

A Dynamic Network Approach to the Assessment of Terrorist Groups and the Impact of Alternative Courses of Action

Kathleen M. Carley

Institute for Software Research International, SCS
Carnegie Mellon University
Pittsburgh, PA 15213
USA

kathleen.carley@cs.cmu.edu

ABSTRACT

Dynamic network analysis (DNA) is an emergent field centered on the collection, analysis, understanding and prediction of dynamic relations among various entities such as actors, events and resources and the impact of such dynamics on individual and group behaviour. DNA facilitates reasoning about terrorist groups as complex dynamic networked systems that evolve over time. An interoperable DNA toolchain for collecting data on, assessing the network of, and forecasting changes in that network is presented. The use of these DNA tools to assess a terrorist group is then demonstrated using open source data. Key techniques are demonstrated using a dataset collected on terror networks. Techniques demonstrated include those for identification of an actor's sphere of influence, emergent leaders, and paths among critical actors, and metrics for assessing the potential immediate and near term impact of various courses of action.

1.0 DYNAMIC NETWORK ANALYSIS

Complex socio-technical systems are furthermore dynamic. The dynamics result from multiple change processes such as natural evolutionary processes including learning, birth and aging as well as intervention processes such as altering the set of individuals who lead a system. Data on these systems is often incomplete, replete with errors, and difficult to collect, which further complicates the understanding and evaluation of these groups which can be extremely large. Traditional analysis approaches, such as Social Network Analysis (SNA) (Wasserman and Faust, 1994), forensic science (Saferstein, 2001), and link analysis (Thelwall, 2004) are limited in their ability to handle the type of multi-plex, multi-mode, large-scale dynamic data that are needed to characterize terror networks. Consequently, tools that go beyond traditional SNA and link analysis are needed.

In response to these needs, a new sub-field, Dynamic Network Analysis (DNA), has emerged (Carley, 2003). In a DNA approach, computational techniques, such as machine learning and artificial intelligence, are combined with social network analytic and link analytic techniques to provide an approach to assessing multi-mode, multi-plex dynamic large-scale systems under varying levels of uncertainty. Dynamic network analysis (DNA) facilitates the collection, analysis, understanding and prediction of dynamic relations among diverse entities and the impact of such dynamics on individual and group behaviour.

2.0 DNA TOOLCHAIN

The integrated CASOS dynamic network analysis toolkit is an interoperable set of scalable software tools for coding, analyzing and forecasting behaviour given relational or “network” data. These tools form a

Carley, K.M. (2006) A Dynamic Network Approach to the Assessment of Terrorist Groups and the Impact of Alternative Courses of Action. In *Visualising Network Information* (pp. KN1-1 – KN1-10). Meeting Proceedings RTO-MP-IST-063, Keynote 1. Neuilly-sur-Seine, France: RTO. Available from: <http://www.rto.nato.int/abstracts.asp>.

toolchain that facilitates the dynamic extraction, analysis, visualization and reasoning about key actors, hidden groups, vulnerabilities and changes in multi-mode multi-plex dynamic networks, such as those formed by terror networks. This toolchain enables analysts to move from raw texts to meta-networks to the identification of patterns in networks to analysis of possible effects of alternative courses of action designed to disrupt or contain these networks. This is accomplished by using tools for text mining, statistical network analysis and computer simulation. Secondary tools such as those for web-scraping, visualization, database management and data editing provide support capabilities. This toolchain is illustrated in Figure 1. Using this toolchain the kinds of questions that can be addressed by the analyst include:

- Who to target (vulnerabilities)
- What groups or individuals stand out
- How to influence
- Are there important connections among actors or groups
- What is the “health” of the organization
- Where might there be missing data
- How different are two groups – or two sources – or the same group at two different times
- What is the immediate and near term impact of various courses of action

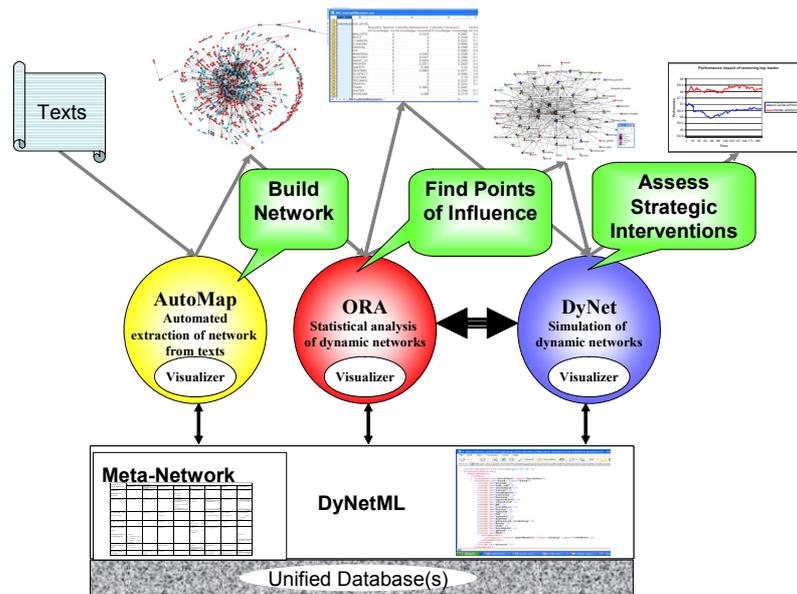


Figure 1: Toolchain for Dynamic Network Analysis

The CMU CASOS group has developed an interoperable suite of tools that acts as a chain to extract networks from texts, analyze these networks, and then engage in what-if reasoning. This tool suite takes into account multi-mode, multi-link, and multi-time period data including attributes of nodes and edges. This toolset contains the following tools: AutoMap for extracting networks from texts, ORA for analyzing the extracted networks, and DyNet for what-if reasoning about the networks (see figure 1). Each of these tools are described in turn.

AutoMap is a semi-automated Network Texts Analysis (NTA) tool for extracting network data from texts (CMU: <http://casos.cs.cmu.edu/projects/automap/>, Diesner and Carley, 2004; 2005). NTA is a specific text analysis method that encodes the relations between words in texts and constructs a network of the

linked words (Popping, 2000). In AutoMap we technique is based on a distance based approach also referred to as windowing (Danowski, 1993). Windowing basically slides a fictitious window over the text and words within the size of that window are linked together if they match the coding rules specified by the analysts (ref Carley). It has been shown in previous research how map analysis (Carley, 1997; Carley and Palmquist, 1992) and its implementation in AutoMap (Diesner and Carley, 2004; 2005) can be applied to systematically extract links between words in texts in order to model the author's "mental map" as semantic networks. Since we implemented the meta-matrix model into AutoMap as a general ontology for classifying concepts as entities of the meta-matrix, adding meta-matrix text analysis as a further type of NTA to AutoMap, the software supports the extraction of the structure of organizations such as covert networks from text collections social and organizational systems (Diesner and Carley, 2004; 2005). The tool also facilitates the comparison of maps generated with AutoMap and the fusion of the networks per texts into a network that represents the structure of a system reflected in a corpus.

ORA is a statistical toolkit for analyzing dynamic network data composed of multiple entities and relations (CMU: <http://www.casos.cs.cmu.edu/projects/ora/>, Carley and Reminga 2004; Kamneva and Carley, 2004). ORA facilitates analyzing the entire meta-network with a series of measures that have been found to be highly valuable in both the command and control and counter-terrorism contexts (Carley, 2004). The metrics in ORA were developed by drawing on state of the art research in organization theory, social networks, communication theory, operations research, economics, and computer science. ORA takes meta-matrix data and generates a series of reports that can be used to identify key actors or organizations, evaluate their sphere of influence and locate who influences them, and identify vulnerabilities in the overall structure of the meta-network for the group. In addition, ORA makes it possible to compare and contrast two different networks and to estimate possible relations between actors based on factors such as relative similarity and expertise. To aid the analyst, in ORA metrics are organized by functions. A report is generated for each function in which figures and all associated metrics are provided. Illustrative reports include:

- Intelligence Report: identifies key actors – individuals and groups – who by virtue of their position in the network are critical to its operation.
- Management Report: identifies over- and under-performing individuals and assesses the state of the network as a functioning organization.
- Context Report: compares measured values against various stylized forms of networks in an effort to characterize the networks topology.
- SubGroup Report: identifies the subgroups present in the network using various grouping algorithms.
- Sphere of Influence Report: for the node in question, identifies the set of actors, groups, knowledge, resources, etc. that influence and are influenced by that node.

DyNet is a multi-agent network simulation package for assessing network change under various conditions of information assurance (CMU: <http://casos.cs.cmu.edu/projects/dynet/>, Carley, 2004). DyNet is built on-top of the Construct simulation engine (CMU: <http://casos.cs.cmu.edu/projects/construct/>, Carley, 1990; 1991; Schreiber and Carley, 2004) Using DyNet the analyst is able to assess the evolution of the network in general or subject to different strategic interventions. The basic engine evolves the network in response to agent interaction and the exchange of information. Two basic mechanisms underlie this diffusion process. The first mechanism is relative similarity whereby actors are more likely to exchange information if they are comfortable interacting with each other and share culturally relevant factors in common. The second mechanism is relative expertise whereby actors are more likely to exchange information if one actor seeks out the other in search of particular information. DyNet can be called directly by ORA in the "Near Term" report.

DyNetML, an XML based interchange language for relational data (CMU: <http://www.casos.cs.cmu.edu/projects/dynetml/>, Tsvetovat, Reminga and Carley, 2004). By using DyNetML as a unified interchange language other tools such as UCINET (Borgatti, Everett and Freeman 2002) can be linked in and data can be easily exchanged. AutoMap exports the coded text in DyNetML. ORA imports and exports meta-network data, and does so in a variety of formats, including DyNetML. We note that for extremely large datasets, an XML inter-change language is unwieldy. Hence, sparse matrix representation schemes, such as DL, are also used. DyNet can also read and write network data in DyNetML.

3.0 APPLICATION TO TERROR NETWORKS

Thousands of open source documents, such as newspaper accounts and web-pages, were scraped and then processed with AutoMap to create a database of information on the actors, organizations, resources, events, and locations. These documents were selected to span the past decade and to cover information discussing events in the mid-east with a focus on Iraq and al-Qaeda. Additional selected documents, approximately 10 per actor, were then located for those actors who were of particular interest in the media such as bin Laden, Zawahiri and Atta.

These documents were then processed with AutoMap. The following entity classes were extracted: agents, knowledge, resources, events, tasks, and locations. Then additional manual cleaning was done to identify the nationality of the various agents. Before the analysis, known U.S. citizens and leaders or political elite of coalition countries were dropped. The goal was to create a reasonable data set of the “red force.” From this the data associated with al-Qaeda was extracted. Clearly additional cleaning could be done. Further, the data could be broken down in to time periods where those individuals, organizations and events that are not active in that period were removed. This was done; however, there are likely to be errors as full data was not available on all actors. Thus the results should be viewed more as illustrating the tools than as providing any details on terror networks.

The resultant data was then analyzed with ORA. The network that emerges shows a decidedly cellular structure with 5-12 members per cell. Drill-down visualizations on this structure reveal the changing roles of key actors. Over time, results show a decrease in the density and communication levels in the terror network but an increase in its congruence. This suggests a movement to a more distributed and more efficient organizational form, possibly with larger cells. Course of action evaluation shows that removal of highly central actors in the communication network will be less effective than removal of key emergent leaders, even though from a purely visualization stand-point the central actors would be thought to be more critical.

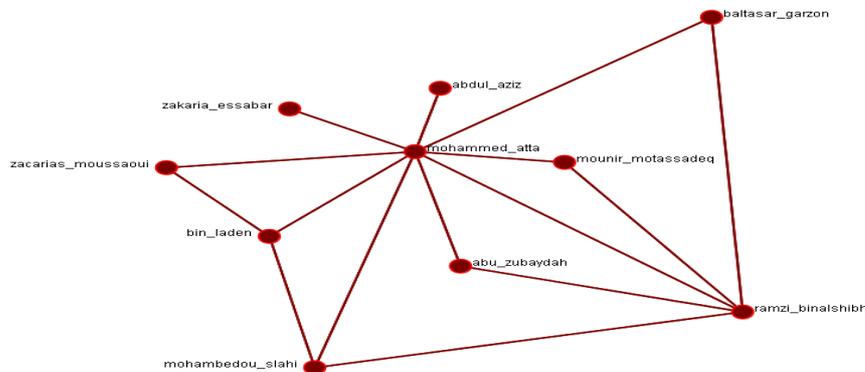
Another result is that actor’s that are critical as emergent leaders are not necessarily the same as those that are the most likely to have access to information or to bind groups together. Table 1 shows an illustrative results for one time period. Additional types of analyses are possible.

Table 1. Key Actors			
Actor Rank	Emergent Leaders	In the Know	Binds Groups Together
1	Riduan Isamuddin	Ramzi Binalshibh	Osama bin Laden
2	Bakar Bashir	Abu Zubaydah	Mohammed Atta
3	Baltasar Garzon	Mohammed Atta	Ramzi Binalshibh
4	Adil Charkooui	Riduan Isamuddin	Abu Zubaydah
5	Zacarias Moussaoui	Osama bin Laden	Ayman al Zawahiri

Essentially, ORA lets the analyst identify different kinds of “hot spots” in the human terrain. Depending on the goal of the operation, different actors may be considered. For example, informants may be chosen from the in the know category; whereas, removal of emergent leaders are most likely to disrupt a cellular organization.

Once a node of interest is identified it might be useful to examine what influences or is influenced by that node. This is done in ORA using the sphere of influence report. This report identifies those actors, event’s, resources, and so on that are associated with the nod of interest. Figure 2 shows the sphere of influence around Mohammed Atta where only the actors are shown. This report also identifies the other node whose sphere is most like the node of interest. In this case, that is the actor Adil Charkaoui. Further analysis reveals that Atta can leverage about 39% of the resources in this data set, but is most likely to be constrained by bin Laden, Moussaoui and Aziz.

pacrim2_aq2



powered by ORA (2006) CASOS Center © CMU

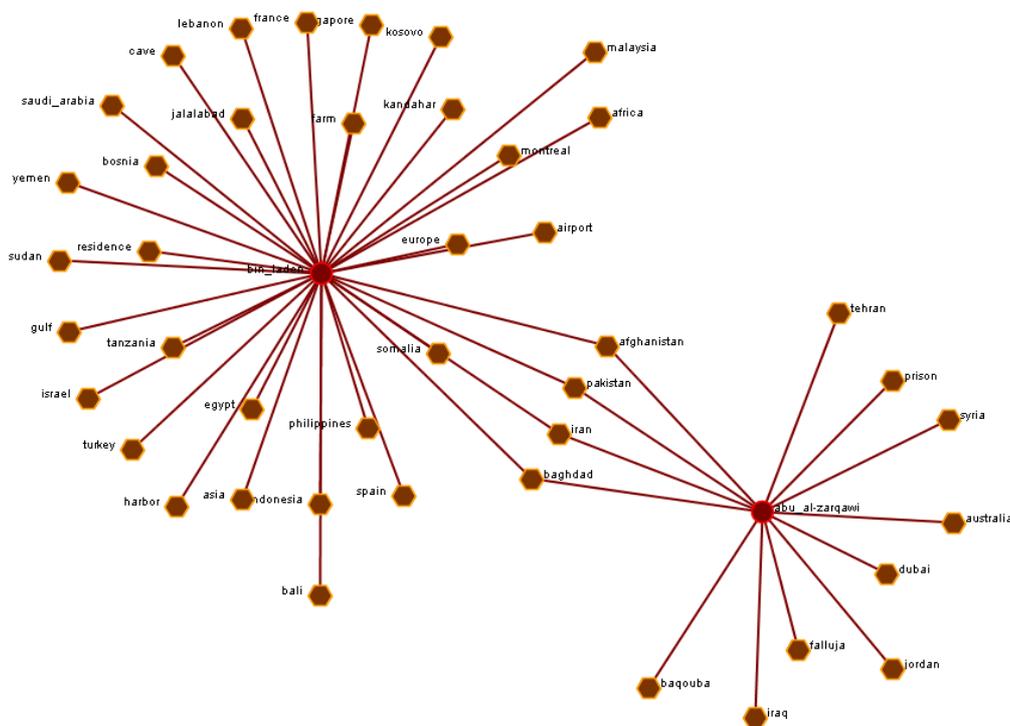
Figure 2. The actor’s who influence and are influenced by Mohammed Atta.

Figure 2 is the sphere of influence level 1 for with Atta as the focal node. It is reasonable, however, to think about ever widening spheres of influence. That is we can ask who does an actor directly and indirectly influence. Thus, the sphere of influence in ORA is defined by the “depth”; i.e., how far from

the focal node the analyst wants to look. Level 1, is simply all nodes directly connected to the focal node (path length is 1). Level 2, includes all nodes that are at most a path length of 2 from the focal node. For example, in the actor-to-actor network this would include ego and those alters that ego is connected to and those in the alter’s immediate network; i.e., my friends and my friends, friends. Level 3, includes all nodes that are at most a path length of 3 from the focal node, and so on.

Location based analyses are also possible. For example, we might want to know what locations saw the most activity (Baghdad). We might also want to know who has been seen in the most locations – Abu al Zargawi and Osama bin Laden. We can see, in Figure 3, that there are a few locations where both have been.

pacrim2_aq2



powered by ORA (2006) CASOS Center @ CMU

Figure 3. Locations associated with Zarqawi and bin Laden

As noted, these tools consider more than just the social network, e.g., they also consider information on the locations, events, resources and so on. As such, social theory about why people interact can be brought to bear to estimate possible missing data. For example, one of the most dominant characteristics of interaction is homophily – the tendency of those who interact to be similar. Based on this tendency, we can estimate who has a high level of homophily yet is not interacting (maybe intelligence missed observing a connection?) or a low level of homophily yes is interacting (possibly suggesting a hidden dimension of interest or a misinterpretation of proximity). In this data set, for example, we see no connection between Rahim al Nashir and Mohamed Mahjoub; however, we would expect there to be one based on homophily.

Once possible nodes of interest are identified, DyNet can be called directly from ORA by using the near term impact report. Here the user specifies the length of the run, which nodes to remove when, and the number of replications. In other words, a virtual experiment is set up. The system then runs that experiment and sends back information on the impact of the removal of said node on the diffusion of information and the expected performance of the network. Diffusion is measuring the general ability of the network to move information. Performance is an ensemble measure estimating the general tendency of the network to make mistakes. Exactly what information will or won't diffuse and what tasks will or won't be done are not forecast.

For example, imagine that an analyst ran an experiment where the baseline was run (no node removal) for five time periods. Then an additional three runs were done such that at time two one of Riduan Isamuddin, Ramzi Binalshibh, or Osama bin Laden is dropped from the network. Illustrative results are shown in Figure 4 (information diffusion) and Figure 5 (performance). These results show that the removal of different actors have very different effects. For example, removal of the top individual who is "in the know" is more likely to speed the rate of information diffusion (actor may have been a bottle neck) and improve overall performance of the network. Removal of the top emergent leader degrades performance the most; whereas, the removal of bin Laden degrades communication more than general performance.

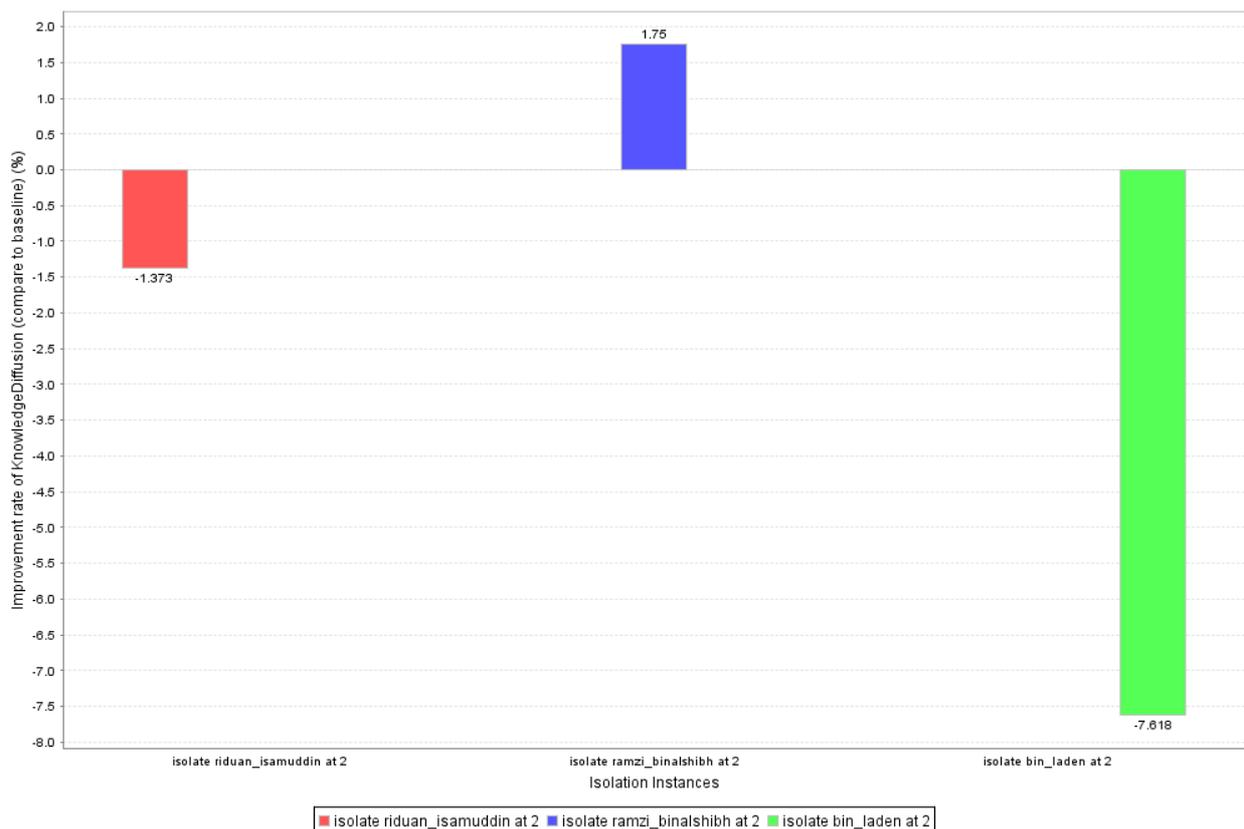


Figure 4. Impact of Isolation Strategy on Information Diffusion in the Network

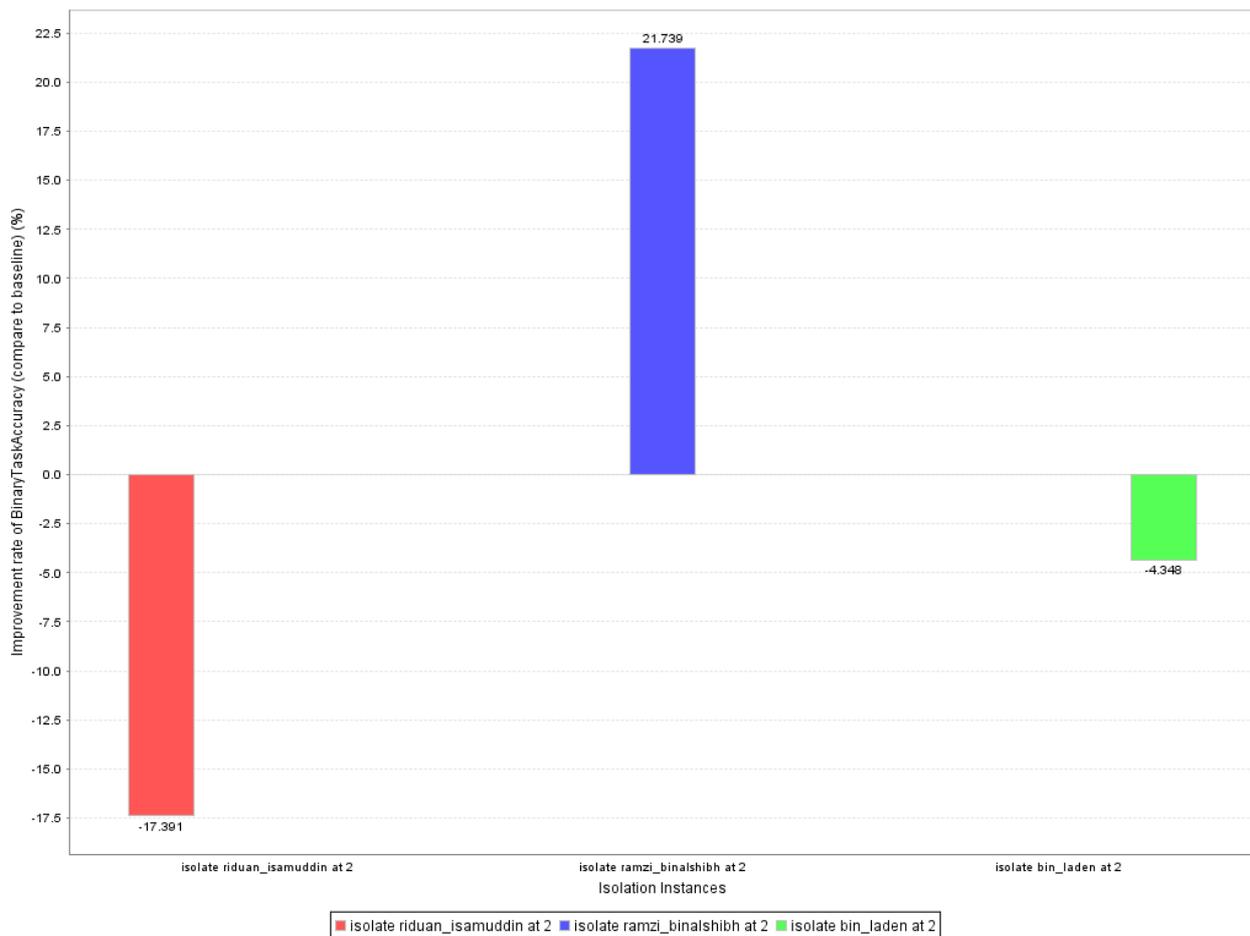


Figure 5. Impact of Isolation Strategy on Expected Performance in the Network

4.0 SUMMARY

Dynamic network analysis (DNA) is an emergent field centered on the collection, analysis, understanding and prediction of dynamic relational data such as who communicates with whom and who knows what. A DNA perspective moves beyond standard social network analysis by focusing not just on who interacts with whom, but also the relations of actors to other entities such as knowledge, resources, tasks, locations and organizations. By combining techniques and ideas from statistics, computer science and organizational theory an integrated tool chain for the extraction, analysis and prediction of relational data is possible.

Network oriented toolchains facilitate better analysis by reducing the time spent in repetitive tasks where little analyst insight is needed. To further support the analyst, however, additional effort needs to be invested in building parsers between the various databases, enabling automated data collection, and identifying common analyses and providing those in a single automated report. With this in mind, it is critical that future generations of DNA tools take into account factors such as the confidence in the data, automated estimates of robustness, data-farming, and tools for user-in-the-loop data testing.

The approach used herein affords analysts with greater analytical power and facilitates assessing large, dynamic and complex socio-technical systems such as terror networks. By taking into account not only the networks among people and organizations but also their relations with resources, knowledge, events, and

locations etc., insights into diverse behaviours are gained. If we look only at the social network then the focus of attention is on hierarchies, communication and other social relations. The addition of events and locations facilitates course of action analysis and enables linkage to various strategic planning tools. The over-arching meta-network approach used in DNA has the promise of enabling effects based operation in areas as diverse as diplomacy, information, military and the economy to be assessed in a relational context.

5.0 REFERENCES

- [1] Borgatti, Everett and Freeman 2002
- [2] Carley, K.M. 1997. Network Text Analysis: the network position of concepts. In Carl W. Roberts (Ed.), Text analysis for the social sciences, (pp. 79-102). Mahwah, NJ: Lawrence Erlbaum Associates.
- [3] Carley, K.M., and Palmquist, M. 1992. Extracting, Representing, and Analyzing Mental Models. *Social Forces*, 70 (3), 601-636.
- [4] Carley, Kathleen M. and Jeff Reminga, 2004. ORA: Organization Risk Analyzer. Carnegie Mellon University, School of Computer Science, Institute for Software Research International, Technical Report CMU-ISRI-04-101.
- [5] Carley, Kathleen M. 2003, "Dynamic Network Analysis" in Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers, Ronald Breiger, Kathleen Carley, and Philippa Pattison, (Eds.) Committee on Human Factors, National Research Council, National Research Council. Pp. 133-145.
- [6] Carley, Kathleen M. 2004, "Estimating Vulnerabilities in Large Covert Networks" In Proceedings of the 9th International Command and Control Research and Technology Symposium. Conference held at Loews Coronado Resort, CA. Evidence Based Research, Vienna, VA.
- [7] Carley, Kathleen M., 1990, "Group Stability: A Socio-Cognitive Approach." Pp. 1-44 in Lawler E., Markovsky B., Ridgeway C. and Walker H. (Eds.) *Advances in Group Processes: Theory and Research* . Vol. VII. Greenwich, CN: JAI Press.
- [8] Carley, Kathleen M. 1991. "A Theory of Group Stability." *American Sociological Review*, 56(3): 331-354.
- [9] Danowski, J. 1982. A network-based content analysis methodology for computer-mediated communication: An illustration with a computer bulletin board. In R. Bostrom (Ed.), *Communication Yearbook*, (pp. 904-925). New Brunswick, NJ: Ransaction Books.
- [10] Diesner, Jana and Kathleen M. Carley, 2005, "Revealing Social Structure from Texts: Meta-Matrix Text Analysis as a novel method for Network Text Analysis," In V.K. Narayanan and D.J. Armstrong (Eds.) *Causal Mapping for Information Systems and Technology Research: Approaches, Advances, and Illustrations*, Chapter 4, Harrisburg, PA: Idea Group Publishing.
- [11] Diesner, Jana and Kathleen, M. Carley, 2004. AutoMap1.2 - Extract, analyze, represent, and compare mental models from texts. Carnegie Mellon University, School of Computer Science, Institute for Software Research International, Technical Report CMU-ISRI-04-100.
- [12] Kamneva, Natasha and Kathleen, M. Carley, 2004. A Network Optimization Approach for Improving Organizational Design. Carnegie Mellon University, School of Computer Science, Institute for Software Research International, Technical Report CMU-ISRI-04-102.
- [13] Popping, R. 2000. Computer-assisted Text Analysis. London, Thousand Oaks: Sage Publications.

- [14] Saferstein, R.E. (2001) *Forensic Science Handbook*, Prentice Hall, NJ.
- [15] Schreiber, Craig and Kathleen M. Carley, 2004. Construct - A Multi-agent Network Model for the Co-evolution of Agents and Socio-cultural Environments Carnegie Mellon University, School of Computer Science, Institute for Software Research International, Technical Report CMU-ISRI-04-109.
- [16] Thelwall, M., 2004, *Link Analysis: An Information Science Approach*. Academic Press.
- [17] Tsvetovat, M., Reminga, J., & Carley, K.M. (2004). DyNetML: Interchange Format for Rich Social Network Data. Carnegie Mellon University, School of Computer Science, Institute for Software Research International, Technical Report CMU-ISRI-04-105. URL: <http://reports-archive.adm.cs.cmu.edu/anon/isri2004/abstracts/04-105.html>
- [18] Wasserman S. and K. Faust, 1994, *Social Network Analysis*. New York, Cambridge University Press.