and degree include minor re-ordering of the top 9 root causes due to their ability to bridge across unrelated causes as opposed to cooccurring with other causes frequently.

Interpretation of Centrality Analysis

The fall network gives very similar indications as the main network in terms of root causes. One of the key factors leading to fall accidents is insufficient safety expenditure by the employer. This is obvious from the high centrality associated with absence of fall arrest systems, guard rails and safety nets, no decking and lack of training which all point to reduced employer spending. Improper supervision and poor safety regulation enforcing are also leading underlying causes for falls where not wearing PPEs and employee misconduct are among the top listed root causes.

Employee training also is a key common factor when looking at construction falls where both the lack of training on correct safe working procedures as well as general health and safety training have high centrality values in causing falls. On site regular inspection is also contributing to many falls where having formwork, scaffold, formwork or equipment poorly assembled on site and having proper inspection of neither worksite nor tools and equipment all point at that more refined inspection can reduce falls.

It also proves that communication problems also play a role in falls which is though not shown by degree centrality, shown by both betweenness and eigenvector centralities. This is shown by the high Bonacich power centrality of poor labeling and the very high flow betweenness of lack of employee knowledge.

4.6.2.2 Struck by Network

As predicted by the density measures, the struck by network shows a less dense structure with considerably less links than those in the fall network as shown in Figure 4.9. The struck by network, however, is the one with the largest number of root causes, where it includes 49 out of the total 66 root causes in the main network. Moreover, the decrease in the number of links between the core and the periphery is less considerable. This predicts the proportion of nodes with higher than moderate centrality to be greater than that observed in the fall network.

Table 4.10 provides a more quantitative perspective of the centrality. As predicted from the graph, 15 out of the 49 causes have higher than moderate beta centrality as opposed to 9 out of the 45 causes in the fall network.

Analysis of Centrality Values

As with the fall network, Table 4.10 shows that the largest value for beta centrality is high (3.4) which indicates that the highest centrality root cause is 3.4 times more central than a moderate centrality cause. This also indicates large variability in centralities which is smaller than that shown by the fall network even though the network is composed of more root causes. The computed value of beta for this network is 0.0499163 meaning that in Bonacich power centrality, centralities of neighboring nodes have about 5% effect. This effect, though much greater than that computed for the fall network, is considered small in magnitude.

Comparing degree centrality to Bonacich power centrality, they both identify the top 15 root causes similarly with the exception of different ordering and difference in only 1 root cause. Beta centrality ranks “equipment malfunction” 11th as opposed to 16th in the case of degree centrality which shows that equipment malfunction cooccurs more frequently with highly central root causes. Degree centrality, on the other hand, ranks “Lack of Inspection for equipment and tools” 14th as opposed to 18th in case of Beta centrality. This particular cause was ranked lower by Beta centrality in both fall and struck by networks which further ascertains that this cause is tied to less central causes than other causes.

Flow betweenness, on the other hand, provides very different centrality rankings from the other two measures with a large difference both in terms of ordering and choice of the top 15 root causes. It places some of the highly-ranked root causes by degree centrality in a much lower rank. This includes “Lack of clear employer instructions”, “No/damaged Cave in protection”,” Damaged/defective equipment”, “Use of unsuitable equipment”, “Noncompliance to equipment manufacturer specifications” and “Operation carried out by noncompetent individual”. This is due to the fact that these causes, though cooccurring frequently, might not cooccur with a wide variety of other normally unrelated causes making them less likely to bridge across a structural gap in the network.

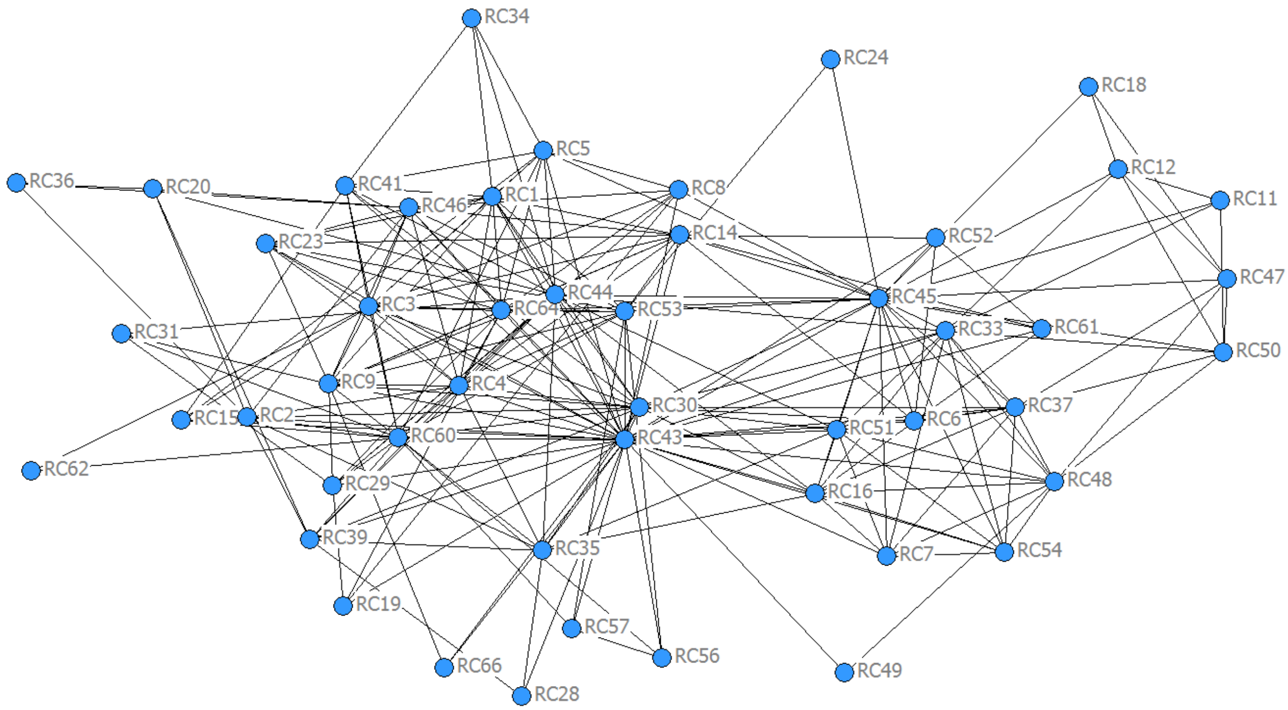

Figure 4.9: Struck by root causes network

Table 4.10: Top centrality root causes for struck by root cause network

Node	Root Cause	PctDegree	nFlowBet	nBeta	Rank by Degree	Rank by Flow FlowBet	Rank by Bon/Eig
RC43	Lack of specific on the job training	10.8621%	13.5863%	3.4046	1	1	1
RC45	No jobsite inspection	6.0345%	9.1944%	2.0478	2	2	2
RC30	Lack of clear employer instructions	4.3103%	2.7799%	1.8245	3	11	3
RC3	Poor Equipment handling	4.1379%	4.1549%	1.6358	4	5	4
RC16	No/damaged Cave in protection	3.6207%	1.6553%	1.5363	5	26	5
RC60	Not following proper work procedures	3.6207%	6.5618%	1.3174	6	3	9
RC44	Lack of general health and safety Training	3.6207%	3.5992%	1.2719	7	7	10
RC64	Noncompliance to equipment manufacturer specifications and recommendations	3.4483%	2.5234%	1.4251	8	13	6
RC1	Damaged/defective equipment	3.2759%	1.9960%	1.3226	9	21	8
RC4	Use of nonsuitable equipment	3.1034%	1.2339%	1.3661	10	29	7
RC9	Poor Labeling	2.7586%	3.1542%	1.0811	11	8	13
RC33	No competent person on site	2.7586%	2.6089%	1.0219	12	12	14
RC37	Operation carried out by noncompetent individual	2.7586%	2.1032%	1.0939	13	18	12
RC46	Lack of Inspection for equipment and tools	2.5862%	2.9454%	0.8862	14	10	18
RC53	Poor housekeeping	2.5862%	4.4690%	1.0184	15	4	15
RC5	Equipment malfunction	2.5862%	2.0869%	1.2076	16	19	11
RC48	Error in design	2.2414%	3.8636%	0.8182	19	6	23
RC39	Employee misconduct	1.8966%	2.4976%	0.7765	23	14	24
RC2	Lack of necessary Equipment	1.5517%	2.3685%	0.5417	26	15	28
RC12	No fall arrest system, guardrails or safety nets	1.2069%	3.0689%	0.2994	30	9	37

Moreover, betweenness centrality places some of the low ranked root causes by degree centrality in a much higher rank. This includes “Poor housekeeping”, “Error in design”, “Employee misconduct”, “Lack of necessary Equipment” and “No fall arrest system, guardrails or safety nets”. Betweenness centrality ranking of these causes can be attributed to the fact that these causes happen to occur with normally non-cooccurring causes bridging a structural cooccurrence gap in the network.

Interpretation of Centrality Analysis

Struck by accidents are the most diverse in terms of root causes. As in the main network, lack of training on the job on the appropriate correct working procedures is the highest centrality cause. “Lack of general health and safety training” is of also high centrality which corroborates the importance of training for reducing struck by accidents. Lack of basic procedural inspections comes also as one of the major organizational defects where “no jobsite inspection” and “lack of inspection for tools and equipment” are both contributing to fatalities due to “struck by” accidents. Another key consideration is the fact that employer is not fully satisfying his role in terms of safety, where the lack of clear instructions given by employer, failing to furnish necessary equipment to carry out the work safely and not providing systems for fall arrest, nets to catch falling objects or guardrails on elevated boundaries of buildings or excavation boundaries are amongst the central causes.

One key consideration in which struck by accidents differ from fall accidents is the contribution of human factor to these accidents more frequently. Struck by fatalities have been attributed more to human-related root causes such as “poor equipment handling”, “work being carried out by incompetent personnel”, “not following proper work procedure” and “employee misconduct”. Those human-related causes can be attributed to both human actions as well as insufficient supervision where one of the key causes was also “lack of competent site personnel”.

Another area where human errors affect “struck by” accidents is technical errors rather than errors pertaining to the tasks being carried out on the jobsite. Technical human errors are those errors in design as well as the relevant engineering controls pertaining to processes, equipment and site. Technical defaults which can result in “struck by” accidents include not abiding by the recommendations of the manufacturers of equipment, utilizing unsuitable equipment for the job being carried out on site, not having protection against cave-ins due to poor calculations of slope stability or using a damaged protection system and design defects which can cause life threatening hazards such as collapse of structures. Design errors might result from miscalculation, insufficient design review or not accounting for particular critical cases which leads to very dangerous struck by hazards as overloading equipment, cave ins or structural failures.

Other causes were related to the work site where poor housekeeping and poor labeling of hazards on site have high centralities. Housekeeping has a very high flow betweenness as it is a root cause which connects many others as it can lead to many different struck by accident scenarios. Poor labelling and no clear employer instructions can be seen as communication issues that need to be addressed from an organizational perspective.

Another very important trend shown by "struck-by" accidents is the fact that they are closely related to equipment where 6 of the 15 key causes for "struck-by" also happen to be relevant to equipment in one way or another. This raises a flag that it should be a top priority for equipment-intensive projects to put in place all possible precautions which can reduce the probability and effect of struck-by accidents.

4.6.2.3 *Caught in Between*

Figure 4.10 shows a network structure that is slightly smaller than the former two where the caught in between network is composed of 34 out of the 66 root cases used in the main network. The network structure is considerably dense with most of the nodes having many edges between them. The difference in density between the core and the periphery is less considerable than in the case of the fall network. This again predicts a larger proportion of the root causes to have a higher than moderate beta centrality.

Table 4.11 provides the centrality values of the highest centrality root causes. As predicted from the graph, 11 out of the 34 causes have higher than moderate beta centrality as opposed to 9 out of the 45 causes in the fall network. This shows that centrality is better divided between root causes in the caught in between network.

Analysis of Centrality Values

The highest value of beta centrality in Table 4.11 is 2.8 which is considerably lower than that for the fall network. This indicates smaller variation in centrality than that in both the fall and struck by networks. The computed value of beta in this case is 0.0605388 which is higher than that for the fall and stuck by networks. This indicates that the centrality of a root cause is increased by a factor of about 6% of the centralities of neighboring root causes which is still a small value.

For this network, the ranking of the top 11 root causes by degree and Bonacich power

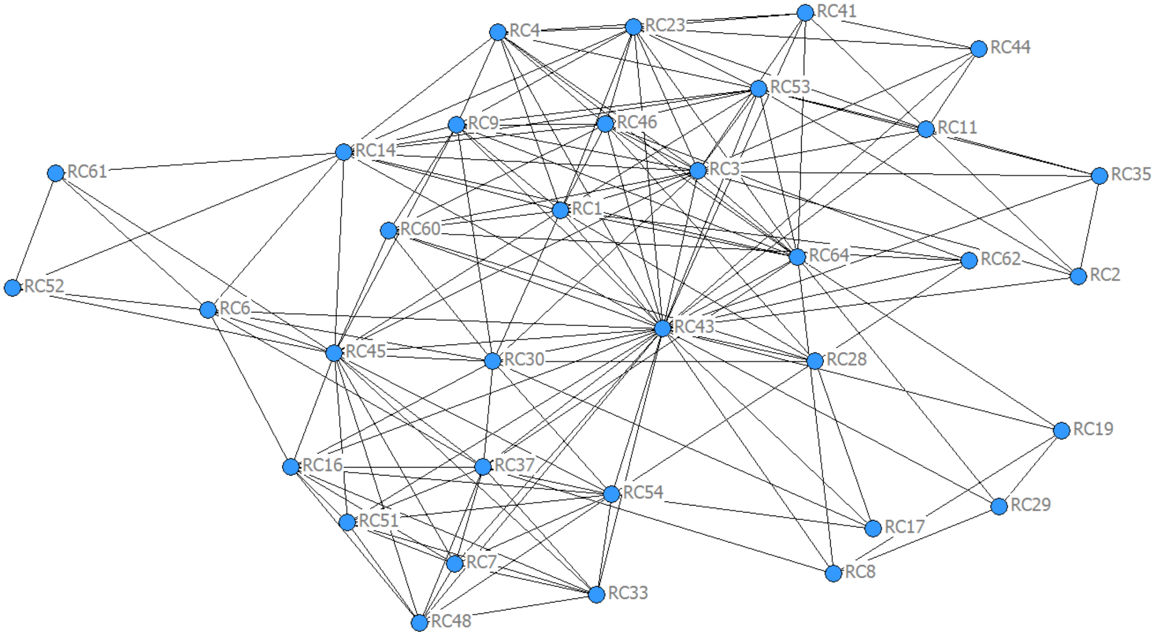

Figure 4.10: Caught in between root causes network

Table 4.11: Top centrality root causes for caught in between root cause network

Node	Root Cause	PctDegree	nFlowBet	nBeta	Rank by Degree	Rank by Flow FlowBet	Rank by Bon/Eig
RC43	Lack of specific on the job training	12.3077%	16.2269%	2.8005	1	1	1
RC3	Poor Equipment handling	6.1538%	5.7583%	1.8086	2	4	2
RC45	No jobsite inspection	5.6410%	6.9590%	1.3995	3	2	4
RC1	Damaged/defective equipment	4.8718%	3.1509%	1.5106	4	11	3
RC64	Noncompliance to equipment manufacturer specifications and recommendations	4.6154%	5.9596%	1.3887	5	3	5
RC37	Operation carried out by noncompetent individual	4.1026%	3.5348%	1.2035	6	9	6
RC9	Poor Labeling	3.5897%	1.2089%	1.1882	7	31	7
RC23	Poor Material Handling	3.5897%	2.7256%	1.1747	8	15	8
RC46	Lack of Inspection for equipment and tools	3.5897%	2.4831%	1.1665	9	17	9
RC53	Poor housekeeping	3.5897%	3.0882%	1.1661	10	14	10
RC30	Lack of clear employer instructions	3.5897%	3.0965%	1.1172	11	13	11
RC14	Not wearing seatbelt	2.4138%	1.9836%	0.9446	12	5	12
RC6	equipment/spoil on excavation edge	1.7241%	1.3932%	0.6203	18	8	20
RC54	No safe exit to site	1.3793%	0.7250%	0.8184	14	6	16
RC11	No PPE	0.6897%	0.8162%	0.8158	16	7	17
RC28	Employer allowed employee to work in an unsafe environment	0.5172%	0.3001%	0.7177	19	10	18

centralities are almost identical. The identified top 11 root causes are the same and have the same order except for the 3rd and the 4th which have their order swapped which is a very minor difference. This means that the effect of centrality of neighbors is almost the same for all the top root causes investigated.

As for flow betweenness centrality, the difference in the ranking using this measure is significant for this network. Flow betweenness ranked some caught in between root causes much higher or much lower than the ranking by degree centrality. The causes “Damaged/defective equipment”, “Poor Labeling”, “Poor Material Handling” and “Lack of Inspection for equipment and tools” are ranked much lower in terms of betweenness even though they are among the top causes by degree centrality. This means that these causes lack the structural role of bridging across non-cooccurring causes discussed earlier. On the other hand, root causes including “Not wearing seatbelt”, “Having equipment or spoil on excavation edge”, “No safe exit to site”, “No PPE” and “Employer allowing employees to work in unsafe conditions” are ranked high in flow betweenness and much lower in degree. This means that these causes though co-occurring frequently, do not occur in different scenarios containing normally non-cooccurring causes.

Interpretation of Centrality Analysis

The most central cause for “caught in between” accidents is lack of specific on the job training which also happens to be the most central cause for “struck by”. However, unlike fall and struck by networks, insufficient generic health and safety training does not belong to the most central root causes for caught in between accidents. What can be deduced from the absence of this cause is that the generic health and safety training is of very limited importance to help workforce avoid caught in between accidents as compared to the more specialized job specific training which helps them recognize and identify hazards relevant to their particular jobs.

Caught in between accidents are, like “struck by”, highly correlated with human related errors. Key root causes related to human errors include handling the equipment and material in an improper manner, not having PPE worn during operations, not using the seatbelt when driving vehicles or equipment on site and having jobs being carried out by personnel who lack the necessary competence. All the aforementioned causes are procedural human errors relevant to following proper work and safety procedures on the jobsite and can be attributed to the slack when enforcing safety rule and insufficient supervision. Not abiding by recommendations of equipment manufacturers as well as poorly designed worksite layout are two technical human related causes.

Lack of worksite housekeeping, lack of hazard labelling on site, insufficient safe site exits and having spoils or equipment at the edges of excavations on site are all examples of improper design, planning and organization of worksite. This improper worksite layout has a significant negative effect on caught in between accidents which raises a flag for work sites which are very congested or have limited work space that not organizing their work properly can potentially result in caught in between accidents. Moreover, it can be deduced that having properly enforced inspection schedules is essential to avoid "caught in between" accidents where insufficient inspection for tools, equipment and the jobsite were all listed amongst the top causes for "caught in between" accidents.

Finally, the data from caught in between network further emphasizes the deficiency in employer culture. Amongst the key causes for caught in between fatalities are employers not giving workforce sufficient instructions, not replacing defective unsafe equipment and allowing their workforce to work regardless of unsafe site conditions.

4.6.2.4 *Electrocution Network*

The electrocution network is composed of 25 interrelated root causes as shown in Figure 4.11. The network has a considerably dense structure where the density is high through the core and doesn't fade as the edges go from core to periphery. This is evident where no nodes have only one edge and only one node (RC63) has only two edges. This predicts a larger proportion of the root causes to have a higher than moderate beta centrality.

Table 4.12 shows the top 8 root causes for each of the centrality measures combined together. In this network, 8 out of 25 root causes for the electrocution network as opposed to 9 out of 45 for fall network have higher than moderate beta centrality. This means that this network has a much more equitable centralization distribution rather than having a few causes dominate the centrality of the whole network.

Analysis of Centrality Values

The highest normalized beta centrality value in this network is 2.53 as shown in Table 4.12 indicating that the highest centrality cause is 2.53 times more central than a moderate centrality cause. Beta value in this network was computed to be 0.0705945 which means that the Bonacich power centrality of a root cause increases by 7.06% of the centralities of neighboring causes.

The top 8 most central causes according to beta centrality are the same as those according to degree centrality with minor reordering. Reordering includes "No PPE" being in the 6^{th} place in

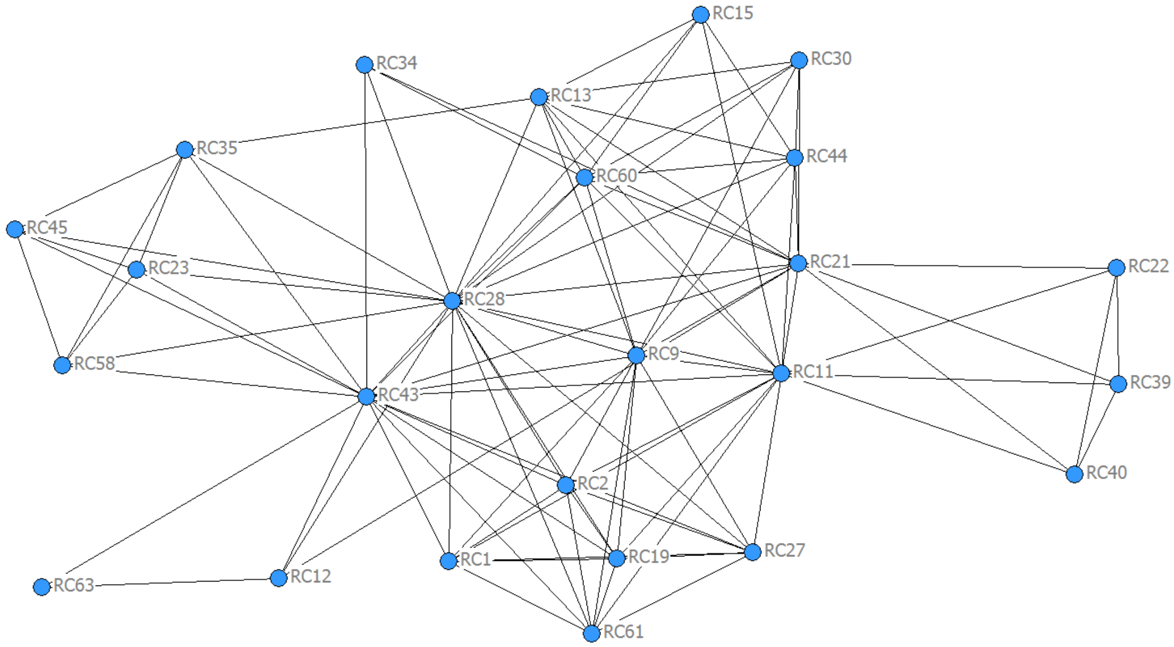

Figure 4.11: Electrocution root causes network

Table 4.12: Top centrality root causes for electrocution root cause network

Node	Root Cause	PctDegree	nFlowBet	nBeta	Rank by Degree	Rank by Flow FlowBet	Rank by Bon/Eig
RC28	Employer allowed employee to work in an unsafe environment	13.8655%	13.3273%	2.5263	1	2	1
RC43	Lack of specific on the job training	9.2437%	17.1265%	1.7198	2	1	2
RC11	No PPE	7.1429%	9.7738%	1.2899	3	4	6
RC21	Poor tool handling	7.1429%	10.3246%	1.5259	4	3	4
RC60	Not following proper work procedures	6.3025%	5.0474%	1.5293	5	5	3
RC9	Poor Labeling	5.8824%	3.3805%	1.2494	6	12	7
RC13	No cable insulation	5.4622%	3.5349%	1.3367	7	11	5
RC44	Lack of general health and safety Training	4.6218%	2.8530%	1.1387	8	16	8
RC12	No fall arrest system, guardrails or safety nets	2.1008%	4.2651%	0.5424	20	6	17
RC22	Inappropriate tools used	1.6807%	4.0015%	0.2335	21	7	22
RC39	Employee misconduct	1.6807%	4.0015%	0.2335	23	8	23

beta centrality rather than the 3rd place in degree centrality. “Not following proper work procedure” being in the 3rd place in beta centrality rather than the 5th in degree centrality. Finally, “No cable insulation” came in the 5th place in beta centrality rather than the 7th in degree centrality.

Flow betweenness provides different indications compared to other centrality measures. The identified top 5 central causes are the same for betweenness and degree centralities even though they are ordered differently. Those ranked from 6 to 8, however, are “No fall arrest systems, guardrails or safety nets”, “Inappropriate tools used” and “Employee misconduct”. These causes are ranked top by betweenness centrality even though they are ranked close to least central by degree centrality. Moreover, each of these causes has only four edges attached to them in the graph. This means that even though each of these causes has a limited number of cooccurrence links, these links bridge across important structural gaps in the network cooccurring with root causes that do not cooccur frequently together. These three causes take the places of “Poor Labeling”, “No cable insulation” and “Lack of general health and safety Training” in the 8 highest centrality causes for degree centrality since these 3 causes are far less important in the network in terms of flow betweenness.

Interpretation of Centrality Analysis

The most central cause both according to degree centrality and according to Bonacich power centrality is that the employers let their workers work in hazardous work zones. Other employer-related root causes of electrocution fatality include failing to ensure the insulation of all power cables and powerlines close to the working area and failure to provide equipment for fall protection like guardrails or nets. This shows that negligence of employers is a key factor subjecting workforce to electrocution and that employers should be extra cautious about their workers when electrocution risks are involved. Training plays a vital role in this network a well, where absence of job-specific training has the largest value of betweenness centrality and absence of generic health and safety training has a very high centrality as well.

Apart from employers caring about their workforce and training them, the other key underlying factors in electrocution accidents are either human-related or relevant to site conditions. Human related factors include not wearing the necessary PPE when dealing with electric hazards, handling the tools used in electric work poorly, use of tools which are not proper for the job like tools which are not sufficiently insulated, or which do not have earth connections and finally poor conduct by employees not following regulations and safe work practices. Site related problems were not having electrocution hazards labelled clearly on site and having a cable or powerline on site which is energized but not insulated.

4.6.2.5 Asphyxia Network

Asphyxia root causes network consists of 14 well connected root causes as shown in Figure 4.12. Being the densest network in terms of weighted density and the second most dense in terms of unweighted density, root causes in this network are well connected to one another. Exceptions to this are only root causes which have 2 connections like RC 25 and RC 44, and root causes which have 3 connections like RC 38 and RC 65. From this network structure, the proportion of causes with higher than moderate beta centrality is expected to be large due to equitable distribution of edges in the network.

The highest 5 causes for each centrality measure are shown in Table 4.13. The beta centrality shows that 5 of the 14 root causes are of higher than moderate centrality compared to 9 of 45 for the fall network. This provides evidence of a much more equitable distribution of centrality in the network.

Analysis of Centrality Values

The highest value of normalized beta centrality in the network is 1.6082 according to Table 4.13 which means that the highest centrality root cause is only 1.6 times as central as a root cause of moderate centrality. This further supports that the network has centrality distributed evenly amongst its nodes. The computed value of beta is 0.101360 which means that for beta centrality the centrality of a root cause increases by 10% of its neighbors' centrality which is a considerable amount. Despite the value of beta computed, the ranking of root causes by degree and beta centralities is exactly identical. This means that the addition of the effect of centrality of neighbors had no effect on the distribution of centrality in the network. In other words, the effect on all nodes due to addition of this effect was similar for all nodes or in line with the node's initial centrality such that it didn't cause any change in the ranking of the nodes.

Flow betweenness centrality, on the other hand, had a different ranking where only 3 of the 5 highest ranking root causes in terms of degree were among the highest-ranking causes in terms of flow betweenness. The other two among the top 5 in degree "Error in design" and "No safe exit to site" were replaced by "High exposure to chemical" and "Lack of general health and safety training". This is one of the cases where flow betweenness centrality is not as useful as other centrality measures as these two causes are only connected to one another as well as the cause "No

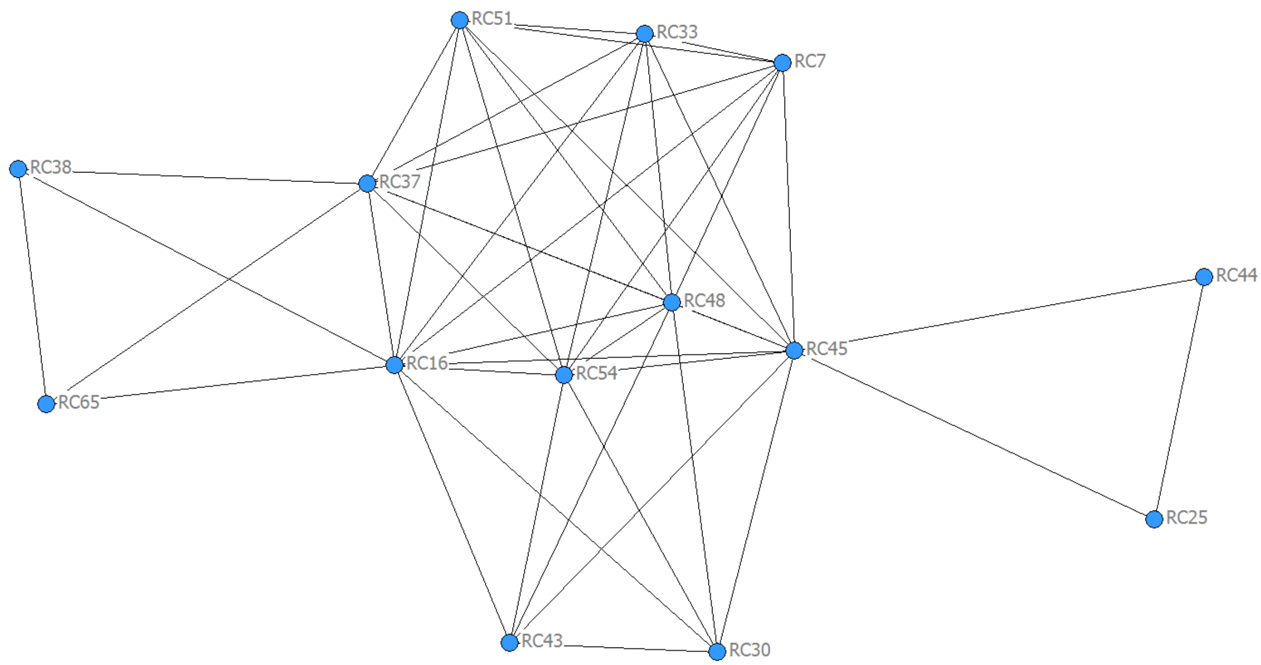

Figure 4.12: Asphyxia root causes network

Table 4.13: Top centrality root causes for asphyxia root cause network

Node	Root Cause	PctDegree	nFlowBet	nBeta	Rank by Degree	Rank by FlowBet	Rank by Bon/Eig
RC16	No/damaged Cave in protection	14.4231%	16.8926%	1.6082	1	2	1
RC45	No jobsite inspection	13.4615%	36.4652%	1.4897	2	1	2
RC48	Error in design	11.5385%	7.6190%	1.4602	3	6	3
RC54	No safe exit to site	11.5385%	7.6190%	1.4602	4	7	4
RC37	Operation carried out by noncompetent individual	9.6154%	11.5079%	1.1191	5	3	5
RC25	High exposure to chemical	1.9231%	7.6923%	0.1693	13	4	13
RC44	Lack of general health and safety Training	1.9231%	7.6923%	0.1693	14	5	14

jobsite inspection”. This means that when flow is directed either to or from one of these causes, it has to pass through the other cause. On the other hand, when the flow is going from any other node to another, due to the high connectivity in the network it can pass in a variety of routes and it is rarely obliged to take a particular route as in this case. This is one of the major loop holes in flow betweenness where in dense highly connected networks as in this case, only nodes on the path to a not well-connected node gain a high centrality. This reflects bridging a structural hole but since the whole network is well connected, this structure hole is only the connection to a single poorly connected node.

Interpretation of Centrality Analysis

Unlike the other networks, the asphyxia network is mainly based on cave in accidents where usually the victims themselves have little to do with the accident, but it is more due to a technical or human related error. Lack of cave in protection or damaged cave in protection comes on top of the list of these causes where cave in protection might not have been put in place due to faulty geotechnical calculations or might have been damaged which could not be captured by inspection. Proper inspection by competent geotechnical engineers for the jobsite can also help identify if structural elements, shoring or cave in protection is in a suitable condition or if soil is stable prior to working under soil surface which is why if it is not present, it can lead to cave in fatalities. Design errors for the suitable shoring, structural elements or soil stable slope can also result in cave ins. Having operations carried out by non-competent personnel whether those assembling the soil bearing system or those carrying out the work under soil level can lead to cave in hazards either due to excessive vibration or poor assembling of soil bearing system. Finally, the site itself must have sufficient safe exits at safe distances to allow evacuation in case cave in starts to occur.

Therefore, asphyxia is mostly due to human errors in design or assembling or inspecting soil bearing systems or soil stable slope. Victims in this case lose their lives mostly due to mistakes of others specially if there are no safe exits to site at a reasonable distance from one another.

4.6.2.6 <u>Burn/explosion Network</u>

The burn network shown in Figure 4.13 consists of 15 nodes which are very well connected. It is the densest network in terms of unweighted density and the second less dense in terms of weighted density. The connectedness in the network is very evident where no given root cause cooccurs with less than four other root causes. The network is composed of two distinct parts where the part at the top has fewer connections compared to the part at the bottom. However, given that root causes in this network have a large number of edges, this network structure is predicted to have the most equitable distribution of centralities of all networks.

The number of nodes with higher than moderate beta centrality in this network is 9 out of 15 root causes compared to 9 out of 45 root causes for the fall network. This number is considerably higher than the second highest (5 out of 14 in case of asphyxia). This is an indicator of equitable distribution of centrality in the network. The 9 highest centrality causes for each centrality measure are shown in the Table 4.14.

Analysis of Centrality Values

As in the asphyxia network, the burn network has the rankings for degree centrality and beta centrality exactly identical as shown in Table 4.14. This phenomenon occurs despite the value of beta being 0.109217 meaning that the centrality of a given root cause is increased by 10.9% of the neighboring root causes. This is a considerable value, yet it doesn't make a difference in the ranking as either it is proportional to the degree of the node or just mostly similar for all values.

The flow betweenness centrality, on the other hand, had only the first four most central causes in common with other centralities. "No hazard identification/communication program" was placed in the 7th place by flow betweenness rather than the 5th by other centrality. Positions 5,6,8 and 9 were occupied by "Failure to locate utility", "Lack of general health and safety Training", "Lack of knowledge by employer about site conditions" and "Lack of clear employer instructions" respectively. These causes are in the upper less dense area of the network diagram and are thus ranked inferior by degree and beta centralities. Flow betweenness offers a different perspective where it ranks them higher as when flow is directed towards or from this upper part of the network, it has fewer paths to choose from. Accordingly, nodes in this area have more flow which must flow through them due to lack of alternative or redundant paths. This is why these nodes are considered to have large flow betweenness where they bridge between causes, which do not cooccur together which is synonymous to bridging to areas with less connection or less dense areas of the network.

Interpretation of Centrality Analysis

Burns and explosions though very little in frequency expose a heavy interaction between different deficiencies of the safety management system which enable their occurrence. These deficiencies are related to employer involvement, communication and training.

Employer is at fault for the majority of the identified burn/explosion fatalities. This

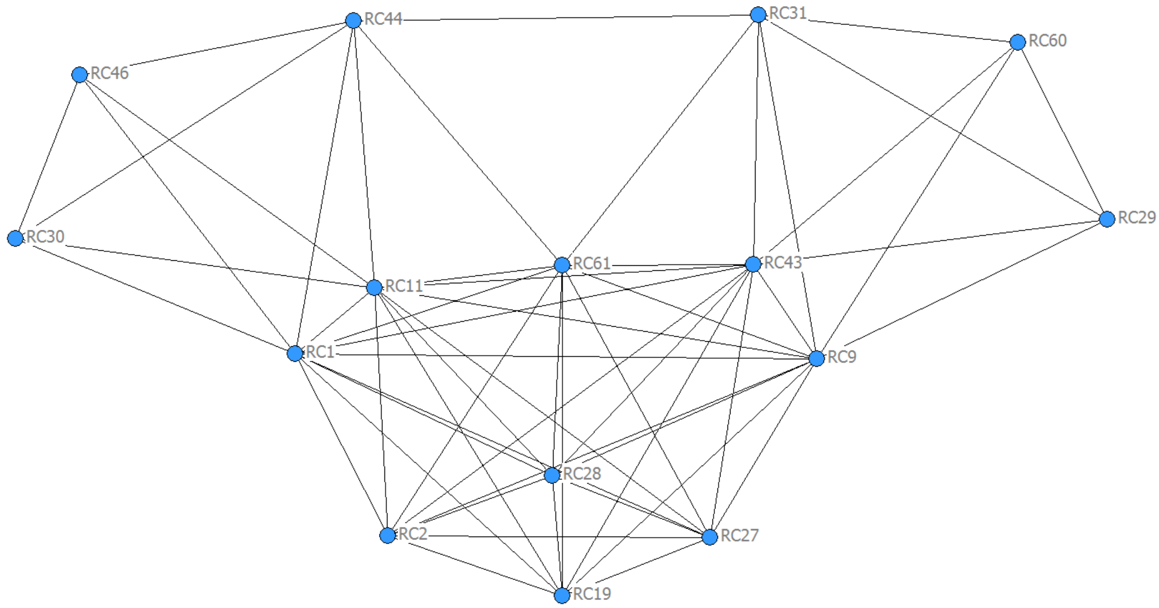

Figure 4.13: Burn/explosion root causes network

Table 4.14: Top centrality root causes for burn/explosion root cause network

Node	Root Cause	PctDegree	nFlowBet	nBeta	Rank by Degree	Rank by Flow FlowBet	Rank by Bon/Eig
RC1	Damaged/defective equipment	10.1695%	12.5092%	1.3640	1	1	1
RC9	Poor Labeling	10.1695%	12.5092%	1.3640	2	2	2
RC11	No PPE	10.1695%	12.5092%	1.3640	3	3	3
RC43	Lack of specific on the job training	10.1695%	12.5092%	1.3640	4	4	4
RC61	No hazard identification/ communication program	8.4746%	6.6850%	1.2059	5	7	5
RC2	Lack of necessary Equipment	6.7797%	3.4615%	1.0893	6	12	6
RC19	Lack of employee knowledge	6.7797%	3.4615%	1.0893	7	13	7
RC27	Employer gross negligence	6.7797%	3.4615%	1.0893	8	14	8
RC28	Employer allowed employee to work in an unsafe environment	6.7797%	3.4615%	1.0893	9	15	9
RC31	Failure to locate utility	5.0847%	9.4139%	0.5866	10	5	10
RC44	Lack of general health and safety Training	5.0847%	9.4139%	0.5866	11	6	11
RC29	Lack of knowledge by employer about site conditions	3.3898%	5.8608%	0.4091	12	8	12
RC30	Lack of clear employer instructions	3.3898%	5.8608%	0.4091	13	9	13

includes not furnishing necessary equipment or having defective equipment to be used on site, giving workers permission to work in an unsafe worksite, gross negligence, poor locating of utilities, not having necessary information about site conditions and not giving clear instructions to workforce. This exposes the fact that explosions or burn fatalities occur on sites were employers do not value their workforce.

Poor communication is also a common cause for burn/explosion fatalities showing a deficiency in safety management system. This is evident where poor labelling of hazards on site, lack of hazard identification and communication program, lack of employee knowledge and lack of clear employer instructions are all causes for burn/explosion fatalities.

Finally, lack of training employees, which also shows how much employer values employees, is a key factor causing fatalities in this network. This is due to the fact that both on the job specific training and general health and safety training are quoted among highest centrality causes for burn/ explosion fatalities. The only factor for which employees can be blamed in this case is not wearing PPEs which is a human error committed by the victims themselves.

4.6.2.7 <u>Health related Network</u>

Unlike the previous networks, the health-related network has two components where one is composed of 4 nodes and the other is composed of an isolated node as shown in Figure 4.14. Among the 5 nodes, 3 nodes are completely connected together, the fourth is connected to only one node and the fifth is isolated. The network structure is simple and self-explanatory.

Of the 5 root causes, the three which are fully connected are those which have higher than moderate beta centrality. The centrality values are given in Table 4.15.

Analysis of Centrality Values

Of all the networks, this simple structure in Figure 4.14 has all centralities providing the same ranking for root causes by all different centrality measures with the “Heart attack” ranked first as shown in Table 4.15. “No fall arrest system, guardrails or safety nets” and “Lack of general health and safety Training” are tied for the second place according to all three centrality measures. The simple geometry of the network and the equal weights are the reason for this phenomenon where betweenness, weighted degree and beta centrality all reflect the same ranking. The beta value computed for this simple network was 0.458507 meaning that the nodes centralities are increased by about 45.9% of the neighboring nodes. However, this still doesn’t change the ranking by beta centrality even though it reduces the variation between the first and the second values. This

Figure 4.14: Health related root causes network

Table 4.15: Top centrality root causes for health-related root cause network

Node	Root Cause	PctDegree	nFlowBet	nBeta	Rank by Degree	Rank by Flow FlowBet	Rank by Bon/Eig
RC42	Heart Attack	30.0000%	25.0000%	1.4990	1	1	1
RC12	No fall arrest system, guardrails or safety nets	20.0000%	5.0000%	1.2800	2	2	2
RC44	Lack of general health and safety Training	20.0000%	5.0000%	1.2800	3	3	3

reduction is evident as the value for degree centrality for the top cause is 1.5 times that of the other two causes whereas the beta centrality value is only 17.11% greater. The highest beta centrality value of 1.499 means that the highest centrality root cause is1.499 times as central as a moderate centrality root cause, which is a relatively reasonable figure showing beta centrality is distributed well in the network.

Interpretation of Centrality Analysis

The data by the network shows that heart attacks are the most common cause for health-related fatalities on construction sites. The network is too small to draw any further conclusions about health-related fatalities in the construction industry.

4.6.3 Network Centralization and Transitivity Comparison

Degree Centralization

The degree centralization is based on the unweighted variations of this measure. This means it only account for the geometric shape of the network and not for tie strength. The reason is that it compares the average difference between the most central cause and the other causes with the maximum possible difference for an n node network. The maximum difference for an n node network is only computable for an unweighted network. In this data, all the networks have large values for weighted degree. This means that the networks all have a considerable difference between the largest centrality cause and other causes. It also reflects that highest centrality root causes in each network have a considerable number of links. This means that their removal from the network leading to a particular direct cause reduces the connectedness in this network considerably. This property can be utilized to prioritize which direct causes to address first.

In this case the largest degree centralization direct cause network is the network corresponding to Caught in between accidents as shown in Table 4.16. This is an indicator that caught in between accidents can be drastically mitigated by addressing the highest centrality root causes. Struck by, fall and electrocution respectively follow with centralization indices above 50%, which is also a very large value for centralization. Asphyxia and Health related networks are centralized as well with above 40% of the maximum possible centralization. Finally, the burn/explosion network has a centralization index of 28.02% which is a mediocre value.

Flow betweenness centralization, on the other hand, resembles the average difference in centrality between the maximum centrality node and the other nodes. A measure is devised in Table 4.16 for a better comparison of betweenness centralization values. This measure simply resembles betweenness centralization, which is the average difference between the maximum and other centrality values, as a percentage of the maximum betweenness centrality value. For example, the highest betweenness centralization ratio belong to the health-related network. The highest flow betweenness centrality value for this group is 25% whereas the centralization is 23% as shown in Table 4.16. This means that the average of all other values of betweenness (including zero centrality nodes) is only 2%. This tendency of the highest centrality node to have most of the centrality in the network leaving other nodes with much lower centralities is the essence of centralization and is reflected efficiently by the ratio devised.

The highest ratio is 92% attained by health-related network yet is let meaningful due to the network being too small and having an isolated node. Struck by, fall, caught in between and asphyxia follow respectively with ratios higher than 80%, which also show they are highly centralized. This is followed by electrocution with a ratio of 77.97% which is still a large centralization ratio. Finally, burn/explosion network also has low flow betweenness centralization relative to others with a ratio of only 42.71%.

4.6.3.1 Transitivity

Transitivity is compared against unweighted density of each network. This comparison is intended to compare the probability two root causes cooccur together given that they cooccur with a third one with the probability that any two root causes cooccur in general. Transitivity phenomenon is not displayed by most of the networks. The largest transitivity was displayed by the struck by network where the probability of two causes cooccurring increases by 4% given that they both cooccur with a third root cause. This increase is not sufficiently large to support that the transitivity property is displayed by the network. Caught in between network follows with transitivity being 1.9% greater than unweighted density. Electrocution, Asphyxia and Health related networks do not display any transitivity with transitivity values only 0.2%, 0.2% and 0% greater than unweighted density. Fall and burn/explosion network display an unexpected property where transitivity values are lower than unweighted densities by 2.2% and 3.1% respectively. This can be interpreted as that if two causes in these networks both cooccur with a third cause, this reduces the probability that these two causes cooccur together.

In essence, all these networks have very slight alterations from unweighted density when it comes to transitivity, so it can be concluded that all these networks exhibit little to no transitivity.

Table 4.16: Comparison of centralization and transitivity of the networks.

	Degree Centralization	Flow Betweenness Centralization	Maximum Flow Betweenness Centrality	Flow betweenness ratio	Transitivity	Unweighted Density
Caught in between	68.9400%	13.5180%	16.23%	83.31%	31.0000%	29.1000%
Struck by	59.7100%	11.6840%	13.59%	86.00%	23.9000%	19.8000%
Fall	56.4500%	11.0750%	13.14%	84.25%	25.7000%	27.9000%
Electrocution	54.3500%	13.3540%	17.13%	77.97%	33.5000%	33.3000%
Asphyxia	41.0300%	30.0430%	36.47%	82.39%	49.7000%	49.5000%
Health related	40.0000%	23.0000%	25.00%	92.00%	33.3000%	33.3000%
Burn/ explosion	28.0200%	5.3430%	12.51%	42.71%	51.2000%	54.3000%

4.7 Result Validation

The validation of centrality results is aimed at comparing the obtained results with those of previous findings in order to identify the areas of agreement with past studies as well as the contributions that resulted from using the social network approach. This was carried out for the main network in addition to the four networks of direct causes of fatalities corresponding to the fatal four direct causes which are the "fall" network, "struck by" network, "caught in between" network and "electrocution" network. This is since these are the main accident direct causes on which previous studies have focused. Since only very few studies tackled fatal construction accidents, the comparison between SNA findings are contrasted against other studies which focused on all accidents both fatal and nonfatal.

4.7.1 Validation of results of main network

A study was conducted by Toole et al. (2002) to propose an assignment of safety responsibilities among designers, contractors and subcontractors. The study identified 8 root causes for construction accidents which are:

- Lack of appropriate training to recognize and avoid work hazards.
- Poor Safety Enforcement.
- Not providing safety Equipment.
- Unsafe working methods or sequencing of work tasks.
- Hazardous site conditions.
- Safety Equipment provided not used by employee.
- Poor employee safety attitude.
- Unexpected sudden isolated employee behavior deviation.

The 7 first causes correspond to top centrality causes already identified and analyzed in the results of SNA for the main network. Their importance was thoroughly investigated and discussed in detail. The only cause which is different is "Unexpected sudden isolated employee behavior deviation" which is not recognized by this study due to the fact that part of the definition of root causes in this study is having countermeasures and this cause has no countermeasures.

Another study conducted by Abdelhamid and Everett (2000) aimed at creating a model for tracing construction accident root causes. Their model identified 3 types of root causes which are:

- Failure to identify a hazardous condition.
- This cause is associated with several of the key root causes identified using SNA including (RC 9, RC30, RC43, RC 44, RC45 and RC46).
- Decision to proceed with work after existing unsafe condition is identified.

- This relates to several of the root causes where the unsafe condition is known, and the decision is to proceed for the sake of money and/or time including (RC1, RC16, RC18 and RC28).
- Decision to act in an unsafe manner regardless of work environment conditions.
- This relates to cases when the decision itself is regarded unsafe and disregards the surrounding working conditions which corresponds to several key root causes including (RC 11, RC 37 and RC 39).

The study suggests 3 key areas in which corrective actions should be directed. These areas are training, improvement of workers attitude as well as modifying management processes and procedures. These 3 areas were thoroughly discussed in the SNA findings. The commonality between the 2 aforementioned studies and the findings of SNA provides validity to the use of SNA as an analytical methodology.

One area where SNA was different from the approaches used by the aforementioned studies is that the two aforementioned studies had their focus on site conditions which are downstream and how they relate to management role. This approach, despite being essential for accident prevention, falls short of locating some areas defect which are more upstream and less tied to site. SNA's utilization of relationships between causes to prioritize them makes SNA results more eye opening identifying other upstream areas where defects occur including areas of engineering controls, design and communication. This is in addition to the proper planning of regular site operations like inspection, maintenance and supervision.

4.7.2 Validation of results of subnetworks

Janicack (1998) investigated the fall accidents and came up with 7 key root causes for falls in construction which lead to the highest frequency of fall accidents. These causes can be listed by order of importance as:

- Absence of fall protection.
- Collapse of structure/ Equipment.
- Slipping/ Falling off a ladder.
- Not attaching fall protection.
- Work surface not in a suitable condition for task performed.
- Fall protection is damaged or defective.
- Scaffold erection/ dismantling.

All these root causes identified by the study have been successfully identified by SNA as well except for "collapse of structure/equipment" and "slipping/falling off a ladder". Again, these two causes do not directly correlate to corrective measures to be taken to avoid their repetition. "Collapse of structure or equipment" can be caused by multiple reasons including design flaws or workmanship, supervision, maintenance and inspection issues which are listed as root causes. “Slipping or falling off a ladder” is similar where ladder might have a defect, ladder might have not undergone suitable inspection, ladder might have been identified as defective but not properly labelled, ladder might have not been used properly by workforce or it might have not been fit for the work being done.

A study suggesting a different accident categorization to be used by OSHA was carried out by Hinze et al. (1998). The study concluded that equipment-related reasons resulted in 53% of “struck by” accidents and 59% of “caught in between” accidents. This trend suggested by the study was replicated in the SNA results which conformed with the study for the “struck by” network and the “caught in between” network. In the “struck by” network, the 20 root causes with highest centralities include 7 which are related to equipment. In “caught in between” networks, on the other hand, the 16 causes with highest centralities included 5 are related to equipment.

Four main root causes of electrocution were identified by Hinze et al. (1998) which are “contact with overhead power lines”, “contact with building power”, “problems with tools” and “problems with wiring”. “Contact with overhead power lines” and “contact with building power” are related to root causes (RC 9, RC 28, RC 43 and RC 60) in the “electrocution” network. “Problems with wiring” is directly related to RC 13. “Problems with tools” is directly related to RC 21.

SNA again was able to replicate most of the findings of previous studies in case of direct cause networks. The use of SNA however provided greater insights where it displayed the importance of job specific training to avoid all direct causes of accidents. In fall accidents, it revealed the potential of poor health and safety training, poor inspection of the jobsite and equipment, employee misconduct and poor labelling to cause accidents. In struck by fatalities, SNA highlighted the importance of clear employer instructions, inspection of the jobsite and health and safety training. For caught in between accidents, SNA found the potential of poor equipment handling and carrying out the job by incompetent personnel to cause accident. Finally, for electrocution accidents, SNA highlighted the causes of unclear directions given by employer, poor labeling as well as poor equipment and tool handling. The root causes which were successfully identified by SNA in spite of the fact that previous studies did not successfully identify them have common characteristics. These characteristics are that these causes cannot single handedly cause a fatal accident. However, combining these causes with other causes can significantly amplify their effect and significantly increase the probability of causing an accident. This is better captured by using SNA.

4.8 Statistical Analysis of Centralities in main network

To provide even further insight on the types of causes which have the largest number of cooccurrences in the main network, different types of root causes are statistically compared together. Root causes are grouped together into four main groups derived from the literature classifications of accident root causes. The first group is human related causes where human factors have the main influence in leading to the cause. The second group was organizational factors where the actions or lack of action within the organization in general or the safety management system in particular lead to the fatal accident. The third group was physical factors including any physical condition including the condition of the site, material, equipment and other conditions necessary for safe job completion which are absent. Finally, technical factors include inaccuracies in designs, drawings, calculations or other technical considerations which can lead to accidents.

To be able to compare weighted degree centralities of these four groups, the distribution for weighted degree centralities had to be identified. The distribution of degree centralities is commonly known as the degree distribution in the network and it is an important property of the network. The aim of finding the network's degree distribution is to confirm whether the centrality values are normally distributed. This is to decide whether the comparison of means should be done by regular analysis of variance or another non-parametric alternative.

4.8.1 Degree Distribution

The first step in this analysis is to achieve a better understanding for the distribution of degrees in the network so as to further understand network behavior. Figures 4.15 and 4.16 compare the network degree distribution to a normal distribution.

The shape of the degree distribution makes it obvious that the weighted degree centralities are not normally distributed. However, the data shows an elaborate curvature where the frequencies of lower weighted degrees are much higher than those at higher weighted degrees. This particular shape is characterized by what is called fat tails and provides indication of the importance of investigating the degree distribution more thoroughly.

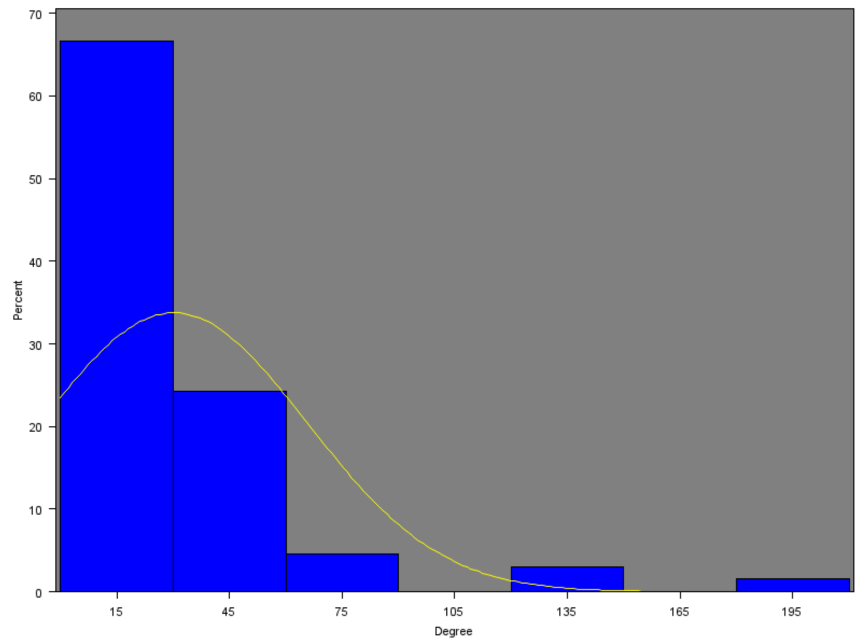

Figure 4.3: Histogram plot for weighted degree centralities

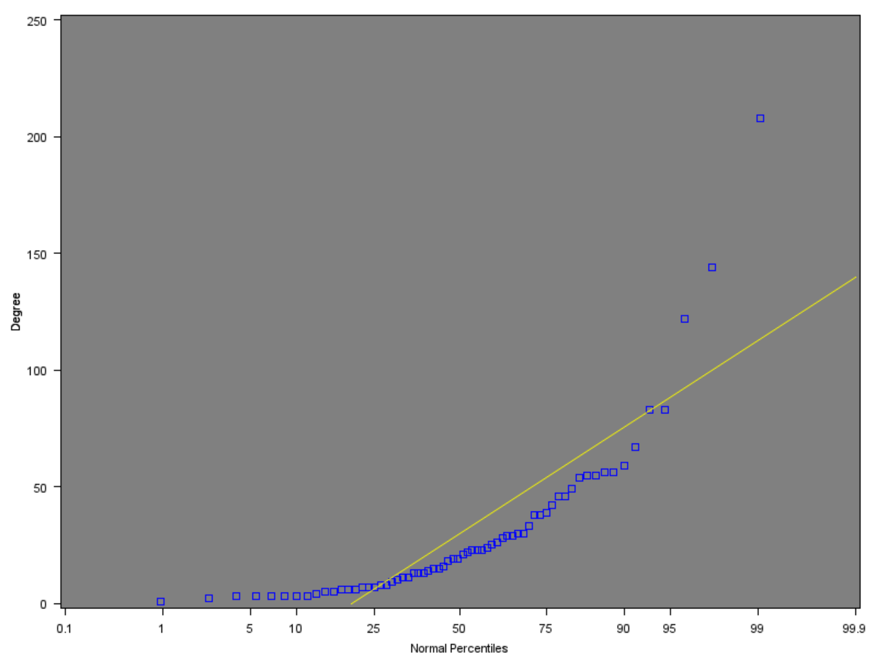

Figure 4.4: Normal probability plot for weighted degree centrality

Further investigation of the degree distribution is carried out by analyzing the distribution of the logs of the weighted degree values rather than the values themselves. This is to be able to know whether the weighted degree follows any power law. The results are shown in Figures 4.17 and 4.18 where when the logs of weighted degrees are analyzed, it is evident that they are normally distributed from the plots. The test for normality in Table 4.17 confirms the result with a large p value showing insufficient evidence to reject lack of fit.

The behavior which is exhibited by the weighted degrees suggest that they follow a scale free distribution and a power law. This behavior is observed in other phenomena in literature including analysis of links on the world-wide web (Albert et al. 1999). The scale free degree distribution can be expressed as:

$$P(d) = k * d^{-s}$$

This expression has P(d) which is the probability of a node (root cause) having a particular degree inversely proportional to the degree itself with a power s. This means that the larger the degree d, the less likely it is to have a node with this particular degree. This equation can be altered by taking the logs of both sides to be:

$$\log\big(P(d)\big) = \log(k) - s * log(d)$$

The expression above was used to fit a linear regression model of log(P(d)) against log(d) to be able to compute the values of log(k) and s respectively. Tables 4.18 and 4.19 show the model fit statistics and the parameter estimates for the degree distribution regression model.

The linear regression model showed a very good fit with R square value of 0.8848. and p values of 0.0354 and 0.0012 for the intercept and log of D respectively as shown in Tables 4.18 and 4.19 respectively. The very high significance of the coefficient of log D provides strong evidence that the hypothesized scale free distribution holds. According to this model, the probability of having a root cause of degree d in the network can be expressed as:

$$P(d) = 17.039088 * d^{-1.40308}$$

This model can be used to predict or the network for any given degree the probability that a node in the network can have this degree. That is, the proportion of the total nodes in the network with a degree similar to this degree. This representation is useful as it provides insight into the network properties and allows comparison with other networks.

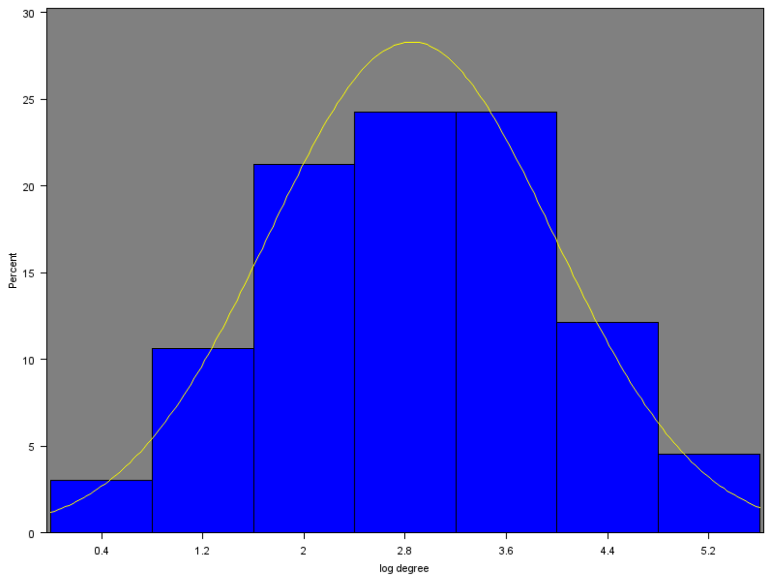

Figure 4.5: Histogram plot for log of weighted degree centralities

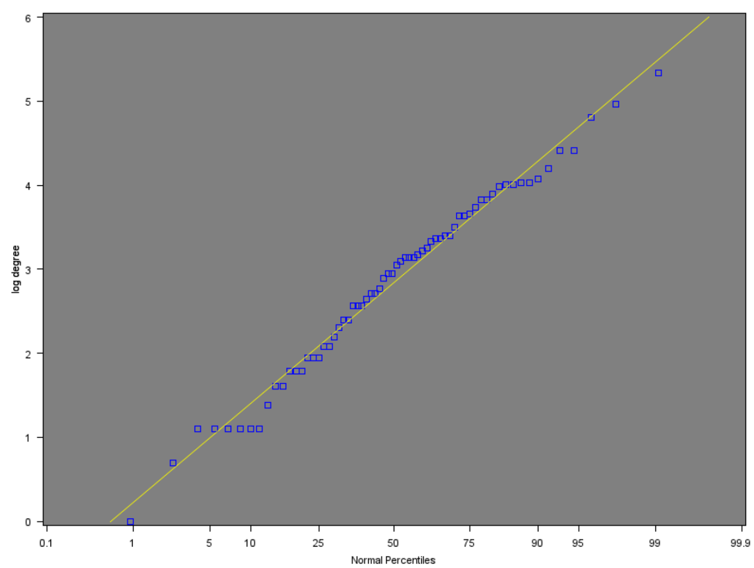

Figure 4.6: Normal probability plot for log of weighted degree centrality

Table 4.17: Test for normality for log of weighted degrees

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	W	0.985368	Pr < W	0.6294
Kolmogorov-Smirnov	D	0.070513	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.057648	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.371166	Pr > A-Sq	>0.2500

Table 4.18: Degree distribution model fit statistics

Root MSE	0.62633
Dependent Mean	-3.024
Coeff Var	-20.712
R-Square	0.8448
Adj R-Sq	0.8189

Table 4.19: Degree distribution parameter estimates

Variable	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	2.83551	1.049	2.7	0.0354
log D	-1.40308	0.24553	-5.71	0.0012

4.8.2 Comparison of means

After identifying the distribution of degrees in the network, the means of the 4 identified groups (human factors, physical factor, organizational factors and technical factors) are compared. Since the weighted degrees follow a scale free distribution, a non-parametric test (Wilcoxon Ranked Sum Test) is used for the comparison of the mean weighted degree for the four groups. The results of the test are shown in Table 4.20.

The Wilcoxon ranked sum test statistics in Table 4.20 shows no difference in means between the 4 groups with a very large p value of 0.3653. This means that there is no sufficient evidence given by the data that the means of all four groups are not equal. The results of the test agree with the boxplot shown in Figure 4.19, where the difference between the means in the four groups are very minor. The conclusion that can be drawn from this analysis is that there is not sufficient evidence given to support that either human, organizational, physical or technical factors have the highest centrality in the network. This is in line with previously mentioned accident causation theories which categorize all these factors as important factors whose interaction leads to accidents.

4.9 Summary of findings

The social network analysis is carried out for the network of root causes for the 100 fatalities investigated as well as the seven subnetworks obtained by breaking down the main network by direct cause. Centrality analysis yields that it is not failures in sophisticated elements of the safety management system which cause fatal accidents but rather failures in the most basic elements.

Centrality analysis yields shortage of job-specific training to be the most central root cause in the main network, struck by network and caught in between network making it a very important cause to address.

Key factors underlying the occurrence of fatal accidents in the main network according to weighted degree centrality are found to be employers not willing to invest in employees, poor enforcement of safety rules, cutting down on safety budget, lack of training, poor communication and deficiency of basic procedural site operations. They also include poor safety culture for the employer and employees, lack of hazard awareness and procedural and technical human errors.

Table 4.20: Wilcoxon ranked sum test statistics

Wilcoxon Scores (Rank Sums)					
	N	**Sum of Scores**	**Expected under H0**	**Std Dev under H0**	**Mean Score**
human	15	544.5	502.5	65.3118	36.3
organizational	26	949.5	871	76.1514	36.5192
physical	12	306	402	60.1102	25.5
technical	13	411	435.5	61.9827	31.6154
Kruskal-Wallis Test					
Chi-Square	3.1759				
DF	3				
Pr > Chi-Square	0.3653				

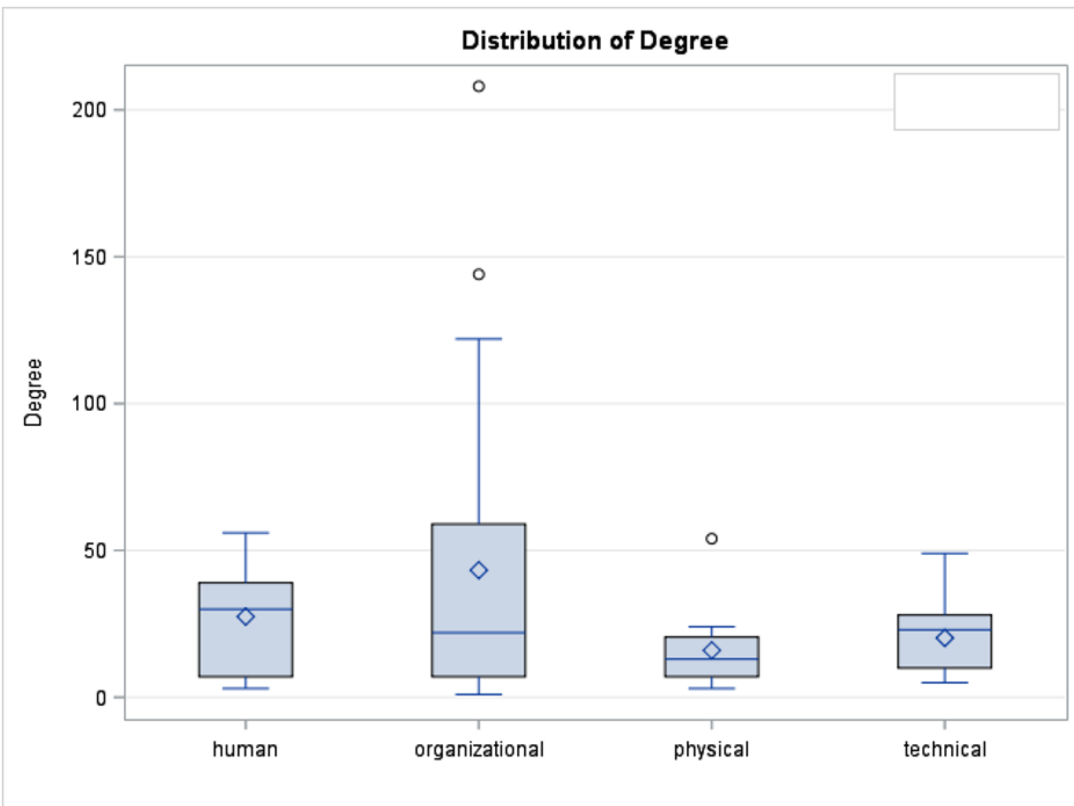

Figure 4.7: Box and whisker plots for weighted degrees of the four groups

Eigenvector centrality yields similar results to degree centrality and betweenness centrality brings focus on site conditions as a key underlying factor that could lead to different scenarios of fatal accidents. The weighted degree is chosen to be the reference centrality where it answers the question if a particular root cause is addressed, how much of the network connectedness is reduced.

For the seven subnetworks, analysis of the fall network yields that reduced employer spending, poor supervision and safety rule enforcement, lack of training, lack of inspection and communication issues are key underlying factors leading to falls. Struck by hazards on the other hand were mostly associated with equipment operations. Key underlying factors leading to struck by fatalities were human errors both procedural and technical. Additionally, lack of training and inspection and poor site conditions were among the factors leading to Struck by fatalities.

Caught in between hazards were similar to struck by hazards where technical and procedural human errors were key underlying factors leading to these hazards. These hazards though put more emphasis on work site condition and layout as a key factor. Reduced employer spending and valuation for the employees is also a deficiency leading to caught in between hazards. Not having regular inspection operations are also a key organizational deficiency leading to struck by hazards.

Electrocution hazards were mainly attributed to human related factors, mainly procedural errors. Employer lack of involvement, lack of training and poor site conditions were also found to be amongst the key factors. Table 4.21 summarizes the key root causes in the main network as well as the four subnetworks corresponding to the fatal four direct causes. The 3 centrality rankings shown in the table are those of degree centrality denoted by D, flow betweenness centrality denoted by FB and Bonacich Power/eigenvector centrality denoted by BP.

The network for asphyxia on the other hand showed that it happens mainly due to human related faults of individuals other than the victim which are either procedural or technical. Work site conditions and inspection are also important factors in this network. Burns and explosions were found to result from a multitude of organizational failures which are mostly attributed to the employer's lack of involvement and lack of spending. These span areas of training and communication. Health related networks was shown to resemble heart attack cases from which not much conclusions could be drawn.

In terms of density, all networks were relatively dense with asphyxia and burn/explosion networks being the densest and stuck by network being the least dense. Moreover, struck by caught in

Table 4.21: Summary of centrality rankings for main network and fatal four direct causes

RC #	Root Cause	Main			Fall			Struck by			Caught in between			Electrocution		
		D	FB	BP	D	FB	BP	D	FB	BP	D	FB	BP	D	FB	BP
43	Lack of specific on the job training	1	1	1	3	2	3	1	1	1	1	1	1	2	1	2
45	No jobsite inspection	5	3	6	8	9	8	2	2	2	3	2	4			
44	Lack of general health and safety Training	4	4	4	4	5	4	7	7	10				8		8
9	Poor Labeling	6	6	5			9	11	8	13	7		7	6		7
11	No PPE	3	9	3	2	4	2					7		3	4	6
3	Poor Equipment handling	8	15	12				4	5	4	2	4	2			
12	No fall arrest system, guardrails or safety nets	2	2	2	1	1	1		9						6	
60	Not following proper work procedures	15	5					6	3	9				5	5	3
39	Employee misconduct	9	11	7	5	6	7		14						8	
30	Lack of clear employer instructions	7	13	14				3	11	3	11		11			
46	Lack of Inspection for equipment and tools	11	10	10	9			14	10		9		9			
28	Employer allowed employee to work in an unsafe environment	10	12	9								10		1	2	1
1	Damaged/defective equipment	12		13				9		8	4	11	3			
64	Noncompliance to equipment manufacturer recommendations							8	13	6	5	3	5			
37	Operation carried out by noncompetent individual		7					13		12	6	9	6			
53	Poor housekeeping		14					15	4	15	10		10			
47	Poor Assembling of equipment/ scaffold/ decking/ formwork	13		8	6	8	5									
18	No decking			11	7	7	6									
21	Poor tool handling													4	3	4
16	No/ damaged Cave in protection		8					5		5						
33	No competent person on site							12	12	14						
13	No cable insulation													7		5
23	Poor Material Handling										8		8			
4	Use of unsuitable equipment							10		7						
35	Lack of supervision	14		15												
19	Lack of employee knowledge					3										
14	Not wearing seatbelt											5				
48	Error in design								6							
54	No safe exit to site											6				
22	Inappropriate tools used														7	
6	Equipment/spoil on excavation edge											8				
5	Equipment malfunction									11						
2	Lack of necessary Equipment								15							

between and fall are the most centralized networks, whereas burn/explosion are the least centralized. All networks do not display any homophily.

An expression for the degree distribution of the main network was found. A statistical comparison of means of human related, organizational, physical and technical factors found no statistically significant difference in centrality between any of the four groups of root causes.

CHAPTER 5:
CONCLUSION AND FUTURE RECOMMENDATIONS

5.1 Conclusion

Fatal accidents are still amongst the key failures impeding the US Construction industry. They are indicators of safety for which other countries are ahead of the US and other industries are ahead of construction. This study harnesses the analytical power of social network analysis to identify key patterns and relationships amongst root causes of fatalities in construction.

SNA utilization facilitated including the interrelations amongst accident root causes which enhanced the analysis enabling the study to identify significant root causes not previously identified using other techniques. The importance of these cause identified solely using SNA stems from that their combination with other causes significantly increases the probability of fatal accidents. SNA allowed inclusion of both root causes stemming from downstream conditions on site as well as others which are more upstream including design related, procedural and organizational root causes. SNA, in that sense, allowed the analysis to be more comprehensive as well as more in depth. It also provided metrics which were objective allowing quantitative methods to be used to prioritize root causes as well as devise the suitable safety management practices to address them efficiently and effectively. Finally, use of SNA enabled identifying 4 root cause pairings which lead to most fatal accidents when they occurred at the same time. All causes in these pairs were among the top centrality causes which shows the effectiveness of SNA in capturing the relationships between causes.

Since centrality is one of the main metrics, comparison is done between the different centrality measures in the context of construction safety. This comparison yields that weighted degree centrality is the most representative of the reduction of network connectedness when a cause is removed from the network which is central to this study. Eigenvector centrality and flow betweenness are indicative of the connectedness of a particular cause with other well-connected causes and its contribution to diverse scenarios of accidents when combined with other causes.

Job-specific training was identified as the cause with the highest centrality in all the analyzed accidents. Additional most central causes are lack of safety nets, guardrails and fall protection equipment, not using PPEs, absence of generic health and safety training and lack of labelling hazards on site.

The analysis of the main network yields that it is a compact structure of cooccurrences between root causes with a small diameter and large density with a scale free degree distribution following a power law or which the equation was provided. Moreover, the main network structure can be described as centralized where most of the centrality in the network is possessed by a small proportion of the causes. This means that the centrality and thus the accident occurrences are easily reducible by addressing the key root causes. This was supported by much of the connectivity in the network being lost by dichotomizing. The network displayed no signs of transitivity.

This analysis is complemented with comparative analysis of subnetworks of root causes contributing to a particular direct cause. The subnetworks, like the main network, showed dense compact structures where "asphyxia", "burn" and "fall" networks had the densest structures by weighted density. This means that these networks on average have each root cause more connected to other causes in the network which suggests that, on average, addressing causes in these networks is more efficient. Such consideration can be particularly useful given time and cost limitation to safety improvement.

"No specific on the job training" was found to be the most prominent cause of fatal accidents in the US, where it is the highest ranked cause in "main" network, "struck by" network and "caught in between" network. This study distinguishes between the generic safety training and the more specialized job specific training to carry out the job safely and effectively. This cause is of particular importance since it is ranked as the top cause in the main network, struck by network and caught in between network. The high centrality of this cause owes to the fact that many of the investigated fatalities involved workers who were not trained to do their jobs in a proper safe manner. This can be attributed to reluctance of employers to train workers on proper safe work practices either out of lack of budget, or time, or simply presuming that they already know the safe job-specific work practices while it is not the case. This cause also contributes to three of the four key root cause pairings causing the majority of accidents. The high centrality of job specific training suggests that workers who haven't had the necessary training to understand their own job and its associated hazards and safety precautions are much more prone to accidents.

"Employer allowing employees to work in unsafe environment" (particularly near energized circuits in this case) is the key cause for electrocution. "Asphyxia" key cause is "no or damaged cave in protection". "Burns/explosions" result from "damaged or defective cave in protection" and "heart attacks" are the main cause for "Health related" fatalities. In terms of degree centralization, “caught in between” network is the most centralized followed by “struck by” then “fall”. The large centralization means there is a large gap in centrality between most important and least important causes in the same networks. In other words, it means that addressing only few of the top causes in “caught in between”, “fall” and “struck by” networks can significantly reduce probability of accident occurrence. This is since these few causes are connected to most other causes in the network. None of the networks shows strong evidence of transitivity.

One important outcome from this analysis is that fatal accidents are mostly related to the simplest and least sophisticated concepts related to safety. This is highlighted by the fact that the root causes of highest centrality have very simple countermeasures. Inferences obtained from the network data indicated that the key root causes in the networks can be linked together with a set if key underlying factors. Reduced safety expenditure, slack of safety regulation enforcement on the worksite, poor coordination and limited visual, verbal and written safety-related communication are among these factors. Another key underlying factor is defective site operations planning and enforcement. This applies to different types of operations which should be systematically carried out but aren’t like supervision, inspection of equipment and tools, jobsite inspection and regular workforce supervision. The key root causes identified by SNA also highlight that one of the underlying factors is deficient safety culture for employers and workers. Employers lack the essential trait of valuing their workforce by spending then necessary money for supplying them with training and with safety equipment to protect them as well as establishing lines of communication focusing on worker welfare with clear instructions. On the contrary, workers also have a poor culture and attitude towards safety. Some of the causes of fatalities included workers not showing proper conduct, failing to follow the correct safe procedure for work execution and intentionally subjecting themselves to hazards which can be life threatening. This raises a flag that laws should be put in place to address minimum requirements for employers concerning the standard of safety on their worksites as well as employers training and educating their workforce. If every employer can educate and train his workforce better, this can enhance the workers’ safety culture and hazard awareness.

SNA successfully identifies key root causes of fatal accidents as well as their relationships and interactions from a holistic perspective. This perspective can complement the current top down approach to trace root causes down from the accidents which relies on accident causation models. This can be achieved since the network approach aggregates those causes found using top down approach in a network and tries to reverse the process by contemplating how they combine to cause accidents. This bottom up approach enriches the analysis by adding to it a different dimension which is the relationships between the causes which is a different dimension added to the analysis. Furthermore, this analysis shows an example of breaking down the network by direct cause which is a versatile technique. The versatility of this technique stems from its ability to break down networks of accident root causes not only by direct causes but other categorization criteria such as relevant safety management policies or accident type for a more in-depth analysis.

5.2 Limitations

Research findings are particularly limited to the US construction industry. This is particularly true since the data gathered is based on the US industry. Moreover, the adopted methodology studies accident root causes through their interrelations rather than through their frequency of occurrence. These interrelations may vary from country to country based on human culture and traditions making it less useful to generalize results from the US construction industry to other parts of the world.

Another key limitation of the SNA model includes that it doesn't include any factors pertaining to social and cultural backgrounds of the workers and their influence on their behavior. The culture in this context does not refer to safety culture but rather to heritage, ancestry and traditions in a broader perspective. This limitation is due to that the fatal accident case files do not provide background information relevant to any social or cultural background of the victim or other workers. Obtaining this information will thus include a significant amount inference and will have high uncertainty associated with it. Accordingly, this information could not be reliably obtained even though its presence could have led to a significant enrichment to the model.

5.3 Recommendations for Future Work

The use of social network analysis for modelling root causes of fatal accidents is a new innovative approach which has not been attempted before. Therefore, there is a lot of room for applying this analytical methodology in different ways to benefit from its analytical power as well as maximize its own potential. This is particularly true since as previously noted, the industry's forward leap in using zero incident techniques and safety leading indicators including near miss reporting should be complemented with compatible analytical tools.

This study was able to identify the degree distribution or the network and express it numerically. However, since it is limited to a single network and other subset networks, this could not be compared with other degree distributions. Future work can include assessing degree distributions in accident networks from various organizations and comparing them to be able to interpret the effect of the degree distribution on the root cause network properties.

Furthermore, some of the SNA techniques which can be useful in safety domain aren’t covered by this research. These techniques include clustering and homophily. Clustering is a metric which can divide the networks to subdivisions which are more strongly connected to one another. Homophily, on the other hand, reveals if causes which have certain characteristics in common are more likely to be connected to each other. Utilizing these techniques in the domain of accident causation analysis can provide a promising area of research.

One other SNA application which can be of value for future research is modelling networks dynamically instead of statically. This can be used as a performance assessment tool in the area of safety management where the network temporal development can be observed and changes in the network can potentially serve as an indicator of performance.

Finally, it is recommended to use social network analysis as a selection, analysis and evaluation tool for safety leading indicators. The benefit of using social network analysis with leading indicators is twofold. First, the use of social network analysis not only identifies but provides a quantitative measure for cooccurrence of particular accident root causes. This can be utilized in the selection stage of leading indicators to select the appropriate leading indicators. This can be done by identifying the number of root causes which the leading indicator addresses, their centralities and the cost for measurement of this leading indicator. This cost could be monetary or non-monetary. Secondly, it facilitates the leading indicator performance evaluation. The leading indicator being attached to particular root causes means that the change in centrality in these particular root causes can be directly attributed to the leading indicator. This is particularly true since most of the literature reviewed about leading indicators assess their performance based on either lagging indicator such as TRIR and FA or expert judgement.

LIST OF REFERENCES

Abdelhamid, T., and Everett, J. (2000) . “Identifying root causes of construction accidents.” *J. Constr. Eng. Manage.*, 126 (1), 52–60.

Albert, R., Jeong, H., and Barabási, A.-L. (1999). "Internet: Diameter of the world-wide web." *nature*, 401(6749), 130-131.

Albert, A., and Hallowell, M. (2014). "Modeling the role of social networks in situational awareness and hazard communication." *Proc., Construction Research Congress*, 1752-1761.

Ale, B. J. M. (2006). “ The occupational risk model. ” *Final Rep.*, ORM, TUDelft, Netherlands.

Aljassmi, H., Han, S., and Davis, S. (2013). "Project pathogens network: New approach to analyzing construction-defects-generation mechanisms." *Journal of Construction Engineering and Management*, 140(1), 04013028.

Alsamadani, R., Hallowell, M., and Javernick-Will, A. N. (2013a). "Measuring and modelling safety communication in small work crews in the US using social network analysis." *Construction Management and Economics*, 31(6), 568-579.

Alsamadani, R., Hallowell, M. R., Javernick-Will, A., and Cabello, J. (2013b). "Relationships among language proficiency, communication patterns, and safety performance in small work crews in the United States." *Journal of Construction Engineering and Management*, 139(9), 1125-1134.

Aneziris, O. N., et al. (2008). “ Quantified risk assessment for fall from height. ” *Safety Sci.*, 46(2), 198 – 220.

Aral, S., and Van Alstyne, M. (2011). "The diversity-bandwidth trade-off." *American Journal of Sociology*, 117(1), 90-171.

Aral, S., and Walker, D. (2012). "Identifying influential and susceptible members of social networks." *Science*, 337(6092), 337-341.

Ayeko, M. "Integrated Safety Investigation Methodology (ISIM)-Investigating for Risk Mitigation." *Proc., Proceedings of the First Workshop on the Investigation and Reporting of Incidents and Accidents (IRIA 2002)*, Department of Computing Science, University of Glasgow, Scotland.

Barrat, A., Barthelemy, M., Pastor-Satorras, R., and Vespignani, A. (2004). "The architecture of complex weighted networks." *Proceedings of the National Academy of Sciences of the United States of America*, 101(11), 3747-3752.

Barthelemy, M. (2004). "Betweenness centrality in large complex networks." *The European Physical Journal B-Condensed Matter and Complex Systems*, 38(2), 163-168.

Beavers, J., Moore, J., and Schriver, W. (2009). "Steel erection fatalities in the construction industry." *Journal of Construction Engineering and Management*, 135(3), 227-234.

Benner, L. (1975). "Accident investigations: Multilinear events sequencing methods." *Journal of safety research*, 7(2), 67-73.

Bird, F. E., and Germain, G. L. (1966). *Damage Controle*, American Management Assoc.

Bonacich, P. (1972). "Factoring and weighting approaches to status scores and clique identification." *Journal of Mathematical Sociology*, 2(1), 113-120.

Bonacich, P. (1987). "Power and centrality: A family of measures." *American journal of sociology*, 1170-1182.

Bonacich, P. (2007). "Some unique properties of eigenvector centrality." *Social networks*, 29(4), 555-564.

Borgatti, S. P., and Everett, M. G. (1997). "Network analysis of 2-mode data." *Social networks*, 19(3), 243-269.

Borgatti, S. P. (2002). "NetDraw: Graph visualization software." *Harvard: Analytic Technologies*.

Borgatti, S. P., Everett, M. G., and Freeman, L. C. (2002). "Ucinet for Windows: Software for social network analysis."

Borgatti, S. P. (2009). "2-Mode concepts in social network analysis." *Encyclopedia of complexity and system science*, 6.

Borgatti, S. P., Mehra, A., Brass, D. J., and Labianca, G. (2009). "Network analysis in the social sciences." *Science*, 323(5916), 892-895.

Borgatti, S. P., Everett, M. G., and Freeman, L. C. (2014). "Ucinet." *Encyclopedia of Social Network Analysis and Mining*, Springer, 2261-2267.

Brookes, N., Morton, S., Dainty, A., and Burns, N. (2006). "Social processes, patterns and practices and project knowledge management: A theoretical framework and an empirical investigation." *International Journal of Project Management*, 24(6), 474-482.

Center for Construction Research and Training (CPWR) (2013). *The Construction Chart Book: The U.S. Construction Industry and its Workers*, Silver Spring, Maryland.

Chan, A. P., Javed, A. A., Wong, F. K., Hon, C. K., Zahoor, H., and Lyu, S. "The application of social network analysis in the construction industry of Hong Kong." *Proc., Abstract Book of 1st International Conference on Emerging Trends in Engineering, Management & Sciences*, City University of Science & Information Technology (CUSIT), 31.

Cheng, Y.-Y., Lee, R. K.-W., Lim, E.-P., and Zhu, F. (2015). "Measuring centralities for transportation networks beyond structures." *Applications of Social Media and Social Network Analysis*, Springer, 23-39.

Chinowsky, P., Diekmann, J., and Galotti, V. (2008). "Social network model of construction." *Journal of Construction Engineering and Management*, 134(10), 804-812.

Chinowsky, P. S., Diekmann, J., and O'Brien, J. (2009). "Project organizations as social networks." *Journal of Construction Engineering and Management*, 136(4), 452-458.

Chinowsky, P., Taylor, J. E., and Di Marco, M. (2010). "Project network interdependency alignment: New approach to assessing project effectiveness." *Journal of Management in Engineering*, 27(3), 170-178.

Cipolla, D., Biggs, H. C., Dingsdag, D., and Kirk, P. J. (2009). *Safety effectiveness indicators project workbook*, Cooperative Research Centre for Construction Innovation, Brisbane, Australia.

Construction Industry Institute (CII) (1993). *Zero injury techniques 32-1*. Austin, TX.

Construction Industry Institute(CII) (2003). *Safety plus: Making zero accidents a reality 160-1*. Austin, TX.

Construction Industry Institute (CII) (2012). *Implementing Active Leading Indicators 284-2*. Austin, TX.

Construction Industry Institute (CII) (2013). *Going Beyond Zero Using Safety Leading Indicators 284-11.* Austin, TX.

Construction Industry Institute (CII) (2014). *Using Near Miss Reporting to Enhance Safety Performance. 301-11.* Austin, TX.

Cook, K. S., Emerson, R. M., Gillmore, M. R., and Yamagishi, T. (1983). "The distribution of power in exchange networks: Theory and experimental results." *American journal of sociology*, 275-305.

Cranmer, S. J., and Desmarais, B. A. (2011). "Inferential network analysis with exponential random graph models." *Political Analysis*, 19(1), 66-86.

D'Exelle, B., and Holvoet, N. (2011). "Gender and network formation in rural Nicaragua: a village case study." *Feminist Economics*, 17(2), 31-61.

Di Marco, M. K., Taylor, J. E., and Alin, P. (2010). "Emergence and role of cultural boundary spanners in global engineering project networks." *Journal of Management in Engineering*, 26(3), 123-132.

El-adaway, I. H., Abotaleb, I. S., and Vechan, E. (2016). "Social Network Analysis Approach for Improved Transportation Planning." *Journal of Infrastructure Systems*, 05016004.

Everett, J. G., and Jr, P. B. F. (1996). "Costs of accidents and injuries to the construction industry." *Journal of Construction Engineering and Management*, 122(2), 158-164.

Ferry, T.S. (1988). *Modern Accident Investigation and Analysis*, John Wiley & Sons, New York.

Fowler, J. H. (2006). "Connecting the Congress: A study of cosponsorship networks." *Political Analysis*, 14(4), 456-487.

Freeman, L. C. (1978). "Centrality in social networks conceptual clarification." *Social networks*, 1(3), 215-239.

Freeman, L. C., Borgatti, S. P., and White, D. R. (1991). "Centrality in valued graphs: A measure of betweenness based on network flow." *Social networks*, 13(2), 141-154.

Friedkin, N. E. (1991). "Theoretical foundations for centrality measures." *American journal of Sociology*, 1478-1504.

Friedman, S. R., Neaigus, A., Jose, B., Curtis, R., Goldstein, M., Ildefonso, G., Rothenberg, R. B., and Des Jarlais, D. C. (1997). "Sociometric risk networks and risk for HIV infection." *American Journal of Public Health*, 87(8), 1289-1296.

Gomez Rodriguez, M., Leskovec, J., and Krause, A. (2010)."Inferring networks of diffusion and influence." *Proc., Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 1019-1028.

Grabowski, M., Ayyalasomayajula, P., Merrick, J., Harrald, J. R., and Roberts, K. (2007). "Leading indicators of safety in virtual organizations." *Safety Sci*., 45(10), 1013–1043.

Guo, B. H. W., and Yiu, T. W. (2016). "Developing Leading Indicators to Monitor the Safety Conditions of Construction Projects." *Journal of Management in Engineering*, 32(1), 04015016.

Hale, A. R., and Glendon, A. I. (1987). *Individual behaviour in the control of danger*, Elsevier Science.

Hallowell, M. R., Hinze, J. W., Baud, K. C., & Wehle, A. (2013). Proactive construction safety control: Measuring, monitoring, and responding to safety leading indicators. *Journal of Construction Engineering and Management*, *139*(10), 04013010.

Hanneman, R. A., and Riddle, M. (2005). "Introduction to social network methods." University of California Riverside.

Heinrich, H. W. (1969). *Industrial accident prevention*, 4th Ed.,McGraw-Hill, New York.

Hinze, J., Pedersen, C., & Fredley, J. (1998). Identifying root causes of construction injuries. *Journal of Construction Engineering and Management*, *124*(1), 67-71.

Hinze, J., and Wilson, G. (2000). "Moving toward a zero injury objective." *Journal of Construction Engineering and Management*, 126(5), 399-403.

Hinze, J. (2005). "Practices that Influence Safety Performance on Power Plant Outages." *Practice Periodical on Structural Design and Construction*, 10(3), 190-194.

Hinze, J., Hallowell, M., and Baud, K. (2013a). "Construction-safety best practices and relationships to safety performance." *Journal of Construction Engineering and Management*, 139(10), 04013006.

Hinze, J., Thurman, S., and Wehle, A. (2013b). "Leading indicators of construction safety performance." *Safety Science*, 51(1), 23-28.

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). "Latent space approaches to social network analysis." *Journal of the american Statistical association*, 97(460), 1090-1098.

Hollnagel, E. (1998). *Cognitive reliability and error analysis method (CREAM)*, Elsevier Science.

Hollnagel, E. (2002)."Understanding accidents-from root causes to performance variability." *Proc., Human factors and power plants, 2002. proceedings of the 2002 ieee 7th conference on*, IEEE, 1-1.

HSE and Chemical Industries Association (2006). "Developing process safety indicators. A step-by-step guide for chemical and major hazard industries." *Health and Safety Executive*.

Ibarra, H., and Andrews, S. B. (1993). "Power, social influence, and sense making: Effects of network centrality and proximity on employee perceptions." *Administrative science quarterly*, 277-303.

Ikpe, E., Hammon, F., and Oloke, D. (2012). "Cost-Benefit Analysis for Accident Prevention in Construction Projects." *Journal of Construction Engineering and Management*, 138(8), 991-998.

Jacinto, C., and Aspinwall, E. (2003). "Work accidents investigation technique (WAIT)–Part I." *Safety Science Monitor*, 7(1), 1-17.

Jackson, M. O., and López-Pintado, D. (2013). "Diffusion and contagion in networks with heterogeneous agents and homophily." *Network Science*, 1(01), 49-67.

Janicak, C. A. (1998). Fall-related deaths in the construction industry. *Journal of Safety Research*, *29*(1), 35-42.

Johnson, W.G. (1980). *MORT Safety Assurance Systems,* Marcel Dekker Inc, New York.

Jones, S., Kirchsteiger, C., and Bjerke, W. (1999). "The importance of near miss reporting to further improve safety performance." *Journal of Loss Prevention in the process industries*, 12(1), 59-67.

Katsakiori, P., Sakellaropoulos, G., and Manatakis, E. (2009). "Towards an evaluation of accident investigation methods in terms of their alignment with accident causation models." *Safety Science*, 47(7), 1007-1015.

Katz, L. (1953). "A new status index derived from sociometric analysis." *Psychometrika*, 18(1), 39-43.

Kim, Y., Choi, T. Y., Yan, T., and Dooley, K. (2011). "Structural investigation of supply networks: A social network analysis approach." *Journal of Operations Management*, 29(3), 194-211.

Kingston, J. (2007). "3CA–Investigator’s manual, NRI-3."

Kirkland, J. H. (2011). "The relational determinants of legislative outcomes: Strong and weak ties between legislators." *The Journal of Politics*, 73(03), 887-898.

Kjellén, U., Larsson, T.J., 1981. Investigating accidents and reducing risks – a dynamic approach. Journal of Occupational Accidents 3, 129–140.

Kjellén, U. (2000). *Prevention of accidents through experience feedback*, CRC Press.

Kjellén, U. (2009). " The safety measurement problem revisited. " *Safety Sci.*, 47(4), 486 – 489.

Krafft, P., Moore, J., Desmarais, B., and Wallach, H. M. (2012). "Topic-partitioned multinetwork embeddings." *Proc., Advances in Neural Information Processing Systems*, 2807-2815.

Krivitsky, P. N., Handcock, M. S., Raftery, A. E., and Hoff, P. D. (2009). "Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models." *Social networks*, 31(3), 204-213.

Leplat, J. (1978). "Accident analysis and work analysis." *Journal of Occupational Accidents*, 1, 331–340.

Levin, D. Z., and Cross, R. (2004). "The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer." *Management science*, 50(11), 1477-1490.

Liu, G., Wong, L., and Chua, H. N. (2009). "Complex discovery from weighted PPI networks." *Bioinformatics*, 25(15), 1891-1897.

Loosemore, M. (1999). "Responsibility, power and construction conflict." *Construction Management & Economics*, 17(6), 699-709.

Manuele, F. (2009). "Leading and lagging indicators." *Professional Safety,* 54 (12), 28–33.

Marks, E., Teizer, J., and Hinze, J. (2014). "Near-miss reporting program to enhance construction worker safety performance." *Proc., Construction Research Congress*.

Marwell, G., Oliver, P. E., and Prahl, R. (1988). "Social networks and collective action: A theory of the critical mass. III." *American Journal of Sociology*, 502-534.

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). "Birds of a feather: Homophily in social networks." *Annual review of sociology*, 415-444.

Mead, S. P. (2001). "Using social network analysis to visualize project teams." *Project Management Journal*, 32(4), 32-38.

Mengolinim, A., Debarberis, L., 2008. Effectiveness evaluation methodology for safety processes to enhance organizational culture in hazardous installations. Journal of Hazardous Materials 155, 243–252.

Moreno, J. L., and Jennings, H. H. (1934). *Who shall survive?,* Nervous and mental disease publishing company, Washington, DC.

Moreno, J. L., and Jennings, H. H. (1960). *The sociometry reader,* The Free Press of Glencoe, Glencoe, Illinois.

Nayak, N. V., and Taylor, J. E. (2009). "Offshore outsourcing in global design networks." *Journal of management in engineering*, 25(4), 177-184.

Newman, M. E. (2001a). "Scientific collaboration networks. I. Network construction and fundamental results." *Physical review E*, 64(1), 016131.

Newman, M. E. (2001b). "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality." *Physical review E*, 64(1), 016132.

Newman, M. E. (2004). "Analysis of weighted networks." *Physical review E*, 70(5), 056131.

Occupational Safety and Health Administration (OSHA) (2002). *Job Hazard Analysis 3071*.

Occupational Safety and Health Administration. (2017). "Commonly Used Statistics." <https://www.osha.gov/oshstats/commonstats.html > (Mar. 30, 2017),

Øien, K. (2001). " Risk control of offshore installations. A framework for the establishment of risk indicators. " Ph.D. dissertation, Univ. of Science and Technology (NTNU), Trondheim, Norway.

Olken, F. (1993). "Random sampling from databases." Ph.D. dissertation, University of California at Berkeley, California.

Opsahl, T., Agneessens, F., and Skvoretz, J. (2010). "Node centrality in weighted networks: Generalizing degree and shortest paths." *Social networks*, 32(3), 245-251.

Otte, E., and Rousseau, R. (2002). "Social network analysis: a powerful strategy, also for the information sciences." *Journal of information Science*, 28(6), 441-453.

Petersen, D. (1971). *Techniques of safety management*. McGraw-Hill, New York.

Pow, J., Gayen, K., and Elliott, L. (2012). "Understanding complex interactions using social network analysis." *Journal of Clinical Nursing*, Vol. 21, No. 19, pp 2772-2779

Priven, V., and Sacks, R. (2015). "Effects of the last planner system on social networks among construction trade crews." *Journal of Construction Engineering and Management*, 141(6), 04015006.

Pryke, S. D. (2005). "Towards a social network theory of project governance." *Construction Management and Economics*, 23(9), 927-939.

Raftery, A. E., Niu, X., Hoff, P. D., and Yeung, K. Y. (2012). "Fast inference for the latent space network model using a case-control approximate likelihood." *Journal of Computational and Graphical Statistics*, 21(4), 901-919.

Rajendran, S. (2012). "Enhancing Construction Worker Safety Performance Using Leading Indicators." *Practice Periodical on Structural Design and Construction*, 18(1), 45-51.

Rasmussen, J. (1987). "Cognitive Control and Human Error Mechanisms." *New Technology and Human Error, pp. 53-61.*

Rasmussen, J. (1997). “Risk management in a dynamic society: A modeling problem*.” Safety Sci.,* 27(2/3), 183–213.

Reason, J. T. (1990). *Human error*, Cambridge University Press, Cambridge, UK.

Reason, J. (1997). *Managing the Risks of Organisational Accidents*, Ashgate Publishing Ltd, Aldershot Hants.

Rubulotta, E., Ignaccolo, M., Inturri, G., and Rofè, Y. (2012). "Accessibility and centrality for sustainable mobility: regional planning case study." *Journal of Urban Planning and Development*, 139(2), 115-132.

Russo, T. C., and Koesten, J. (2005). "Prestige, centrality, and learning: A social network analysis of an online class." *Communication Education*, 54(3), 254-261.

Sabidussi, G. (1966). "The centrality index of a graph." *Psychometrika*, 31(4), 581-603.

SAS Institute Inc. Copyright (c) 2002-2012. SAS for Windows, Version 9.4. Cary, NC: SAS Institute Inc.

Schoen, M. W., Moreland-Russell, S., Prewitt, K., and Carothers, B. J. (2014). "Social network analysis of public health programs to measure partnership." *Social Science & Medicine*, 123, 90-95.

Shappell, S. A., and Wiegmann, D. A. (1997). “ A human error approach to accident investigation: The taxonomy of unsafe operations. ” *Int. J. Aviat. Psychol*., 7(4), 269 – 291.

Shappell, S. A., and Wiegmann, D. A. (2000). “ The human factors analysis and classification system — HFACS. ” *Rep. No. DOT/FAA/AM-00/7*, Federal Aviation Administration, Washington, DC.

Skriver, J., Haukenes, H., Alme, I. (2003). “Accident investigation at Norwegian State Railways: a sociotechnical methodology.” Proc., JRC/ESReDA Seminar on Safety Investigation of Accidents, Petten, Netherlands, 170–176.

Sparrowe, R. T., and Liden, R. C. (2005). "Two routes to influence: Integrating leader-member exchange and social network perspectives." *Administrative Science Quarterly*, 50(4), 505-535.

Stackman, R. W., and Pinder, C. C. (1999). "Context and sex effects on personal work networks." *Journal of Social and Personal Relationships*, 16(1), 39-64.

Sundararajan, A., Provost, F., Oestreicher-Singer, G., and Aral, S. (2013). "Research commentary-information in digital, economic, and social networks." *Information Systems Research*, 24(4), 883-905.

Taylor, J. E., and Levitt, R. (2007). "Innovation alignment and project network dynamics: An integrative model for change." *Project Management Journal*, 38(3), 22-35.

Taylor, J. E., and Bernstein, P. G. (2009). "Paradigm trajectories of building information modeling practice in project networks." *Journal of Management in Engineering*, 25(2), 69-76.

Thorpe, T., and Mead, S. (2001). "Project-specific web sites: Friend or foe?" *Journal of Construction Engineering and Management*, 127(5), 406-413.

Toellner, J. (2001). "Improving safety & health performance: identifying & measuring leading indicators." *Professional Safety*, 46(9), 42.

Toole, T. M. (2002). Construction site safety roles. *Journal of Construction Engineering and Management*, *128*(3), 203-210.

Turner, J. R., and Müller, R. (2003). "On the nature of the project as a temporary organization." *International Journal of Project Management*, 21(1), 1-8.

U.S. Bureau of Economic Analysis (2015). “GDP by industry."< http://www.bea.gov/iTable/iTable.cfm?Re

qID=51&step=1#reqid=51&step=51&isuri=1&5114=a&5102=1> (Mar. 30, 2017).

U.S. Bureau of Labor Statistics (2017). “Construction: NAICS 23”< https://www.bls.gov/iag/tgs/iag23.htm> (Mar. 30, 2017).

U.S. Census Bureau (1997). “Geographic Area Summary – 1997 Economic Census, Construction, Subject Series, Definition of Terms, Building Construction,” Appendix A, pg A-5 – A-7, Issued April 2002.

Wagenaar, W. A., Groeneweg, J., Hudson, P., and Reason, J. (1994). "Promoting safety in the oil industry." *Ergonomics*, 37(12), 1999-2013.

Wambeke, B. W., Liu, M., and Hsiang, S. M. (2011). "Using Pajek and centrality analysis to identify a social network of construction trades." *Journal of Construction Engineering and Management*, 138(10), 1192-1201.

Wasserman, S., and Faust, K. (1994). *Social network analysis: Methods and applications*, Cambridge university press, Cambridge, UK.

Wreathall, J. (2009). “ Leading? Lagging? Whatever! ” Safety Sci., 47(4), 493 – 494.

Zachary, W. W. (1977). "An information flow model for conflict and fission in small groups." *Journal of anthropological research*, 452-473.

Zheng, X., Le, Y., Chan, A. P., Hu, Y., and Li, Y. (2016). "Review of the application of social network analysis (SNA) in construction project management research." *International Journal of Project Management*, 34(7), 1214-1225.

VITA

Seifeldeen Eteifa is a graduate second year PhD student and Chancellor’s Fellow at University of Tennessee, Knoxville at the department of Civil and Environmental Engineering. Eteifa earned his B.Sc. degree from the American University in Cairo where he was awarded the president’s cup and Mohamed El-Beleidy Academic award. Currently, Eteifa’s research at UT focuses on construction safety and accident causation utilizing Social Network Analysis. Other fields of research interest for Eteifa include statistical analysis, modelling, simulation and optimization. To date, Eteifa has two publications in ASCE journals.