CHAPTER XX

Dynamic Network Analysis (DNA) and ORA

Kathleen M. Carley, Jürgen Pfeffer

Carnegie Mellon University
Pittsburgh PA, USA
kathleen.carley@cs.cmu.edu, jpfeffer@cs.cmu.edu

ABSTRACT

Dynamic network analysis can be used to assess complex socio-cultural systems from a network perspective. Key elements of this approach include: 1) dynamic meta-network representation of the who, what, where, how, why; representation of data from both a trail and network perspective where the nodes are attributed and the links probabilistic; extension of social networks analytics to the geo-spatial network analytics; techniques for assessing and forecasting change in networks; infrastructure tools to support data extraction, analysis and forecasting ranging from machine-learning models for network extraction to agent-based models for assessing the impact of the co-evolution of networks in various domains on human socio-cultural behavior.

Keywords: social networks, dynamic networks, geo-spatial networks

1 INTRODUCTION

Reasoning about human socio-cultural behavior requires reasoning about individuals in context. In other words, it requires understanding not just the people (who), but answering the classic journalistic queries of who, what, why, where, how and when. While social network analysis supports identification of critical actors and so answering who is important and to an extent why, it does not support the full assessment. For that, dynamic network analysis is needed. That is, dynamic network analysis places the social network within context.

Two factors underlie dynamic network analysis. First the use of both meta-network and trail representation. Second the use of an interoperable tool suite that supports moving from text to network and visual analytics to simulation and forecasting. At the heart, is the network and so the techniques for assessing and visualizing this network – and for this ORA can be used.

2 META-NETWORKS AND TRAILS

One of the key insights underlying dynamic network analysis is that networks are samples of relations formed through the intersection of individual trails. A trail is the path that an individual follows over their life course expressed in terms of who was where when doing what with whom, why and how. Within any trail there are sets of interaction events, such as Joe sends email to Helen. By specifying a slice through defined in terms of a range on a time window, a geo-spatial window, and a set of actors a network can be identified; e.g., the sets of relations that can be extracted by considering all employees of the World Bank, in Afghanistan, during 2011. Dynamic network analysis uses the duality of trails and networks to generate novel grouping algorithms (FOG), trails assessments of changes in networks, and algorithms for visualizing and grouping trails based on meta-network node linkages.

Most socio-cultural systems can be represented as meta-networks linkage Agents and Organizations (Who), Resources and Knowledge/Expertise (How), Tasks/Activities and Events (What), Beliefs (Why), Locations (Where) through time (When). These ontological classes provide a way of categorizing and segmenting nodes. Nodes in a network are instances of an ontological class; e.g., Joe and man are nodes in the Agent ontology class. Most analysts find it beneficial to have a second level in this ontology segmenting agents, organizations, locations and events into specific entities and generic entities. Trails weave patterns between any three entity classes; however, the typical trails involve who, how, or what by where and when.

Given two ontology classes there exists one ore more networks composed of the relations of this type between the nodes in this ontology class. Some of these networks are uni-modal such as social networks connecting agents to agents. Whereas, other networks are bi-partite connecting nodes in one ontology class to nodes in another ontology class such as an attendance network indicating who attended what event. For dynamic network analysis the power comes from being able to extract, assess and forecast change in both types of networks at the same time, and in exploiting the constraints implied by the entire meta-network.

3 METRICS

ORA contains over 150 metrics for assessing various aspects of networks and several dozen algorithm based tools to support various activities such as finding local patterns of interest, comparing networks, and characterizing groups. The metrics include both the standard social network metrics, bi-partite and multi-network metrics. Standard network metrics such as degree centrality, betweenness centrality, closeness and eigenvector centrality are useful for identifying nodes with undue structural influence. Bi-partite metrics, particularly those assessing specialization and redundancy are useful for characterizing the roles of agents or organizations relative to their environment. For example, imagine a people by knowledge network. In this cases agents who are specialists have exclusive or near exclusive access to knowledge other agents do not; whereas, redundancy captures the extent to which there are many individuals with access to the same areas of expertise.

Multi-network metrics such as congruencies, dependencies and load are useful for assessing key performance related aspects of the overarching system. Congruency metrics are measures of fit. They assess the extent to which there is a fit between who or what is needed and who or what is assigned or available. In general, the higher the congruence the better the overall performance of the system. Cognitive demand measures, from a meta-network perspective, the relative “business” of the agent by assessing the extent to which the agent has many others they interact with, many tasks, complex tasks, many resource to juggle, many others that they need to coordinate with and so on. In general, cognitive demand is a reasonable indicator of emergent leadership; not who will become a formal leader, but who is likely to be directing traffic behind the scenes. Finally, the various loads, such as work-load are alternative measures of overall system performance – as assessed from a network perspective.

4 CORE TECHNOLOGIES

As noted, dynamic network analysis is supported through a suite of interoperable tools. This is shown in Figure 1. Additional tools can be added and old ones dropped as technologies evolve. Each of the core technologies can be used alone or as part of an overarching process. Data extraction and cleaning can be done for raw text by using a combination of AutoMap and ORA. Analysis and forecasting of change in networks can be accomplished by using ORA and Construct. ORA, the network analysis engine is at the heart of the dynamic network analysis process. The tools described here support analysis of networks ranging in size and scope from a few nodes to 10^6 nodes per ontology class. Many of the same metrics work on much larger data sets; however, that is not the focus of these technologies.

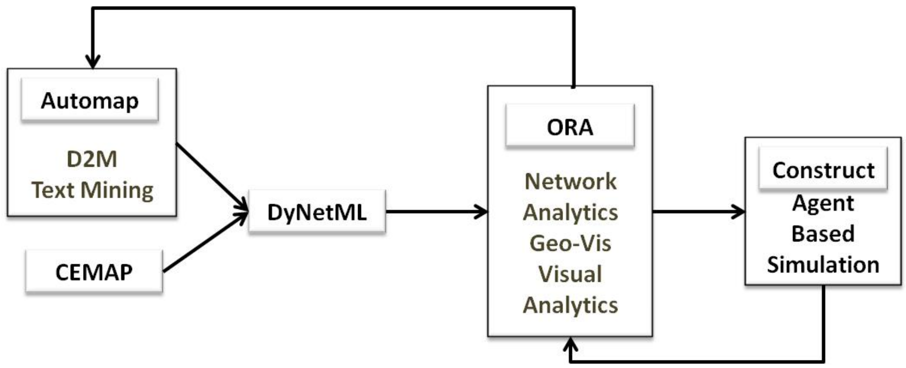

Figure 1 Illustrative interoperability of Core Technologies Supporting Dynamic Network Analysis

AutoMap (CMU: http://casos.cs.cmu.edu/projects/automap/, Carley et al, 2011a) is a mixed-initiative system for the extraction of nodes and relations from raw unformatted texts. Using advanced machine learning techniques and basic thesauri construction techniques AutoMap can be used to support content analysis (extraction of concepts and frequencies), semantic network analysis (extraction of network of concepts), dynamic-network analysis (extraction of ontologically cross-classified nodes and relations), and aspects of sentiment analysis. AutoMap uses a windowing process for link identification (Danowski, 1993) in conjunction with conditional random fields (Diesner and Carley, 2004) for concept classification and thesauri application. Specialized features include sub-tools for extraction of email and phone numbers and post-processers to augment the extracted data with the latitude and longitude of key locations. AutoMap can be used in batch or GUI mode; however, for those unfamiliar with text processing, the best mode is through the data-to-model (D2M) process in which AutoMap and ORA are used together in a predefined and optimized sequence to clean and structure the extracted meta-network data. In general, the techniques for extracting and classifying agents, organizations, and locations are more accurate than those for knowledge, resources, tasks and beliefs.

CEMAP is a mixed-initiative system for the extraction of nodes and relations from semi-structured texts, such as blogs or email. Specialized parsing tools admit the extraction of meta-networks from header data. For most semi-structured data only the agents, organizations, and tasks are extracted.

ORA (CMU: http://casos.cs.cmu.edu/projects/ora/, Carley et al, 2011b) is a powerful network analysis tool, capable of handling large 10^6 networks, and supporting meta-network data, geo-spatial network data, and dynamic network data. ORA is capable of generating over 150 metrics including all standard social network metrics for uni-modal and bi-partite networks and specialized metrics for multi-mode data including measures of loads and demands such as cognitive demand.. Relatively unique features include trail, network, and geo-network visualization, classical and fuzzy grouping algorithms, multi-mode network assessment, built in network simulators, and special reports for semantic network data. User guides, tool tips and integrated help support the user. ORA can import and export data in a large number of formats including direct imports for CSV and UCINET and export of images in png, jpg, pdf, and svg.

Construct (CMU: http://casos.cs.cmu.edu/projects/construct/, Carley, 1990, 1991; Carley, Martin and Hirshman 2009) is an agent-based dynamic-network model for assessing the co-evolution of social and knowledge networks through fundamental learning, information diffusion and belief dispersion processes. Using Construct the impact of various interventions can be assessed at the individual, group or network level under alternative communication media environments. Interaction logics based on homophily and expertise seeking govern individual choice of interaction partners, and knowledge masks accounting for general and work related knowledge and beliefs impact message construction, while forgetting, attention limits and social interaction spheres are used to characterize the boundedly rational agent. Construct gains its power for evolving change in the networks by accounting for the influence of bi-partite networks in constraining the development of the uni-modal networks.

DyNetML is an XML based language for the interchange of relational data (CMU: http://casos.cs.cmu.edu/projects/dynetml/ , Tsvetovat, Reminga and Carley, 2004). Developed as a way to extend graphml, DyNetML supports the representation of geo-temporal meta-network data, meta-data, and attributes on nodes and relations. DyNetML is used to exchange data between AutoMap and ORA with AutoMap exporting DyNetML and ORA importing and exporting DyNetML.

5 ILLUSTRATION OF CRITICAL NETWORK ANALYTIC CAPABILITIES

Critical network analytic capabilities include key entity identification, visual analytics, group identification, path finding, dynamic analytics, geo-network analytics, and forecasting. Aspects of these techniques are illustrated using data gathered on Afghanistan. This is not a complete cataloguing of all capabilities, but a taste of the overall system.

Open source data on Afghanistan were extracted from Lexis-Nexis and others provided as part of an overarching HSCB SNARC effort. The resulting 282,000 texts covered the period December 1999 to August 2011 were processed by AutoMap and a full meta-network extracted. Cleaning resulted in a merging of alternate spellings of the same words, and collapsing. The resulting data is a set of meta-networks one per year for multiple years. Over time, the networks became increasingly dense. Americans and friendly foreign leaders were removed.

Key entity identification was used to find those agents, organizations, locations, and so on that stand out on one or more network dimensions. For agents, for example, metrics indicating the criticality of the individual in the social network are used such as degree, betweenness, eigenvector and so on. A composite look across these metrics suggests that those individuals who are important on many dimensions include key leaders such as Hamid Karzai – see Figure 2. Top ranked key agents are often such leaders. This assessment is done using the key entity report in ORA.

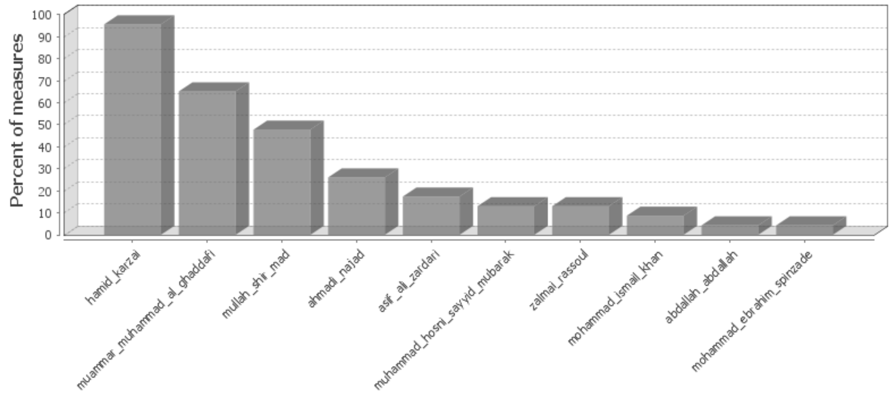

Figure 2. Most Recurring Agents as Key on Numerous Network Metrics

In large networks, these top ranked individuals may be less interesting than the "seconds", those individuals who are not top but still critical. Such individuals might be "the power behind the throne" or at least "actors of interest" to watch. In Afghanistan two of these "seconds" are Abdul Rasid Dostum and Mohammad Ismail Khan. For such individuals, the analyst might want to explore their sphere of influence. The sphere of influence is an extension of the of the ego-net concept to the meta-network. It is the set of entities regardless of ontology class that are directly connected to ego and the connections among them. Khan has a more elaborated sphere of influence in this data than does Dostum indicating a higher level of integration and ability to control more resources. These spheres of influence can be found directly in the visualizer, or through using the sphere of influence report in ORA.

Over time, the power of various agents can increase and decrease. In Figure 3 the over time profile of Dostum and Khan is shown for the network metric cognitive demand. Individuals who are high in betweenness are thought to be more influential. In this case, the analyst has selected to use ORA's over time charting capability to examine the over time changes in the betweenness of Dostum and Khan. After their appointment to the ministry their influence dropped, which was the intent of the appointment. Dostum's further dropped as he left Afghanistan after the Akbar Bai incident, rising only briefly when he returns to support Karzai's re-election. In contrast, Khan's influence is seen to be rising steeply in recent years.

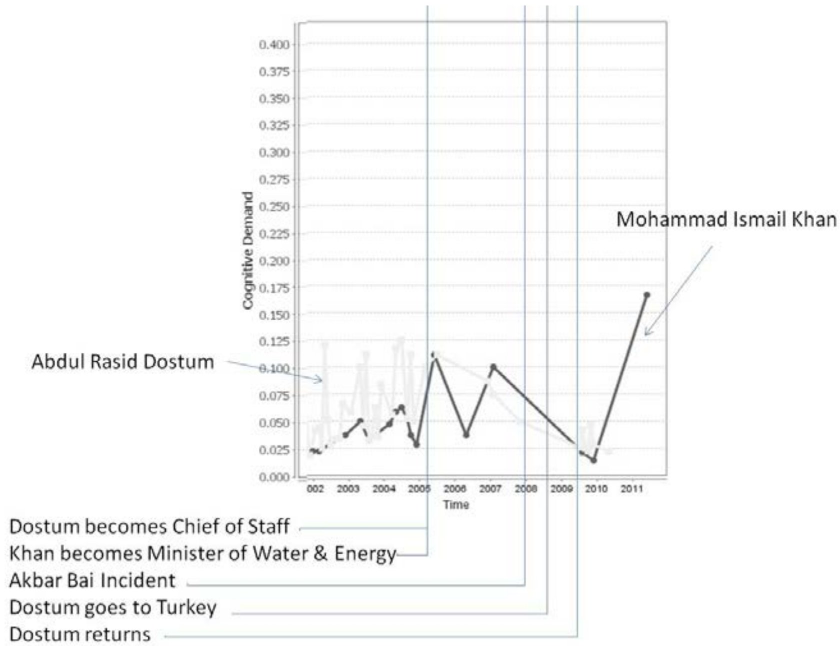

Figure 3. Dynamics of the Betweenness of Dostum and Khan

Spatially, a key issue is what is the region where an actor has influence. For this, we can examine the polygon within Afghanistan that contains the activities of the agents of interest. ORA supports geo-spatial visualization and exports using any of NASA World Wind, ArcGIS, or Google Earth. Figure 4 shows the region of influence calculated for Dostum and Khan. As can be seen, both cover most of Afghanistan, which as former warlords and then ministers in the Karzai government is not too surprising.

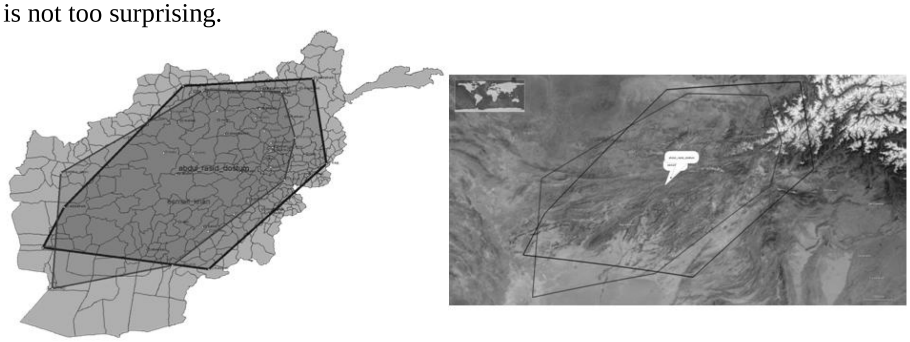

Figure 4. Regions of Influence for Dostum and Khan

Moving on to the network, we might ask where are the agents? Using the agent by location subnetwork ORA can color the regions by frequency and overlay the network. Here we see in Figure 5, that the density of agents is highest near the areas of Kunduz and Kandahar.

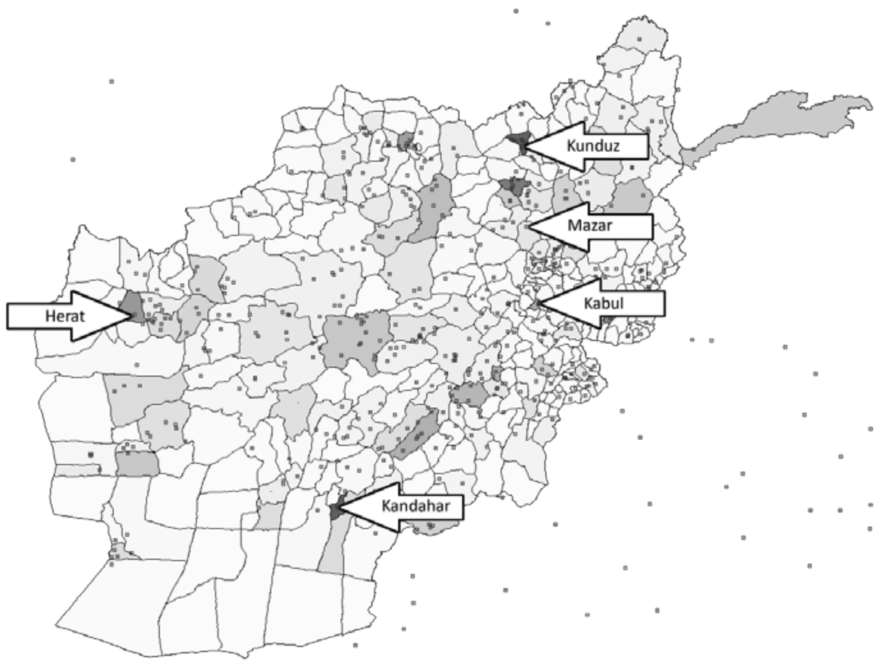

Figure 5. Regions of Concentration of Activities by Political Elite

Analyses such as those just described characterize and provide an in-situ understanding of the existing socio-cultural environment. To go beyond this is to engage in forecasting. ORA provides two distinct types of forecasting technology. The first, immediate impact, enables the analyst to examine the immediate impact on a network or meta-network of removing one or more nodes or links. Figure 6 shows the impact on Karzai's power when Dostum and Khan are removed. As can be seen, these agents are to an extent holding Karzai in check and the removal of either agent strengthens Karzai's power base. However, networks can heal themselves. That is, individuals interact and in doing so build new connections and maintain existing one. Agent-based dynamic-network modeling, via Construct, is used to assess the changes due to this "healing." Through ORA Construct can be called by using the Near Term Impact report.

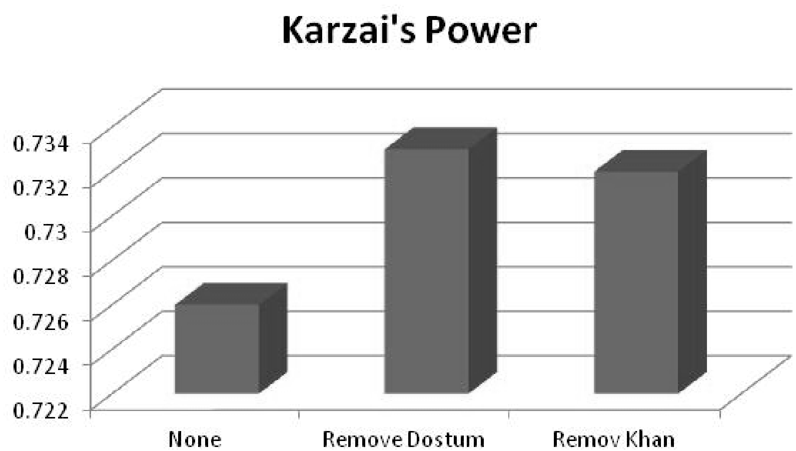

Figure 6. Impact on Karzai's Power if Course of Action Removing Either Dostum or Khan is Followed

6 CONCLUSIONS

Dynamic network analysis is a valuable approach for understanding human socio-cultural behavior and for forecasting the space of future possibilities. Importantly, these techniques can be used to reason about both own and adversarial groups. These techniques can be used to identify weaknesses and strengths in any group, and to suggest ways of overcoming those. The breadth of application is one of the core strengths of this approach.

Core challenges, that if solved would further increase the utility of these techniques include automatic interpretation of metrics; auto-identification and visualization of critical features, and improved data extraction and fusion techniques. Key advances supporting improved dynamic and spatial analysis are likely to occur in the near future. Even without meeting these challenges, ORA and Dynamic Network Analysis can support overall reasoning about individuals, groups and their activities, course of action analysis, and improve overall situation awareness. Simply moving to meta-network reasoning and the ability to focus on geo-spatially embedded networks and dynamic networks improves the analysis and enables a more contextual based understanding.

ACKNOWLEDGMENTS

This work was supported in part by the Office of Naval Research – ONR N000140811223 (SORASCS), N000140811186 (Ethnographic), the AFOSR FA9550-05-1-0388. Additional support was provided by the center for

Computational Analysis of Social and Organizational Systems (CASOS). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Office of Naval Research, the Air Force Office of Scientific Research, the Department of Defense or the U.S. government.

REFERENCES

Carley, K. M. J. Reminga, J. Storrick and D. Columbus, 2011b, "ORA User's Guide 2011," Carnegie Mellon University, School of Computer Science, Institute for Software Research, *Technical Report*, CMU-ISR-11-107.

Carley, K. M., 2002, Smart agents and organizations of the future. In L. Lievrouw & S. Livingstone (Eds.), *The Handbook of New Media: Social Shaping and Consequences of ICTs* (pp. 206–220). Thousand Oaks, CA: Sage.

Carley, K. M., D. Columbus, M. Bigrigg and F. Kunkel, 2011a, "AutoMap User's Guide 2011," Carnegie Mellon University, School of Computer Science, Institute for Software Research, *Technical Report*, CMU-ISR-11-108.

Carley, K.M. 2011. Dynamic network analysis and modeling for CANS. In Belinda Bragg and George Popp (eds.), A Guide to Analytic Techniques for Nuclear Strategy Analysis. NSI, inc.

Carley, K.M., M. K. Martin and B. Hirshman, 2009, "The Etiology of Social Change," Topics in Cognitive Science, 1.4:621-650.

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. Greenwhich, CN: JAI Press.

Carley, Kathleen M.1991. "A Theory of Group Stability." *American Sociological Review*, 56(3): 331-354.

Diesner, J. and K. M. Carley, 2008, "Conditional Random Fields for Entity Extraction and Ontological Text Coding," *Computational and Mathematical Organization Theory*, 13.3: 248-262.

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.

Pfeffer, J. & K. M. Carley, Forthcoming, Rapid Modeling and Analyzing Networks Extracted from Pre-Structured News Articles. Computational and Mathematical Organization Theory.

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.

Wasserman, S., & Faust, K., 1994, *Social Network Analysis: Methods and Applications*: Cambridge University Press.