2014-09-16

A Complex Social Network Analyses of Online Finanical Communties in Times of Geopolitcal Military and Terrorist Events

James Usher
Technological University Dublin

A Complex Social Network Analyses of Online Financial Communities in times of Geopolitical Military and Terrorist events

James Usher

Dublin Institute of Technology

A thesis presented to Dublin Institute of Technology,

Faculty of Computing

For the Msc in Computing (Information & Knowledge Management)

September 2014

I certify that this project which I now submit for examination for the award of Msc in Computing (Information & Knowledge Management) is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

This work has not been submitted in whole or in part for an award in any other Institute or University.

The Institute has permission to keep, to lend or to copy this document in whole or in part, on condition that any such use of the material of the project is duly acknowledged.

Signed ___________________________ Date ______________

ABSTRACT

Given the advances in technology the field of social network analysis has very much hit the forefront in recent years. The information age harnesses the use of social network analysis for multiple industries and for solving complex problems. Social network analysis is an important tool in the world of the military and counter intelligence, whether it's the capture of Osama Bin Laden or uncovering hidden Al Qaeda terrorist networks, the world around us is built on networks, be that hidden or otherwise.

Online social networks give new information in the world of intelligence agencies similarly online financial communities such as Yahoo Finance gives intelligent information to knowledge hungry investors. This thesis is concerned with the exploration and exploitation of online financial community dynamics and networks using social network analysis (SNA) as a mechanism. Social network analysis measurement techniques will be applied to understand the reaction of online investors to military and terrorist geopolitical events, the stock market's reaction to these events and if it is possible to predict military stock prices after military and terrorist geopolitical events.

ACKNOWLEDGEMENTS

I am eternally grateful and thankful to my supervisor and great friend, Pierepaulo Donio, whose encouragement, advice, direction, inspiration and expertise has enabled me to produce this thesis on a subject matter that has significant importance to everyday day life in the financial arena.

Damien Gordon for new insights and methods and the enthusiasm you bring.

I dedicate this work to my parents Mary and Martin Usher, who have always been the driving force behind my achievements.

Dad, I knew there was merit in watching all of those Sunday afternoon Channel 4 war movies with you and Paddy…..

Paddy for being the single most creative and innovative person I know.

Natasha for all the support during this phase of my life.

Martin and Niall for the joy you bring into my life, you are forever my inspiration.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

1 INTRODUCTION

This thesis is concerned with establishing military and terrorist geopolitical events that have a measureable influence on military stocks and what influence online financial communities have in conjunction to military stock price movements utilising current online financial market community communication. The model will seek to gauge the reaction of online investors to military and terrorist geopolitical events, the stock market's reaction to these events and based on the gathered intelligence, a prediction model will be utilised to try and determine future military stock movements for geopolitical events measured against the S&P 500. The ultimate aim is to establish a framework that aims to understand how military and terrorist geopolitical events are reflected in financial markets using Social Network Analysis (SNA) as a knowledge management tool.

This document is broken into a number of sections. The thesis to this dissertation and the number of experiments conducted. Chapter 2 will give a brief history of some important military and terrorist geopolitical events and feature military and terrorist geopolitics within the financial arena by providing a background of what these events consist of and how they are defined within the thesis. Some of the key questions posed by social network analyses and its value will be addressed in this section. Chapter 3 will consist of the literature review and will look at the concepts associated with social network analysis and the evolution of same thus including a number of disciplines' and techniques, which incorporate the following: Betweeness, Centrality, Clustering Coefficient, Degree, Density, Diameter and Modularity. The element of data mining to this thesis is not the fundamental core objective of the experiments it is a mechanism used herein for data extraction and discovery. Data extracts consisting of online communication will be taken from Yahoo Finance forums for each of the relevant stocks. Section 4 will outline the statistical and data mining methods. Section 5 deals with the experiment approach and implementations for each experiment and the analysis of the results. Section 6 will summarise and conclude the thesis.

The scientific belief of this thesis is that the applied social network analyses and prediction model may help investors understand how military and terrorist geopolitical events can affect their investments i.e. portfolio stock prices and how online financial communities react and relate to geopolitical events in addition to understanding the effect of these events on future military stock price movement. Moreover by providing this research strategists may know what to expect in terms of precedent set forth by previous geopolitical military and terrorist events.

1.1 Research Questions

1. How do online financial communities react to military and terrorist geopolitical events?

2. How does the stock market react to military and terrorist geopolitical events?

3. Can military stocks prices be predicted after a military or terrorist geopolitical event?

1.2 Experimental Design

The experimental design follows a procedure that can be described in the following steps:

1. Establish a research base for social network analysis and online financial communities.

2. Define the information and data points needed to support each variable for each of the experiments.

3. Utilise a parser to mine the Yahoo Finance online financial community communication.

4. Establish, define and classify quantitative SNA metrics exclusive to the experiments.

5. Compare and contrast the SNA metrics against online communities before and after geopolitical events.

6. Compare and contrast the prices of military stocks before and after geopolitical events.

7. Compare and contrast military stock price movements against the S&P 500 index before and after geopolitical events.

8. Perform a concluding analysis to the outcome of each experiment.

CHAPTER 2

2 MILITARY AND TERRORIST GEOPOLITICS ON THE RISE…

Chapter two looks at the rise of military and terrorist geopolitics in the modern day world and what effect geopolitics has on nations. It also looks at military spending and defines geopolitical categories that are in scope for this thesis. The chapter also focuses on some of the past and most recent military and terrorist geopolitical events and what effect they had on the market. Fundamental social network analysis questions key to this thesis are also addressed.

Geopolitical events are somewhat quite difficult to predict and often investment decisions rarely incorporate geopolitical risk. Geopolitical risks are always a potential pitfall for any investment decision process, with geopolitical risk on the rise, one can conceive the world is an unsafe place as is the world of investment decision making. With the end of the Cold war, technology brought about a sense of globalisation, privatisation and an increase in state owned industries and state intergovernmental generated propaganda. For a while global economies focused on their internal infrastructures whilst the wage of war subsided, a paradigm shift in the decrease of defence spending begun to have longitudinal consequences. The world appeared to be more peaceful. The subprime machine was in full swing moving from country to country, continent to continent which encouraged a glorious sense of prosperity and economic indulgence.

Since the crumble of the tsunami of cheap credit and in some cases sovereign debt economic and financial market activity has become increasing more volatile. Events should as the Arab spring, 9/11 terrorist attacks and the rise of nuclear weapon capabilities have had a profound effect on global markets and localised manufacturing output. Evidence has demonstrated that geopolitical events shape economies, generate wealth, even if that is just short term gains. Geopolitics can potentially shape investment decision making thus enhancing countries prosperity. These events can give rise to shrew real time investment decisions thus determining how markets behave. In the current political environment tensions are high between countries as a result of natural resource scarcity, oil, gas, nuclear energy capabilities, have the ability to create wide scale controversy and militant events.

Nations have resorted to more draconian measures to kick-start their economies. The threat of war is a method that generates volatility within the financial markets. Nathan Rothschild, one of the founders of the Rothschild banking dynasty, is probably the single investor most associated with profiting from conflict and the related geopolitics. The network of agents and investments he built across Europe in the early 1800s enabled him to learn of Wellington's victory at the Battle of Waterloo in 1815 ahead of other investors. He sold heavily, driving the market down and encouraging others to sell, too, only to quietly buy again before the victory became known and markets rose. He is said to have coined the maxim: "*Buy on the sound of cannons, sell on the sound of trumpets*." Dyson (2013)

The proliferation of low and high intensity warfare features prominently in this thesis, such that the chosen geopolitical warfare events can be split up into two categories: A third category exists as deemed below although all of the data has been sourced, it is not in scope for this project because of time limitations. Table 1 illustrates same below. However, it can be used for a future experiment.

1	Natural Resource and Energy Geopolitics: Access to oil, gas and countries that want to extract and mine foreign state energy reserves. Examples include the Libyan Revolution and the current Ukrainian Crises.
2	Income and Ideological Geopolitics: Refers to terrorist activity and political attacks from terrorist groups and rebel insurgency outfits against domestic governments. The Syrian conflict would embody this ideology.
3	National Strategic Geopolitics: Consists of nuclear activity ambitions, such as plutonium and Uranium manufacturing for the use in nuclear weapons. The Iran Nuclear crises and the North Korean nuclear crises personifies same.

Table 1 Geopolitical categorisation

Military expenditure has been on the rise since the end of the cold war. World military expenditure in 2012 is estimated to have reached $1.756 trillion. The USA with its massive spending budget has long been the principal determinant of the current world trend, often accounting for close to half of the entire world’s military expenditure. The effects of global financial crisis and the post Iraq and Afghanistan military operations have seen a decline in its spending; now accounting for 39% of spending in 2012. (Global Issues). Figure 1 illustrates the World Military expenditure. To this end, the stock selection focuses on American military stocks. Interesting in today's world strategic power between heavy weight nations are less defined, the great superpower giants of yesteryear are now focusing their battle for supremacy by aligning their military sovereignty, technology and science with the emerging nations and rebel insurgency groups to not only increase their personal GDP, but also to quash the rise of religious beliefs.

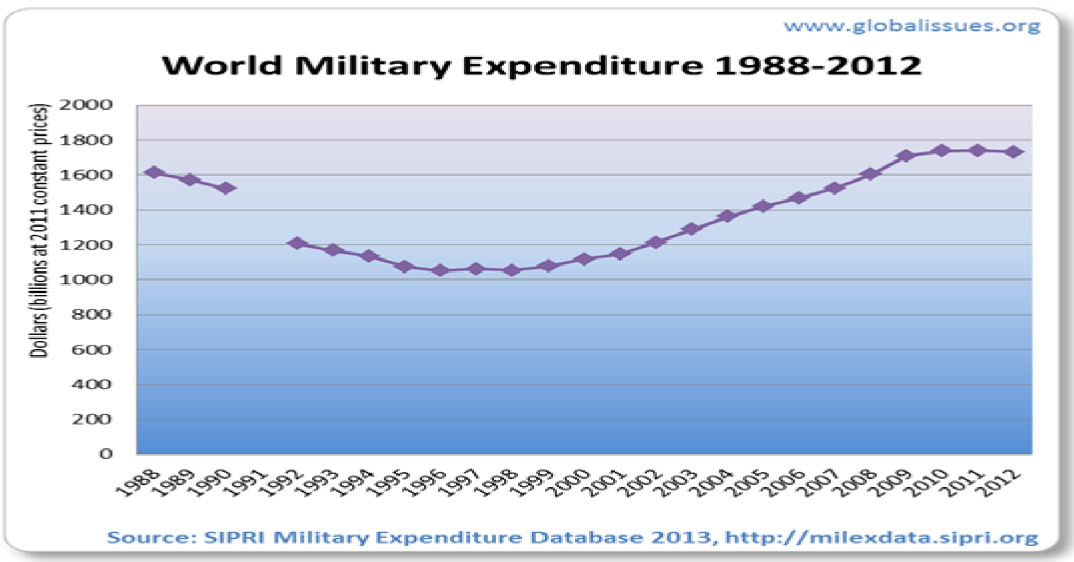


Figure 1 World Military spending 2003 - 2012

2.1 MILITARY AND TERRORIST GEOPOLITICS AND THE FINANCIAL MARKETS

Sam Stovall, chief equity strategist at S&P Capital IQ, conducted a broad study last year of how U.S. stocks reacted since World War II to anything ranging from wars, near wars, assassinations, assassination attempts, terrorist attacks and financial collapses. In the 14 examples he cited, he found a running theme: Stocks initially sold off, but then didn't take long to rebound and recover those losses. The chart below, courtesy of S&P Capital IQ, expands upon those examples (WSJ)

SHOCKS TO THE SYSTEM: Market Declines and Recoveries Since WWII

	Closing Levels			Bottom			Days to
Market Shock Events	Prior Day	Next Day	% Chg.	Level	Days	% Chg.	Recover
Japanese Tsunami: 3/11/11	1304.28	1296.39	(0.6)	1256.88	3	(3.6)	6
Flash Crash: 5/6/10	1165.87	1128.15	(3.2)	1110.88	1	(4.7)	4
Lehman Bankruptcy: 9/15/08	1251.70	1192.7	(4.7)	676.53	121	(46.0)	285
Madrid bombing: 3/10/04	1140.58	1123.89	(1.5)	1093.95	10	(4.1)	18
Terrorist Attacks: 9/11/01	1092.54	1038.77	(4.9)	965.80	5	(11.6)	19
Collapse of LTCM: 9/23/98	1066.09	1042.72	(2.2)	959.44	11	(10.0)	9
Iraq's Invasion of Kuwait: 8/2/90	355.52	351.48	(1.1)	334.43	2	(5.9)	30
Program Trading: 10/19/87	282.70	224.84	(20.5)	223.92	33	(20.8)	223
Reagan shooting: 3/30/81	136.30	134.7	(1.2)	134.70	1	(1.2)	4
Nixon Resignation: 8/8/74	82.65	81.57	(1.3)	62.28	39	(24.6)	143
OPEC oil embargo: 10/17/73	111.30	110.05	(1.1)	109.16	6	(1.9)	10
Kennedy assassination: 11/22/63	71.62	69.61	(2.8)	69.61	1	(2.8)	2
Cuban missile crisis: 10/22/62	54.96	53.49	(2.7)	53.49	1	(2.7)	5
Pearl Harbor Attack: 12/7/41	9.38	8.97	(4.4)	8.37	18	(10.8)	257
Medians			(2.4)		6	(5.3)	14

Source: S&P Capital IQ. Past performance is no guarantee of future results.

Figure 2 Shocks to the system: Market declines and recovery since WWII

Focusing on terrorist and warfare events, it took the S&P 500 18 and 19 days to recover from the 9/11 and Madrid attacks respectively. The Pearl Harbour attack took the biggest single warfare S&P 500 recovery with 257 days. What is striking is the bottomed out % change in the 9/11 attack and the Pearl Harbour attack. Percentage wise its quite similar 11.6% and 10.8% even though the days to recovery gap appear to have no correlation. It could potentially appear that modern day investors have adapted to military geopolitical events as a result of their alarming frequency and predetermined precedents. However when the world appeared to be on the brink of disaster during the Cuban missile crises', the market only took a mere 5 days to recover from the bottomed out position. Crises not related to terror and warfare attacks such as the Lehman bankruptcy have had even more ephemeral resonance for the market, nevertheless this evidence illustrates that terrorism and warfare have initial negative effects on the markets.

Company stock prices and the stock markets in general can be influenced by world events such as war, civil unrest and terrorism. These influences can be direct and indirect and they often occur in chain reactions. For example, the social uncertainty and fear generated by the terrorist attacks on September 11th 2001 affected markets directly as they caused many investors in the United States to trade less and to focus on stocks and bonds with less risk. From my own personal experience whilst working on the Convertible trading desk for an American derivate firm on that ill-fated day, the direction was to sell all stocks correlated to the attack (e.g. airlines stocks) and shift the capital to debt instruments such as US Government bonds and Treasury Bills. Indirectly, you can expect the stocks of military equipment companies and weapons manufacturers to rise in value as a nation gears up for armed conflict due to increased demand.

War affects the value of assets above all else. Even the suggestion of a war in the Middle East is often enough for the price of crude oil to sky-rocket due to the region being such a major oil exporter. Oil stocks have a negative correlation so rising oil prices due to political unrest usually signals falling stock prices, especially those stocks denominated in dollars and energy stocks.

It is perhaps not surprising that some economists point to a potential economic upside to war. War at times can kick-start a struggling economy especially its manufacturing base when forced to concentrate its efforts on war time production. Think of the United States in World War II. The U.S.'s entry into the war following the attacks on Pearl Harbour almost instantly pulled the country out of the grips of the Great Depression Banc De Binary (2014)

Dyson (2013) looked at analysis conducted by Deutsche Bank, which was circulated to its institutional clients, which studied 12 conflicts over the past 30 years. The first was the missile strike ordered by President Reagan on Libya in April 1986. This was a response to the bombing, 10 days earlier, of a Berlin nightclub in which American soldiers were killed. The most recent of the 12 conflicts analysed by Deutsche was the March 2011 deployment of aircraft over Libya i.e. Operation Ellamy, when a coalition of Western states including Britain intervened in the Libyan civil war. Between these two conflicts Deutsche Bank looked at 10 other events involving US and Western forces, including the 1991 Gulf War, the 1996 Iraqi Kurdish conflict, the Afghanistan war in late 2001 and the 2003 Iraq war. With each event, analysts sought to identify a precise date on which the US entered the conflict, and then study the market either side of this date by examining the commonly followed S&P 500 index of leading US companies. Three data points were collected for each event: the highest point in the three months preceding the strike date, the strike date itself and the market's position one month after the strike date. In all 12 cases the market on the strike date was below the high point of the three previous months, although the fall varied. The smallest drop was 1pc (Bosnia, August 1995) and the largest 15pc (Afghanistan, 2001).

The data relating to the weeks after the strike was more dramatic. Although many of those conflicts would drag on for months or years, the S&P 500 tended to rise sharply after intervention. In every single case the index rose during the month following the strike, by as much as 18pc (the 1991 Gulf War). In seven cases out of 12 the S&P 500 reached a higher point one month after the strike date than the highest point in the three months before it. Overall, across the 12 conflicts the S&P 500 fell by an average of 5.9pc from a high point in the three months preceding the strike to the strike day itself. In the following month the average rally was 7pc. Revisiting Nathan Rothschild famous words from 1815 "*Buy on the sound of canons*" suggests that this investment strategy would appear still be widely practiced.

Bremmer and Keat (2009) state "*a growing number of investors and policymakers understand the importance of political risk. Yet they also know that they lack a comprehensive and systemic set of tools for evaluating these risks*" With the application of SNA to online financial communities in times of geopolitical events some key questions arise. Can historical SNA online financial community mined data and historical geopolitical military and terrorist events be used to form future prediction geopolitical military models for Investment banks? Can future geopolitical military prediction models be classed as a business asset for the Investment banking Industry? Is there value in providing an automated algorithmic application that will mine historical SNA, geopolitical military and terrorists events to buy and sell military associated stocks once intelligent systems realise pending geopolitical events are unfolding, for example the application of Nathan Rothschild's rule in alliance with the parameters set forth from the Deutsche Bank report. Furthermore could this become a commercially viable product?

CHAPTER 3

3 LITERATURE REVIEW

Chapter 3 contains the literature review. The literature review of this thesis will seek to provide a basis for the experimental approach by doing so the literature review will start by introducing the world of SNA inclusive of social capital, small world and SNA theory and provide a review of existing methodologies and papers in the field. Thereafter existing approaches to handing and exploring social network analysis techniques will be discussed.

3.1 SOCIAL NETWORK ANALYSIS

Given the advances in technology the field of social network analysis has very much hit the forefront in recent years. The information age harnesses the use of social network analysis for multiple industries and for solving complex problems. Whilst my project is concerned with the use of SNA for online financial communities discussing stocks, there are many areas that the SNA science can be applied to. Wasserman and Faust state "*Social network analysis is inherently an indisplincinary endeavour*" SNA has its roots associated with many disciplines these include social sciences, applied mathematics, statistical and computing methodology

Social media services, such as Facebook, Twitter, Wikipedia, among others, create a wealth of data that has the potential to provide new insights to marketers, social scientists, community administrators, and system developers. Much of this data comes in the form of networks: people connected via friend and follow relationships, websites connected via hyperlinks, and books connected to other books based on shared purchasing patterns. Over the past decades, researchers have developed sophisticated techniques for analysing and visualising network data, particularly social network data Hansen et al (2009). Communication patterns were modelled in the "*Analysis of Facebook Social Network*" the intention was to offer a unique perspective of how society functions and to uncover hidden relationships in the Facebook network Akhtar et al (2013).

As pointed out in the Gartner study, this huge development in social networks leads to the growing need for social network mining and SNA methods in order to provide deeper comprehension of the networks and to detect communities. SNA has found applications in various areas like computer science, life sciences, law enforcement agencies, civil society organisations, network operators Gartner (2008).

Typically the SNA will look to analyse all connections between the nodes (or vertices). In Wen-Chih Chang "*Applied Social Network Analysis to Project Curriculum (2010)"* paper he used a model based on six degrees of separation theory. The aim and objective in the social network analysis was to represent the strength of relationships between team members but also to illustrate the individual role in his (/her) own network. Thus, the team member can observe what the role they play in the team and how to strengthen communicative skills with their partners. From an organisational perspective this can be used as a knowledge management tool to understand the strengths and weakness of each of the nodes on the network.

Work such as "*Do you know the way to SNA*" undertook an experiment to understand how graduates could apply SNA to data from online communities. What becomes clear is the fact that SNA tools are very much in their infancy in terms of ease of use, unlike tools such as Google Analytics. Hansen et al (2009) demonstrates through their experiments that SNA tools graduates can effectively grasp the working of SNA within a short space of time; however the experiment does make reference to the fact that the SNA practise would benefit more if there was a toolset that reduced processing complexity. It could be perceived that the SNA toolset or metrics outlined in this thesis could have potential merit in terms of a standard SNA framework.

Kristin Forbes Robert Rigobon (2002) "*No Contagion, Only Interdependence: Measuring Stock Market Co-movements*," paper aimed to understand contagion coefficients in times of market crisis during the events of the Mexican peso collapse, the 87 US stock Market Crash and the Asian crisis. The measurement of same appeared to be quite manual and labour intensive. It could be argued that use of a SNA tools for this particular experiment may have uncovered a greater degree of correlation or indeed a greater discovery of high market co movement.

Antweiler W, Frank MZ (2004) researched the idea that the markets are influenced by negative and positive news in times of crisis and this has a strong correlation to geopolitical events. More than 1.5 million messages posted on Yahoo! Finance and Raging Bull about 45 companies in the Dow Jones Industrial Average and the Dow Jones

Internet Index, by measuring bullishness. They found that stock messages helped to predict market volatility both on a daily base and also within the same trading day. More specifically, they found that higher message postings predicted negative subsequent returns

More recently Bollen J, Mao H, Zeng X (2011) found that Twitter mood predicted more than 80% of daily volatility of closing values of the Dow Jones Industrial Average. The use of SNA in both of these experiments highlights the value of the SNA toolset in terms of financial and monetary gain. Gloor et al (2009) demonstrated that useful trends can be identified using SNA by calculating the betweenness's centrality for online communities.

Khonsari, K.K. et al (2010) found that political groups also find use for SNA. In the social network analysis of" *Iran's Green Movement Opposition Groups using Twitter*" paper the SNA aimed to identify the structure of the Green party using the twitter application to voice their disapproval and opposition that they perceived to be a fraudulent election.

In the world of counterterrorism, social network analysis can effectively become the social "terror" network analysis. Governments and intelligence agencies now use SNA as a tool to prevent terrorism. The SNA methods deployed are no different from the methods applied elsewhere in this literature review. Since the September 11 terrorist attacks, many researchers have started to apply social network analysis in the fields of counter-terrorism, which has become a new and popular tool for counter-terrorism research. Such information as social relationships among terrorists, correlations of terrorists during planning, organising and conducting terrorist activities as well as time and locations of terrorist activities provides a basis for analysing terrorist networks Julei Fu et al (2012) Similarity Krebs (2002) "*Mapping Networks of Terrorist cells"* used publically available data to map the 9/11 terrorist network. To capture the data for the SNA Krebs collected data pertaining to the hijacker's links and relationships from data pertaining to major newspaper articles for a few weeks after the attack.

3.2 Social Capital Theory

One of the most fundamental aspects of social capital theory is that SNA represents value; all social ties in various networks can be conceived as channels for the flow of data, information and knowledge. The structure represents the value of the network. A number of SNA metrics which represent the value will be discussed herein. Colman (1998) outlines that "*Two main characteristics are common for all forms of social capital: (i) they comprise a part of the social structure and (ii) they facilitate the actions of individuals in that structure"* Theoretically social capital is normally associated with capital in monetary terms, however in a metaphorical sense social capital in SNA refers to value and strength of the ties and entities of the network. Moreover Mark Granovetter work on in his widely cited 1973 article "*The Strength of Weak Ties*," argues that "*weak ties*" are more important than "*strong ties*" and that includes your relationships with family and close friends when trying to find employment. Granovetter's article and subsequent research extended this argument by positing that more disperse, non-redundant, open networks have greater access to information and power than smaller, denser, and more interconnected networks because they supply more diversity of knowledge and information.

Chung et al (2013) relays a classic example of trading in the diamond market in New York where bags of diamonds often worth thousands of dollars were exchanged amongst merchants frequently to other merchants for them to inspect at their own leisure. Considering that this was done without any formal insurance, an objective observer might think it to be risky as there could be potential for opportunities of fraud and theft. However, the market was extremely successful and efficient. Coleman argued that the market worked because of closeness, high degree of trust and trustworthiness amongst the merchants; thus attributing the success of the markets to high levels of social capital.

Considering we will be reviewing terrorist networks in terms of SNA metrics Perliger, et al (2011) definition of social capital becomes appealing "*One of the main advantages which SNA provides for students of terrorists groups is the capability to uncover the informal division of influence and social capital within the group, which, in turn, influences the group's internal political and social processes and the outcome of its activities"*

Likewise within online financial communities social capital exists as does weak and strong ties and as per Colman's theory social structures will emerge from the shadows.

3.3 It's a small world

A distinguished gentleman is sitting in a side walk café in Venice, needing a light for a cigarette. He asks the man next to his table for a light. They fall into conversation; the man at the table is an American who spent time in Ireland during the Celtic tiger years in the financial district. *"This may appear to be a silly question"* says the distinguished gentleman *"however did you ever hear of a person called Martin O Farrell Usher?*" "*Martin O Farrell Usher from Oldcastle, tall, handsome chap"* said the American*. "Yes"* said the gentleman. *"I don't believe it!"* said the American; *"I know him well and his family!" "That's incredulous"* replied the gentleman, "***Good Lord it's a small world"***

Almost all of us have had this experience, where we have encountered a stranger far from home and much to our surprise they share a mutual acquaintance. This type of experience occurs with sufficient frequency such that almost everyone says *"It's a small world"*

One of the best known and earliest SNA experiments was conducted by professor Stanley Milgram in 1967.The experiment consisted of mailing 160 letters to randomly chosen individuals in Nebraska. Once each recipient received the letter the objective was then to get the letter to a Stockbroker in Boston using intermediaries known to one another on a first name basis. In all 42 letters reached the target destination. This experiment became known as the "*Small world effect*". Milgram's experiment was designed to explore the properties of social networks: interconnecting bonds of friendship among individuals in a society. Amongst the letters that arrived at the desired location, it was derived that the average path length was 6; this led to the phrase, "*the six degrees of separation"*. What was of significance was the empirically gained insight into the configuration aspects, of social networks which laid the foundation for most of the concluding work. Milgram (1967) points to the following:

(i) In those chains, participants were three times more, likely to contact people of the same gender.

(ii) Many chains passed through the same few people – stars – as the last step before reaching the target person. (iii) Although physical distance is important in the chains of intermediaries, social distance seems even more important.

The Small World Effect describes that network size does not have any effect on length of ties among nodes. In other words, the distance between any pair of nodes is much smaller than the size of the network

3.4 Social network theory

One way we can think about social networks is to use the mathematical discipline of graph theory. Graphs are defined as a collection of points called vertices that are connected in pairs by lines called Edges Oram (2001). When sociologists borrowed this way of graphing things from the mathematicians, they renamed their graphs as "sociograms". There are a number of variations on the theme of sociograms, but they all share the common feature of using a labelled circle for each actor in the population we are describing, and line segments between pairs of actors to represent the observation that a tie exists between the two. Visualisation by displaying a sociogram as well as a summary of graph theoretical concepts provides a first description of social network data Mohsen Jamali and Hassan Abolhassani (2012).

"*…the workers have – whether aware of it or not – formed themselves into a group with appropriate customs, duties, routines, even rituals*" (Mayo, 1971)

Elton Mayo most famous for the Hawthorn experiments stated that the industrial world at the beginning of the twentieth century was more technologically advanced than ever before while being more socially incompetent than ever Bendix & Fisher (1949). In his studies of factory work and interactions between workers, Mayo pioneered the use of sociograms to depict interpersonal relations and group structure and was thus among the first to systematically depict the importance of spontaneous cooperation being organised by workers. Moreover, Mayo noticed that managers who had some level of understanding of the social processes such as group solidarity among workers had a greater ability to control and influence worker behavior Scott, J., (2000). Finally, Mayo explicitly linked socio-emotional issues with productivity through cooperation and job satisfaction Mayo (1971). This created the point of departure for not only the Human Relations school but also of the study of the informal structures and the focus on relations as the unit of analysis Mayo, (1939), which is the most central issue of SNA.

Berkowitz (1982) states that social network analysis is a set of research procedures for identifying structures in systems based on the relations among actors. Grounded in graph and system theories, this approach has proven to be a powerful tool for studying networks in physical and social worlds, including on the web. Faust (1997) declares that SNA focuses on relations and ties in studying actor's behavior and attitudes. Thus the positions of actors within a network and the strength of ties between them become critically important. Social positions can be evaluated by finding the centrality of a node identified through a number of connections among network members. Such measures are used to characterise degrees of influence, prominence and importance of certain members. Networks are interchangeable and consonantly evolving, network activity may change from day to day as each node activity may increase or decrease over the lifespan of the network. Social network analysis is a distinct research perspective within the social and behavioral sciences, as the social network perspective is based on the assumption of the importance on relational concepts or processes between individuals. Wasserman and Faust (1994) noted four characterising principles of social network analysis.

(1) Actors and their actions are viewed as interdependent rather than independent, autonomous units.

(2) Relational ties (linkages) between actors are channels for transfer or '*flow'* of resources (either material or nonmaterial)

(3) Network models focusing on individuals view the network structural environment as providing opportunities for or constraints on individual action.

(4) Network models conceptualise structure (social, economic, political, and so forth) as lasting patterns of relations among actors.

Jacob Levi Moreno, was heavily involved in advancing the notion of sociometry (i.e. the measurement of social relationships) this involved the plotting of social relations, this had been a project that had been progressing since the First World War According to Moreno. The brutality and meaninglessness of the war prompted social scientists to reconsider their views of society and the place of the individual within it and in an effort to understand the emancipation and dispersement of a common structured civil society. Moreno created actors as nodes and denoted the ties with lines. Moreno presents the following definition: "*Sociometric procedures try to lay bare the fundamental structures within a society by disclosing the affinities, attractions and repulsions, operation between persons and persons and between persons and objects"* Moreno (1937).

3.5 SOCIAL NETWORK METHODS AND METRICS

Theoretically social network activity depends on the movement of its actors. An actor or otherwise known as a node on the network can refer to a person or entity that has an interaction with another person or entity. The most important property of an actor is the measure of centrality in the network. Any relationships between actors are referred to as a tie, which typically links the pair together. A pair of actors with interconnecting ties between them is known as a *"Dyad".* Similarly where three actors exist with interconnecting ties, this is known as a "*Triad".* A "*Subgroup* "can consist of dyads and Triads in any social network. Wasserman and Faust (1994) refer to a group as *"a collection of all actors on which ties are too be measured"* further to the definition they state that "*once one decides to gather data on a group, a more concrete meaning of the term is necessary. A group then consists of a finite set or sets of actors who for conceptual, theoretical or empirical reasons are treated as a finite set of individuals on which network measurements are made".* The strength of the tie is denoted by the frequency and duration of contact of use over a period of time. In the case of this thesis the groups will consists of all of the actors associated with a single stock from the online financial communities.

There are two types of variables that can be included in a network data set: *structural and composition.* Structural variables are measured on pairs of actors and are the cornerstone of social network datasets. Structural variables measure ties of a specific kind between pairs of actors. For example structural variables can measure business transactions between corporations, friendships between people or trade between nations. Actors comprising these pairs usually belong to a single set of actors.

Composition variables are measurements of actors attributes, Composition variables, or actor attribute variables, are of the standard social and behavioral science variety and are defined at the level of individual actors. For example we might record gender, race or ethnicity for people, or geographical location, after tax profits or numbers of employees for corporations Wasserman and Faust (1994). The social networks that form part of the experiments contained herein will be largely based on structured variables, that form a *"two mode network"* which can be described as a network that focuses largely on one set of actors (i.e. the Yahoo Finance online financial community for a particular stocks) and one set of events (the actual listed geopolitical war or terrorist listed in chapter 5)

3.5.1 *Graph Theory*

Graph theory is essential and at the core of SNA. Graph theory is used in the identification of the most influential actors in the network. Metrics that can be derived from the graph theory are outlined herein. The prominence of actors can be measured in terms of centrality within the network. When a graph is used as a model of a social network, points (called nodes) are used to represent actors, and lines connecting the points are used to represent the ties between the actors. In this sense a graph is a model of a social network in the same way that a model train set is a model of a railway system Wasserman and Faust (1994). Nodes or actors can also be referred to as vertices or points, lines are commonly known as edges and sometimes arcs. Social network graphs consists of two sets of information, a number or set of nodes N = {n1, n2 n3} and a set of lines L = {l1, l2, l3} between pairs of nodes where N = Nodes and L = Lines. Ties are either present or nonexistent between actors, thus indicating an undirected dichotomous relation. Non directional relations include informal groupings. Wasserman and Faust (1994) state *"In a graph of a social network with a single non-directional dichotomous, relation, the node represent actors, and the lines represent the ties that exist between pairs of actors on the relation"* Graphs are essential to display the network variables for visualisation purposes. To display the output of the experiments in graph format the open source product Gephi will be used. *SNA methods and Metrics* (2011) refer in the most common sense of the term, a graph is an ordered pair G = (V, E) comprising a set V of vertices or nodes together with a set E of edges or lines, which are 2-element subsets of V (i.e., an edge is related with two vertices, and the relation is represented as unordered pair of the vertices with respect to the particular edge). To avoid ambiguity, this type of graph may be described precisely as undirected and simple. In graph theory, a vertex (plural vertices) or node is the fundamental unit out of which graphs are formed: an undirected graph consists of a set of vertices and a set of edges (unordered pairs of vertices), while a directed graph consists of a set of vertices and a set of arcs (ordered pairs of vertices). From the point of view of graph theory, vertices are treated as featureless and indivisible objects, although they may have additional structure depending on the application from which the graph arises; for instance, a semantic network is a graph in which the vertices represent concepts or classes of objects. The two vertices forming an edge are said to be its endpoints, and the edge is said to be incident to the vertices. A vertex w is said to be adjacent to another vertex v if the graph contains an edge (v, w). The neighborhood of a vertex v is an induced subgraph of the graph, formed by all vertices adjacent to v. The degree of a vertex in a graph is the number of edges incident to it. An isolated vertex is a vertex with degree zero; that is, a vertex that is not an endpoint of any edge. A leaf vertex (also pendant vertex) is a vertex with degree one. In a directed graph, one can distinguish the out-degree (number of outgoing edges) from the in-degree (number of incoming edges); a source vertex is a vertex with in-degree zero, while a sink vertex is a vertex with out-degree zero.

3.5.2 *Centralisation*

The Centralisation concept refers to the difference between the number of ties for each node divided by maximum possible sum of differences. A centralised network will have many of its links dispersed around one or a few nodes, while a decentralised network is one in which there is little variation between the numbers of ties each node possesses. The idea of centrality as applied to human communication was introduced by Bavelas in 1948. Chung et al (2013) uses the following definition *Centralisation explains the extent to which the connectedness is focused around a particular node. To measure centralisation in a network, we need to observe the differences in the centrality values of the most central nodes and all the other nodes. Then, to arrive at the centralisation value, we calculate the ratio of the sum of actual differences and the sum of the maximum possible differences. Centralisation is thus defined as:*

$$c = \frac{\sum_{i=}^{g} 1\{\max(Di) - Di\}}{(g-1)(g-2)}$$

Where (Di) is the number of people in the network that are directly linked to a person? The number of actors is represented by g in this equation.

Perliger et al (2011) looked at In-betweenness as the division of power within a group. His paper focused on the Jewish Terror organisation who were a group of religious zealots who were responsible for terrorist attacks against Arabs in the West Bank and East Jerusalem In defining the actors on the network that have unique characteristics, First is the classic hub, or an actor that has a significantly higher number of ties than other members (high level of centrality). Such a member is, in many cases, responsible for coordinating the group's activities, for recruitment of members, and has the ability to manipulate the flow of information within the group more easily. Figure 3 illustrates what Perliger refers to as the classic hub, in other words the hive of activity. To illustrate, in the case of the Israeli Jewish Underground terrorist group the hubs intentionally did not reveal the long-term highly extreme goals of the group to some members in order not to alienate them from participating in short-term less extreme operations. The hubs were able to do that not just because of their high level of centrality, but since entire sections of the network were dependent on them for connection with other parts of the network (high level of betweenness). Since a relatively high number of members were tied to the network via a specific hub, this also implies that the hubs were responsible for recruitment processes. Finally, the importance of detecting the hubs stems not just from the fact that this allows us to uncover the power division within the network and who possesses more social capital (Bourdieu 1986) but also to better understand the motives beyond the group's actions. As demonstrated by Pedahzur and Perliger, (2006), by looking at the characteristics and history of the hubs, the researchers were able to detect the motivations behind the attacks of the studied networks of Palestinian suicide bombers, epitomising that the "*hubs*" will, in many cases, use the networks to promote personal or local political interests.

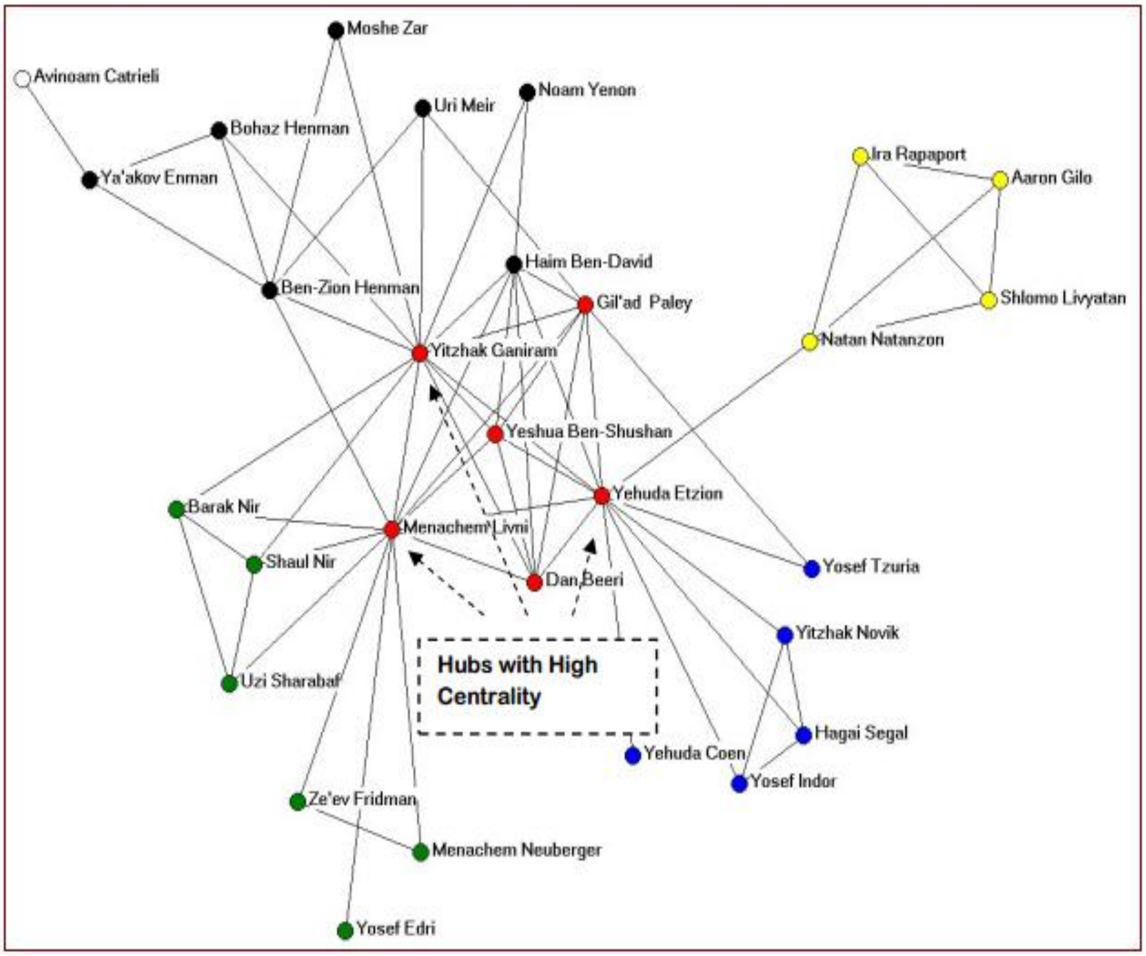


Figure 3 The Jewish Underground Terrorist group

The case of the Jewish Underground also exemplifies that a high number of ties is not the only criteria for detecting informal leaders. While some groups pose a dense structure, others are constructed as bundle of connected subgroups. In this case, the actors who are in strategic locations and serve as connectors between the different subgroups possess significant power and are crucial for the survival of the network. While they do not have to be connected to high numbers of members, they can veto almost any operation that needs the cooperation of the different subgroups. Finally, there are those who do not have a particularly large number of ties, nor are they connectors between different parts of the network, but they are situated in a strategic location in terms of their proximity to hubs or to large numbers of members within the network; hence, they have high level of access to information and resources. Chug et al (2013) refer to online social networks and the way in which people socially engage with each other**.** They go further by stating **"** *In network parlance, this pattern is suggestive of assortativity - the tendency in networks where individual actors connect with other actors who are similar to themselves based on certain characteristics . Thus, in the example of OSNs and its users, assortativity can be defined in terms of attributes such as willingness, frequency of usage or use per se of the OSNs for community building. From a networks perspective, one could also assess network assortativity, or the homophily of the OSN based on the number of messages one sends and/or receives. For instance, in the Twitter social network, it has been shown that happy users tend to connect to happy users whereas unhappy users tend to be predominantly connected to unhappy users measured by the Subject Well-being index. This was attributed to two mechanisms: (i) homophily -where "birds of a feather flock together" and (ii) mood contagion - where unhappy actors converged to other unhappy actors thus making them unhappier"*

Therefore it's arguable that assortabiity and centralisation have a correlation.

(i) Homophily - i.e. terrorist groups, online financial communities: Each has a common goal, to receive, retrieve, capture and execute information and knowledge

ii) Mood shared interest: Depending on the severity of the geopolitical event, this could potentially affect an online community's behavior.

3.5.3 *DEGREE CENTRALITY*

To understand the importance of nodes on a graph, Degree centrality is used. The centrality measurement is concerned with finding influential people on the network; however it is also concerned with the number of degrees of the user. The centrality of a node depends on the number of nodes attached to it directly. It can be best described as where most of the activity is occurring on the network and is commonly used to understand the potency or might of the information authority on the network. Kretschmer (2010) outlines: The original used measures of social network analysis (SNA) are related to Wassermann & Faust

The Degree Centrality (DC_A) of a node A is equal to the number of nodes (or ties) to which this node is connected. For example, in collaboration networks in armies the degree centrality of a node A is equal to the number of node collaborators or senior army personnel. An actor (node) with a high degree centrality is active in collaboration. That node has collaborated with many armies. In correspondence with Wassermann and Faust the *Group Degree Centralisation* quantifies the variability or dispersion of the individual Degree Centralities of the nodes. Centralisation describes the extent to which the links (ties) are organised around particular focal nodes, i.e. it provides a measure on the extent to which a whole network has a centralised structure. There are several degree based measures of graph centralisation. One of them is as follows:

$$\text{GDC} = \sum_{1=1}^{V} \frac{(DCL-DCi)}{(v-1)(v-2)}$$

$(v-1)(v-2)$ reaches its maximum value of 1 when one actor (node) has collaborated with all other v-1 actor, and the other actors interact only with this one, central actor. This is exactly the case in a star graph. The index attains its minimum value of 0 when all degrees are equal. Therefore, representing the weight graph is very necessary in finding the influential people. In summary the *unweighted measure* means the ties (or nodes) are counted independently from the strength of the ties. If the network is directed (meaning that ties have direction), then we usually define two separate measures of degree centrality, namely in-degree and out degree. In-degree is a count of the number of ties directed to the node, and out-degree is the number of ties that the node directs to others. Kretschmer method could be potentially suitable for the case of online communities, as it focuses on the number of interactions between nodes inclusive of reply interactions.

Referring to Roy, R.B.; Sarkar, U.K (2011), the Degree centrality method was used to identify influential stocks in the global market using various centrality measures and examine change in their ranks following the collapse of Lehman Brothers in the USA. The findings from the social network concluded that there was dominance in terms of geographical positioning. European stocks were seen to be more influential than global stocks in terms of stock market behavior based on a degree centrality. Rachman et al (2013) used degree centrality method to discover the highest frequency of tweets; retweets and replies in the twitter application to locate the most important influential person on the network. Koochakzadeh et al (2012) created a social network of financial expert's in an effort to allow novice investors model their investment behavior on similar likeminded financial experts. Behavior similarities are taken into account and the social network is then used to recommend an appropriate managed portfolio to non-professional investors based on their behavioral similarities to the expert investors. The degree centralities of each expert investor were measured, the highest centralities represented the low risk investors and investors that shared the same appetite for risk were aligned accordingly to that portfolio investment strategy. Zheng Chen and Xiaoqing Du (2013) use degree metrics to understand the interactions from the Shanghai/Shenzhen stock exchange, and the online Chinese stock forum *Guba.com.cn.* The paper is concerned with understanding online behavior in relation to stock movements. Their experiment uses two degree methods: the average degree which represents the strength of the users and the standard degree which shows how much variation or dispersion exist from the average node.

3.5.4 *DENSITY*

Network density represents the actual number of ties in a network as a ratio of the total maximum ties that are possible with all the nodes of the network. A fully dense network has a network density value of 1, which indicates that all nodes are connected to each other. A network with a density value near 0 indicates that it is a sparsely knit network. Hence, *density* is a measure of network cohesiveness. For a directed graph with n nodes, density D is defined as

$$D = \frac{\sum_{i\,j=1}^{n} x_{i}j}{\mathrm{n(n-1)}}$$

Krebs (2002) "*Mapping Networks of Terrorist cells"* used publically available data to map the 9/11 terrorist network. To capture the data for the SNA Krebs collected data pertaining to the hijacker's links and relationships from major newspaper for a few weeks after the attack. The SNA illustrated how sparse the network was and how distant i.e. ***"low density"*** many of the hijackers on the same team were from each other. Many pairs of team members were beyond the horizon of observability from each other. A strategy for keeping cell members distant from each other, and from other cells, minimises damage to the network if a cell member is captured or otherwise compromised. Osama bin Laden even described this plan in his infamous videotape, which was found in Afghanistan. In the transcript (U.S. Department of Defence, 2001) Osama bin Laden mentions:

"Those who were trained to fly didn't know the others. One group of people did not know the other group." Judging from this statement its clear the terrorist group had a clear strategy for managing the control of information and knowledge via actors on the network. What is interesting here is, did the terrorist group use a SNA model to ensure the distribution of knowledge and information was kept to a minimum?

In the world of espionage and terrorist activity we can now see a mechanism that can be used to counter terrorism. In cases where military teams or investigation teams have suspicions or indeed receive intelligence regarding certain groups the SNA inclusive of the networks to map will form a basis of knowledge to act upon. This will allow for the identification of key personnel and target opportunities to prevent the transfer of knowledge with the terrorist group. Taking figure 4 into consideration, hypothetically speaking if the CIA used the SNA appropriately (as it is widely believed that the CIA had ground intelligence on all of the hijackers) it's quite possible that if this knowledge was managed correctly the CIA could have potentially used the *"six degrees of separation*" theory to remove Mohamed Atta from the network. Because he had the highest degree of centrality, closeness and in-betweeness, this could have potentially prevented the attack from taking place.

Let's assume the 9/11 attack did not take place, Figure 4 and figure 6 below could be leveraged to provide a graphical illustration of the current communication and knowledge transfer between the terrorist group members. Where new members have been identified as part of the terrorist network each can be added as a node to the network ultimately bringing the geography of the terrorist group to life. By managing this knowledge base this in itself becomes a real asset in counter intelligence. The same method could be applied to all online communities.

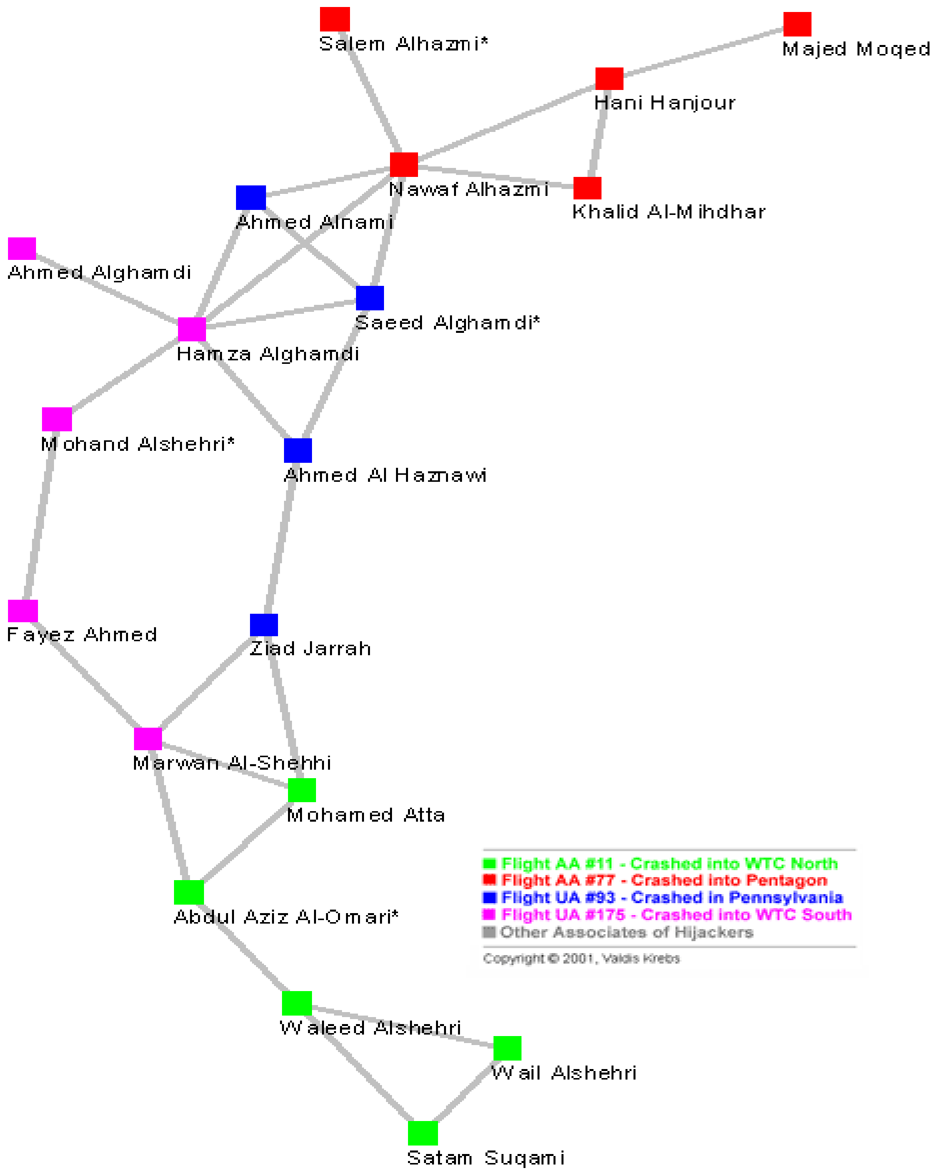


Figure 4 Krebs (2002) 9/11 SNA model

3.5.5 *Clustering Co-efficient*

To understand that how the network is likely to behave, it is necessary to segregate nodes into cliques. It can be a very important aspect of social structure. A clique is simply a sub-graph in which all nodes are more closely tied to each other than they are to nodes who are not part of the graph, similarly it can depict the closeness of groups within social networks. The clustering coefficient of a node i, denoted by c (i), is defined as the number of directed links that exist between the node's neighbours, divided by the number of possible directed links that could exist between the node's neighbours. Thus, if a node i's neighbours have n directed links between them, then the clustering coefficient of I is defined as

$$C(i) = \frac{n}{\mathrm{d}i(\mathrm{d}i - 1)}$$

The clustering coefficient of a graph is the average clustering coefficient of all its nodes, and can be denoted as follows C (G) or

$$C(G) = \frac{E_{v \in V}\, c(v)}{|v|}$$

Thus, the clustering coefficient of a graph ranges between 0 and 1, with higher values representing a higher degree of "cliquishness" between the nodes. In particular, a graph with clustering coefficient of 0 contains no "triangles" of connected nodes, whereas a graph with clustering coefficient of 1 is a perfect clique. Aggarwal (2011) refers to the probability that the two related nodes are associated with each other, or the level to which nodes are "*friends*" to each other. This coefficient illustrates the strength of connection in the network neighbourhood of the observed node. It is calculated as the ratio of the number of connections between the neighbours of the observed node and potential connections unestablished by its neighbours. Watts and Strogatz (1998) found that high clustering and short characteristic chain length are the distinctive properties of many small-world networks. They defined a "*clustering coefficient*" (C) to represent the amount of clustering. Wang (2012) states: the clustered is the important features of social systems and economic systems. In his securities network stocks or bonds have links connected with other nodes (i.e. other stocks and bonds). If the nearest neighbor of the initial node (stock or bond) is part of the entire group then there is links among them to calculate a cluster coefficient. Bai et al (2011) puts forward a community identification method based on clustering coefficient, which identifies the hidden community structure in the network, but also can reasonably process boundary nodes that simultaneously belong to several communities. Interestingly the cluster co efficient would appear to be an ideal technique in uncovering hidden dark networks as described by Raab and Milward (2003). The goal of his paper is to identify actors and organisations that operate in a two dimensional environment consisting of illegal and legal activities. They further the hypothesis by stating "*that there maybe connections between the illegal networks, dark networks in our terms and the legal networks striving to destroy them. In a world of illegal activity a CIA agent who penetrates a terrorist network is called a double agent.*" Figure 5 below illustrates reciprocal relationships between amongst illegal network linked to legal organisations trying to control or curtain the illegal activity. According to Raab and Milward (2003) it also captures the possibility of interpretation between illegal and legal networks. Forgetting the requirement to be functionally non-judgmental in the social sciences, the actors and organisations that cooperate in the problem space are called dark networks, in that their activities are both covert and illegal. Figure 5 illustrates makes full use of the clustering coefficient by providing a visualisation of the cliques with the network.

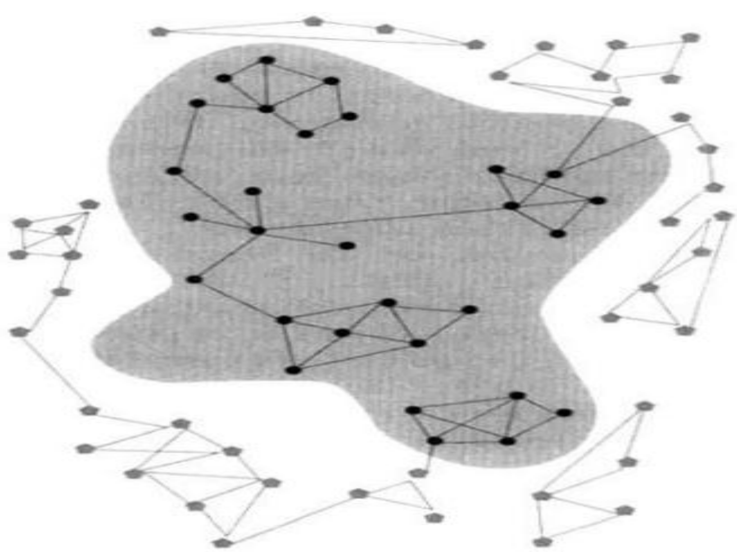


Figure 5 Raab and Milward (2003) Reciprocal relationships between amongst illegal network linked to legal organisations

Krebs (2002) denotes: for a small network of less than 20 nodes, we see a long average path length of 4.75 steps. Several of the hijackers are separated by more than 6 steps. From this metric and from Bin Laden's comments above we see that covert networks trade efficiency for secrecy. This detection method can be used to discover hidden secret networks within online financial communities. A hidden network could be a network whose sole objective is to prevent liquidity for particular stocks on the market to aid a geopolitical terrorist or warfare events.

	Clustering Coefficient	Average Path Length
Contacts	0.41	4.75
Contacts + Shortcuts	0.42	2.79

Table 2: Small-World Network Metrics

There is a constant dynamic between keeping the network hidden and actively using it to accomplish objectives (Baker and Faulkner, 1993) to this end Krebs (2002) identifies that six shortcuts were added to the network temporarily in order to collaborate and coordinate. These shortcuts reduced the average path length in the network by over 40% thus improving the information flow in the network as illustrated in table 2. When the network is brought closer together by these shortcuts, all of the pilots ended up in a small clique the perfect structure to efficiently coordinate tasks and activities. Figure 6 illustrates the pilot network. The gold bars indicate the clustering co-efficient, the coming together of the pilot "clique" in the network.

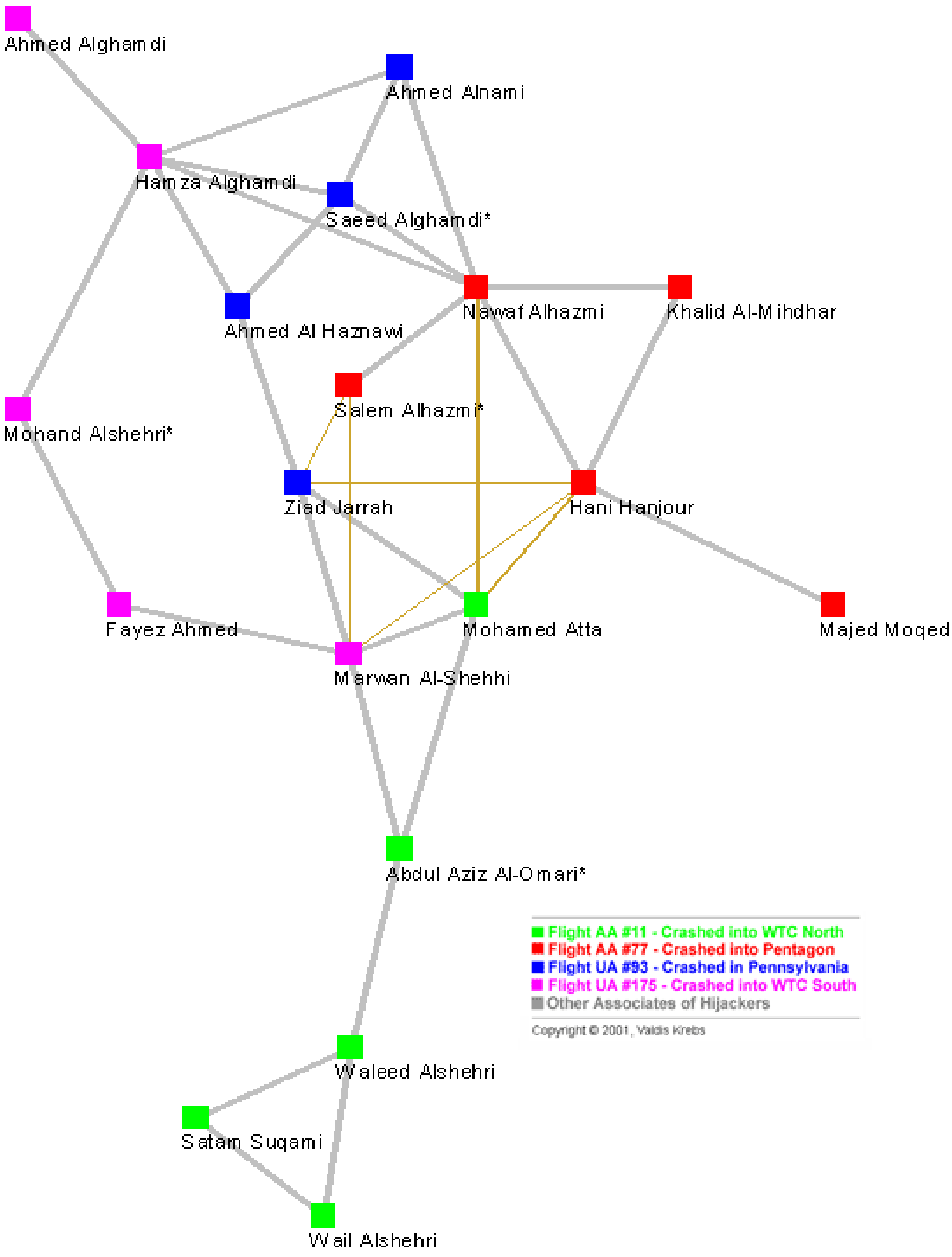


Figure 6 9/11 Pilot clique identification

3.5.6 *BEWTEENNESS*

Narayanan (2005) defines Betweenness in his thesis. He defines the following "*Betweenness is a centrality measure of a vertex within a graph Vertices that occur on many shortest paths between other vertices have higher between's than those that do not"* We are concerned with a measure called betweenness centrality. The vertex betweenness centrality BC(v) of a vertex v ∈ V is the sum over all pairs of vertices u, w ∈ V , of the fraction of shortest paths between u and w that pass through v:

$$BC(v) = \sum_{\substack{u,w \in V \\ u \neq w \neq v}} \frac{\sigma_{uw}(v)}{\sigma_{uw}}$$

Where σuw(v) denotes the total number of shortest path between u and w that pass through vertex v and σuw denotes the total number of shortest paths between u and w.

$$BC(e) = \sum_{\substack{u,w \in V \\ u \neq w}} \frac{\sigma_{uw}(e)}{\sigma_{uw}}$$

Where σuw(e) denotes the total number of shortest path between u and w that pass through edge e and σuw denotes the total number of shortest paths between u and w.

In reference to Kinsella (2011) he focuses on the illicit arms trade and draws on a database named the "*Illicit Arms Transfer Database*", which systematises information contained in journalistic reports on illicit small arms transfers. The purpose of which is to reveal high profile positions occupied by former Soviet bloc countries in the illicit arms trade network. He examines the Betweenness centrality calculations for Liberia's Illicit Arms Trade using UN reports documenting arms embargo violations. Individuals and transactions involved in four arms transfers from 1992 to 2002 are collated and examined. The SNA network contains 38 people that include arms brokers, arms transportation vendors, arms purchasers including Liberian president Charles Taylor and his son Chuckie. The Inbetweenness network illustrated in figure 7 below clearly shows Viktor Bout, the high-profile Russian, arms broker and transporter now in U.S. prison for arms trafficking as the actor with the highest *Betweenness centrality* score. He states that "*Brokers"* or "*transport agents*" have a highest level of Betweenness centrality. Furthermore he elaborates on that by stating *"A closely related SNA concept useful for the study of illicit arms trade networks is brokerage. Brokers, in network analytic terms, are nodes positioned on directional path network. Analysts have gone on to specify particular brokerage roles based on the actors 'membership in groups or other attribute categories. For instance, a node occupies a coordinator role when it is interposed between nodes within its same group or organization; when the three nodes are members of different groups, the broker acts as a liaison .Other brokerage roles are defined when the broker and one actor are members of one group and the other actor is a member of a second group: brokers that mediate inflows into their group are gatekeepers; those that mediate outflows from their own group are representatives. Identifying important brokers in a social network involves counting the number of triads in which that node is positioned as an intermediary between nonadjacent nodes. Naturally, they tend to have high betweenness scores.*

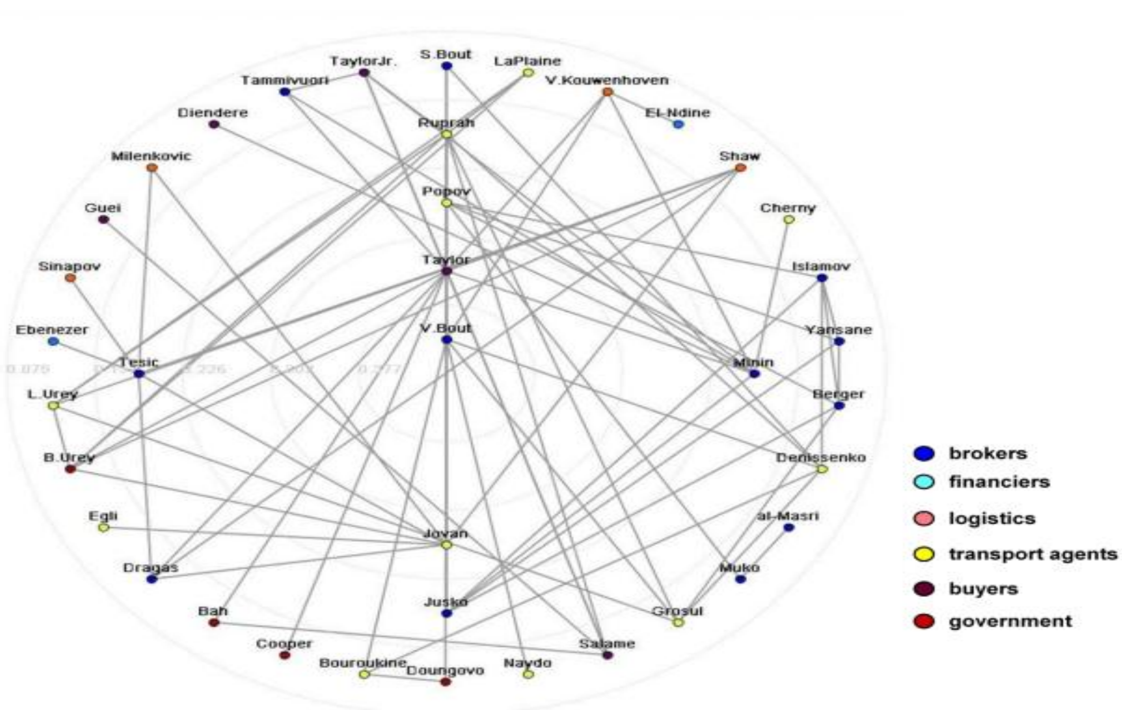


Figure 7 Liberia's Illicit Arms Trade betweeness centrality Kinsella (2011)

Inserting a concealed financial broker into an online financial community now becomes a interesting prospect. The impact of which may potentially cause a ripple effect in terms of stock price movement from brokered online chat or communication.

Looking at figure 8 and the results in table 3 calculated by Krebs (2002) in terms of betweeeness, it came as no great shock that once the degrees, betweenness and closeness were calculated Mohamed Atta was identified as the leader. Looking at the analysis written by Thomas A. Stewart in his article "*The Six Degrees of Mohamed Atta"* he states "*According to three measures. One is "degrees," or activity, which measures the number of times someone contacts others in the network. A second is "betweenness." For example, there appears to have been no direct link between Abdulaziz Alomari and Ziad Jarrah; Atta and Marwan , Al-Shehhi stood between them. The more often someone is in that "between" position, the more control he exercises in the network. The third attribute is "closeness," which measures the extent to which a person has direct contact with others, with no go between, this is another clue to how important an individual is to a network.*

3.5.7 *CLOSENESS*

Closeness is preferred in network analysis to mean shortest-path length, as it gives higher values to more central vertices, and so is usually positively associated with other measures such as degree. In the network theory, closeness is a sophisticated measure of centrality. It is defined as the mean geodesic distance (i.e., the shortest path) between a vertex v and all other vertices reachable from it:

Morris T (2012) defines "*Closeness is a measure to understand where individual nodes lie between other nodes in the network. Closeness reflects another way of understanding the relationship between diagnostic and prognostic frames in each text and over time. Examining the closeness between nodes takes into account the connectivity of the node's neighbor and assigns a higher value for nodes that have the shortest paths to other nodes in the network"*

The closeness metric does not feature in any of the experiments.

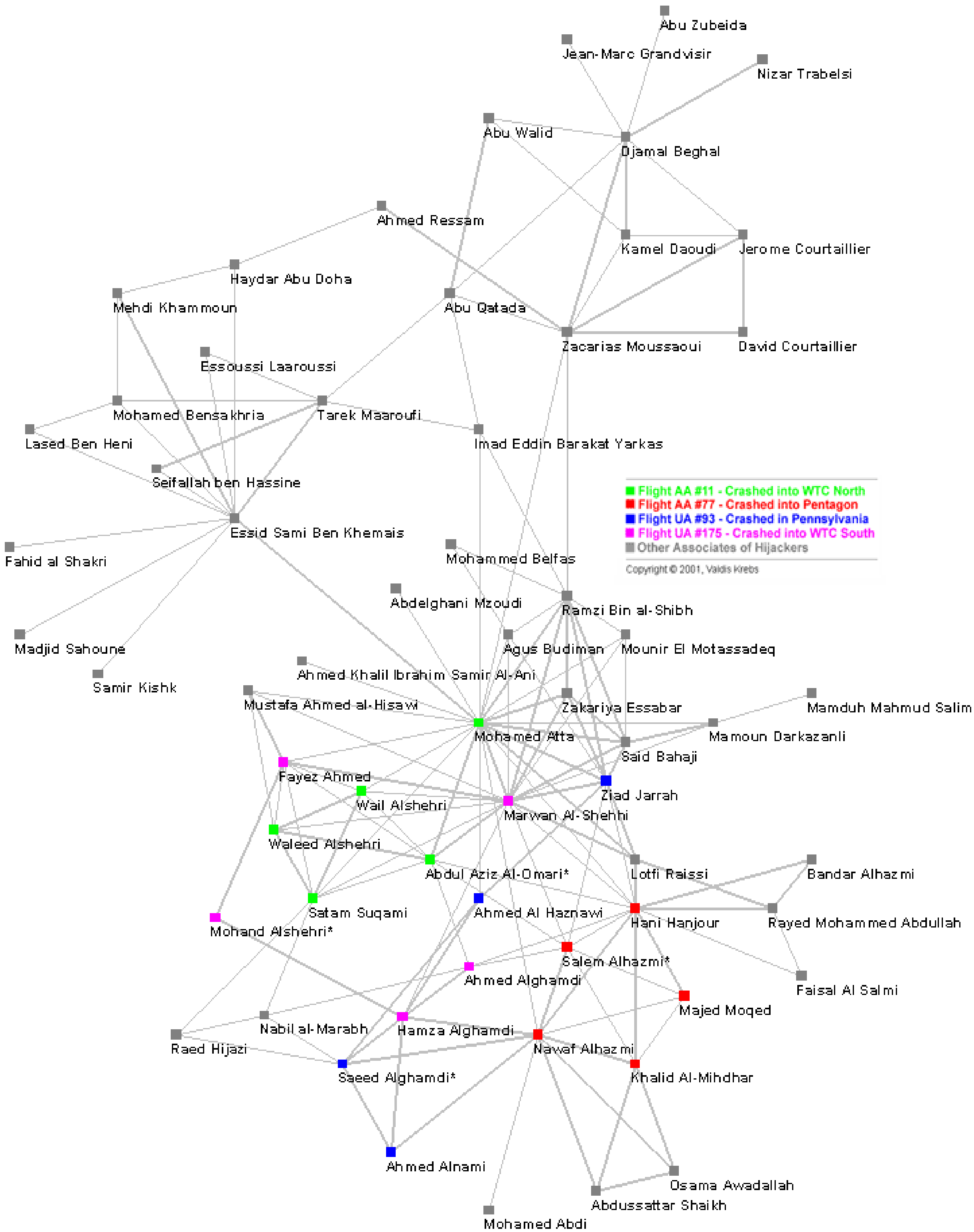


Figure 8 Detailed 9/11 Hijacker network (Krebs 2002)

Table 3 shows the Betweeness and Closeness SNA calculations below.

Degrees * possible false ID		Betweenness		Closeness	
0.361	Mohamed Atta	0.588	Mohamed Atta	0.587	Mohamed Atta
0.295	Marwan Al-Shehhi	0.252	Essid Sami Ben Khemais	0.466	Marwan Al-Shehhi
0.213	Hani Hanjour	0.232	Zacarias Moussaoui	0.445	Hani Hanjour
0.180	Essid Sami Ben Khemais	0.154	Nawaf Alhazmi	0.442	Nawaf Alhazmi
0.180	Nawaf Alhazmi	0.126	Hani Hanjour	0.436	Ramzi Bin al-Shibh
0.164	Ramzi Bin al-Shibh	0.105	Djamal Beghal	0.436	Zacarias Moussaoui
0.164	Ziad Jarrah	0.088	Marwan Al-Shehhi	0.433	Essid Sami Ben Khemais
0.148	Abdul Aziz Al-Omari*	0.050	Satam Suqami	0.424	Abdul Aziz Al-Omari*
0.131	Djamal Beghal	0.048	Ramzi Bin al-Shibh	0.424	Ziad Jarrah
0.131	Fayez Ahmed	0.043	Abu Qatada	0.409	Imad Eddin Barakat Yarkas
0.131	Salem Alhazmi*	0.034	Tarek Maaroufi	0.409	Satam Suqami
0.131	Satam Suqami	0.033	Mamoun Darkazanli	0.407	Fayez Ahmed
0.131	Zacarias Moussaoui	0.029	Imad Eddin Barakat Yarkas	0.404	Lotfi Raissi
0.115	Hamza Alghamdi	0.026	Fayez Ahmed	0.401	Wail Alshehri
0.115	Said Bahaji	0.023	Abdul Aziz	0.399	Ahmed Al

			Al-Omari*		Haznawi
0.098	Khalid Al-Mihdhar	0.022	Hamza Alghamdi	0.399	Said Bahaji
0.098	Saeed Alghamdi*	0.017	Ziad Jarrah	0.391	Agus Budiman
0.098	Tarek Maaroufi	0.015	Ahmed Al Haznawi	0.391	Zakariya Essabar
0.098	Wail Alshehri	0.013	Salem Alhazmi*	0.389	Mamoun Darkazanli
0.098	Wail Alshehri	0.013	Salem Alhazmi*	0.389	Mamoun Darkazanli
0.098	Waleed Alshehri	0.012	Lotfi Raissi	0.389	Mounir El Motassadeq
0.082	Abu Qatada	0.012	Saeed Alghamdi*	0.389	Mustafa Ahmed al-Hisawi
0.082	Agus Budiman	0.011	Agus Budiman	0.372	Abdelghani Mzoudi
0.082	Ahmed Alghamdi	0.007	Ahmed Alghamdi	0.372	Ahmed Khalil Al-Ani
0.082	Lotfi Raissi	0.007	Ahmed Ressam	0.365	Salem Alhazmi*
0.082	Zakariya Essabar	0.007	Haydar Abu Doha	0.361	Hamza Alghamdi
0.066	Ahmed Al Haznawi	0.006	Kamel Daoudi	0.343	Abu Qatada
0.066	Imad Eddin Barakat Yarkas	0.006	Khalid Al-Mihdhar	0.343	Tarek Maaroufi
0.066	Jerome Courtaillier	0.004	Mohamed Bensakhria	0.339	Ahmed Alghamdi
0.066	Kamel Daoudi	0.003	Nabil al-Marabh	0.335	Waleed Alshehri

0.066	Majed Moqed	0.002	Jerome Courtaillier	0.332	Djamal Beghal
0.066	Mamoun Darkazanli	0.002	Mustafa Ahmed al-Hisawi	0.332	Khalid Al-Mihdhar
0.066	Mohamed Bensakhria	0.002	Said Bahaji	0.332	Saeed Alghamdi*
0.066	Mounir El Motassadeq	0.002	Wail Alshehri	0.328	Majed Moqed
0.066	Mustafa Ahmed al-Hisawi	0.001	Abu Walid	0.324	Ahmed Ressam
0.066	Nabil al-Marabh	0.001	Mehdi Khammoun	0.323	Ahmed Alnami
0.066	Rayed Mohammed Abdullah	0.001	Mohand Alshehri*	0.323	Nabil al-Marabh
0.049	Abdussattar Shaikh	0.001	Raed Hijazi	0.321	Haydar Abu Doha
0.049	Abu Walid	0.001	Rayed Mohammed Abdullah	0.319	Mohamed Bensakhria
0.049	Ahmed Alnami	0.001	Waleed Alshehri	0.316	Essoussi Laaroussi
0.049	Haydar Abu Doha	0.000	Abdelghani Mzoudi	0.316	Jerome Courtaillier
0.049	Mehdi Khammoun	0.000	Abdussattar Shaikh	0.316	Kamel Daoudi
0.049	Osama Awadallah	0.000	Abu Zubeida	0.316	Seifallah ben Hassine
0.049	Raed Hijazi	0.000	Ahmed Alnami	0.314	Rayed Mohammed Abdullah
0.033	Ahmed Ressam	0.000	Ahmed Khalil Al-Ani	0.313	Raed Hijazi

0.033	Bandar Alhazmi	0.000	Bandar Alhazmi	0.311	Abdussattar Shaikh
0.033	David Courtaillier	0.000	David Courtaillier	0.311	Bandar Alhazmi
0.033	Essoussi Laaroussi	0.000	Essoussi Laaroussi	0.311	Faisal Al Salmi
0.033	Faisal Al Salmi	0.000	Faisal Al Salmi	0.311	Mohand Alshehri*
0.033	Lased Ben Heni	0.000	Faisal Al Salmi	0.311	Osama Awadallah
0.033	Mohammed Belfas	0.000	Jean-Marc Grandvisir	0.308	Mehdi Khammoun
0.033	Mohand Alshehri*	0.000	Lased Ben Heni	0.308	Mohamed Abdi
0.033	Seifallah ben Hassine	0.000	Madjid Sahoune	0.307	David Courtaillier
0.016	Abdelghani Mzoudi	0.000	Majed Moqed	0.307	Mohammed Belfas
0.016	Abu Zubeida	0.000	Mamduh Mahmud Salim	0.305	Lased Ben Heni
0.016	Ahmed Khalil Al-Ani	0.000	Mohamed Abdi	0.303	Fahid al Shakri
0.016	Fahid al Shakri	0.000	Mohammed Belfas	0.303	Madjid Sahoune
0.016	Jean-Marc Grandvisir	0.000	Mounir El Motassadeq	0.303	Samir Kishk
0.016	Madjid Sahoune	0.000	Nizar Trabelsi	0.281	Mamduh Mahmud Salim
0.016	Mamduh Mahmud Salim	0.000	Osama Awadallah	0.264	Abu Walid

0.016	Mohamed Abdi	0.000	Samir Kishk	0.250	Abu Zubeida
0.016	Nizar Trabelsi	0.000	Seifallah ben Hassine	0.250	Jean-Marc Grandvisir
0.016	Samir Kishk	0.000	Zakariya Essabar	0.250	Nizar Trabelsi
0.081	Average	0.032	Average	0.052	Average
0.289	Centralization	0.565	Centralization	0.482	Centralization

Table 3: 9/11 Hijacker Betweenness and closeness

3.5.8 *MODULARITY*

The modularity measurement is basically calculating the number of edges within the communities minus the expected number of such edges Modularity is defined on a partition of the nodes in a graph. Let $G = (V.E)$ be an undirected graph modeling a social network with n nodes and m edges. Assume that each node v belongs to community *cv*. We define the indicator function $\delta(cu, cv) = 1$ if and only if $cu = cv$, i.e., u, v are in the same community, and otherwise $\delta(cu, cv) = 0$ for two nodes $u, v \in V$. The modularity Q ‘of this specific community partition is

$$Q = \frac{1}{2m} \sum_{u,v} \left(A_{uv} - \tfrac{d_u d_v}{2m} \right) \delta(c_u, c_v),$$

Where A is the adjacency matrix of G with $A_{uv} = 1$ if $(u, v) \in E$ and 0 otherwise, du and dv are degrees of u and v respectively.

According to Newman (2006) many networks of interest in the sciences, including a variety of social and biological networks, are found to divide naturally into communities or modules. The problem of detecting and characterising this community structure has attracted considerable recent attention. One such metric is the modularity measure proposed by Newman. Modularity is then a measure of the fraction of intra-community edges minus the expected value of the same quantity in a network with the same community divisions, but with edges placed without regard for communities. Mislove

(2009) defines the following: "*Modularity therefore ranges from -1 to 1, with 0 representing no more community structure than would be expected in a random graph, and significantly positive values representing the presence of community structure. In practice, modularity over 0.3 or higher is observed in real-world networks with significant community structure".* Merely the finding that a network contains tightly-knit groups at all can convey useful information: a network contains tightly-knit groups at all can convey useful information: One of the objectives Weinstein et al (2009) aimed to achieve was to employ social network analysis (SNA) algorithms as a filtering step to divide the Jihad group *Jemaah Islamiyah* into distinct communities from the September 2004 bombing of the Australian embassy in Jakarta atrocity. They used the Newman modularity community detection SNA algorithms. The experiment used community detection on the simulated graph, detected communities were searched and the community with the highest number of terror cell actors, who, for the purposes of this experiment, were known in advance. This detected a number of terrorist clutter actors in a particular community. Each clutter actor represents a false alarm, and each terrorist actor represents a positive detection. Given those counts, this allowed to precision calculations and recall measures on the detected community Results also indicated that for smaller graphs the community detection performs quite well with high precision scores. However, as the graph gets larger, precision scores begin to drop dramatically. However when examining communities in networks, one often requires an objective metric to evaluate how "good" a particular division of the network into communities is.

3.5.9 *Diameter and Radius*

Mislove (2009) states in his thesis, *the radius and diameter of a graph, which represents how far away nodes, are from each other in the network. First, the eccentricity of a node v is the maximal shortest path distance between v and any other node. The radius of a graph is then the minimum eccentricity across all vertices, and the diameter is the maximum eccentricity across all vertices. Thus, the radius represents the maximal distance from the most "central" node in the graph to all other nodes, and the diameter represents the maximal distance from the least "central" node in the graph to all other nodes. Due to the computational complexity associated with determining the actual radius and diameter, the* *radius and diameter of a graph is often estimated by calculating the eccentricity of a large random sample of nodes in the network. In such cases, the diameter should be viewed as a lower bound of the true diameter, and the radius as an upper bound of the true radius*.

3.6 CONCLUSION

SNA metrics provide a useful insight into how networks evolve, who the main influencers are as it focuses on interaction rather than behaviour. The literature identifies an array of disciplines the methodology can be applied too. Whilst It is helpful to understand how an online financial community and terrorist community network evolve and be able illustrate how it can be destabilised using the six degrees of separation theory. It maybe more helpful, however, to understand how a network conscripts participants and why people wish to join terrorist or online financial community's networks.

From the literature review the main limitation of social network analysis is the same that applies to any new and innovative technology: social network analysis is just one tool that can be used to understand networks and communities, and is just one piece of the jigsaw. Text mining is another tool that can be used in conjunction with same It is true to state that networks be that online financial networks or terrorist groups, share striking similarities in regards to centrality inbetweenness and modularity. Despite their non-hierarchical approach, most networks are not completely organised in a network structure.

CHAPTER 4

4 EXPERIMENTAL METHODOLOGY: STATISTICAL TEST AND DATA MINING MODELS

The literature review considered the SNA metrics existing techniques and research used within the field. Chapter 4 will portray the data mining and prediction techniques used to formalise the approach taken. These methods were used to perform the basic operations of data mining such as predictive analysis and descriptive analysis. The predictive analysis involves looking at the past history with the intent to predict future behavior. Description analysis looks at deviation and similarity based analysis. The process used for the data extraction follows the basics Knowledge Discovery in databases (KDD) which is concerned with finding useful information and patterns in databases and the use of algorithms is used via the Weka application to extract patterns derived from the KDD process.

4.1 GINI INDEX

The Gini coefficient is a measure of inequality of a distribution, and in this thesis is applied to the distribution of messages posted by users. It is defined as a ratio with values between 0 and 1: the numerator is the area between the Lorenz curve of the distribution and the uniform distribution line; the denominator is the area under the uniform distribution line. It was developed by the Italian statistician Corrado Gini and published in his 1912 paper "Variabilità e mutabilità" ("Variability and Mutability").

The Gini index is the Gini coefficient expressed as a percentage, and is equal to the Gini coefficient multiplied by 100. (The Gini coefficient is equal to half of the relative mean difference.) The Gini coefficient can also be used to measure wealth equality and inequality. This use requires that no one has a negative net wealth. It is also commonly used for the measurement of discriminatory power of rating systems in the credit risk management. The Gini coefficient is defined as a ratio of the areas on the Lorenz curve diagram. The Lorenz curve is a graphical representation of the cumulative distribution function. If the area between the line of perfect equality and Lorenz curve is A, and the area under the Lorenz curve is B, then the Gini coefficient is A/(A+B).

If the Lorenz curve is represented by the function Y = L(X), the value of B can be found with integration and

$$B = \int_0^1 L(X)dX.$$

4.2 *Paired Test*

A paired t-test compares two samples in cases where each value in one sample has a natural partner in the other. A paired t-test looks at the difference between paired values (the mean values) in two samples, takes into account the variation (Standard deviation) of values within each sample, and produces a single number known as a *t-value*. Considering that the experiments are taking pre and post data sets before geopolitical events the pair test will outline the longitudinal differences between both sets of data. The paired test formula is as follows:

$$t_{value} = \frac{\overline{\mu_1} - \overline{\mu_2}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}}$$

Where $\overline{\mu_1}$ is the mean of the first dataset, $\overline{\mu_2}$ is the mean of the second dataset, σ_1 is the standard deviation of the first dataset, σ_2 is the standard deviation of the second dataset and n is the size of the dataset. In our experiments we will have n=154 elements (i.e.22 geopolitical events multiplied by 7 stocks = 154 elements). In order to understand the t-value corresponding to a certain confidence level the student's t-value distribution (see Table 4) is used. For a 90% confidence rating, referring the table 4, a t-value of 1.301 is required when N > 45, while a t-value of 1.679 is needed for a confidence level of 0.95% and 2.41 for a 0.99% confidence level. The computation of the paired testes will be calculated using Network X.

v	0.10	0.05	0.025	0.01	0.005	0.001
1.	3.078	6.314	12.706	31.821	63.657	318.313
2.	1.886	2.920	4.303	6.965	9.925	22.327
3.	1.638	2.353	3.182	4.541	5.841	10.215
4.	1.533	2.132	2.776	3.747	4.604	7.173
5.	1.476	2.015	2.571	3.365	4.032	5.893
6.	1.440	1.943	2.447	3.143	3.707	5.208
7.	1.415	1.895	2.365	2.998	3.499	4.782
8.	1.397	1.860	2.306	2.896	3.355	4.499
9.	1.383	1.833	2.262	2.821	3.250	4.296
10.	1.372	1.812	2.228	2.764	3.169	4.143
11.	1.363	1.796	2.201	2.718	3.106	4.024
12.	1.356	1.782	2.179	2.681	3.055	3.929
13.	1.350	1.771	2.160	2.650	3.012	3.852
14.	1.345	1.761	2.145	2.624	2.977	3.787
15.	1.341	1.753	2.131	2.602	2.947	3.733
16.	1.337	1.746	2.120	2.583	2.921	3.686
17.	1.333	1.740	2.110	2.567	2.898	3.646
18.	1.330	1.734	2.101	2.552	2.878	3.610
19.	1.328	1.729	2.093	2.539	2.861	3.579
20.	1.325	1.725	2.086	2.528	2.845	3.552
21.	1.323	1.721	2.080	2.518	2.831	3.527
22.	1.321	1.717	2.074	2.508	2.819	3.505
23.	1.319	1.714	2.069	2.500	2.807	3.485
24.	1.318	1.711	2.064	2.492	2.797	3.467
25.	1.316	1.708	2.060	2.485	2.787	3.450
26.	1.315	1.706	2.056	2.479	2.779	3.435
27.	1.314	1.703	2.052	2.473	2.771	3.421
28.	1.313	1.701	2.048	2.467	2.763	3.408
29.	1.311	1.699	2.045	2.462	2.756	3.396
30.	1.310	1.697	2.042	2.457	2.750	3.385
31.	1.309	1.696	2.040	2.453	2.744	3.375
32.	1.309	1.694	2.037	2.449	2.738	3.365
33.	1.308	1.692	2.035	2.445	2.733	3.356
34.	1.307	1.691	2.032	2.441	2.728	3.348
35.	1.306	1.690	2.030	2.438	2.724	3.340
36.	1.306	1.688	2.028	2.434	2.719	3.333
37.	1.305	1.687	2.026	2.431	2.715	3.326
38.	1.304	1.686	2.024	2.429	2.712	3.319
39.	1.304	1.685	2.023	2.426	2.708	3.313
40.	1.303	1.684	2.021	2.423	2.704	3.307
41.	1.303	1.683	2.020	2.421	2.701	3.301
42.	1.302	1.682	2.018	2.418	2.698	3.296
43.	1.302	1.681	2.017	2.416	2.695	3.291
44.	1.301	1.680	2.015	2.414	2.692	3.286
45.	1.301	1.679	2.014	2.412	2.690	3.281

Table 4: T – Value confidence rating table

4.3 DECISION TREES USED FOR PREDICTION MODELS

Sas (2012) states that Decision trees are a simple, but a powerful form of multiple variable analyses. They provide unique capabilities to supplement, complement, and substitute for a variety of data mining tools and techniques and statistical intelligence. Decision trees are produced by algorithms that identify various ways of splitting a data set into branch-like segments. These segments form an inverted decision tree that originates with a root node at the top of the tree. The object of analysis is reflected in this root node as a simple, one-dimensional display in the decision tree interface. The name of the field of data that is the object of analysis is usually displayed, along with the spread or distribution of the values that are contained in that field. The display of this node reflects all the data set records, fields, and field values that are found in the object of analysis. The discovery of the decision rule to form the branches or segments underneath the root node is based on a method that extracts the relationship between the object of analysis (that serves as the target field in the data) and one or more fields that serve as input fields to create the branches or segments. The values in the input field are used to estimate the likely value in the target field. The target field is also called an outcome, response, or dependent field or variable. To create a decision tree a list of variables for input is needed this is located in appendix B. The experiment will use the open source platform Weka using the J48 algorithm.

4.4 *THE CONFUSION MATRIX*

A confusion matrix (Kohavi and Provost, 1998) contains information about actual and predicted classifications done by a classification system. Performance of such systems is commonly evaluated using the data in the matrix. The following table shows the confusion matrix for a two class classifier.

actual class (t/f)		
predicted class (p,n)	**tp** (true positive) Correct result	**fp** (false positive) Unexpected result
	fn (false negative) Missing result	**tn** (true negative) Correct absence of result

Table 5: The Confusion Matrix

Based on the data contained in the confusion matrix, several performance indicators can be defined. The most common performance indicators used are precision, recall and the F-measure, defined as follows:

$$\text{Recall} = \frac{tp}{tp+fn}$$

$$\text{Precision} = \frac{tp}{tp+fp}$$

$$F = 2.\ \frac{Precision\ Recall}{Precision+Recall}$$

In order to explain the meaning of these three metrics, we refer to how they are used in the field of information retrieval. If we have a set of records in a database and a set of records to be retrieved by a search engine, in most of the cases the set returned by the search engine may not fully match the set of relevant records.

Recall in information retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved. That means that we have maximum recall if all the relevant documents were retrieved. However, even non-relevant documents might have been retrieved. In order to test the quality of the retrieval, the precision is also needed.

Precision is the fraction of retrieved documents that are relevant to the find that means that in order to have high precision, if 10 documents are retrieved, all of the 10 are relevant (but the recall could be low, meaning that there were more than 10 relevant documents).

The F measures accuracy mixing the statistics precision p and recall r. Precision is the ratio of true positives (tp) to all predicted positives (tp + fp). Recall is the ratio of true positives to all actual positives (tp + fn). Thence the F score is

$$F = 2.\ \frac{Precision}{\text{Precision+Recall}} \times 100 \quad \text{where } p = \frac{tp}{\text{tp+fp}} \qquad r = \frac{\text{tp}}{\text{tp+fn}}$$

4.5 PYTHON PARSER

A Python parser is used to get a list of messages from the online financial communities within Yahoo Finance. This is used to create the networks and compute the SNA metrics of the associated actors from each stock forum on Yahoo Finance. A Python library known as Network X is used in the computation.

4.6 WEKA

Weka is open source software that was used for the predictive modeling in this thesis. Weka uses a collection of machine learning algorithms for data mining and data prediction. The algorithms can either be applied directly to a dataset or called from your own Java code. The application requires a preformatted file pertaining to the experimental data for input into the application. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualisation. Appendix A contains the file used for the experiments.

4.7 GEPHI

Visualisation plays an important role in SNA, for illustration and exploration purposes alone, the visualisation of SNA content allows the examiners to analyse and manipulate the network in the most effective manor and means. Gephi is open source software for graph and network analysis. It uses a 3D render engine to display large networks in real-time and to speed up the exploration. A flexible multi task architecture brings new possibilities to work with, complex data sets and produce valuable visual results Bastian (2009) Pajek is another open software application, however it's widely believed that the Gephi application is a more powerful application. Typically results are displayed in graphic format with graphs that consist of nodes to represent actors and lines on the network. The Gephi application was used to create the network diagrams on this thesis.

4.8 NETWORKX

NetworkX is a python language software package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks. With NetworkX you can load and store networks in standard and nonstandard data formats, generate many types of random and classic networks, analyse network structure, build network models, design new network algorithms, draw networks, and much more (NetworkX)

CHAPTER 5

5 EXPERIMENT INTRODUCTION

This chapter presents a set of experiments performed to test the reaction of online communities and the stock market to geopolitical events related to conflicts and international crisis.

Three experiments were undertaken. The first experiment tests how online communities of investors talking about stocks operating in the military sector change *before* and *after* a geopolitical event. In order to do so, a build of online investors' networks is required and the metrics of different networks such as density, modularity and centralisation will be compared by mean of a statistical test.

The second experiment is concerned with the examination of how prices of military related stocks change in reaction to geopolitical events. In particular, we compare the price of these stocks *before* and *after* the event in order to understand if stock prices are sensitive to such events, and we also compare the price change with the S&P 500 index to understand the behavior of such stocks relatively to the market benchmark.

Finally, in the last experiment a decision tree predictive model is used to investigate if the prices of military stocks a week after a geopolitical event could be predicted by using a set of features encompassing the past prices of each stock and a set of SNA metrics gathered from online community's activity relevant to each stock.

5.1 DATASET

This section describes the dataset collected and used in the three experiments of this chapter. The naming convention and notation are also introduced.

5.1.1 *MILITARY STOCKS*

The inception point was to select a number of stocks operating in the defense and military sector. These are mainly medium and big capitalisation manufacturing companies with a strong R&D department.

We call S the set of stocks identified. S contains 7 US military stocks, each of them identified by its market ticker. For instance, we use the symbol S_{UTX} to refer to the stock

United Technologies. The stocks are listed in table 6. The column capitalisation contains the capitalisation of each stock in billions of dollars, while the percentage in parenthesis is the relative size of each stock over the total capitalisation of all the seven stocks. The seven military stocks have a total capitalisation of about $300 billion, which represents about 2% of the total capitalisation of the S&P 500 index, estimated at about $15 trillion in July 2014.

Stock	**Ticker**	**Capitalisation**	**Description**
Honeywell International	HOC	74.7B $ (24.84%)	Honeywell International Inc. operates as a diversified technology and manufacturing company worldwide. Its Aerospace segment provides turbine propulsion engines, auxiliary power units, environmental control and electric power systems, engine controls, flight safety, communications, navigation, radar and surveillance systems, and aircraft lighting products for aircraft manufacturers, airlines, business and general aviation, military, space, and airport operations, as well as offers management and technical, logistics, aircraft wheels and brakes, and repair and overhaul services
United Technologies	UTX	100.34B $ (33.37%)	United Technologies Corporation provides technology products and services to the building systems and aerospace industries worldwide. It also offers electronic security products, including intruder alarms, access control systems, and video surveillance systems; and monitoring, response, and security personnel services.. Its Pratt & Whitney segment supplies aircraft engines for commercial, military, business jet, and general aviation markets, as well as provides fleet management services for commercial engines. The company's UTC Aerospace Systems segment supplies aerospace products. Its Sikorsky segment manufactures military and commercial helicopters, as well

			as provides aftermarket helicopter and aircraft parts and services.
L-3 Communication Holding	LLL	9.42B $ (3.14%)	L-3 Communications Corporation, provides command, control, communications, intelligence, surveillance, and reconnaissance (C3ISR) systems; aircraft modernisation and maintenance; and national security solutions in the United States and internationally. The company operates in four segments: Aerospace Systems, Electronic Systems, Communication Systems, and National Security Solutions. It offers a range of products and services, including components, products, subsystems, and systems, as well as related services to military and commercial customers in various business areas.
Lockheed Martin Corporation	LMT	55.7B $ (18.51%)	Lockheed Martin Corporation, a security and aerospace company, is engaged in the research, design, development, manufacture, integration, and sustainment of advanced technology systems, products, and services for defense, civil, and commercial applications in United States and internationally. The company operates in five segments: Aeronautics, Information Systems & Global Solutions, Missiles and Fire Control, Mission Systems and Training, and Space Systems. The Aeronautics segment offers military aircrafts, such as combat and air mobility aircrafts, unmanned air vehicles, and related technologies. The Missiles and Fire Control segment offers air and missile defense systems; tactical missiles and air-to-ground precision strike weapon systems; logistics and other technical services; fire control systems; and manned and unmanned ground vehicles.
Alliant	ATK	4.1B $	Alliant Techsystems Inc. develops and

Technologies		(1.4%)	produces aerospace, defense, and commercial products to the U.S. government, allied nations, and prime contractors in the United States, and internationally. It supplies ammunition, firearms, and shooting accessories. The Aerospace Group segment offers rocket motor systems for human and cargo launch vehicles, conventional and strategic missiles, missile defense interceptors, small and micro-satellites, satellite components, structures and subsystems, lightweight space deployables, solar arrays, decoy and illuminating flares, and aircraft countermeasures. The Defense Group segment provides military small, medium, and large caliber ammunition; propulsion systems for tactical missiles and missile defense applications; strike weapons; precision munitions; gun systems; aircraft survivability systems; fuses and warheads; energetic materials; and special mission aircraft.
Northrop Grumman Corporation	NOC	26.5B $ (8.81%)	Northrop Grumman Corporation provides systems, products, and solutions in aerospace, electronics, information systems, and technical service areas to government and commercial customers worldwide. The company's Aerospace Systems segment designs, develops, integrates, and produces manned aircraft, unmanned systems, spacecraft, high-energy laser systems, microelectronics, and other systems and subsystems. This segment sells its products primarily to government agencies for use in various mission areas. Its Electronic Systems segment offers solutions for sensing, understanding, anticipating, and controlling operating environment to military, civil, and commercial customers. This segment's solutions comprise defense electronics and

			systems, airborne fire control radars, situational awareness systems, early warning systems, airspace management systems, navigation systems, communications systems, marine power and propulsion systems, space systems, and logistics services.
Raytheon Co.	RTN	30.1B $ (10.1%)	Raytheon Company develops integrated products, services, and solutions in the areas of sensing; effects; command, control, communications, and intelligence; mission support; and cyber and information security worldwide. It operates in four segments: Integrated Defense Systems; Intelligence, Information, and Services; Missile Systems; and Space and Airborne Systems. The Integrated Defense Systems segment provides integrated air and missile defense; radar solutions; naval combat and ship electronic systems; command, control, communications, computers, and intelligence solutions; and air traffic management systems. The Intelligence, Information, and Services segment offers a range of technical and professional services, intelligence, surveillance and reconnaissance, navigation, DoD space and weather solutions, cybersecurity, analytics, training, logistics, mission support. The Missile Systems segment develops and supports a range of weapon systems, including missiles, smart munitions, close-in weapon systems, projectiles, kinetic kill vehicles and combat sensor solutions.

Table 6: List of military stocks

5.1.2 *GEOPOLITICAL EVENTS*

A number of geopolitical events were collected related to global conflicts, war or political instability of recent years. We call the set of events $\mathcal{E}$, and a specific event $e \in \mathcal{E}$. Each event is represented by a date, the conflict it belongs to and the text of the news associated with the event. News is supposed to be market-sensitive, especially for military stocks, and supposed to trigger a reaction on the online communities of investors. There was a large amount of geopolitical events gathering pertaining to 10 conflicts. The data set was condensed down to 22 geo-political events from three recent conflicts: the Libyan rebellion during the Arab Spring, the Syrian civil war and the Ukraine crisis. The time interval of the events goes from the beginning of 2011 to April 2014. The British Broadcasting Corporation (BBC) News was used as the source for collecting the data.

ID	**Date**	**Conflict**	**Event News**
1	16/02/2011	Libya	Libyan protesters clash with police in Benghazi Arrest of human rights activist triggers demonstrations in Libya's second largest city
2	22/02/2011	Libya	NATO to take control in Libya after US, UK and France reach agreement Britain, France and the US have agreed that NATO will take over the military command of the no-fly zone over Libya in a move that represents a setback for Nicolas Sarkozy, who had hoped to diminish the role of the alliance.
3	24/03/2011	Libya	Libyan plane and tanks destroyed by allied jets French fighter jets have destroyed a Libyan plane in the coastal city of Misrata in the first enforcement of the no-fly zone imposed by the UN to try to halt Muammar Gaddafi's anti-rebel offensive.
4	21/04/2011	Libya	Drones can be used by NATO forces in Libya, says Obama
5	27/06/2011	Libya	War crimes court issues Gaddafi arrest warrant ICC orders Libyan leader and his son Saif al-Islam Gaddafi to stand trial on charges of torturing and killing civilians and rebels
6	30/06/2011	Libya	Boeing projects $300M overrun on tanker project.
7	29/06/2011	Libya	France admits arming rebels. NATO is reviewing the conduct of its military campaign in Libya after France admitted arming rebel fighters in apparent defiance of the UN mandate.

8	03/03/2014	Ukraine	Russia's parliament approves Vladimir Putin's request to use force in Ukraine to protect Russian interests. Pro-Russian rallies are held in several Ukrainian cities outside Crimea, including the second-biggest city Kharkiv. Barack Obama tells Mr Putin to pull forces back to bases. Ukraine's interim PM Yatsenyuk says Russia has effectively declared war. US says Russia is in control of Crimea. "Black Monday" on Russian stock markets as reports suggests Russia's military had issued a deadline for Ukrainian forces in Crimea to surrender. The reports are later denied. Russia's UN envoy says toppled President Yanukovych had asked the Russian president in writing for use of force.
9	07/03/2014	Ukraine	7/8/9/10th March 2014: Russia says it will support Crimea if the region votes to leave Ukraine. Russia's state gas company Gazprom warns Kiev that its gas supply might be cut off. The US and France warn of "new measures" against Russia if it does not withdraw its forces from Ukraine. Warning shots are fired at international monitors trying to enter Crimea. Armed men seize a military hospital in Simferopol.
10	11/03/2014	Ukraine	The European Commission offers Ukraine trade incentives worth nearly 500m euros ($694m; £417m). Ukrainian MPs ask the US and UK to use all measures, including military, to stop Russia's aggression.
11	13/03/2014	Ukraine	Barack Obama pledges to stand with Ukraine during a meeting with interim Prime Minister Arseniy Yatsenyuk at the White House.
12	30/04/2014	Ukraine	Acting President Olexander Turchynov reinstates conscription, warning Ukraine is on "full combat alert". Pro-Russians take over the regional prosecutor's office in eastern Donetsk.
13	15/03/2011	Syria	Activists call for a "Day of Rage" across Syria, inspired by other popular uprisings across the Arab world. In February, several youths were arrested in the southern town of Daraa for writing graffiti calling for the downfall of the regime of President Bashar Assad.
14	18/03/2011	Syria	Activists say five people were killed as security forces dispersed crowds in Daraa — one of several demonstrations across the country — in the first deadly violence reported in the uprising. Unrest spreads in coming months.

15	26/04/2011	Syria	Thousands of soldiers backed by tanks and snipers open fire on civilians in Daraa and two other locations, according to witnesses. Armed security agents conduct house-to-house sweeps. Neighborhoods are sectioned off and checkpoints are erected. Electricity, water and cellphone services are cut. At least 11 people are killed and 14 others lay in the streets, either dead or gravely wounded
16	18/05/2011	Syria	U.S. imposes sanctions on Assad and senior Syrian officials for human rights abuses.
17	05/08/2011	Syria	The United States, Britain, France and Germany and the European Union demand Assad resign, saying he is unfit to lead.
18	6/10/2011	Syria	Russia and China veto a European-backed U.N. Security Council resolution that threatens sanctions against Syria if it doesn't immediately halt its military crackdown against civilians.
19	23/12/2011	Syria	Back-to-back car bombs near Syria's intelligence agencies in Damascus kill at least 44 in first major attack in the heart of the capital. Syria's state-run TV blames Al-Qaeda militants.
20	04/02/2012	Syria	Russia and China veto a resolution in the U.N. Security Council that backs an Arab League plan calling for Assad to step down. The double-veto outrages the U.S. and European council members who fear it will embolden Assad regime.
21	20/08/2012	Syria	Obama says U.S. will reconsider its opposition to military involvement in Syria if Assad's regime deploys or uses chemical or biological weapons, calling such action a "red line" for the United States
22	3/12/2012	Syria	Speaking of chemical weapons, Obama says Assad should know "if you make the tragic mistake of using these weapons, there will be consequences and you will be held accountable."

Table 7: List of military stocks

5.1.3 *Stock Prices*

Historical prices of the stocks were collected in the set S and the value of the S&P 500 index, used as the market benchmark in our experiments. Closing prices adjusted by dividends and splits were collected via a Bloomberg terminal.

Given a stock x and an event $e_i \in \mathcal{E}$, we use the following notation: $P_{e_i}(x)$ is the price of stock x at the date of event e_i, while $P_{e_i}^{\pm d}(x)$ is the closing price of stock x after (or before) d days from the event. For instance, referring to the *event id* of table 7 and the tickers of table 6, $P_{e_1}^{+5}(UTX)$ is the closing price of the stock UTX five days after the event e_1 (a Libyan-related event happened on the 16th of February 2011).

The return of each stock (also called the *gain* of a stock) is expressed as a percentage in [0,1] and it is denoted by G. For istance, $G_{e_i}^{\pm d}(x)$ is the gain of the stock x after (or before) d days from the event e_i. By definition it is:

$$G_{e_i}^{+d}(x) = \frac{P_{e_i}^{+d}(x) - P_{e_i}(x)}{P_{e_i}(x)}\ , \qquad G_{e_i}^{-d}(x) = \frac{P_{e_i}(x) - P_{e_i}^{-d}(x)}{P_{e_i}^{-d}(x)}$$

Stock prices follow a log-normal distribution, while stock returns follow a normal distribution.

5.1.4 *Online Communities Data*

Online message board data was collected about online investors' interactions from 2010 till 2014. Our source of online communities' data is represented by Yahoo! Finance Message Boards.

Yahoo! Finance keeps a message board for each stock quoted on the US market. Each message board is a stream of threads opened by registered users. Each thread is a stream of messages posted by users. A user can decide to add a new message to a thread, answer to a message or open a new thread.

A preexisting DIT parser using *Python 2.7* programming language, the *urllib* library and regular expression was used for the message board extraction from the Yahoo Finance website. The parser collected the discussions regarding the seven military stock of interest from 2010 till July 2014. Data was collected about the list of threads, the list of messages for each thread, the content of each message, time of the message, users and the citations between messages. Data collected by the parser was stored in a MySQL database. Table 8 describes in details the data collected and it is an exact mapping of the database table used.

Field	Description
Message ID	An incremental unique ID of each message posted
Thread ID	An incremental unique Id of each Thread opened
Thread Title	The title of the thread
Message Timestamp	Timestamp when the message was posted
Message Content	The text of the message
Stock	The stock message board the message refers to (one of the seven military stocks considered)
User	A unique username of the author of the message
Message Quoted	If the message is a response to another message, the field contains the ID of the message quoted.

Table 8 Data collected from Yahoo! Finance Message Boards

There was approximately 85,000 messages regarding the 7 stocks examined, written in about 9,500 threads by about 3,850 users.

5.1.5 *BUILDING A NETWORK OF INVESTORS*

Using the Message Board data described in Table 8, I was able to define a social network for each stock and a specific interval of time$[t_1, t_2]$. The nodes of the networks were represented by users positng a message in the interval of time, while an edge is drawn from node a to node b if user a quoted at least one message written by user b.

The notation followed is the following. We call $N^{[t_1,t_2]}(x)$ the network of online users for the stock x built considering all the messages posted in the interval $[t_1, t_2]$. Since we are interested in the behaviour of online communities *before* and *after* a geopolitical event e_i, the social network of investors d days *before* an event e_i will be represented by the time interval $[t_{e_i} - d, t_{e_i}]$, where t_{e_i} is the timestamp of event e_i. In the same way, the network

d days *after* an event e_i is identified by the timestamp $[t_{e_i}, t_{e_i} + d]$. We use the shorter notation $N_e^{-d}(x)$ and $N_e^{+d}(x)$ to identify the network d days before and after an event e_i.

In order to describe each network N, the SNA metrics described in Table 9 were computed.

Indicator	Description
$N_d(N)$	Number of nodes in the networks, equal to the number of users active on the network
$E_d(N)$	Number of edges of the networks, equal to the number of users cited
$\mathcal{D}(N)$	Density of the network N
$\mathcal{C}(N)$	Freeman's centralization of the network N
$\mathcal{B}tw(N)$	Average in-betweennes centrality of the network N
$Mod(N)$	Newmann's modularity of the network N
$Gini(N)$	The Gini index of the distribution of the number of messages posted by each user in the network. Low value of the Gini index are an indicator of messages equally spread over users and viceversa. Note how this indicator is not strictly speaking an SNA indicator
Diameter(N)	Diameter of the network N

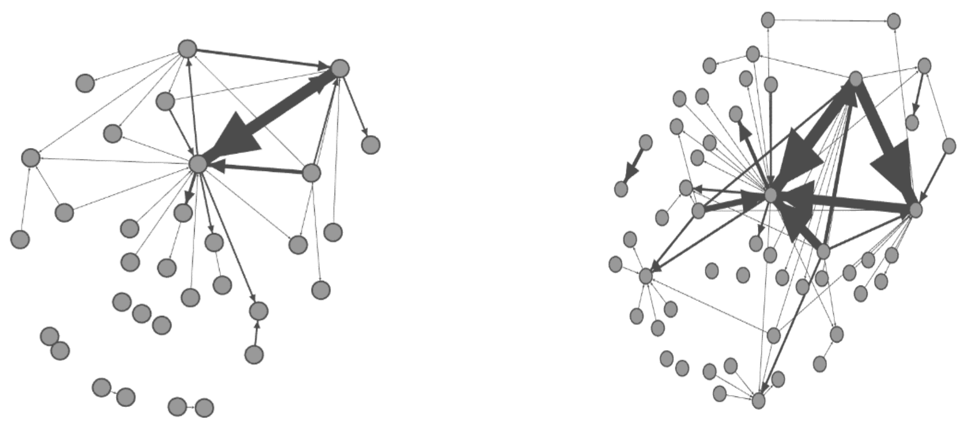
Table 9 SNA indicators computed for each network.

5.1.6 *Experiment 1: Testing online investors reaction to geopolitical events*

This experiment wants to verify if the structure of the network of online investors is significantly modified by geopolitical events. In order to investigate the issue we defined, for each stock s in S and each event e_i a pair of networks: one including all the online community activity d days before the event (called the *before* network) and one including all the activity d days after the event (called the *after* network). Scope of the experiment is to verify if there is a statistical difference between the SNA metrics of the *before* and *after* network. Therefore, we will perform a paired t-test for each metrics.

In the experiment, we used a value of the number of days d equal to 2, 5 and 20 trading days (equal to 2 days, one week and one month), to test the reaction of the community at different time interval.

Since we have a dataset of 22 events, 7 stocks and three different time interval (2,5 and 20 days), we have collected a total of 462 *before* and 462 *after* networks. Figure 9 depicts one of the most common patterns identified between *before* and *after* networks. Usually, the *after* network is more centralised, with higher number of users and quotations and it is distributed around few central actors, while the *before* networks appears less united with a higher number of isolated nodes.



Before **After**

Figure 9: The network of online investors talking about UTX the week before and after Obama said U.S. will reconsider its opposition to military involvement in Syria (20th August 2012). The graph on the right has a higher number of nodes and links, it is more connected and the core actors of the network make it more centralised.

The following tables show the results of a paired t-test performed between the *before* and *after* networks. We divided our results according to the size of the time interval considered (2, 5 and 20 days).

Indicators	Mean Difference	$\frac{\sigma_{diff}}{\sqrt{n}}$	T value	Significance
$N_d(N)$	3.64	1.28	2.84	+++
$E_d(N)$	3.21	1.6	1.99	++
$\mathcal{D}(N)$	0.014	0.023	0.74	=
$\mathcal{C}(N)$	0.048	0.032	1.74	++
$\mathcal{B}tw(N)$	0.64	0.413	1.55	+
$Mod(N)$	-0.164	0.104	-1.58	-
$Gini(N)$	-0.037	0.056	-0.66	=
$Dia(N)$	1.12	0.82	1.36	+

Table 10 Results of the paired t-test between the before and after networks with d=2

Table 10 illustrates the results of the paired t-test between the before and after networks with d=2 days. N= 154. The symbol +++ means statistical significance with a 0.99% confidence level and that the values of the after network are greater than the values of the before networks, ++ and + corresponds to a significance level of 0.95% and 0.90% respectively. The symbol ---, -- and – denotes statistical difference at 0.99, 0.95 and 0.90 confidence level but where values of the before network are greater than ones of the after network.

Indicators	Mean Difference	$\frac{\sigma_{diff}}{\sqrt{n}}$	T value	Significance
$N_d(N)$	2.62	1.41	1.85	++
$E_d(N)$	1.72	1.83	0.93	=
$\mathcal{D}(N)$	0.015	0.025	0.6	=
$\mathcal{C}(N)$	0.047	0.031	1.52	+

$\mathcal{B}tw(N)$	0.625	0.452	1.38	+
$Mod(N)$	-0.04	0.153	-0.26	=
$Gini(N)$	-0.028	0.071	-0.39	=
$Dia(N)$	0.89	1.01	0.88	=

Table 11 Results of the paired t-test between the before and after networks with d=5 days

Indicators	Mean Difference	$\frac{\sigma_{diff}}{\sqrt{n}}$	T value	Significance
$N_d(N)$	-0.46	1.38	-0.33	=
$E_d(N)$	-0.86	1.94	-0.44	=
$\mathcal{D}(N)$	0.013	0.028	0.45	=
$\mathcal{C}(N)$	0.028	0.054	0.518	=
$\mathcal{B}tw(N)$	0.34	0.493	0.689	=
$Mod(N)$	-0.104	0.194	-0.536	=
$Gini(N)$	-0.059	0.0626	-0.94	= (- with cl 80%)
$Dia(N)$	-0.43	0.89	-0.48	=

Table 12 Results of the paired t-test between the before and after networks with d=20 days

5.1.7 *Experiment 1: Analysis of Results*

Tables 10, 11, and 12 show a clear trend. When we consider the period immediately after a geopolitical event (d=2 days), the majority of the SNA metrics except density, modularity and Gini index are significantly higher in the *after* network than in the *before* network. Based on experimental results, we can conclude that during the two days following a geopolitical event the following happens:

- More online users are talking about military stocks, as evidenced by the high increase in the number of nodes in the *after* network (t-value = 2.84, 0.99 confidence level)

- These users are interacting more, as evidenced by the high increase in the number of edges in the after network (t-value=1.99). This behaviour could be interpreted as an increased collective effort of the users to join strengths together and try to discuss and make sense of the consequences of the recent event. Casnici et al (2014) call this behaviour *joint-attention* of online investors.

- The centralisation and average in-betweennes centrality of the *after* networks increase significantly. This means that in the *after* networks there are a group of actors with high importance in the network, acting as hubs and central point of reference during the discussion. This could be interpreted as the presence of a small group of authoritative users that take centre stage when there is something potentially critical to discuss

- In accordance with this, the modularisation of the network is significantly lower in the *after* network. This means that the network is less segregated in sub-communities and it is more centralised into one big discussion. However, on an absolute scale the values of modularity *before* and *after* are low (less than 0.25) meaning that in general the networks of online investors show little segregation.

- The graph density is not significantly changed, even if it is higher in the *after* networks, while the Gini index of the distribution of messages per user is not significantly different but it is *lower* in the after network. We remind how the Gini index measures the inequality of a distribution, and a low Gini value is an index of a more uniform and "democratic" distribution. Therefore, there is a tendency in the *after* network to host a more democratic discussion where more users have the possibility to interact.

When the period of observation increases, (d=5 or d=20 days), this effect fades quickly. 5 days networks the number of nodes, the centralisation and in-betweennes are still significantly higher in the *after* network, meaning that the discussions after an event are still bigger and more centralised after 1 week. In the 20 days network there is no statistical difference, meaning that after 1 month from the event online discussions do not differ significantly. The highest t-value is represented by the Gini index, lower in the *after* network, suggesting again how discussions after an event have a tendency to be more democratic.

5.1.8 *EXPERIMENT 2: MARKET REACTIONS TO GEOPOLITICAL EVENTS*

In this experiment we test if the price of the seven military stocks after a geopolitical event differs significantly from the price before the event. Moreover, we also compare the price change of the military stocks with the S&P 500 index, in order to understand if the military stock behaves in a different way than the overall market trend, and to what extent the military stocks under-performs or outperforms the market benchmark.

We follow a similar methodology as experiment 1. We consider the price change (i.e. the gain) of each stock 5 days and 1 day before the event and we compare it with the gain after the event using a paired t-test. Using our notation, we check the statistical difference between the gain $G_{e_i}^{-1}(x)$ and $G_{e_i}^{+1}(x)$ for all the stocks x and event e_i ; and the same for $G_{e_i}^{-5}(x)$ and $G_{e_i}^{+5}(x)$. We therefore include all the stocks in the comparison without performing different experiments for different stocks. In this way we check if the set of military stocks are overall changing their price in reaction to geopolitical events. Table 13 shows the results of the price comparisons. Table 14 shows the same experiment with the S&P 500 prices, to check how the S&P 500 is also reacting to geopolitical events.

In the last part of the experiment I examine if the price change of the military stocks differs significantly from the S&P 500 price change. The comparison is valid since, even if all the seven military stocks are part of the S&P 500 index, their total capitalisation is about 2% of the index capitalisation, and therefore they are too small to manipulate the price. In order to do a meaningful comparison with the S&P 500 index, we compute an aggregated index for our seven stocks in the same way the S&P 500 is computed. In fact, the S&P 500 is a weighted average of each stock value based on the capitalisation of each stock. We followed the same approach. The price changes of each military stock are weighted by the capitalisation of the stock relative to the total capitalisation of all the seven stocks. In table 6, the column *Capitalisation* reports the percentage of each stock over the total. For instance, HON represents almost 25% of the total capitalisation of the seven military stocks. Finally, Table 15 shows the comparison between the S&P 500 index and our military stocks index 1 and 5 days after an event. G_{ml}^{+d} is used to refer to the gain of our military stocks index after d days from the event.

Indicators	Mean Difference	$\frac{\sigma_{diff}}{\sqrt{n}}$	t-value	Significance
$G^{-1} vs. G^{+1}$	0.25%	0.002158	1.17	= (+ with cl=0.85)
$G^{-5} vs. G^{+5}$	1.18%	0.005196	2.27	++

Table 13 Comparisons between military stocks price before and after a geopolitical event

Indicators	Mean Difference	$\frac{\sigma_{diff}}{\sqrt{n}}$	t-value	Significance
$G_{sp}^{-1} vs. G_{sp}^{+1}$	0.089%	0.003158	0.281	=
$G_{sp}^{-5} vs. G_{sp}^{+5}$	0.69%	0.0044	1.53	+

Table 14 Comparisons between S&P 500 price before and after a geopolitical event

Indicators	Mean Difference	$\frac{\sigma_{diff}}{\sqrt{n}}$	t-value	Significance
$G_{ml}^{+1} vs. G_{sp}^{+1}$	0.042% (avg G_{ml}^{+1}=0.075% avg G_{sp}^{+1}=0.033%)	0.001609	0.281	=
$G_{ml}^{+5} vs. G_{sp}^{+5}$	0.38% (avg G_{ml}^{+5}=2.15%, avg G_{sp}^{+5}=1.77%)	0.00303	1.254	= (+ with cl=0.85)

Table 15 Comparisons between military stock prices v S&P 500 index prices before and after a geopolitical event.

5.1.9 *EXPERIMENT 2: ANALYSIS OF RESULTS*

Table 13 confirms that, one week after the geopolitical event, the price of military stocks is higher than the price before the event (confidence level 0.99). However, this effect is much smaller after one day immediately after the event, where the price is still higher but a t-value of 1.17 guarantees statistical significance only for a confidence level of 0.85.

The though process then shifted to if this increase is specific to military stocks or it also affects the market, represented by the S&P 500 index. Table 14 shows that actually the entire market increased after one week from an event, even if with smaller confidence level (confidence level=0.9, t-value=1.53), while there is no statistical difference for the price after 1 day from the event.

Finally, given that both the market and the military stocks significantly increased their value after an event, an examination took place to ascertain if there is a statistical difference between the increase of the seven military stocks and the increase of S&P 500. Table 15 shows how there is no statistical difference 1 day after the event, while there is a low statistical difference in favour of the military stocks one week after the event (confidence level = 0.85, t-value=1.25).
We conclude how both military stocks and the market index significantly increase one week after geopolitical events, and there is a low tendency of military stocks to outperform the market during that week.

5.1.10 *EXPERIMENT 3: PREDICTING STOCK PRICE MOVEMENTS AFTER GEOPOLITICAL EVENTS*

In this experiment we investigate if it is possible to predict the price of a military stock one week after a geopolitical event based on a set of features including SNA metrics, historical prices and S&P 500 index prices.

The set of features includes the SNA metrics of the *before* networks at 2, 5 and 20 days, plus the SNA metrics of network the day of the event. Regarding historical prices, we include the price of 1 week before the event, the price of the day before the event and the closing price of the day of the event. We also include the same historical prices for the S&P 500.

Using a J48 decision tree algorithm, our goal is to predict G^{+5}, i.e. the future movement of the price of a stock after 5 days form the event. We model the prediction problem as a binary classification problem, meaning that we aim to predict if the value of G^{+5} will be above or below a certain threshold th. Even if the most obvious choice is $th = 0$, we actually chose the value $th = 0.01$, that means that we aim to predict if the price of the stock will rise above 1% in the next following days. The reasons why we have set this threshold are the following.

First, there is a reason linked to our dataset. Our dataset covers a period of time from 2011 to 2014, a period where the US stock market was mainly bullish with a steady positive trend. In our dataset 86.4% of cases the stock price increased a week after an event. Therefore, by using a threshold of 0 the two classes to be predicted would be too unbalanced.

On the other side, the median value of the weekly stock price change in our dataset is equal to 1.21%. Therefore, by setting a threshold at 1%, (quite close to the median value), our dataset includes about 62% of positive case and 38% of negative ones.

Moreover, there is a technical trading consideration. Predicting if the price will rise or decrease is not enough for sustaining a profitable trading strategy after commission costs are considered. A gain of 1% is regarded as a solid psychological threshold for a successful weekly trading strategy.

The following table describes the list of features used in our classifier. Note how features are divided into market-based (mainly historical prices of stocks and S&P 500 index) and SNA-based features.

Feature	**Description**
In all the features, the apex d represent the time interval and it takes the values -10, -5, -1 and 1.	
$N_d^d(N)$	Number of nodes in the networks, equal to the number of users active on the network
$E^d(N)$	Number of edges of the networks, equal to the number of users cited
$\mathcal{D}^d(N)$	Density of the network *N*
$\mathcal{C}^d(N)$	Freeman's centralisation of the network *N*
$\mathcal{B}tw^d(N)$	Average in-betweeness centrality of the network *N*
$Mod^d(N)$	Newmann's modularity of the network *N*
$Gini^d(N)$	The Gini index of the distribution of the number of messages posted on the network. Note how this indicator is not strictly speaking an SNA indicator
$Dia^d(N)$	Diameter of the network *N*
S	Stock considered, belonging to the set *{UTX, HON, NOC, ATK, LLL, LMT, RNT}*
$G_{sp}^{-10}, G_{sp}^{-5}, G_{sp}^{-1}, G_{sp}^{0}$	Return of the S&P 500 index 10, 5 and 1 days before the event and the day of the event
$G_s^{-10}, G_s^{-5}, G_s^{-1}, G_s^{0}$	Return of the stock S 10, 5 and 1 days before the event and the day of the event
$G_s^{+5} \geq th = 0.01$	Binary class to be predicted

Table 16 List of Features used

Considering that each feature is considered with a time interval of 10, 5, and 1 day before the event and 1 day after, there are 41 features and 1 predictive class to be considered. All of them are numeric, except the feature stick that is nominal.

The database contains 462 elements, and it has been divided into training and a testing set with a 70/30 split. Using the open source software Weka, we trained three different J48 decision tree model with a pruning factor of 0.02 (results are not changing significantly for other choices of the factor). The first model (called $\mathcal{M}_{all}$ has been trained using all the features, the second model, called $\mathcal{M}_{price}$, has been trained using market-related features (stock and index historical prices) and the third model ($\mathcal{M}_{SNA}$) has been trained using SNA-related features.

The following figures describe the results for the three classifiers. The key metrics used to compare the models are based on the analysis of the confusion matrix, such as precision, recall, F-measure. Although trials were performed for each model, the performance of the models did not show high variance and therefore the data presented are good representative of a typical performance for each model.

```
=== Summary ===
Number of Leaves  :       13
Size of the tree :  20

Correctly Classified Instances          120               83.5052 %
Incorrectly Classified Instances        24                16.4948 %
Kappa statistic                          0.6468
Mean absolute error                      0.1755
Root mean squared error                  0.3797
Relative absolute error                 37.1309 %
Root relative squared error             78.1412 %
Total Number of Instances               144

=== Detailed Accuracy By Class ===

               TP Rate    FP Rate    Precision    Recall   F-Measure    ROC Area
Class
               0.757     0.117      0.8        0.757     0.778      0.861    0
               0.883     0.243      0.855      0.883     0.869      0.861    1
Weighted Avg.  0.835     0.195      0.834      0.835     0.834      0.861

=== Confusion Matrix ===

  a   b   <-- classified as
 42   12  |  a = 0
 11   78  |  b = 1
```

Figure 10 Typical performance of the $\mathcal{M}_{\mathrm{all}}$ model.

```
=== Summary ===
Number of Leaves  :       8
Size of the tree :  15

Correctly Classified Instances          119               82.4742 %
Incorrectly Classified Instances        25                17.5258 %
Kappa statistic                          0.6267
Mean absolute error                      0.1954
Root mean squared error                  0.4008
Relative absolute error                 41.3502 %
Root relative squared error             82.4754 %
Total Number of Instances               144

=== Detailed Accuracy By Class ===

               TP  Rate    FP  Rate    Precision    Recall   F-Measure    ROC  Area
Class
                 0.757     0.133      0.778     0.757     0.767      0.827    0
                 0.867     0.243      0.852     0.867     0.86       0.827    1
Weighted Avg.    0.825     0.201      0.824     0.825     0.824      0.827

=== Confusion Matrix ===

  a    b   <-- classified as
 41    13 |  a = 0
 11    78 |  b = 1
```

Figure 11 Typical performance of the $\mathcal{M}_{\text{price}}$ price model.

```
=== Summary ===
Number of Leaves  :       19
Size of the tree :  32

Correctly Classified Instances          93               64.5833 %
Incorrectly Classified Instances        51               35.4166 %
Kappa statistic                          0.2415
Mean absolute error                      0.3531
Root mean squared error                  0.5444
Relative absolute error                 74.7013 %
Root relative squared error            112.0202 %
Total Number of Instances               144

=== Detailed Accuracy By Class ===

               TP  Rate    FP  Rate    Precision    Recall   F-Measure    ROC  Area
Class
                 0.486     0.25       0.545     0.486     0.514      0.68     0
                 0.75      0.514      0.703     0.75      0.726      0.68     1
Weighted Avg.    0.649     0.413      0.643     0.649     0.645      0.68

=== Confusion Matrix ===

  a     b    <-- classified as
 28    26 |  a = 0
 21    68 |  b = 1
```

Figure 12 Typical performance of the $\mathcal{M}_{\text{SNA}}$ model.

Model	Correct predictions	Precision	Recall	F-Measure
$\mathcal{M}_{all}$	120/144	0.834	0.835	0.834
$\mathcal{M}_{price}$	119/144	0.824	0.825	0.827
$\mathcal{M}_{SNA}$	93/144	0.643	0.649	0.645

Table 17 Summary of Precision, Recall and F measure results

5.1.11 *Experiment 3: Analysis of Results*

There is a clear distinction among the two models that includes market-related features $\mathcal{M}_{all}$ and $\mathcal{M}_{price}$ and the $\mathcal{M}_{SNA}$ model. The $\mathcal{M}_{all}$ and the $\mathcal{M}_{price}$ model has high and very similar performance, with an accuracy of about 82%-83%. In general, this represents an extremely high value considering the fact that stock market prices forecast is a hard task and there is a strong consensus in literature that stock market prices behave like random walks and therefore cannot be predicted. However, in period of sustained market trend (such as 2011-2014), similar numbers have been reported in literature. For instance, Bollen (2010) studied the predictive capability of Twitter discussions and he reported an accuracy of 83.3% in predicting the closing price of the S&P 500 index in October 2010.

However, the main result of our analysis is not the absolute level of accuracy of the predictors, but the fact that the addition of SNA metrics to the feature set did not add any value. $\mathcal{M}_{all}$ and the $\mathcal{M}_{price}$ have comparable performance, meaning that a classifier based on historical prices only has the same predictive capabilities of a classifier based on historical prices and SNA metrics. Moreover, the $\mathcal{M}_{SNA}$ model, only based on SNA metrics, has an accuracy of around 65%, underperforming the other two classifiers by about 18%. Moreover, an accuracy of 65% is still significantly higher than a naïve zero-rule classifier (62% positive case, 38% negative case).

Therefore, it can be concluded that online communities' metrics has little or no predictive value in helping predicting the future price of stocks.

5.1.12 *Conclusions*

In this chapter we have presented a set of experiments to analyse how online financial communities and the market react to geopolitical events related to conflicts, war and political instability.

Our experiment on online communities of investors showed how the dynamics of such communities do change in the 2 days immediately after a geopolitical crisis. The discussions are bigger, more intense, centralised and they gravitate around a few group of individuals that appear to be authoritative users leading the discussion and connecting the remaining actors in the networks. This effect fades rapidly one week from the event and it disappears after 1 month.

Our price analysis revealed how military stocks significantly increase their price one week after a geo-political event. However, the market itself, represented by the S&P 500 index, also increased its price significantly. Both of the effects are present 1 week after the event, while they are present with a smaller effect 1 day after. The comparison between the S&P 500 and seven military stocks considered shows how military stocks outperformed the market, but the gap is significant only with a confidence level of 0.85%.

Finally, experiment 3 showed how SNA metrics before and during a geo-political event do not add predictive value to the task of predicting future price movements 1 week after an event. A classifier based only on historical prices outperformed a classifier based on SNA metrics by 18%, while a classifier obtained by merging historical prices and SNA metrics exhibited similar performance to the price-based classifier.

CHAPTER 6

6 CONCLUSION

This chapter will look to summarise the findings of the experiments in an attempt to answer the research questions.

6.1 RESEARCH DEFINITION AND RESEARCH OVERVIEW

As described in the outset of this thesis the research question is concerned with the following questions and is broken down into the following subsections.

6.1.1 *HOW DO ONLINE FINANCIAL COMMUNITIES REACT TO MILITARY AND TERRORIST GEOPOLITICAL EVENTS*

The frequency range examined was a period of 2, 5 and 20 days. The SNA metrics taken into consideration were nodes, edges, density, centrality, betweenness, modularity, Gini index and diameter. The results established that after a period of two days the SNA metrics are considerably higher. In most cases geopolitical events generate an intensified communication, with an increase number of actors within the online community and a more centralised network.

6.1.2 *HOW DOES THE STOCK MARKET REACT TO MILITARY AND TERRORIST GEOPOLITICAL EVENTS*

The aim of the test was to establish if the stock price of the seven US military stocks after geopolitical events differs significantly from the price before the events. The results established that after one week of geopolitical events, the price of military stocks is higher than the price before the event using a confidence level of 0.99%.

6.1.3 *CAN MILITARY STOCK PRICES BE PREDICTED AFTER A MILITARY OR TERRORIST GEOPOLITICAL EVENT*

The objective was to establish if it was possible to predict the price of military stocks one week after a geopolitical event based on a set of features including SNA metrics, historical prices and S&P 500 index prices. The evidence concluded that online community's metrics has little or no predictive value in helping predicting the future price of stocks.

6.2 *Limitations*

There was a huge data set collected for this thesis. This included the Somali Pirate crises, French intervention in Mali, the Iraq 2003 war, the Iran nuclear crisis and North Korean nuclear crises. The collected data also included the news for each day per stock in conjunction with a news sentiment rating. Further work identified where various military stocks were linked to associated military stocks on the same Bloomberg stock news page. It was deducted that recent military operations in the Middle East and Eastern Europe would serve as a foundation for these experiments as these geopolitical events are quite recent and familiar to the general population.

6.3 *Future work and Research*

There are many research question open for future work such as from a real-time perspective can an investment bank influence a single actor or a cluster of actors on the network in terms of spreading good sentiment to boost single military stock activity, or in the opposite case decrease military stock activity? Is it possible to identify an actor or influencer on the network that is profitable to the Investment Bank? Is it possible to uncover hidden networks that have hidden agenda's to influence stock market prices? What influence the collected sentiment has on the stock prices, geopolitical events and the S&P 500. Future research could also use text mining with predictive mining for geopolitical events.

BIBLOGRAPHY

Aggarwal (2011) Aggarwal C.C, Social Network Data Analytics, IBM T. J. Watson Research Center, Yorktown Heights, New York, USA, 2011

Akhtar et al (2013) Akhtar, N.; Javed, H.; Sengar, G., "Analysis of Facebook Social Network," *Computational Intelligence and Communication Networks (CICN), 2013 5th International Conference on* , vol., no., pp.451,454, 27-29 Sept. 2013

Antweiler W, Frank MZ (2004) Is all that talk just noise? The information content of internet stock message boards. Journal of Finance 49(3): 1259–1269 Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. Journal of Computational Science 2: 1–8

Bai et al (2011) Bai Jinbo; Li Hongbo; Chu Yan, "Community identification based on clustering coefficient," *Communications and Networking in China (CHINACOM), 2011 6thInternationalICST,Conference* ,vol.,no.,pp.790,793,17-19Aug.2011
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6158261&isnumber=6158103 {Accessed August 12, 2014}Jörg Raab and H. Brinton Milward, ―Dark Networks as Problems. Journal of Public Administration Research and Theory 13 (2003), pp. 413-39

Baker and Faulkner (1993) Wayne E. Baker and Robert R. Faulkner, 1993. "The Social Organization of Conspiracy: Illegal Networks in the Heavy Electrical Equipment Industry," *American Sociological Review*, volume 58, number 6 (December), pp. 837-860

(BancDeBinary2014)BancDeBinary2012Availableat: https://bancdebinary.com/education-centre/learn-binary-options/political-events-affecting-stocks/ {Accessed August 13, 2014}

Bendix and Fisher (1949) Bendix, R., & Fisher, L. H. (1949) "The Perspectives of Elton Mayo" The Review of Economics and Statistics, 31(4), 312-319

Bremmer and Keat (2009) The Fat Tail: The Power of Political Knowledge in an Uncertain World

Bastian M.,Heymann S.,Jacomy M. (2009) Gephi: an open source software for exploring and manipulating networks. International AAAI Conference

Bollen, J.,Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market.Journal of Computational Science, 2(1), 1-8.

Budva, Montenegro (2013) Social Network Analysis, Methods and Measurement calculations

Casarin R, Squazzoni F (2013) Being on the Field When the Game Is Still Under Way. The Financial Press and Stock Markets in Times of Crisis. PLoS ONE 8(7): e67721. doi:10.1371/journal.pone.0067721

Chung et al (2013) Chung, K.S.K.; Piraveenan, M.; Levula, AV.; Uddin, S., "Assessing online community-building through assortativity, density and centralization in social networks," *System Sciences (HICSS), 2013 46th Hawaii International Conference on* , vol., no, pp.1993,2002, 7-10 Jan. 2013 {Accessed 26 August 2014}

(Dyson 2013) What will happen to shares if the West strikes Syria? Available at http://www.telegraph.co.uk/finance/personalfinance/investing/10290802/What-will-happen-to-shares-if-the-West-strikes-Syria.html {Accessed July 13, 2014} (Global Issues)

Faust K(1997) Centrality in affiliation networks. Social Networks, 19:157–191, April 1997

Gartner. Hype Cycle for social software, (2008) G00158239, 2008

Gloor et al (2009) Gloor, P.A.; Krauss, J.; Nann, S.; Fischbach, Kai; Schoder, D., "Web Science 2.0: Identifying Trends through Semantic Social Network Analysis," *Computational Science and Engineering, 2009. CSE '09 International Conference on* , vol.4, no., pp.215,222, 29-31 Aug. 2009

Granovetter (1973) Mark Granovetter, “The Strength of Weak Ties,” *American Journal of Sociology* 78, no. 6 (1973): 1360-1380{Accessed 3rd August 2014}

Hansen et al (2009) Hansen, D.L.; Rotman, D.; Bonsignore, E.; Milic-Frayling, N.; Rodrigues, E.M.; Smith, M.; Shneiderman, B., "Do You Know the Way to SNA?: A Process Model for Analyzing and Visualizing Social Media Network Data," *Social*

Informatics (SocialInformatics), 2012 International Conference on , vol., no., pp.304,313,14-16Dec.2012

Julei Fu; Jian Chai; Duoyong Sun; Shouyang Wang, "Multi-factor Analysis of Terrorist Activities Based on Social Network," *Business Intelligence and Financial Engineering (BIFE), 2012 Fifth International Conference on* , vol., no., pp.476,480, 18-21 Aug. 2012

Khonsari, K.K. et al (2010) Khonsari, K.K.; Nayeri, Z.A.; Fathalian, A.; Fathalian, L.; , "Social Network Analysis of Iran's Green Movement Opposition Groups Using Twitter," *Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on* , vol., no., pp.414-415, 9-11 Aug. 2010

(Kristin Forbes Robert Rigobon) Forbes, Kristin J. and Roberto Rigobon "No Contagion, Only Interdependence: Measuring Stock Market Co-movements," Journal of Finance, 2002, v57(5,Oct), 2223-2261.

(Krebs 2002) Uncloaking Terrorist Networks:
http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/941/863 {Accessed August 23 2014)

(Krebs 2002) Mapping Networks of Terrorists Cells
http://www.aclu.org/files/fbimappingfoia/20111110/ACLURM002810.pdf {Accessed August 23, 2014}

Kretschmer, Hildrun. Kretschmer Theo. (2010) A New Centrality Measure for Social Network Analysis Applicable to Bibliometric and Webometric Data. Humboldt-University Berlin : Department of Library and Information Science, 26-D-10117.

Koochakzadeh et al (2012) Koochakzadeh, N.; Kianmehr, K.; Sarraf, A; Alhajj, R., "Stock Market Investment Advice: A Social Network Approach," *Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on* , vol., no.,pp.71,78,26-29Aug.2012
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6425782&isnumber=6425553{{Accessed August 12, 2014}

Kinsella (2011) Kinsella, David Todd, "Former Soviet Bloc Locales in Africa's Illicit Arms Trade Network"(2011) Political Science Faculty Publications and Presentations. Paper 11http://pdxscholar.library.pdx.edu/polisci_fac/11 {Accessed August 23, 2014}

Mayo, E. (1971). Hawthorne and the western electric company. In Pugh, D. S. (Ed.), Organizational Theory - Selected Readings (2nd ed.) London: Penguin Business.

Mayo, E. (1939). "Routine interaction and the problem of collaboration". American Sociological Review, 4(3), 335-340

Moreno, J. L. (1937). "Sociometry in Relation to Other Social Sciences" Sociometry, 1(1/2), 206-219

Measurement Metrics: http://train.ed.psu.edu/WFED-543/SocNet_TheoryApp.pdf

Mohsen Jamali and Hassan Abolhassani (2012) Jamali, M.; Abolhassani, H.; "Different Aspects of Social Network Analysis," *Web Intelligence, 2006. WI 2006. IEEE/WIC/ACM InternationalConference* ,vol.,no.,pp.66-72,18-22Dec.2006,
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4061343&isnumber=4061322

Morris, T., "Extracting and Networking Emotions in Extremist Propaganda," *Intelligence and Security Informatics Conference (EISIC), 2012 European*, vol, pp.53, 59, 22-24Aug.2012,
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6298813&isnumber=6298809
{Accessed August 28, 2014}

Newman (2006) Modularity and community structure in networks

NetworkX http://networkx.github.io/documentation/networkx-1.9/overview.html#whonetworkx {Accessed 29 August 2014}

N. Casnici, P. Dondio, R. Casarin, F. Squazzoni. (2014) Online investor reactions to market crashes: a study case on the Italian market. Submitted to PNAS, August 2014

Oram (2001) Peer to Peer: Harnessing the Power of Disruptive Technologies

Perliger, et al (2011) Perliger, Arie and Pedahzur, Ami, "Social Network Analysis in the Study of Terrorism and Political Violence"(2010), Working Papers, http://opensiuc.lib.siu.edu/pn_wp/48 {Accessed August 25, 2014}

Perveen et al (2012) Perveen, N.; Gupta, S.; Verma, K., "Facial expression recognition using facial characteristic points and Gini index," *Engineering and Systems (SCES), 2012 Students Conference*,vol.,no.,pp.1,6,16-18March201 http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6199086&isnumber=6199003 {Accessed August 28, 2014}

Rachman et al (2013) Rachman, Z.A; Maharani, W.; Adiwijaya, "The analysis and implementation of degree centrality in weighted graph in Social Network Analysis," *Information and Communication Technology (ICoICT), 2013 International Conferenceof* ,vol.,no.,pp.72,76,20-22,March2013 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6574552&isnumber=6574528{{Accessed August 19, 2014}

Roy, R.B.; Sarkar, U.K., (2010) Roy, R.B.; Sarkar, U.K., "A Social Network Approach to Examine the Role of Influential Stocks in Shaping Interdependence Structure in Global Stock Markets," *Advances in Social Networks Analysis and Mining (ASONAM), 2011 InternationalConference* ,vol.,no.,pp.567,569,25-27,July2011 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5992668&isnumber=5992565 {Accessed August 19 2014}

Sas (2012) http://support.sas.com/publishing/pubcat/chaps/57587.pdf {Accessed 29 August 2014}

Scott, J. (2000) Social network analysis. SAGE Publications, London.

S.D. Berkowitz (1982) An Introduction to Structural Analysis: The Network Approach to Social Research. Toronto: Butterworths,

Shivaram Narayanan (2005) The Betweeness of Centrality of Biological Networks, (http://scholar.lib.vt.edu/theses/available/etd-10162005-200707/unrestricted/thesis.pd {Accessed August 19, 2014}

Smu.edu (2013) http://faculty.smu.edu/tfomby/eco5385/lecture/Confusion%20Matrix.pdf {Accessed August 28, 2014}

SNA methods and Metric source: http://train.ed.psu.edu/WFED-543/SocNet_TheoryApp.pdf {Accessed July 16, 2014}

Tom Stewart, 2001. "Six Degrees of Mohamed Atta," *Business 2.0"* http://www.business2.com/articles/mag/0,1640,35253,FF.html, {Accessed 23 August, 2014}

Wang (2012) Peng Wang, "Analysis of the Topology Structure of Network of Chinese Stock Market," *Management of e-Commerce and e-Government (ICMeCG), 2012 InternationalConference* ,vol.,no.,pp.403,407,20-21Oct.2012
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6374951&isnumber=6374868

Watts and Strogatz (1998) D.J. Watts, and S.H. Strogatz, "Collective dynamics of 'small-world' networks", Nature, 393, (1998), 440-442

Wasserman and Faust (1994) Wasserman, Stanley, Faust, Katherine, (1994), "Social network analysis - methods and applications", Cambridge University Press, ISBN: 0-521-38707-8

Wen-Chih Chang (2010) Wen-Chih Chang; Hsiao-Wen Lin; Lai-Chi Wu; , "Applied Social Network Anaysis to Project Curriculum," *Networked Computing and Advanced Information Management (NCM), 2010 Sixth International Conference on* , vol., no., pp.710-715, 16-18 Aug. 2010

Weinstein et al (2009) Weinstein, C.; Campbell, W.; Delaney, Brian; O'Leary, G., "Modeling and detection techniques for Counter-Terror Social Network Analysis and Intent
http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4839642&isnumber=4839294
{Accessed August 26, 2014}

World Military Spending: Available at http://www.globalissues.org/article/75/world-military-spending {Accessed July 13, 2014}

(WSJ) How Stocks React to Geopolitical Uncertainty Available at:
http://blogs.wsj.com/moneybeat/2014/03/03/how-stocks-react-to-geopolitical-uncertainty/
{Accessed July 14, 2014}

Zheng Chen and Xiaoqing Du (2013) Zheng Chen; Xiaoqing Du, "Study of Stock Prediction Based on Social Network," *Social Computing (SocialCom), 2013 International Conferenceon* ,vol.,no.,pp.913,916,8-14Sept.2013
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6693438&isnumber=6693296{Accessed August 12, 2014}

APPENDIX A: WEKA FILE

2.arff

APPENDIX B: MILITARY AND TERRORIST GEOPOLITICAL EVENTS AND STOCK PRICES

Military and Terrorist
geopolitical events ar

APPENDIX C: S&P 500 HISTORICAL INDEX PRICES WITH STOCK PRICE FREQUENCY BREAKDOWN

S&P500 HISTORICAL
INDEX PRICES WITH