The Pennsylvania State University

The Graduate School

College of Engineering

SOCIAL NETWORK ANALYSIS AND SIMULATION

OF THE DEVELOPMENT OF ADVERSARIAL NETWORKS

A Thesis in

Computer Science and Engineering

by

Razvan Orendovici

Submitted in Partial Fulfillment
of the Requirements
for the degree of

Master of Science

December 2011

The thesis of Razvan Orendovici was reviewed and approved* by the following

Frank E. Ritter
Professor of Information Sciences and Technology, Computer Science and Engineering, and Psychology
Thesis Adviser

Raj Acharya
Professor of Computer Science and Engineering
Department Head

Wang-Chien Lee
Professor of Computer Science and Engineering
Committee Member

*Signatures are on file in the Graduate School

Abstract

This paper presents a new way to monitor and look at adversarial networks through the Adversarial Network Analyzer (ANA) and supporting mathematical calculations. Recent work on social networks has begun to be used in the analysis of adversarial networks and their underlying structure with hopes of detecting and preventing future activity. Adversarial networks are abstracted to be network subgroups that work against the interests of the group studying it. ANA, the software package created by this work, can be a tool for network visualization and a teaching exercise for students in the field of security analysis. It can portray the structure of such networks and allow analysts to find patterns and key players in the network while watching the network evolve to carry out a goal. Finally, this thesis uses output from ANA to study how a simulated adversarial scenario grows in structure. I compare this network against more traditional social networks on standard measures and also analyze the change in time of this network on those same dimensions.

Table of Contents

List of Figures

List of Tables

List of Equations

Acknowledgements

1. Introduction

I present a novel way to monitor, analyze, and look at an adversarial social network through a visualization tool and supporting mathematical analysis. This is accomplished by presenting a simulated adversarial network that is used to guide the development of the visualization tool. Finally, I analyze how this network evolved and present the implications for education, network science, social network analysis and measurement.

i. Motivation

Social network analysis and visualization of adversarial networks is a very powerful method of understanding the network and keeping track of relevant information about the network as it evolves and becomes more defined. Recent events and political pressures have put a lot of research effort into the understanding of adversarial networks and their identification. This combined with recent popularity of social network analysis has made a network centric analysis of these networks interesting and useful. Throughout this work I look at allowing users to take mock intelligence gatherings on the communications and individuals involved in a possible adversarial network and keep track of this information while visualizing it and performing statistical analysis.

I present a visualization tool tailored specifically at the style of data that intelligence agencies capture that allows the user to store the information and visualize it as it changes. Once this information is entered and visualized, the tool provides additional features to allow this data to be analyzed over its evolution and new patterns detected. In Chapter 5, I compare the adversarial network created by intelligence data with analysis found on more traditional networks seen in the literature.

ii. Data

Intelligence data on adversarial networks is a close kept secret of the various government agencies who work with it, so we substitute it with a set of mock intelligence reports. These reports are for an educational simulation on analyzing terrorist networks and stopping a plot. The reports are not based on any real agencies, affiliations or subjects but based on years of experience in the intelligence field. They provide a very interesting use case for data handling, visualization, and analysis of a niche network type. The mock simulation created by Donald Shemanski (2011) is used as a guiding use case through the development of the visualization platform and drives the data analysis later in this paper.

iii. Visualization for Analysis

The visualization platform ANA described in Chapter 3 provides the ability to enter in the data provided by Shemanski as it appears. This data entry is important for bookkeeping and later analysis but the tool provides additional features. Rather than just maintain the current data as it exists ANA keeps a history of all the entered data as a series of events. This feature is very useful in visualizing the simulated plot 'play-out.' An interface to allow the intelligence reports to be rolled back to any particular state in time and view its evolution from there provides a movie like visualization of the network. This gives a rich animation of actors being added, communications between actors showing up on the network, and a deeper understanding of how ties in the network develop. Finally, as Chapter 4 describes ANA facilitates social network analysis calculation to show how various values about the network have evolved

iv. Preview of Contributions

This work makes several contributions to network science, education, data entry, data visualization, and data analysis.

A review of work done in network science with mathematical measures on individual nodes as well as aggregate measures of the entire network is presented in Chapter 2. This goes into detail of studies done in criminology and similar fields to look specifically at adversarial networks and their characteristics.

I present a new way to look at these types of networks and how to separate them from standard social networks. Rather than just comparing standard network measures of these networks this work finds it more important to analyze the change in these network measures over time.

To better suit education and the field of risk analysis this paper provides the ANA simulation tool capable of allowing users to input and visualize information about an evolving social network. Students are able to play with such data and gain real world experience in an education environment with simulated intelligence information. The tool allows them to best analyze their data and get involved with intelligence reports and unknown data.

The analyses run on this data provide contributions to the field of network science that has been studying adversarial networks such as this for the past few years. Using standard values of social network analysis alongside new time based measures to look at these networks is a starting point for future research looking to differentiate networks on more dynamic features. The difference in the growth of an adversarial network and that of a standard social network is a very interesting topic presented in this work.

Finally, I present some limitations in network science and a list of future work both in network science as well as future features for ANA that would improve its capabilities and ease of use.

v. Summary

Chapter 2 provides an overview of the work done in network theory and useful measure that will be used to analyze the network. Chapter 3 introduces the visualization tool, its features, limitations and use cases. Chapter 4 uses this tool to analyze the evolution of the network based on the measures and common analysis introduced in Chapter 2. Finally, in Chapter 5 I provide an overview of the work and contributions presented in this paper as well future extensions and uses of the ideas here.

2. Background

Social networks and network theory have been discussed in the literature since the 1930's and earlier. This was originally vague ideas about networks, the individuals and their relationships in the early 1930's when Moreno (1934; 1953) introduced the first notion of a Sociogram, a graphical way to represent people and their relationships. His work began the foundation of sociometry and network science; with sociometry later evolving into social network analysis. His work of the sociogram formalized a way to graphically connect entities according to their relationships and laid the foundation for directionality in networks and the reciprocity of relationships. Sociogram studies evolved into social networks when Barnes (1954) made the first mention of a social network. Since the first social network, this type analysis has become a much more quantitative science.

i. Networks and Their Measures

The early work in social network analysis was the binding element of a lot of academic fields and allowed mathematics and social studies to have applications to multiple areas. Drug adoption and spread through the social network was measured by Coleman, Katz, and Menzel (1957), while Laumann and Pappi (1973) studied the networks of business leaders; in the early 1970s the economic recession gave rise to a number of studies that analyzed how people find jobs and get new careers through the development of weak ties and information dissemination (Granovetter, 1973; 1974). These ties were notied to have strong uses in information across a connected population. Karate clubs (Wayne, 1977) and family networks (Milgram, 1967) have also been studied to detect patterns and underlying structures of groups. These various applications of social network analysis have shown its strength in finding patterns and understanding the actions of individuals within the context of a group.

More recently with the advances of the world wide web and larger connecting ties, new studies have emerged such as studies about co-authorship of collaborative papers in the academic

community (Newman & Girvan, 2004; Barabasi, Jeong, Nelda, Ravasz, Schubert, & Vicsek, 2002). Online communities such as Second Life have been analyzed from a socioeconomic viewpoint (Bakshy, Simmons, Huffaker, Teng, & Adamic, 2010). Most recently social networks have been combined to help predict music recommendations (Ioannis, Vassilios, & Joemon, 2009), webpage bookmarking (Heymann, Koutrika, & Garcia-Molina, 2008), and even the stock market (Bollen, Mao, & Zeng, 2011).

The performance of members of inside of a social network has recently been a topic of interest such as Baldwin, Bedell, and Johnson (1997) who performed a study to see the performance of students based on their social network. This study was expanded to study online learning of students where positive and negative relationships exist between the students (Yang & Tang, 2003). This helps simulate an adversarial network and evaluates the member's performance based on positive and negative connections.

Adversarial Network in these studies has a slightly different connotation from those in texts about terrorist cells. Their definition of adversarial networks relates to a network in where there are positive and adversarial relationships. For this work I abstract this notion and consider the collection of nodes connected by negative relationships (from the viewpoint of an intelligence agency) to be considered an adversarial network in its entirety.

The adversarial network, a group of actors connected together against the interest of another group, has been a topic of literature interest for the past two decades. Such networks can be terrorist cells (Krebs, 2002) organized crime money flow networks (Klerks, 2001), and communication networks (Carley, Lee, & Krackhardt, 2002) all joined together for a common goal. These networks have a variety of forms and have changed significantly in the course of the last century.

The adversarial networks of the first half of the century were Mafia families organized in a hierarchical fashion with a very common goal (Klerks, 2001). Following the September 11 attacks, mappings of the actual terrorist network were shown in academia (Krebs, 2002) off data gathered from news sources and published intelligence reports (Figure 1). This work showed each of the groups involved in the plot along with their communications and ties. The plot was much more of a network and less of a hierarchical tree. Farley (2003) argued that such adversarial networks should be considered a hierarchy and used mathematical order theory to try to disrupt the network and stop its function. He attempted to reorder an adversarial network into a tree structure and remove the root nodes such that it becomes a set of disconnected tree stumps unable to coordinate and carry out any attacks or goals.

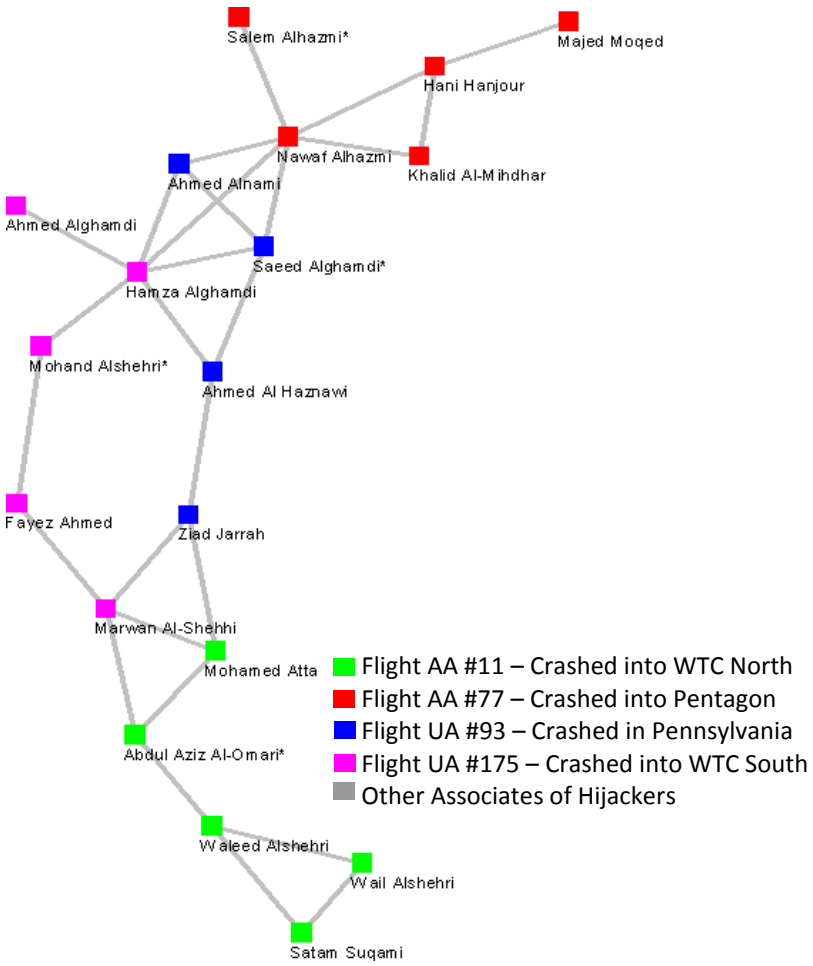


Figure 1 September 11th Network Diagram (in color in original article) (Krebs, 2001)

Identifying key targets in a tree, or network of an adversarial network is very difficult, and the effectiveness of removing that node is hard to measure (Carley, Lee, & Krackhardt, Destabilizing Networks, 2002). This work put forth a number of calculation of how disrupted a network is along with tools (Organization Risk Analyzer) to mathematically calculate these values for such networks.

These networks are very different and constantly changing making it difficult to a mathematical function that can apply to all networks and properly identify key actors and remove them from the network. This task is made even more difficult by the evolution of the network as it may continue to change and recover the links it had originally lost (Carley, Lee, & Krackhardt, 2002).

For this reason this work does not attempt to provide a mathematical solution but rather a way to visualize each individual network and analyze key players in the network. We allow the analyst to look at all the data and make the correct decision about the network. Once a potential node is removed from the adversarial network it can continue to be monitored to view its continued evolution. Repeated acts of interfering with the network can allow analysts to see how it dealt with disruption.

ii. Static Network Measures

Since the original proposal of social networks, many different researchers have identified quantitative methods to compare and analyze networks. Common measures analyze to see how connected a graph is and whether the graph contains one central area that connects everything or if nodes are equally connected all around.

Degree Centrality

Moreno (1934) proposed one of such measures, *the degree measure of centrality,* for directed graphs by measuring the in degree and outdegree of the nodes as in Equation 1; indegree are edges that point at the particular node, and outdegrees point from the node in question out towards other nodes. This measure is called a *prestige* or *status* measure of nodes. The *degree*

measure of centrality is useful for networks where orders are given, advising happens, or information is relayed in a singular direction.

$$C_d = \frac{d(n_i)}{g-1}$$

Equation 1 Degree Centrality of node n_i (Wasserman & Faust, 1999, p. 179)

In non-directional networks, the measure of can be calculated simply by the number of connections that node has (Proctor & Loomis, 1951; Shaw, 1954; Freeman, 1979). The more connections a node has, the higher a measure of centrality it has. As graphs become larger, nodes can reach more and more members thus having very large centrality values. To make this measure comparable across networks of different sizes we can normalize the score by dividing it by the number of nodes in the graph. This measure becomes a ratio of how many nodes in the graph the interesting node is connected to, compared to the total number of nodes.

The node specific measure described is general enough to allow it to be applied to an entire graph or just a subgroup of the graph. For each node we compare its degree centrality with that of the most central node (n^*) in the graph and compare with the maximum connections possible (Equation 2). The domain is bounded by [0,1]; at 0 it represents a graph where all nodes are equally connected to each other while at 1 it represents a graph with one highly centralized node that connects everybody together.

$$C_D = \frac{\sum_{i=1}^{g}[C_d(n^*) - C_d(n_i)]}{(g-1)(g-2)}$$

Equation 2 Aggregate Degree Centrality (Wasserman & Faust, 1999, p. 180)

Distance Centrality

Graph centrality can also be measured by seeing how close all the nodes in the network are to each other. By measuring the minimum distance between the node of interest N_i and all other nodes in the graph we can measure how close this node is to every node in the graph (Sabidussi,

1966) and normalize this value by dividing by the number of nodes in the graph (Beauchamp, 1965). This centrality measure is again on a [0,1] domain; 0 represents graphs where a node is not connected to the rest of the graph, and 1 represents graphs when the node is directly connected to every node in the graph. Similar to other centrality measures of individual nodes, distance centrality can also be averaged over all pairs of nodes in the graph to obtain a global measure of the network centrality.

Betweenness Centrality

Shimbel (1953) proposed a communication theory approach to studying the centrality of nodes inside a graph. In this measure we compare the shortest paths between all pairs of two nodes n_i and n_j in the graph. A third actor n_k has high betweenness centrality if a large portion of the paths between $<n_i, n_j>$ must past through node n_k. This places node n_k in very high power because it has a lot of control over the communication between the previous nodes (Anthonisse, 1971; Freeman, 1977).

This can represent a node that can corrupt the data or cut communication, it can also represent a node that when removed from the graph could make communication between nodes $<n_i, n_j>$ difficult. Freeman (1979) extended this to be a measure of not just a particular node, but an aggregate measure over the entire graph as a group *betweenness centrality measure* by averaging it over all the nodes n_k and their influence on any pair of nodes $<n_i, n_j>$.

Each of these centrality measures predict how centralized and connected a graph is but do it in slightly different measures. They differentiate between a network where all nodes are connected to each other, a fully connected graph, and a star like network where there is one central node that connects everybody together.

Power Law of Networks

The previous measures of networks are useful for seeing just how closely packed the social network is and how well each of the people involved in it are connected to each other. These models and statistics were unable to properly consider two important factors of social networks, growth and preferential attachment (Barabasi & Albert, 1999). From here the power-law was proposed to study how attached nodes are. The proportion of nodes in a network with k-links actually follows what had been a power law distribution in mathematics and statistics; this value can be calculated as proportional to c/k^2 (Equation 3).

$$f(k) = \frac{c}{k^2}$$

Equation 3 Proportion of nodes with k-links

This value showed up in studies of the connections on the World Wide Web as early as 2000 (Broder, et al.). This aspect was also analyzed mathematically by Albert and Barabasi (2002) across a variety of networks like the World Wide Web, movie actors, and co-authorship in various fields. Newman (2003, p. 188) repeats these and other findings in a review of collaboration networks, citation networks, power grid, the internet, and World Wide Web networks. The power law was shown to be representative of preferential attachment that is seen in real networks that earlier models were unable to account for.

More commonly this can be looked at from an economics standpoint and be referred to as a *Rich-get-richer* scheme (Easley & Kleinberg, 2010). Nodes that are already in the graph and very popular will continue to become more popular. The exposure given to them by the many other nodes they are connected to will help them become even more connected. This group of nodes that are very highly connected is very limited and generally represents celebrities, or webpage hubs where all the connections link to. However, the majority of the nodes have much fewer connections and have a harder time growing to popularity and developing as massive a following.

Extending the power law out to economics in a different manner again was a study not of social networks, but retailers and the products they sell (Anderson, 2004). The primary question being related to where stores earn most of their money, does it come from a few very popular products that attract all the attention and spread virally, or do they earn their money from the 'tail' of the power law of many different niche products that sell a few items. This coined the term 'Long Tail' by Anderson (2004). While his works is not fully applicable to social networks, it describes situations where the mass of the graph is not in the few very popular nodes, but in the large collection of nodes that are lightly connected.

The application of the long tail to adversarial networks can provide an interesting use case in the active deception of the network. The adversarial network is a social network where an active deceptive power is trying to hide the network and its structure from outside groups. Having a long tail allows the network to have a lot of '*niche*' actors performing distinct tasks with limited connectivity. Not having many very popular nodes keeps the important leaders of the network hidden from external observation and separated from the other member by levels of communication hierarchy.

Transitivity and Clustering Coefficient

Transitivity is a view of networks that looks at groups of three members, a triad, together rather than the usually analysis of pairs of two network members. A Triad in the graph is a group of three members that are connected in a triangular shape. The idea behind triad analysis is to understand if relationships in a graph spread from one friend to another.

If nodes n_a and n_b are connected, and nodes n_a and n_c are connected, will a transitive relationship appear between actors n_b and n_c as they meet through a and become connected (Wasserman & Faust, 1999). This relationship described in Equation 4 can be quantified mathematically by counting the number of triangles available in the network compared to the total

number of possible triangles. This allows us to compare networks and see how transitive relationships inside it are.

$$\boldsymbol{if}\ n_a \leftrightarrow n_b\ and\ n_a \leftrightarrow n_c\ \boldsymbol{then}\ n_b\ \leftrightarrow n_c$$

Equation 4 Clustering Coefficient Requirements

Original work on transitivity (Holland & Leinhardt, 1971; Holland & Leinhardt, 1972) showed how in social networks this property pushes the increase in ties and closeness of graphs. Pairs of actors that share a mutual connection are likely to themselves become connected. Recent work has shown this property to exist in modern online social networks like twitter (Golder & Yardi, 2010).

The easiest way of measuring this value is by analyzing the number of triangles, fully connected triples, in the graph (Wasserman & Faust, 1999).

iii. Dynamic Network Measurements

The only problem with all of the centrality, power law and similar analyses lies in the fact that they are static measures. At a snapshot of the network we can view these numbers and compare it to a different network. They do not do anything to address the evolution of the graph. To compare how similar or different two graphs are it is important to see how they reached their end state and how this path compares. There exist networks whose evolutions do not to follow standard social network evolution and keep members from becoming highly connected. Such networks provide interesting study cases especially with applications to intelligence and risk analysis. The distinction between the evolution of a standard social network or subgroup compared with that of an adversarial or cell network is an interesting topic analyzed throughout this paper.

Centrality

Over time the centrality of the entire network may change differently from a corporate network, online social network, or small scale social network. Not all nodes in the network have equal interest in developing connections with other players.

When comparing an adversarial network against a corporate leadership network, the evolution of connections of key players in this network will not be similar. In a corporate environment, leaders develop increasingly more connections over time while in adversarial networks there is a danger that keeps leaders disconnected from the network. The static measures of centrality and connectivity, such as degree, distance, and betweenness of a graph can be analyzed over a time interval to uncover such uncommon patterns.

Power Law of Networks

More patterns can also be seen in the power law evolution of the adversarial networks. Standard social networks evolve to fit the power law over time and reach a stable state; the state where an adversarial network stabilizes to is of equal interest. Does the network not follow a power distribution at all, is its slope significantly different from a standard social network, or does it simply follow a power law pattern like all other networks.

Transitivity and Clustering Coefficient

The static measure of transitivity and clustering provides some interest for the user in terms of seeing just how interconnected the graph is. As shown by social network analysis (Golder & Yardi, 2010; Holland & Leinhardt, 1972) these graph properties tend to increase the connectivity of a social network over time due to their mutual connections. This value becomes interesting to analyze from a dynamic measure to see how as the graph grows and evolves existing members of the graph form more ties with each other.

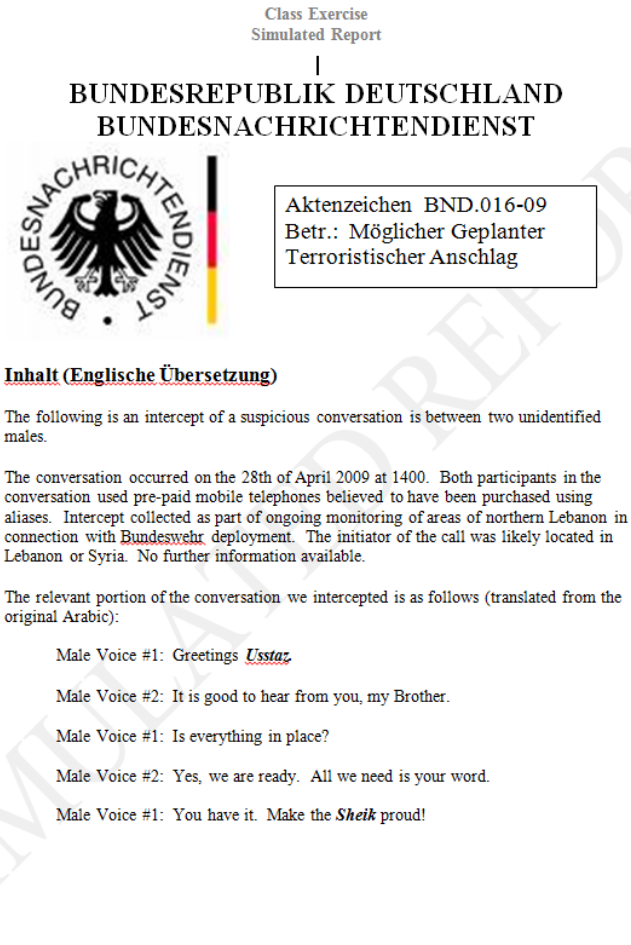
iv. Data Set

Due to difficulties with data collection, most work done on social networks relies on the static network measures presented in this Chapter. To use the proposed dynamic measures we must find a social network that has all of the history of the social network and the evolution of the ties between the members. This network must also be an adversarial network to demonstrate the

importance of using time-based measures to show nonstandard social network patterns. Simulated intelligence reports of an adversarial plot provided by Shemanski (2011) provide such a social network.

The dataset is a series of intelligence reports similar to Figure 2 and includes conversations between two or more actors. In this particular report, the conversation is between the leader of the plot, giving the final go-ahead, and the buffer person placed between him and the 'playmaker.' At this point the plot is ready to come to an end and most of the social network of actors is visible by those who have studied the intelligence reports. There are 73 communications that create a network of 30 members connected by 46 established routes of communication.

The entire list of communications provides the implementation by an adversarial group to strike at a major international event after all finances, logistics, weapons and recruitment details are figured out. This plot is similar to what analysts and students in risk analysis might look at and need to keep track of. All of the pertinent plot information can be recorded in ANA to be reviewed and analyzed at different times.



Class Exercise
Simulated Report

BUNDESREPUBLIK DEUTSCHLAND
BUNDESNACHRICHTENDIENST

Aktenzeichen BND.016-09
Betr.: Möglicher Geplanter
Terroristischer Anschlag

Inhalt (Englische Übersetzung)

The following is an intercept of a suspicious conversation is between two unidentified males.

The conversation occurred on the 28th of April 2009 at 1400. Both participants in the conversation used pre-paid mobile telephones believed to have been purchased using aliases. Intercept collected as part of ongoing monitoring of areas of northern Lebanon in connection with Bundeswehr deployment. The initiator of the call was likely located in Lebanon or Syria. No further information available.

The relevant portion of the conversation we intercepted is as follows (translated from the original Arabic):

Male Voice #1: Greetings ***Usstaz***.

Male Voice #2: It is good to hear from you, my Brother.

Male Voice #1: Is everything in place?

Male Voice #2: Yes, we are ready. All we need is your word.

Male Voice #1: You have it. Make the ***Sheik*** proud!

Figure 2 Sample Intelligence Report

v. Graph Visualization and Animation

Many tools exist for visualizing a social network loaded from different file format and social network matrices. These tools can display groups, allow for automatic layout and they can spread out the network as much as possible to make the visualization easy to understand. Unfortunately very few of these tools provide ways to create this network in an intuitive way. They usually present the user with a large matrix of all of the actors and force to establish the connections between members in a complicated way.

Even fewer graph visualization toolkits provide the ability to visualize the network as it changes over time. They rarely keep track of the order of events and when nodes and edges were added to the graph. Without such information it is very difficult to do any time based analysis of the social network or to be able to visualize the network evolve.

While the visual evolution of the social network is not an empirical method to understand social networks it is a useful tool for a basic understanding of social networks, and also useful to explain unexpected patterns. Spikes in communication or connectivity within nodes can be seen by the animations and explained by specific actions done by the members in the network.

vi. Summary

The introduction provided in this Chapter outlines the measures of social network analysis that I will be using to analyze the given adversarial network and what patterns are expected out of these measures. The patterns described are for standard social networks, later in the paper we compare the patterns of this adversarial network against what is expected. The ANA toolkit described in Chapter 3 meets the requirements of creating, visualizing, and playing back a social network in a user friendly fashion that other social network analysis tools are unable to do.

3. Adversarial Network Analyzer (ANA)

To best see these networks evolve and meet the requirements of the data set, this paper presents a graph visualization tool tailored to handling the simulated adversarial network. The Adversarial Network Analyzer (ANA) is a Java application that allows users to input new connections about the graph and visualizes the state of the graph at all-time intervals. To provide powerful graph visualizations ANA is written on top of the Prefuse visualization toolkit (Heer, Card, & Landay, 2005). This library provides a visualization library for many datasets and full of features. The feature most commonly used in ANA is the use of graph visualization through Nodes and Edges. Prefuse provides library functions to properly render and lay out the network structure.

The application was made as a Java UI applet so that it could best be used by developers, analysts, and students interested in network evolution or adversarial networks. ANA 1.0 is planned to be used in the spring of 2012 by a security and risk analysis course taught in the College of Information Science and Technology at the Pennsylvania State University. The students who will be using this software will be studying simulations like the Shemanski dataset in order to understand an adversarial plot and identify it.

The UI shown in Figure 3 is broken down into four distinct sections, each providing a useful interface to managing the graph. The Content Pane, The Edit Pane, The Expand Graph Pane, and the Playback Pane all contribute to provide the functionality of ANA. The panels map almost perfectly onto the list of tasks ANA supports, described in Table 1.

Table 1ANA Supported Tasks

ANA Supported Tasks
Visualize graph
Add nodes to existing graph
Add edges to existing graph
Modify existing edges and add more details
Modify existing nodes and add more details
Playback of graph expanding from the first node
Save and reload graph
Export state of graph to standard XML for mathematical analysis by ORA

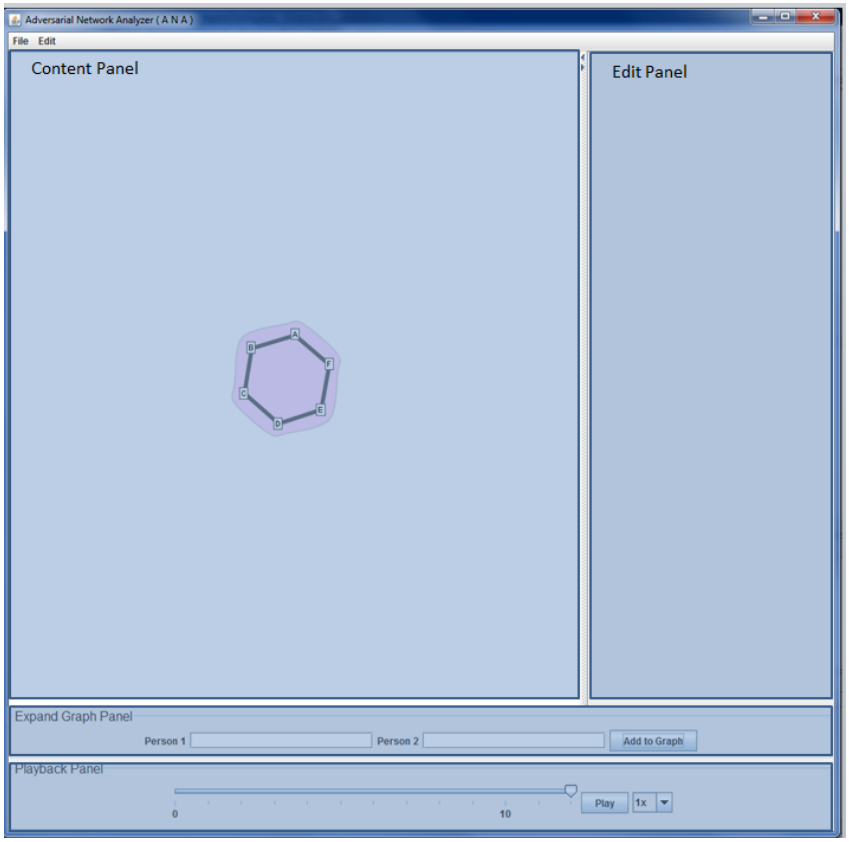


Figure 3 Adversarial Network Analyzer (ANA version 1.0) User Interface

The largest section of the UI is the *Content Panel* it provides a big open area where the user can see the graph evolved and manage it. This allows the user to reorder nodes if they believe there is a better layout, and to zoom in and out to see more detail on particular parts of the graph or get an overall view of the entire network. The content pane shows all nodes and edges and provides colored outlines around defined subgroups of the graph.

Multiple subgroups can be defined in the graph, and nodes can belong to one or more of these subgroups. For the Shemanski dataset this is particularly important in allowing users to visualize the different sections of the adversarial network. Some members might be in a *weapons group* while others are the *financiers* of the network. By outlining these groups in different colored sections it makes it much easier to get an overview of the network.

The *Edit Panel* is a contextual UI element that allows for more details to be added to the graph. When the user clicks on a Node it allows them to edit this particular node of the graph as shown in Figure 4. The user is be able to enter biographical information known about the actor like previous records, known associates, special skills, know locations or any group affiliations. The location and group affiliation are kept separate from the rest of the information to be used as methods for filtering data. All other details a user feels are important about a particular actor can go in the *details* text area. By having all information about a node be editable, ANA allows users to name actors of a plot as unknown with the ability to later go in and edit their information once an actual identity is tied to them.

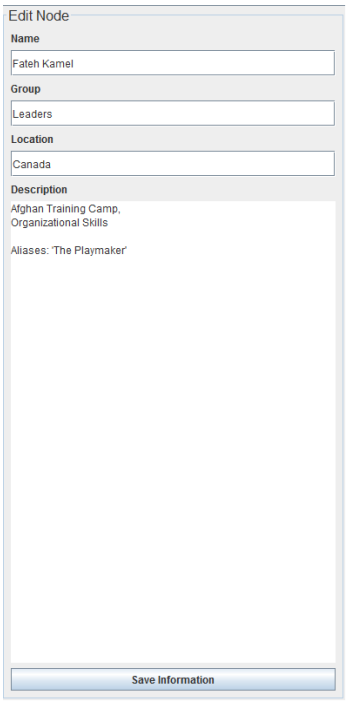


Figure 4 ANA Person Edit Panel

When the user clicks on an edge, the contextual edit panel changes yet again to allow the user to label the communication between the two players of the adversarial plot (Figure 5). Any number of details can be put about the connection between two nodes. Based on the available data set, users can enter multiple communications between the nodes, with short summary of the communication, as well as dates, locations, and other pertinent details that may help the plot unravel.

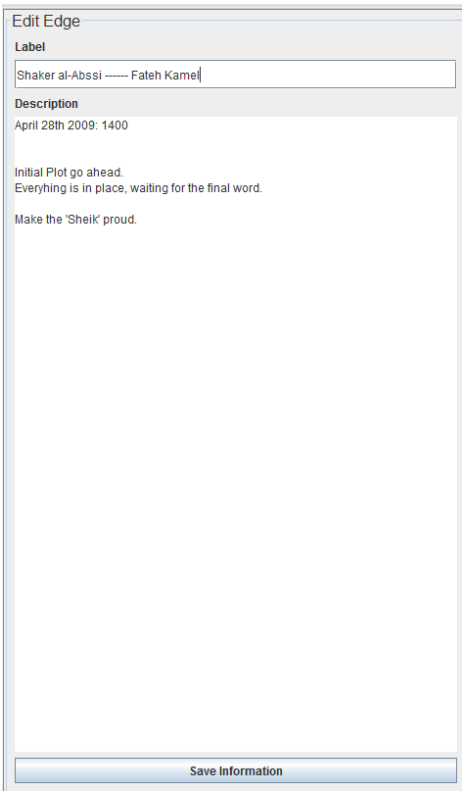


Figure 5 ANA Edge Edit Panel

The most important part of the interface is the *Expand Graph Panel in* Figure 6 to add new nodes and edges to the graph. Here by typing the name of any two nodes they will be added to the graph and connected with an edge. If either of the names already exists in the graph ANA will find them in the layout and connect them together.

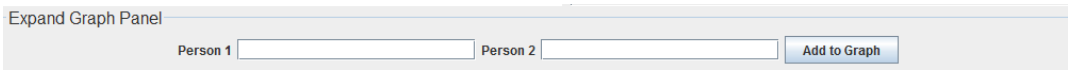


Figure 6 ANA Expand Graph Panel

The bottom panel is the *Playback Panel,* shown in Figure 7 that allows us to rewind the information and go see how the graph evolved. This panel gives the user frame-by-frame control to see all changes made to the graph such as node additions, edge additions or any edits of the information in these nodes or edges. The playback panel is used to quickly present to others how the vision of this network has evolved and to mark key players that are making the graph more interconnected.

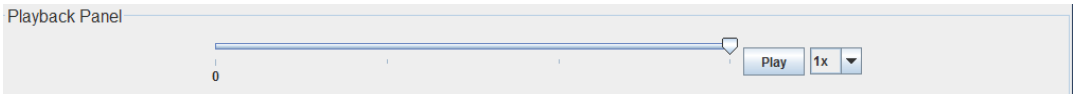


Figure 7 ANA Playback Panel

i. On-Line Network Creation

An important part of the visualized data set is the requirement to handle unknown and incomplete data. As communications are intercepted and found by various intelligence organizations nobody knows the full layout of the cell network or the identities of everybody involved. The simulation has to be able to accept incomplete data and allow us to later fill in the blanks as more information became available. This is the reason behind allowing all information to be edited, and why nodes are added to the graph one at a time.

Graphs do not just appear with hundreds of nodes and interconnections; they evolve slowly from nodes connecting to each other and new communications between actors showing up. We monitor this slowly from intelligence gatherings and intercepted communications.

ii. Change Detection

To be able to animate the evolution of the graph, ANA Keeps track of how the network evolves. It detects changes like addition of nodes and edits of those nodes through the various UI elements and records them as events. A Java class stands between the UI elements and the underlying graph that they represent. This Java class writes everything to a list of actions that can be saved to an XML file. By saving everything to XML users have the ability to save the state of their network and all of the actions that have been taken.

Each action that can occur in the graph is represented by a Java class responsible for keeping track of identification of nodes being added or edited and all of the new details of these graph elements. Through the JAXB Java utilities each of these classes can be automatically

marshaled to and from Java Objects to XML string representation. This file can be written out by the program and then saved to a file.

Classes such as the 'AddEdgeAction' in Figure 8 present fields for all the pertinent details of an edge, the ID's of the nodes it connects, the label on this edge, the description, an ID for this edge, and the order of this event. Order is a stack rank of actions used to keep actions ordered and avoid inconsistencies in the graph. The Edge ID is a parameter used to refer to this edge at a later date in case we modify the details about it such as the description and label. This class gets automatically converted through JAXB into the XML representation in Figure 9 and prints the values of all of its parameters. Because this class is nothing more than a data container access rights to most of the variables are kept as public for easy marshaling and unmarshalling by JAXB.

```
import
javax.xml.bind.annotation.XmlRootElement;

@XmlRootElement(name="AddEdgeAction")
public class AddEdgeAction extends
GraphAnimation {

        private int order;

        public int source;
        public int destination;
        public String description;
        public String label;

        public int edgeID;

}
```

Figure 8 Event Capture Internal Class

```
<AnimationList>
  <GraphAnimation>
    ……
    <AddEdgeAction>
      <orderID>2</orderID>
      <source>1</source>
      <destination>2</destination>
      <description></description>
      <label></label>
      <edgeID>1</edgeID>
    </AddEdgeAction>
    ……
  </GraphAnimation>
</AnimationList>
```

Figure 9 Event Capture Output XML

When reading and animating these various animations ANA knows how to playback each possible action (AddNodeAction, AddEdgeAction, EditEdgeAction, EditNodeAction) and keep track of it while making changes to the visualization. Additionally, ANA knows how to undo each of these possible actions for when users are scrolling backwards in time. It can undo edits on nodes by restoring the previous state of the node from either the most recent edit or the most recent addition of that node. This ability to play forward and backward allows the user to treat his network animation as a movie and skip to any particular time of interest.

Currently, all actions are treated as *frames* and ordered using a stack order based on when the user made the change in the interface. As future feature of the application I propose to add an additional ranking where events can be given timestamps. This ranking would allow users to associate more directly a specific time with events in the graph. Specifically it can allow the graph to record when an actor came into a graph, or when a communication appeared.

These actual times could be different from when the intelligence report is created and presents a source of error in the network. This underlying time of the event is a difficult situation to model and presents a source of error for social network as described in the next session.

iii. Underlying Social Network

Analysis of this data and network reveals a previously existing problem of network science and social network analysis. For most of the social network analysis works created networks are analyzed at a particular point in time based on the static measures we present above. Because these works analyze the networks at a particular point in time they suffer from one type of error. There are estimation errors because the network they observed of connections does not capture the actual social network in its entirety. New ties and actors may be emerging in the existing network that may not be caught by the research methods. For snapshot research however this is a minor source of error.

When analyzing a networks evolution over time, this error shown by the difference between the on the ground network, and how it is being reported and observed externally is compounded. At every time interval this error adds because we are unable to point specifically to when a new actor entered the network, and when a tie between two actors shows up. This is a problem of analyzing the knowledge about a particular network, (e.g., what intelligence reports reveal about the actors knowing each other, or when people become friends on an online social network) against analyzing the evolution of the underlying network (e.g., when two people actually met before they began

communicating via telephone, or when two people met physically before becoming friends on an online social network).

This problem needs to be analyzed more in depth in future work. For the purposes of this paper, the simulated data presents cases where people are in fact communicating for the first time in the intercepted intelligence reports. These were connected via a previous player in the network and are establishing initial contact. In such cases the knowledge of the network matches perfectly with the underlying social network. This is not always the case and presents a useful study to be applied to other networks such as online social network, publication networks, forums, and small communities.

iv. Stress Testing

In this section I test the ability of ANA to handle networks larger than the 30 node simulation used throughout this paper. The library that ANA is built on top of, Prefuse, allows a large quantity of data to be stored in its table structure. The original release of the library provided the visualization of a 600,000 node dataset that could show and hide sections based on user interaction. This is an extreme case because nodes in the visualization were lightly connected and the entire graph was not visible at the same time. For the graphs displayed by ANA much smaller sizes are needed but the graphs are denser and require more layout calculations.

I timed ANA to see how long it would take to create graphs of varying sizes with an average nodal degree of 5. Each time the graph would create the nodes and connected them with 5*n number of edges between randomly generated nodes. These timing values are displayed in Figure 10 and shows that ANA can scale mostly linearly. Figure 11 presents the same data but divides the time by the number of nodes that had to be inserted. With graphs smaller than 1,000 nodes and 5,000 edges, the creation of a node and layout is a constant time operation. The time to create the entire network becomes a linear operation in terms of the required network size.

Once the graph becomes too large the speed decreases. At this stage there is a lot of memory being used to store all the data and heavy processing is going in to laying out such a large and cross connected graph. With 500 nodes and 2500 edges the visualization was still fully interactive. The program was completely responsive as I rewound it and allowed the animations to play forward. At 2000 nodes however the graph became difficult to interact with, it became very busy on the screen with nodes and edges overlapping, and became difficult to make adjustments to the viewing of the graph.

For educational and risk analysis purposes such sizes of graphs are large enough. ANA is not meant to visualize and display social networks the same scale as Facebook or Twitter, but rather smaller adversarial networks, classrooms, or colleges. Such dimensions it can handle quite well and remain fully interactive.

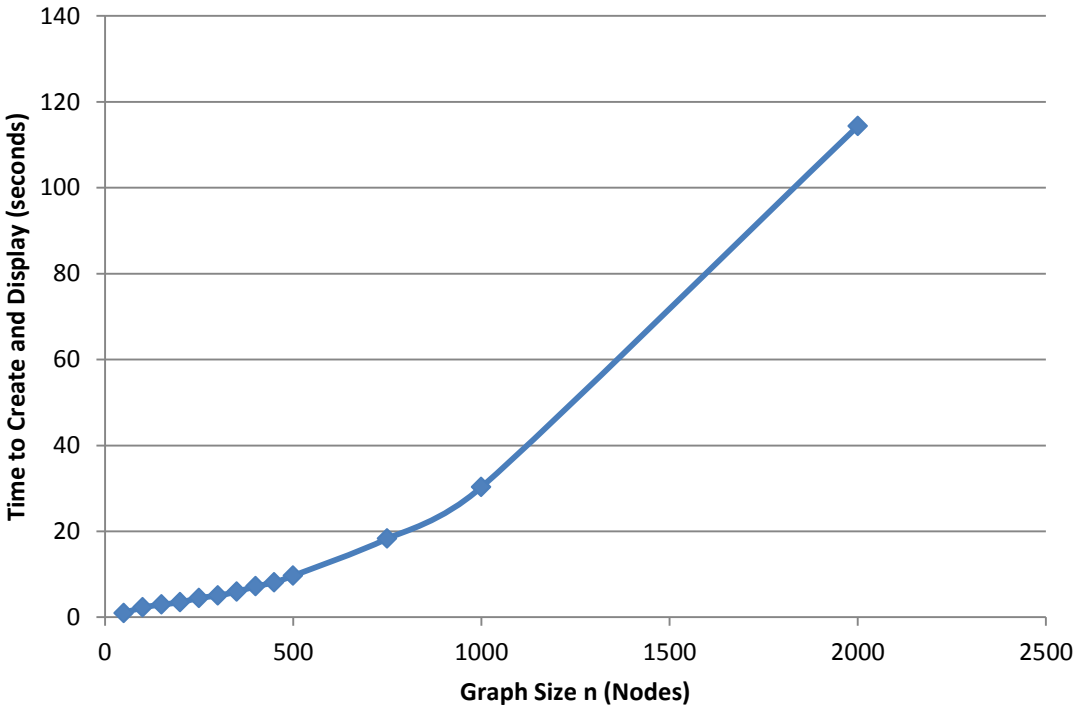


Figure 10 Scaling of ANA based on graph size. Run on a Core 2 Duo, 2.0 GHZ with 2GB of memory.

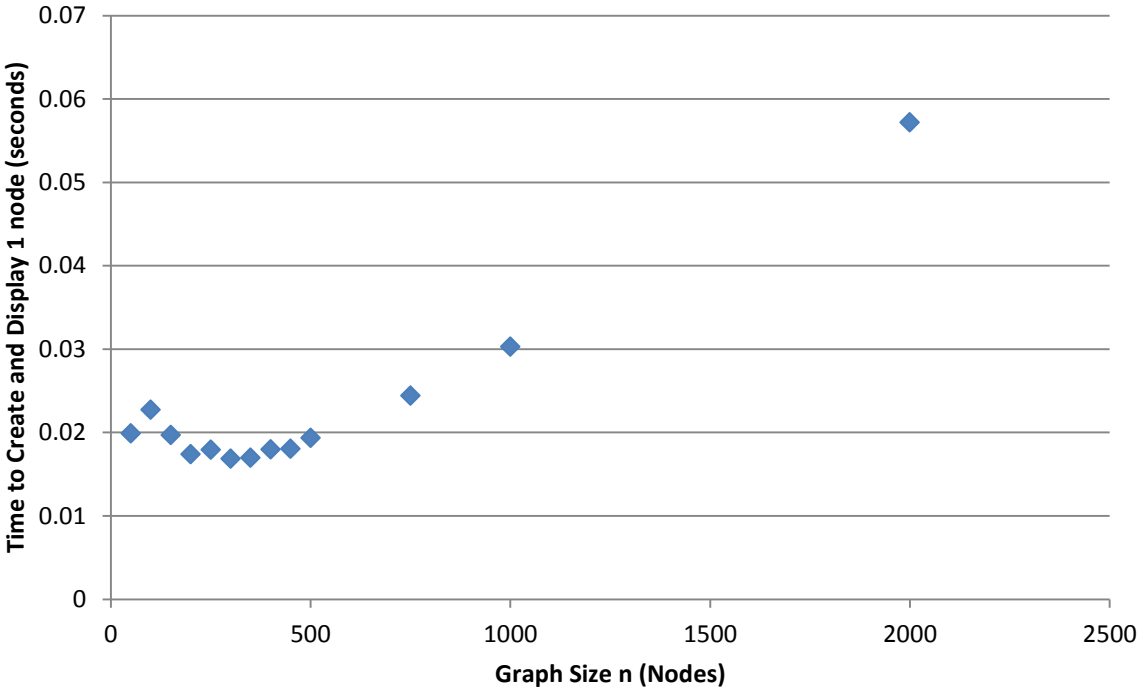


Figure 11 Scaling of ANA based on graph size. (Time/insert operation)

v. Mathematical Analysis

Performing mathematical analysis on a social network is a task currently done by many pieces of software. Some of the more well-known packages for such work are Organization Risk Analyzer ORA (Carley & Reminga, 2004; Carley, Columbus, DeReno, Reminga, & Moon, 2008) and UCINET (Borgatti, Everett, & Freeman, 2002), Graphviz (Ellson, Gansner, Koutsofios, North, & Woodhull, 2002), and many others available both commercially and open source or shareware. Each piece of software provides its own implementation of the calculations discussed in Chapter 2 and output these standard measures.

To avoid duplicating work and recreating another piece of social network analysis software that can calculate basic social network measures ANA provides an exporting function. This function can format the graph being visualized to a format that ORA can easily read and analyze. This allows ANA to remain lightweight network evolution and playback software. More intense calculations can

be done by existing software that have been improved and expanded for over 6 years of open source development.

In the file menu of ANA exists an export submenu with options such as ORA. This outputs the graphs list of nodes and edges to a format known as the DyNetML, an XML format that can hold all of the necessary data to be imported into ORA as a snapshot of the graph. When exporting the data for ORA some information is lost such as the timeline of all events leading to that state of the graph. For this reason ANA maintains its own XML format that is used to represent the graph, and at any time frame of the evolution the user can export the graph.

This allows users, analysts, students, and others in the intelligence community a simple and user friendly interface to enter data into a social network and evolve it with incoming communication. This interface is simpler to use than the more power but complex tools such as ORA and UCINET. Once the simple ANA UI is used to create the graph any analyst or user of the software may perform more mathematically intensive analysis inside ORA without having to recreate the network there.

4. Network Analysis

With the easy to use ANA software available I began analyzing the *Stop the Terrorist!* dataset (Shemanski, 2011) to visualize it and perform mathematical analysis. The final network visualized by ANA can be seen in Figure 12 and presents all of the actors along with the groups they belong in. The network displays features of social networks and contains many different subgroups. In the center of it all is Fateh Kamel who is easily confused to be a leader or the most central player in the network. He is however just a playmaker, and like similar adversarial networks the leader 'Shaker al-Abssi' is very lightly connected to the network.

Each of the subgroups acts as a cell only connected to the rest of network through only one or two connections allowing each task of the network to function independently and not be compromised if other groups are caught or disabled.

The center of the graph is show in Figure 13 to show the number of connections of Fateh Kamel, Shakel al-Abssi, as well as the relationship between these two players in the network.

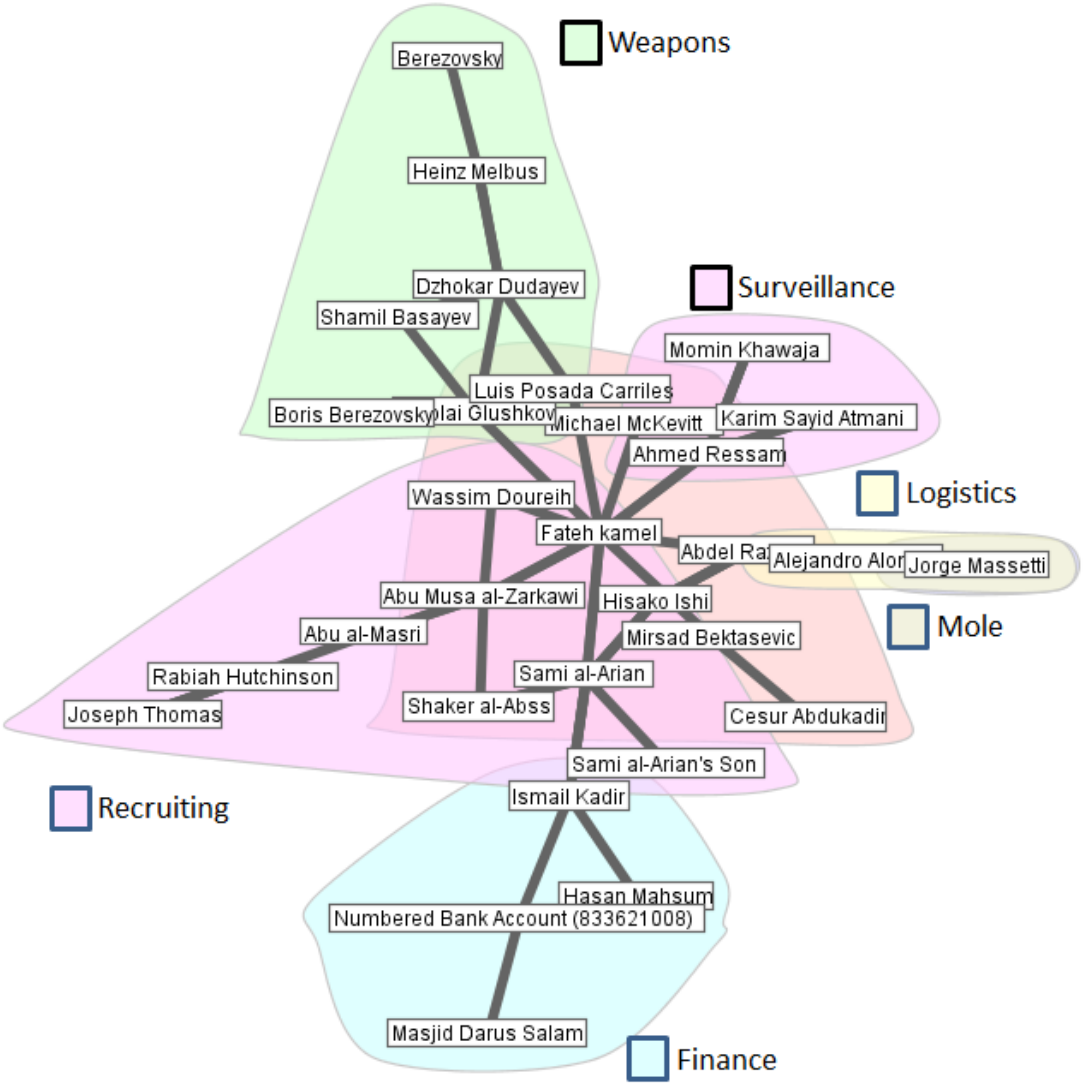


Figure 12 Final Visualized Network of Actors. Screenshot from ANA 1.0 (legend added by hand for b&w printing)

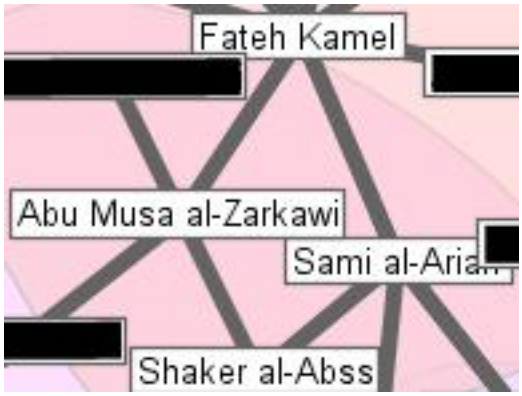


Figure 13 Center of final social network. Modified for emphasis on leader, playmaker, and their connections.

i. Static Measures

With the graph fully created and exported into ORA I can run a number of interesting mathematical analysis to see just how tightly coupled the network is and to identify key players. The measures from Chapter 2 are presented in **Error! Reference source not found.** with their average values and the values of key players along with the average distance between any two actors, and the clustering coefficient. The various centrality values are all low, measure less than 15% and show that the network is spread out and not very highly connected.

In most situations the leader, Shaker al-Abssi has a very low centrality value, and clustering coefficient. As a leader of an adversarial network, he remains hidden from the public and does not connect with all of the various subgroups of players in the plot. On the other hand, Fateh Kamel is the playmaker and he is highly connected to all of the members of the graph. In all but distance centrality, Fateh Kamel is most central to the graph and highly visible to all members. He also has the highest clustering coefficient for forming triads with nearby actors.

For centrality distance the leader has a higher value than expected because he is, in fact, central to the network. While he might not be connected directly to many of the members of the network, by being directly connected with the playmaker, Shaker al-Abssi places himself within small distance of most of the other actors. Edge actors, who are small time recruits for very specific function, however increase this value for the overall graph by being very distant from the other functions of the network.

Additionally, we can see that the distance of this graph is surprisingly small, at 3.440 simply because this is a small plot and the only connections included in the graph are relevant members to this particular task. The transitivity of this graph is also very low showing that only 4.7% of triads that could exist in the graph actually exist. Connectivity across this graph does not spread; members of the network who share a common relationship do not become connected. Communication

proceeds through the original hierarchy and ties rather than expanding to create direct communication channels that could expose members of the network.

Table 2 Static network Measures Summary

Measure	Graph Average	Fateh Kamel	Shaker al-Abssi
Degree Centrality	.053	.207	.034
Distance Centrality	.146	.305	.225
Betweenness Centrality	.048	.430	.012
Clustering Coefficient	.080	.042	.000
Distance	3.440		
Transitivity	.047		

The clustering coefficient, average distance, and average degree are measures that are frequently analyzed by network scientists. Albert and Barabasi (2002) did an in-depth study of a number of different networks from their own work as well as those presented in other papers. Some of the networks include the World Wide Web, movie actors, and co-authorship in many different fields. The values they put forth (Table 3) in that paper can be compared to values obtained from this adversarial network and compared for interestingness.

In Table 3$<k>$ represents the average degree of each node, l represents the average path length, and C represents the clustering coefficient of this particular network. Those values subscripted with *rand* are values that would exist in a random network of similar node size. These values of interest are extracted into Table 4 along with the adversarial network for comparison.

The adversarial network is of a very different nature from other networks, both social and nonsocial that are looked at in network science and social studies. While the average distance in an adversarial network is very comparable to many other networks (World Wide Web, Movie Actors, and some co-authorship networks), the clustering coefficient and average degree of these networks are very different. Clustering coefficient is a magnitude smaller than most comparable networks, and in degree is also smaller than other networks.

These features of the networks help them achieve their primary goals of remaining covert. Most actors are only connected to one, maybe two other actors so they can not reveal each other, and the chain of command remains large and hard to interfere with. The clustering coefficient remains small to keep actors in the network from communicating with each other and being found in large groups. An adversarial plot is much easier to detect when a weapons person, a financer, and a logistics person are all related and connected to each other. Without such connections the network will have a harder time carrying out its plot, but it will be harder to detect and interfere with.

Table 3 Network measure summary table duplicated from Albert and Barabassi (2002)

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998

Table 4 Combined network measures. Shemanski Adversarial Network shown in bold and key values extracted from and Albert, Barabasi (2002) work shown for comparison.

Network	Size	Average Degree	Average Distance	Clustering Coefficient
Shemanski Network	**30**	**1.53**	**3.44**	**.080**
WWW	153,127	35.21	3.10	.1078
Movie Actors	225,226	61	3.65	.79
LANL co-authorship	52,909	9.7	5.90	.43
MEDLINE co-authorship	1,520,251	18.1	4.60	.066
SPIRES co-authorship	56,627	173	4.00	.726
NCSTRL co-authorship	11,994	3.59	9.70	.496
Math. Co-authorship	70,975	3.9	9.50	.59
Neurosci. Co-authorship	209,293	11.5	6.00	.76
Power Grid	4,941	2.67	18.70	.08

To analyze the connectivity of these graphs in details I chose to compare its distribution of connections to more regular social networks by looking at how well this adversarial network fit a power law pattern. For each node in the graph I output the number of edges it has and turned the plot into the log-log histogram in Figure 14.

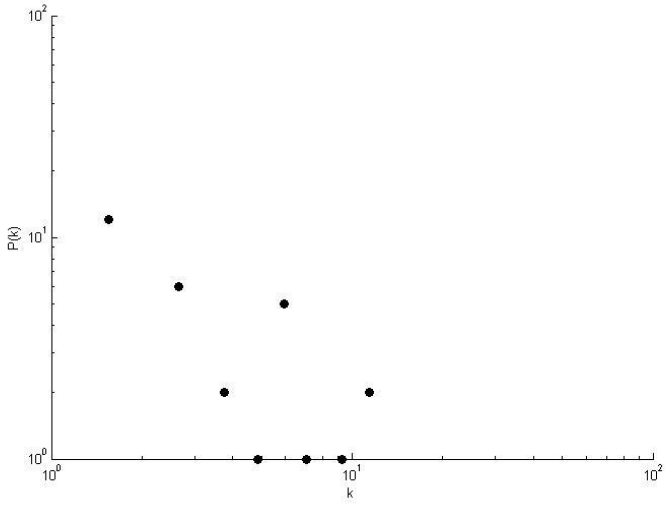


Figure 14 Log-log histogram of degree distributions

Unfortunately this network is very small and does not lend itself to such analysis very easily. The pattern seen here is that most nodes have one or two connections, and there are a few nodes that have many connections (the playmaker, Fateh Kamel). This pattern does not allow for many parallels to be drawn to other social networks but the slope presented by this plot is similar to power law analysis published in previous work (Barabasi, Jeong, Nelda, Ravasz, Schubert, & Vicsek, 2002) shown in Figure 15. Similar to the adversarial networks these have a decreasing slope, but allow for much more detail in the line due to a much larger network.

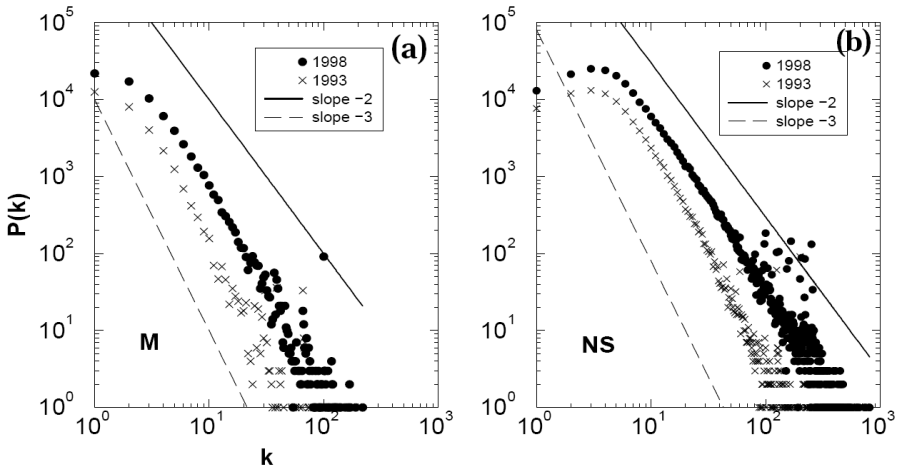


Figure 15 Log-log histogram of degree distributions. Co-authorship networks in mathematics (a) and neuroscience (b) (Barabasi, Jeong, Nelda, Ravasz, Schubert, & Vicsek, 2002)

ii. Dynamic Measures

Static measures of the network showed us that this particular social network differs significantly from the normal social networks we observe in everyday life. As individual numbers those were interesting but they present more interest when they are evaluated over the time period of the network developing.

To create these analyses I exported the graph at many different time intervals from ANA to the ORA format. Because it took 111 total frames of graph changes to achieve the final graph in Figure 12, I then broke down the network into 10 different snapshots. At every 10 frames I exported the state of the graph to an XML file that could be easily loaded into ORA for their time series analysis. The final frame was at frame 111 rather than 110 because the last 20 frames contain only changes to add members to specific subgroups rather than adding more edges or nodes to the graph.

These 11 different snapshots of the graph were then analyzed on the same values as presented in Table 2 but over the time course of the network. Not only do I now present the results for the average measures presented above but I also compare the two important members of the network, the leader Shaker al-Abssi and the playmaker Fateh Kamel.

Centrality Measures

Figure 16 presents the results for the change in centrality values over the time course of the network evolution. These values are already small, less than 50%, for all of the measures but present an interesting pattern of decreasing over the time of the simulation. Degree centrality and distance centrality show how both values decrease once the graph achieves a size of greater than 5 nodes, and then stabilize at a very small value of measurement. The deception forces inside this network keep it from becoming too centralized so that it may maintain its cell like structure remains hard to detect or infiltrate.

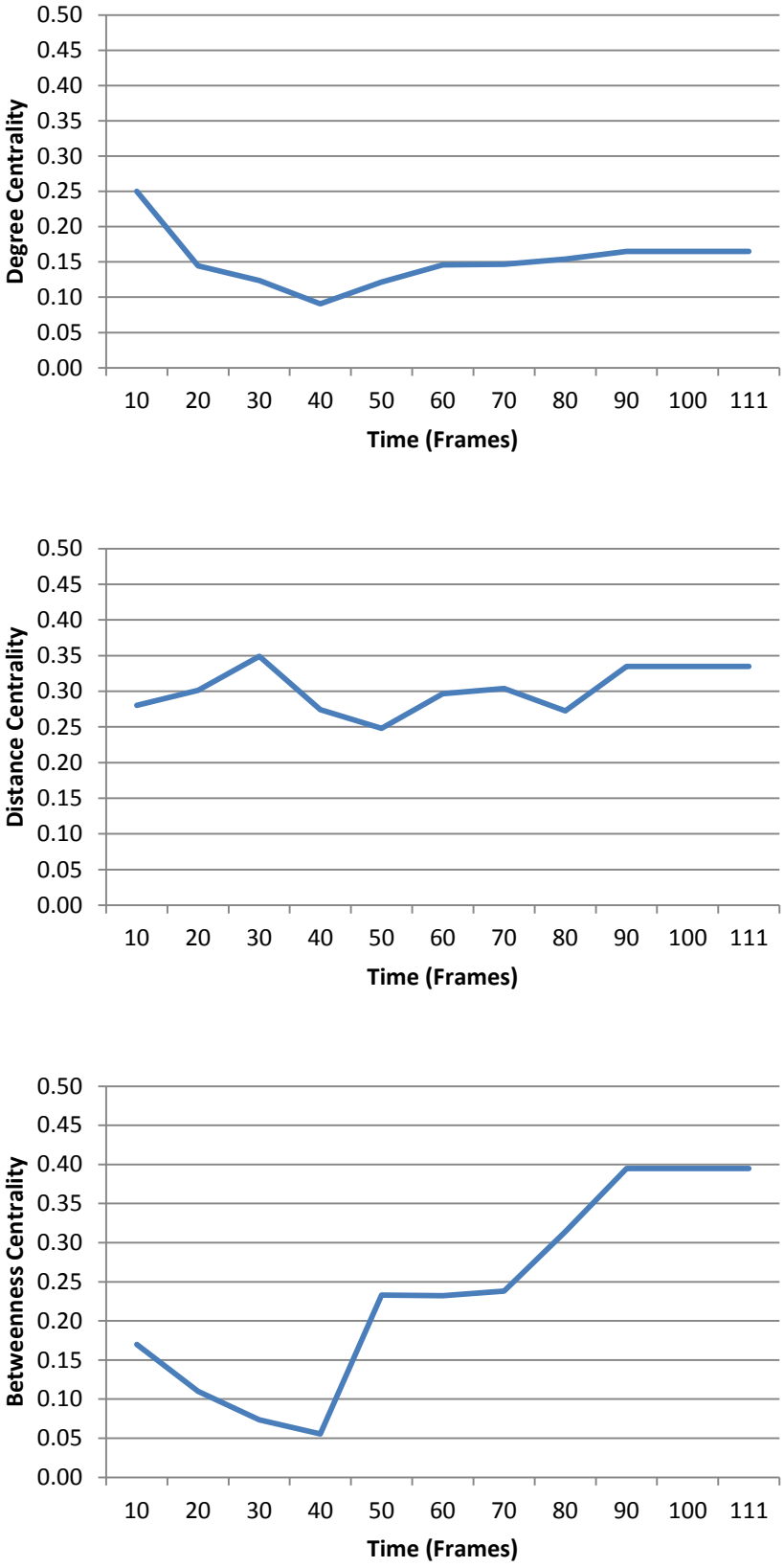


Figure 16 a) Degree Centrality b) Distance Centrality, c) Betweenness Centrality over time for adversarial network simulation.

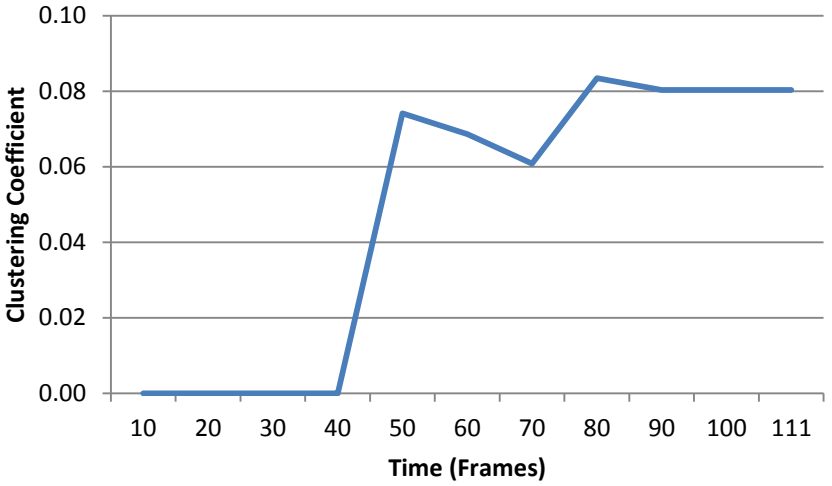
Additional Measures

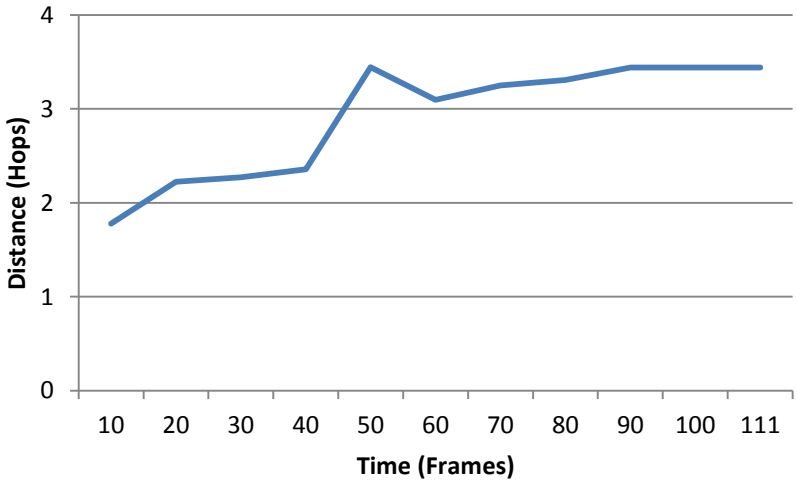
For the non-centrality based measures, clustering coefficient, average distance, and transitivity, presented in Figure 17 we see very similar patterns of the graph aiming to be more spread out and less tightly connected as more actors are brought into the network.

Clustering coefficient does increase between the 40 and 50 frames due to a few connections developing. These connections bring the finance and weapons subgroups closer together so they can more effectively work inside of their subgroup. Over time this clustering coefficient does not continue to grow and the groups grow farther apart.

Distance across the network continues to increase over the entire time course of the network as more nodes are added to the far ends without connecting them to the center of the graph for quick communication paths. This keeps the two far ends of the network very far apart and allows one end of the network to remain safe if anything were to compromise the other end.

Finally, transitivity only increases when weapons and finances subgroups become connected but decreases afterwards similar to the clustering coefficient. This shows that very few ties are created between triads of actors and instead the network chooses to communicate through the longer pre-existing chains of command and communication.





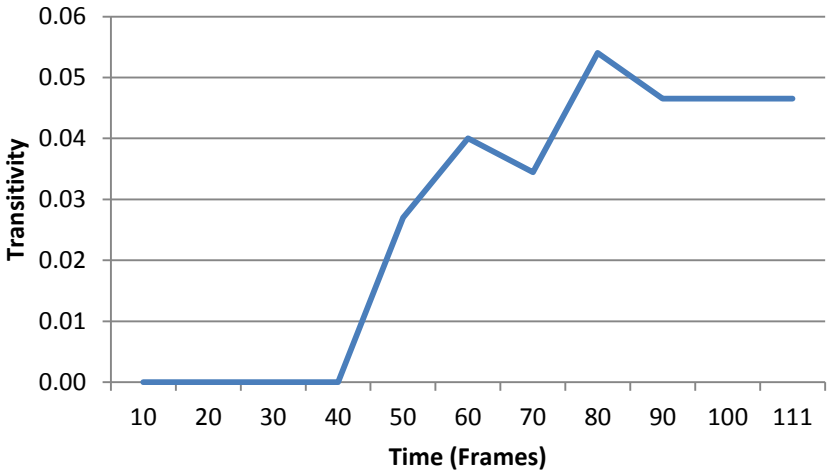


Figure 17a) Clustering Coefficient, b) Average nodal distance c) Transitivity over time for adversarial network simulation

Actor vs. Actor Measures

Many of the values calculated in the previous sections can be calculated for particular actors. In Figure 18 I show a comparison between centrality and clustering of the two key players of the network, Fateh Kamel and Shaker al-Abssi. The important pattern seen in all of these graphs is that while both actors have low values for all of these network measures, the leader tries to remains less central and less visible compared to the playmaker. The leader Shaker is always looking to be more obscured by the surrounded network, while the playmaker Fateh Kamel is at times looking to grow more connections so that he can more quickly work with the various subgroups and leaders of those subgroups. Surprisingly, the leader has a lower *distance centrality* in the graph by the end of the simulation.

The betweenness centrality, degree centrality, and clustering coefficient for the playmaker actually increase in the graph as he becomes more tightly coupled to some of the people he is directing and organizing. This allows him to be effective, and yet leave the actual leader undetectable.

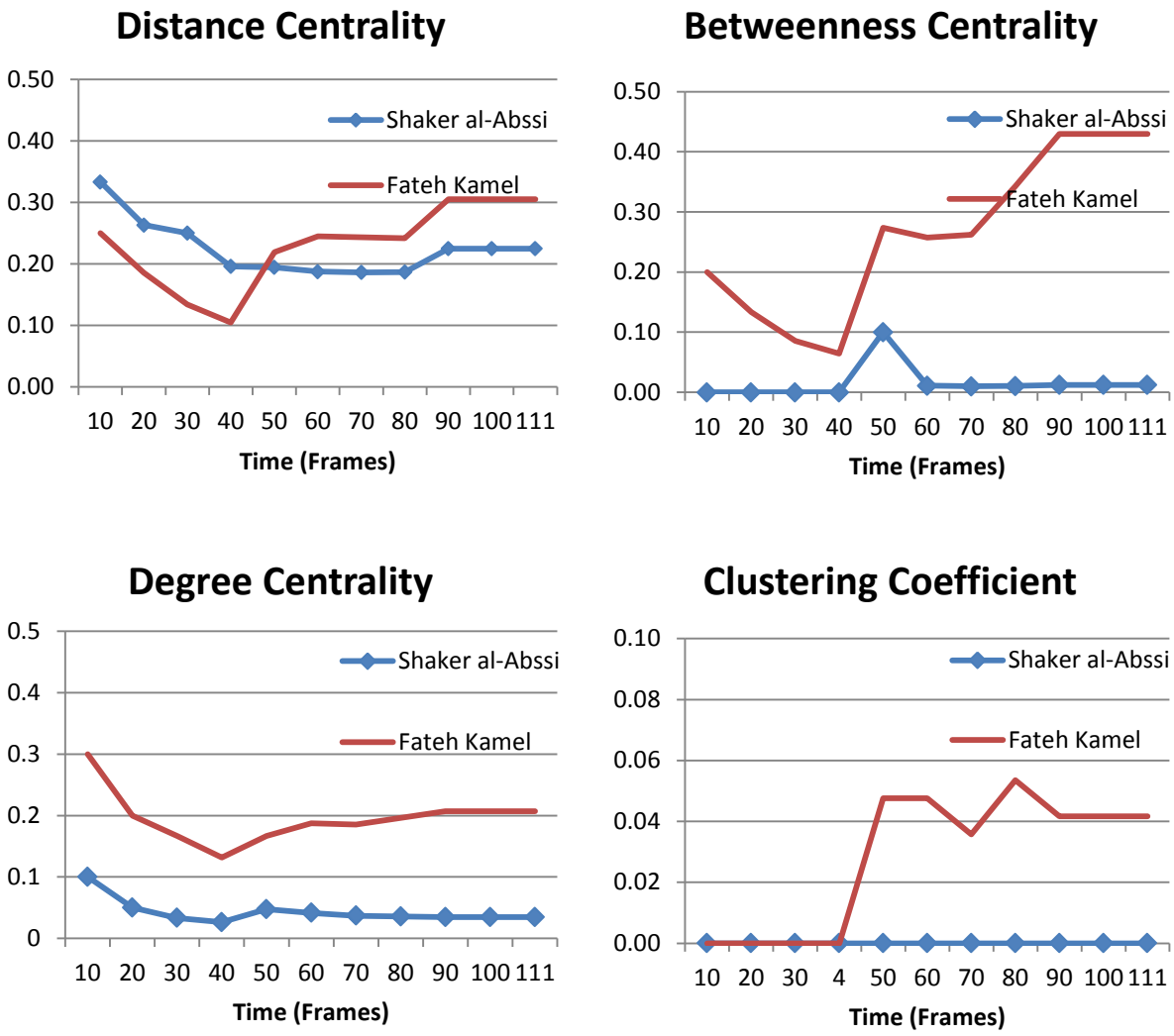


Figure 18 a,b,c,d Centrality and clustering measures of actors over time course of network

iii. Animations

The animations provided by ANA are useful in detailed analysis of the network and the patterns exhibited by it. By playing back frames 40-50 I was able to quickly see that sections of the graph changed in that time frame. This allowed me to see how the graph became more connected and explained the patterns seen in Figure 17a and c as well as in Figure 18d when the graph grew more tightly connected as opposed to spread out as expected. While ORA provided great mathematical toolbox to look at this network it could not allow for a more in depth look at particular events noticed in the graph. Besides being a learning exercise for students or analysts who use ANA, the playback feature can provide analytical purposes such as these.

iv. Summary

By looking at an adversarial network through both static and dynamic network analysis methodologies we can see that empirically see how these networks differ from what is a standard social network of friends, online communities, or co-authorship. Adversarial networks maintain a much smaller degree centrality than other networks. They are not interested in having each person be connected to as many other people in the network as possible. Each person is only connected to one, maybe two other people that are strictly necessary to accomplish the task at hand. Additionally the clustering coefficient is small to minimize triangles of connections between users.

In terms of dynamic measures these networks do not follow pattern of increasing connectivity, centrality, and clustering. These values actually decrease over the time of the network, and the average distance between members of the network increases. The networks push to be more spread out and increase the distance between recruits with high risk of being compromised and the leaders and playmaker of the network that hold everybody together and carry out key actions.

The leader of such a network remains hidden from almost all members of the network. While this dataset does not specifically say how many people inside the network could identify who the leader is, our information and analysis of this shows that almost nobody would be able to identify the leader. Shaker al-Abssi remains nearly invisible from all but one or two members of the network. These members are the playmaker who actually carries out all of his orders and missions, and an additional buffer person. These covert strategies allow the playmaker to be the most visible person in the network. If he was to be compromised, he could be easily replaced by another member to carry out the orders of the leaders and bring the various functions together for a successful plot.

These networks are an interesting case study in social networks that are not standard, and motivate new work to be done that can maintain, visualize and analyze them. More measures of these networks could be performed and analyzed to compare against different types of networks.

5. Conclusions

This paper introduced a new way to look at social networks, in particular adversarial networks and to analyze them for new patterns. These networks are seen through the eyes of a new animated visualization tool ANA that can build the network as information about its actors and connections emerge. One such network, the Shemanski simulation of an adversarial plot is visualized through ANA, and using ANA's interface to standard file formats it is mathematically analyzed through social network measures.

The next sections provide a summary of the contributions provided by this paper relating to the following topics.

- Review of social network analysis and network science foundations (i, iii, iv)
- Easy to use social network editing and visualizing tool for students and analysts (ii)
- Comparison of adversarial network to standard social networks (iii)
- Time based analysis of evolution of adversarial networks (iii, iv)

i. Contributions to Network Science

To best understand the social networking implications behind adversarial networks, the requirements of the ANA application, and the analysis conducted in this paper it is important to have a strong background on network science. For this reason Chapter two of this paper provided a strong review of network science, graph theory, and important concepts.

Once I introduce social networks, the measures of centrality, and clustering I present some of the shortfalls of these measures. Most work done in network science using these measures analyzes these values as static variables at a snapshot of the network. They are not enough to see

the features of a social network. Chapter 2 provides an introduction of dynamic versions of these measures by looking at how they change over the time course of the network.

Dynamic measures of centrality, transitivity, and clustering allow users to view if the network is evolving like normal networks to become more connected, at what rate, and what unnatural patterns exist. For adversarial networks these patterns are of interest.

ii. Contributions to Education

The introduction of network science in Chapter 2 gives new students of this field a good introduction of the field as well as its foundations. Additionally, ANA provides an interactive way to study and analyze social networks for students and analysts looking at adversarial networks. The tool allows people to add information to a social network as they receive it and instantly visualize the changes in the network.

ANA is simple to use in terms of adding nodes and edges to an existing and modifying information about them. It allows users to keep as much detail as they want about actors and communications between them in an internal details section. ANA also allows different members of a network to be grouped in subgroups and displayed as an aggregate group.

The last feature of ANA is a playback feature of showing the network from the very beginning to the current state. By detecting every minor change to the network, ANA is able to play forwards and backwards through these changes to show in an visually animated fashion the changes and evolution of the graph. This feature is important for data analysis to explain the patterns and abnormal events seen by mathematical analysis of the network.

ANA is a Java application available for free for anyone interested in examining how networks change and visualizing this process. It uses a simple XML format to save and load data and provides interfaces to connect to other more complex social network analysis software packages. Export

functions will save the network to DyNetXML formats for software such as ORA to run mathematical analysis on.

iii. Measures of Time Evolution of Networks

The strength of ANA is displayed in Chapter 4 of this paper when its implementation is used to visualize and analyze a simulated adversarial network. This network is a plot of a terrorist attack and shows all the communication between the members involved in it. All of the data is inserted into ANA one communication report at a time and the network is evolved from a simple one person idea to a fully connected adversarial plot.

This finalized graph is visible in Figure 12 and is exported for mathematical analysis inside the ORA analysis framework. The results show that this network differs from standard social networks by being farther spread out, less central, and having less clustering.

The important actors, the leader, and the playmaker, of an adversarial network show interesting patterns. The playmaker is the most connected member of the graph because he brings together the leaders of various functional subgroups. He is important to the network, but only receives orders and carries them on without much decision making power. The leader of the plot remains lightly visible and separated from the playmaker to maintain the covertness of the network and to protect himself from outside interference.

In a timewise analysis of the social network these two actors show many more differences between them. The leader, as expected, avoids building more connections and rather allows himself to be less and less central to the network as the network evolves. He is always less central and connected than the playmaker. The playmaker while not being heavily connected must build a number of connections between actors so the functional subgroups can work together.

These time based analyses of adversarial social networks show concretely how these differ fundamentally from normal social networks. The network does not evolve to be more connected, nor does it evolve to be more central. The actors remain spread far apart so that each subgroup is separated and protected from problems that may occur in other parts of the network.

iv. Limitations

Social Network analysis of these types of networks is a challenging task due to the nature of the networks. The network itself is adversarial and remains covert or tries to hide its underlying structure to improve its own performance. This causes error in the data that is obtained about the network and can complicate the analysis of such a network.

The observation methods used to record and look at social networks suffer from an inherent lag that can cause error in the analysis. All analysis is done on the network evolution of when we observe connections to be created. This is an analysis of our understanding of the social network rather than an analysis of the underlying evolution of the social network. Not all of the connections that are appearing through communications between actors are the first interaction between them. Many of these connections could have been formed days, months or even years earlier but only been called into action when we observed it.

This inherent difference between our view of the network and the underlying structure of the network presents a source of error and remains as something to be looked at in future time based analyses of social networks. Even work that does not study adversarial networks suffers from such a lag. Connections on popular social networks (Facebook, Twitter, LinkedIn) are not formed in a vacuum and are usually representative of an earlier interaction between them. Same limitation applies to studies of publication networks that are frequent in network science. These publication databases suffer from a lag between the time when a collaborative paper is published, and when those researchers met each other and began sharing ideas and working together.

v. Future Work

ANA in its current state does a good job of fulfilling a use case for students and intelligence analysts with a very simple interface and easy to use features. Future improvements can add more strength to ANA by providing support for more types of networks. The three additional types of networks that should be supported by future versions of ANA include networks with positive and negative relationships, directional relationships, and multi-mode networks to support the inclusion of events, and multi-person meetings. These features would allow more complete modeling of the interactions of an adversarial network but would have to be carefully implemented as to not overly complicate the interface and visualization of the network.

Graph Segmentation

Networks with positive and negative relationships between actors (e.g., Actor A dislikes Actor B, and Actor B likes Actor C) provide an interesting use case for adversarial networks. When extending the network of players in such a simulation to include the members of the international intelligence community a new type of network can emerge. Various members of the intelligence community can have negative connections to those people they are monitoring such as terrorist subjects. At the same time the intelligence community contains positive ties within it as hopefully the members have favorable impressions of each other. A mathematical balance theory (Heider, 1946; 1963; 1967) can be applied to such a network to determine if the network is able to be segmented into two distinct communities that contain only negative ties between themselves, but positive ties within the community.

For future work we propose to use this simulation with additional data that includes the various intelligence agencies as actors in the simulation. Agencies that monitor particular people of interest can be considered to have negative relationships to the person of interest, and agencies who share information amongst each other can be considered to have positive relationships. Such a

network should then be mathematically segmented to show that the entire network is split into two communities working against each other.

Directional Graphs

Adversarial and hierarchical networks are not always composed of bidirectional communication. There are many instances when rather than two actors communicating; one actor is actually just giving an order to the other actor. This extension to ANA will provide even more power to the analysis and visualization provided by this tool. It is important that this feature be implemented in a user friendly way so it does not overly complicated the interface. This visualization becomes even more complex when two actors have multiple interactions between them. If some of their interactions are bidirectional communications but one interaction contains a one directional order, the visualization of such situations can become complicated and difficult to understand.

Multi-Mode Networks

Networks can evolve to be more complex than relationships between two actors by including events, locations, modes of communication, and other objects in the network. These multi-mode networks allow multiple types of agents to be related to each other. Instead of actors having an edge between them for being connected, they connect to a common event where they both participated (Qiu, Ivanova, Yen, Liu, & Ritter, 2011). These types of networks are more complex, but they do have more flexibility in expressing how and why members are connected. Each communication between two agents in the system can be a separate object to show more detail about what happens between two actors.

Besides just allowing two actors to connect to each other through multiple communications, Multi-mode networks allow us to have communications between three or more actors at the same time. Interactions where three members participated are important and need to be differentiated from cases where three members only communicate in a pairwise fashion rather than as a unity

(Breiger, 1974). Multi-mode networks can display this type of relationship and allow for an extension to the social network of adversarial networks.

With the added strength of such network visualization, unfortunately there is added complexity for the user, and the students who might be trying to learn network science and social network analysis. In future versions of the ANA toolkit I propose the introduction of multi-mode networks in a fashion that is easily understandable and usable. The human computer interaction and cognitive implications of this addition are very complex but would give much more power to filtering methods and mathematical analysis of the networks.

vi. Summary

This paper provided a novel way to monitor, analyze, and look at adversarial social networks through a visualization tool and supporting mathematical analysis. The analysis exposes some shortfalls of social network analysis that are caused by intrinsic difficulties in obtaining data about various social networks. Analyses that look at how centrality, clustering, and connectivity of social networks change during their evolution are interesting and can show patterns that may match standard social networks, or diverge from them in adversarial networks. This type of time based data is difficult to obtain, and suffers from a lag between the recording system, and the underlying connections being formed between the actors.

These social network limitations do not interfere with this papers ability to provide a new interface that can handle such data for future networks. The ANA toolkit provided in this work allows us to take incoming data about members joining a social network and new connections being created in the network and record them for visualization purposes. The visualization of the network is constantly evolving as the network itself is changing.

As this network changes, we can easily export snapshots of the network to standard file formats to be analyzed quickly in mathematical frameworks. This type of analysis presented in

Chapter 4 showed interesting patterns in adversarial networks and how they remain covert by not developing additional connections and remaining a widely spread graph with limited connections. Social network analysis and visualization continues to be a methodology useful for looking at the activities of groups, in particular adversarial networks, by considering the context and structure of their connections.

References

Albert, R., & Barabasi, A.-L. (2002). Statistical Mechanics of Complex Networks. *Review of Modern Physics, 74*, 47-97.

Anderson, C. (2004, October). The Long Tail. *Wired, 12*(10).

Anthonisse, J. M. (1971). *The Rush in a Graph.* Amsterdam: Mathematische Centrum.

Bakshy, E., Simmons, M. P., Huffaker, D. A., Teng, C.-Y., & Adamic, L. A. (2010). The Social Dynamics of Economic Activity in a Virtual World. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.* Washington DC.

Baldwin, T. T., Bedell, M. D., & Johnson, J. L. (1997). The Social Fabric of a Team Based M.B.A. Program: Network Effects on Student Satisfaction and Performance. *Academy of Management Journal, 40*(6), 1369-1397.

Barabasi, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*, 509-512.

Barabasi, A.-L., Jeong, H., Nelda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A 311*, 590-614.

Barnes, J. A. (1954). Class and committees in a Norwegian island parish. *Human Relations, 7*, 39-58.

Beauchamp, M. A. (1965). An improved index of centrality. *Behavioral Science, 10*, 161-163.

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

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

Breiger, R. (1974). The Duality of Persons and Groups. *Social Forces, 53*, 181-190.

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., et al. (2000). Graph Structure in the Web. *Proceedings of 9th International World Wide Web Conference*, *33*, pp. 309-320.

Carley, K. M., & Reminga, J. (2004). *ORA: Organization Risk Analyzer.*

Carley, K. M., Columbus, D., DeReno, M., Reminga, J., & Moon, I.-C. (2008). *ORA User's Guide.*

Carley, K. M., Lee, J.-S., & Krackhardt, D. (2002). Destabilizing Networks. *Connections 24, 3*, 79-92.

Coleman, J., Katz, E., & Menzel, H. (1957). The Diffusion of an Innovation among Physicians. *Sociometry, 20*, 253-270.

Davis, J. A. (1963). Structural balance, mechanical solidarity, and interpersonal relations. *American Journal of Sociology, 35*, 444-462.

Davis, J. A. (1967). Clustering and structural balance ing raphs. *Human Relations, 20*, 181-187.

Easley, D., & Kleinberg, J. (2010). *Networks, Crowds, and Markets: Reasoning about a highly connected world.* Cambridge University Press.

Ellson, J., Gansner, E., Koutsofios, L., North, S. C., & Woodhull, G. (2002). Graphviz - Open Source Graph Drawing Tools. *Lecture Notes in Computer Science, 2265*, 594-597.

Farley, J. D. (2003). Breakign Al Qaeda Cells: A Mathematical Analysis of Counterterrorism Operations (A Guide for Risk Assesment and Decision Making). *Studies in Conflict & Terrorism, 26*, 399-411.

Freeman, L. C. (1977). A Set of Measures of Centrality Based on Betweeness. *Sociometry, 40*, 35-41.

Freeman, L. C. (1979). Centrality in Social Networks: I. Conceptual Clarification. *Social Networks, 1*, 215-239.

Golder, S. A., & Yardi, S. (2010). Structural predictors of tie formation in Twitter: transitivity and mutuality. *Proceedings of the Second IEEE International Conference on Social Computing*, (pp. 88-95). Minneapolis, MN.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology, 78*, 1360-1380.

Granovetter, M. (1974). *Getting a Job: A Study of Contacts and Careers.* Cambridge, MA: Harvard University Press.

Heer, J., Card, S. K., & Landay, J. A. (2005). Prefuse: A toolkit for interactive information visualization. *SIGCHI conference on Human factors in computing systems* (pp. 421-430). ACM.

Heider, F. (1946). Attitudes and Cognitive Organization. *Journal of Psychology, 21*, 107-112.

Heymann, P., Koutrika, G., & Garcia-Molina, H. (2008). Can Social Boomarking Impove Web Search. *First ACM International Conference on Web Search and Data mining.* Stanford, CA.

Holland, P. W., & Leinhardt, S. (1971). Transitivity in structural models of small groups. *Comparative Group Studies, 2*, 107-124.

Holland, P. W., & Leinhardt, S. (1972). Some evidence on the transitivity of positive interpersonal sentiment. *American Journal of Sociology*(72), 1205-1209.

Ioannis, K., Vassilios, S., & Joemon, M. (2009). On social networks and collaborative recommendation. *Proceedings of the 32nd International ACM SIGIR Conference*, (pp. 195-202). New York, NY.

Klerks, P. (2001). The Network paradigm Applied to criminal Organizations: Theoretical nitpickign or a relevant doctrine for investigators? Recent developments in the Netherlands. *Connections 24, 3*, 53-65.

Krebs, V. E. (2002). Mapping Networks of Terrorists. *Connections 24, 3*, 43-52.

Laumann, E., & Pappi, F. (1973). New Directions in the Study of Elites. *American Sociological Review, 38*, 212-230.

Milgram, S. (1967, May). The Small World Problem. *Psychology Today, 1*(1), 61-67.

Moreno, J. L. (1934). *Who Shall Survive? : Foundations of Sociometry, Group Psychotherapy, and Sociodrama.* Washington DC: Nervous and Mental Disease Publishing Co.

Moreno, J. L. (1953). *Who Shall Survive?: Foundations of Sociometry, Group Psychotherapy, and Sociodrama* . Beacon, New York: Beacon House Inc.

Moreno, J. L. (1978). *Who Shall Survive?: Foundations of Sociometry, Group Psychotherapy, and Sociodrama.* Beacon, NY: Beacon House, Inc.

Newman, M. E. (2003, June). The Structure and Function of Complex Networks. *SIAM Review, 45*, 167-256.

Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69.

Proctor, C. H., & Loomis, C. P. (1951). Analysis of Sociometric Data. *Research methods in social Relations*, 561-586.

Qiu, B., Ivanova, K., Yen, J., Liu, P., & Ritter, F. E. (2011). Event-driven modelling of evolving social networks. *Int. J. Social Computing and Cyber-Physical Systems, 1*(1), 13-32.

Sabidussi, G. (1966). The Centrality Index of a Graph. *Psychometrika, 31*, 581-603.

Shaw, M. E. (1954). Group Strcture and the Behavior of Individuals in Small Groups. *Journal of Psychollogy, 38*, 501-507.

Shemanski, D. R. (2011). *Stop the Terrorists! Team-based Simulation of an International Terrorist Plot to Acquire and Use a Weapon of Mass Destruction.* ACS Tech Report 2011-1.

Shimbel, A. (1953). Structural Parameters of Communication Networks. *Bulletin of Mathematical Biophysics, 15*, 501-507.

Wasserman, S., & Faust, K. (1999). *Social Network Analysis: Methods and Applications.* Cambridge: Cambirdge University Press.

Wayne, Z. (1977). An information flow model for conflict fission in small groups. *Journal of Anthropological Research*, 452-473.

Yang, H.-L., & Tang, J.-H. (2003). Effects of Social Network on Students' Performance: A Web-Based Forum Study in Taiwan. *Journal of Asynchronous Learning Networks, 8*(3), 93-107.