

Dynamic Networks: Rapid Assessment of Changing Scenarios

Nadya Belov[†], Michael K. Martin^{*}, Jeff Patti[†], Jeff Reminga^{*}, Angela Pawlowski[†] and Kathleen M. Carley^{*}

[†]<nbelov, jpatti, apawlows>@atl.lmco.com, Lockheed Martin Advanced Technology Laboratories, Cherry Hill, New Jersey

^{*}<mkmartin, reminga, kathleen.carley>@cs.cmu.edu, Center for Computational Analysis of Social and Organizational Systems, Institute for Software Research, Carnegie Mellon University, Pittsburgh, Pennsylvania

Abstract As events unfold, the underlying networks change. Most network science tools, however, assume analysts have a single snapshot of the data, or at most, a second snapshot at a different time. The underlying network representation schemes, assessment technologies, and visualizations do not lend themselves naturally to dynamic networks. Herein, we identify key criteria for network representation of dynamic, uncertain information and present a technology enabler that combines information fusion and dynamic network analysis. The value of technological solutions for network analytics in a dynamic environment are discussed.

1 Introduction

New challenges have arisen as network scientists move to apply their analytical techniques to military and other real-world environments. The target environments involve large amounts of relational data that characterize dynamic, often distributed, decision-making situations. These situations require time-sensitive network analyses to support command decision-making as real world events alter the networks and, sometimes, prudent courses of action. Analysts need to capture and assess the impact of these changes, with minimal computational overhead; they require Network Representation Languages (NRLs) that adequately support the continuous ingestion of battlefield data, kinetic and non-kinetic, into dynamic network analysis packages such as ORA [1]. In the following, we identify core requirements for NRLs to meet this need, and sketch the value of rapid network assessment.

We believe the core requirement for continuous network analysis of battlefield data is an NRL that adequately handles dynamic, uncertain meta-network data. For many xml schemes—such as GraphML—it is possible to do all the things we are suggesting in terms of edges and meta-data. However, that does not mean the approach is easy or there is a network analysis tool that can understand that information. Our goal is to

identify what changes are needed to handle real time data, extending DynetML to produce CoreNetML and DyNetML 2.0. We further extend our goal to re-use tools such as ORA, originally design for static analysis, to assist in dynamic network analysis. In this paper we discuss how this is accomplished through CHANS. The majority of well-known NRLs (e.g., CSV, edge-list, UCINET-native [2], and GraphML [3] are very good at representing static networks. The representation of uncertain observations, however, can be difficult if not impossible. Three key improvements are needed to handle the battlespace: Enhanced Edge Representation (EER), Enhanced Node Representation (ENR), and Enhanced Meta-Data Representation (EMR).

In Section 2, we define criteria for network representation of distributed, dynamic, uncertain environments, and provide a high level score card that compares representation capabilities of well-known NRLs, alongside enhanced representations such as CoreNetML—developed by Lockheed Martin Advanced Technology Laboratories (LM ATL) as an extension of DyNetML 1.0 (see [4] for an initial description of DyNetML and <http://www.casos.cs.cmu.edu/projects/dynetml/> for the 1.0 specification). Next, in Sections 3 and 4, we illustrate how enhanced NRLs can enable dynamic battlefield assessment in a notional system where battlefield observations are input to LM ATL’s Core HumAn Network System (CHANS) prototype and updated DyNetML is exported to ORA for rapid analysis of network changes. Section 5 concludes.

2 Criteria for Dynamic Distributed Network Analytics

The dynamic, distributed, uncertain character of real world settings often yields incomplete data. For example, examination of the Tanzania embassy bombing indicates the data arrived piecemeal. A group of terrorists worked together to build, deploy and explode a bomb. Early attention to this group might have revealed some work related to bomb manufacturing, but only as the events unfolded would information start appearing on associated surveillance and transportation activities. Even as data became available, it may not have been reliable or complete because it would have depended on leads from informants, tracking via sensors, and so on.

Data from real situations has these features: 1) one or more sources with characteristics such reliability and pedigree, 2) potential delays in positive identification of unknown nodes (entities), and 3) nodes and edges (relations) that may appear, disappear, or change in strength over time as new information arrives. In addition, the networks of concern are “meta-networks” that are multi-modal (different types of nodes, e.g., people, places, assets), multi-link (different types of relations),

Dynamic Networks: Rapid Assessment of Changing Scenarios

and multi-level (multiple types of networks, e.g., variations in the granularity of representation such as individuals versus groups).

To analyze such data, a simple solution is to divide it into time periods and then assess a sequence of networks, each defined at a particular level of reliability. However, from a computational perspective such a solution does not scale. Computationally, the key criteria are: 1) scalability, 2) minimization of data IO costs, 3) facilitation of temporal analysis, and 4) facilitation of sensitivity analysis.

Real-world data characteristics and computational criteria imply a need for attributed data. Three types of attribute information need to be represented:

- 1) Edge Attributes: Details including strength, direction, reliability, existence, and change in these factors (DyNetML 1.0 with EER).
- 2) Node Attributes: Details including existence, attributes about the state of the node such as level of education or age, and so on (DyNetML 1.0 with ENR).
- 3) Meta-Data Attributes: Details about who generated the data, why, when, how, coding constraints, information pedigree, credibility, level of classification, etc. (DyNetML 1.0 with EMR).

CoreNetML, the LM ATL enhancement of DynetML 1.0, addresses the features of real-world data outlined above by encoding varying and often conflicting information, along with temporal and geo-spatial markers; encoded confidence values based on source and pedigree identifiers enable de-confliction by outside data normalization services. CoreNetML exemplifies aspects of EER, ENR, and EMR in the representation of real-world data. In a similar vein, a new DyNetML 2.0 with EER, ENR, and EMR as standard features, and a new version of ORA using these features, will be released November 2008.

Table 1 scores common NRLs, alongside enhanced NRLs, in terms of their capability to represent attributed data. For each feature the NRL is rated as $\sqrt{}$ =yes, D=difficult, or U=unknown/to be determined; a blank cell means no. Additionally, $\sqrt{*}$, means that when there are multiple edges between the same pair of nodes, an equivalent way of handling classes of such edges is with attributes on the network for that edge class rather than on the edge itself. This form of representation is more compact.

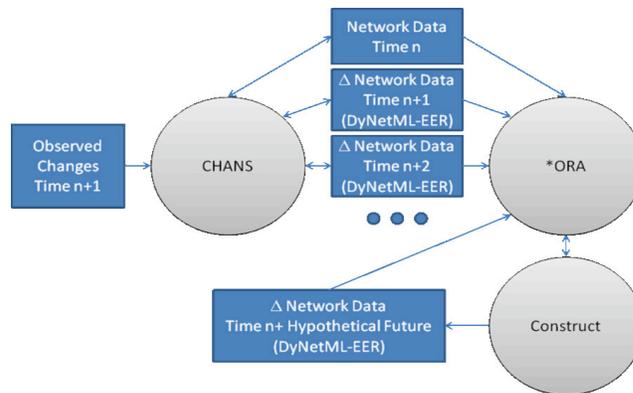
As can be seen in Table 1, the various enhancements each focus on a section of the feature-set of a network. A wide range of capabilities are now enabled by attribute-enhanced representations (e.g., CoreNetML, DyNetML 2.0). In the area of network dynamics, these include, but are not limited to:

- 1) Ingestion and assessment of real time changes.
- 2) Characterization of change in groups by change in members attributes.
- 3) Visualization of and detection of change in networks with minimal IO.
- 4) Tighter link to simulation and forecasting tools as ability to handle evolving networks including node attributes such as age and beliefs are possible.
- 5) Ability to reason about, locate patterns in, and track trail data over time.

3 Enhanced NRLs and Dynamic Battlefield Analysis

Battlefields require timely collection and assessment of new information in the context of prior information. This new information may be confirmations of suppositions, completely new intelligence, reports on the impact of a course of action and so on. From a network standpoint, intelligence changes the network by altering the attributes of nodes or edges and the presence of nodes or edges. Attribute-enhanced NRLs support the tracking of such changes, and open possibilities for dynamic battlefield assessment systems comprised of dynamic network analysis technologies. Information fusion of soft, non-kinetic and potentially conflicting data has only recently been recognized as a core requirement for dynamic network analysis [5]. Figure 1 illustrates such a system using Core HumAn Network System from LM ATL, and ORA and Construct from CMU.

Figure 1. Attribute-enhanced NRLs facilitate network analysis of newly observed changes and prediction of expected changes using simulation tools such as Construct [6].



CHANS is designed to bring together disparate information, data sources and network analysis technologies to produce effective classification and forecasting results. It includes a central data store called the Human Network Core (HNC), where all human network and communication data is persisted, a Data Fusion Service responsible for maintaining data integrity, and Forecasting and Classification Interfaces that provide network analysis tools with access into the HNC.

The HNC is central in the CHANS. It persists data represented in CoreNetML from various sources including, but not limited to imagery, human generated reports and open data sources such as blogs and news feeds. The Data Fusion Service addresses significant hurdles to effective network analysis of evolving battlefield situations—data normalization—by assessing confidence in the accuracy of incoming data based on its source, pedigree, occurrence and receipt time, and then normalizing incoming data to reflect confidence based on existing data.

CHANS provides interfaces that facilitate the retrieval of information stored in the HNC for use by Forecasting and Classification services such as ORA and Construct. Edge enhancement in CoreNetML, for example, facilitates faster ingest, and assessment due to reduced IO, for both observed and forecasted changes. Consequently, techniques for dynamical analysis such as network change detection—a statistical process control approach to detecting substantive changes in networks—can be applied to real and simulated data ([7] [8], respectively). Expected changes due to network adaptation and healing can be forecast using simulation tools such as Construct [6].

4 Scenario

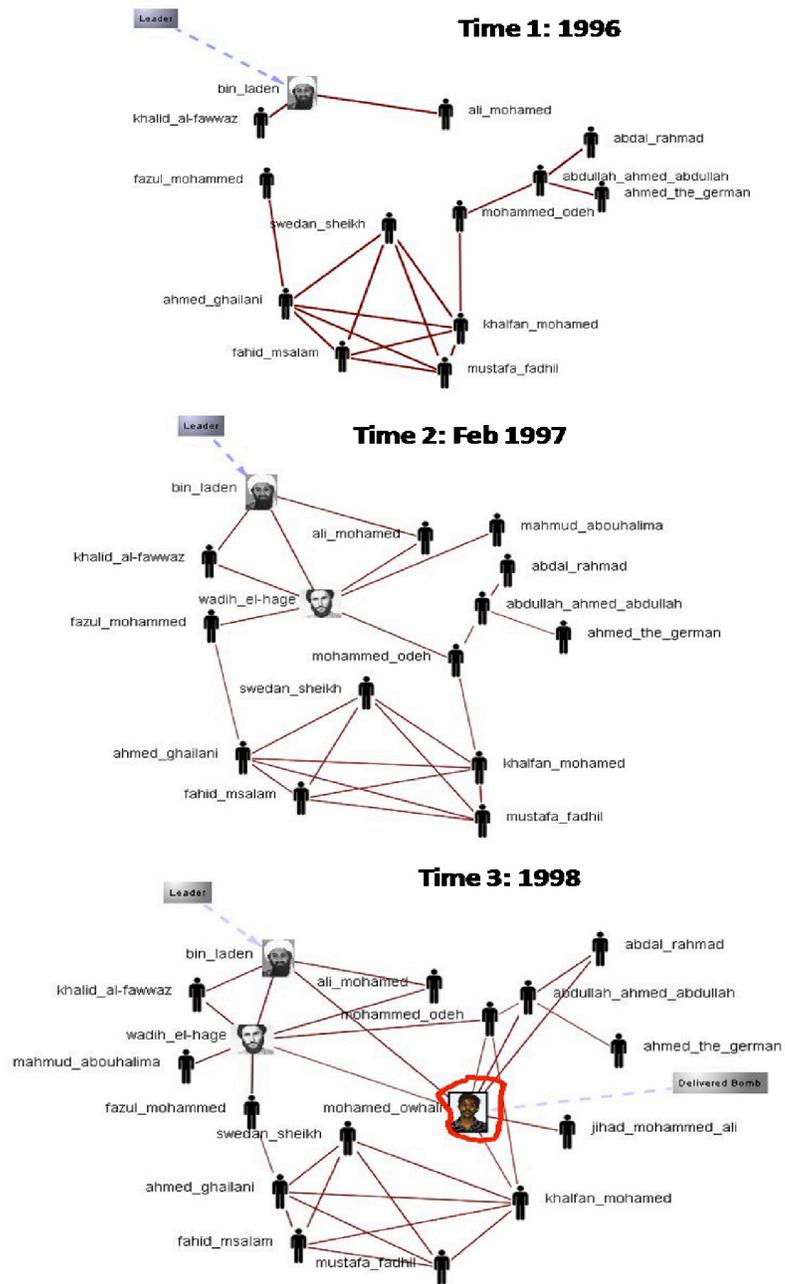
How would the process have worked if we were able to conduct dynamic network analytics at the time of the Tanzania bombing in 1998? To illustrate, we use a subset of the Tanzania bombing data collected from open sources by Connie Fournelle at ALPHATECH and expanded by the CMU CASOS team [9]. We can think of this data—represented as a meta-network describing who, what, where, how and, to a lesser extent, why—as intelligence that has been collected from diverse sources and fused together by CHANS. For brevity’s sake, we show three periods of change in Figure 2 using items from the ORA Key Entity Report, and in the images show only people though other factors are tracked. El Hage is an FBI photo; Owhali is a CNN photo. We focus on three key actors identified in ORA’s Key Entity Report: the emergent leader of the group (highest in cognitive demand), the person most likely to be in the know (highest in degree centrality), and the person most likely to connect groups (highest in boundary spanning—high betweenness, low degree centrality).

As we start the scenario, the brigade staff has been using CHANS and ORA to continuously update their understanding of the evolving situation. Late 1996: trusted informant warns al Qaeda may attack embassies; known cell structure is in Figure 2 as Time 1. Analyses show Khalfan Mohamed to be in-the-know and the emergent leader; Abdullah Ahmed Abdullah is the boundary spanner. February 1997: Wadih el Hage returns to Kenya and phone monitoring picks up el Hage in discussions with terror cell. This is picked up and processed by CHANS. Network analyses show a shift in key actors, with el Hage identified as in-the-know and the emergent leader, and Mohammed Odeh as the boundary spanner. May 1998: image analysis reveals that Mohamed Owhali is connected to Bin Laden—leading to another shift in structure as the group nears the August bombing, with el Hage in-the-know, Khalfan Mohamed again as the emergent leader, and Owhali as the boundary spanner.

At each time period, CHANS ingests new reports, fuses these reports with the existing network using the Data Fusion Service and produces a revised DyNetML that

Dynamic Networks: Rapid Assessment of Changing Scenarios

Figure 2.
Change in
Networks.



reflects reported changes in the network. This is imported into ORA, and the results are immediately compared with previous data. As new observations arrive, the relative standing of key actors is impacted—as are the implied courses of action. Until now updates to social networks based on incoming information were in a time consuming, manual manner. Therefore, a fast impact analysis due to rapid changes in the network was not possible. CHANS facilitates timely and automated fusion of newly available data into a cohesive, up-to-date network picture for analysis by tools such as Construct and ORA. This scenario and the interaction between CHANS, ORA and Construct showcases the potential enabled by attribute-enhanced NLRs.

5 Concluding Remarks

CHANS facilitates continuous social and human network monitoring through the collection and fusion of information such as reports from observers in the battlefield who report on events, locations of persons of interest, suspicious containers, and other relevant information. While there is always some degree of uncertainty associated with the reports, CHANS attempts to address this by estimating information confidence through its Data Fusion Service. Through the Core Network System component, the CHANS merges incoming observables into an evolving network representation. The outputs of CHANS are an updated network that can be readily imported into a network analysis tool such as ORA for rapid analysis of substantive changes in the network, which can be fed back to commanders in the field on short order. CHANS provides the capability to rapidly ingest, assimilate, and analyze battlefield and provide feedback regarding important current and potential future network changes. Modifications to standard NRLs enable this capability by enhancements to the representation of dynamics and uncertainty, along with reduced computational overhead. Dynamic network analytics can then take advantage of observed changes in the data to engage in trail analysis and change detection.

References

1. Carley K, Columbus D, DeReno M, Reminga J, Moon I (2008) ORA User's Guide 2008. Technical Report, Carnegie Mellon University, School of Computer Science, Institute for Software Research, CMU-ISR-08-125.
2. Borgatti S, Everett M, Freeman L (1999) UCINET 5 for Windows: Software for Social Network Analysis. Analytic Technologies, Inc., Natick, MA.
3. Brandes U, Eiglsperger M, Herman I, Himsolt M, and Marshall M.S (2002) GraphML Progress Report: Structural Layer Proposal, Proceedings of the 9th International Symposium on Graph Drawing, Lecture Notes of Computer Science, Volume 2265, pp. 501-512.

Dynamic Networks: Rapid Assessment of Changing Scenarios

4. Tsvetovat M, Reminga J, Carley K (2004) DyNetML: Interchange Format for Rich Social Network Data. CASOS Technical Report. Carnegie Mellon University, School of Computer Science, Institute for Software Research International, CMU-ISRI-04-105.
5. Nick Morizio (2008) Context Selection for Linguistic Data Fusion, in the Proceedings of the 11th International Conference on Information Fusion, Cologne, Germany, June 30-July 3, 2008.
6. Schreiber C, Carley K (2004) Construct - A Multi-agent Network Model for the Co-evolution of Agents and Socio-cultural Environments. Construct - A Multi-agent Network Model for the Co-evolution of Agents and Socio-cultural Environments. Technical Report, Carnegie Mellon University, School of Computer Science, Institute for Software Research International, CMU-ISRI-04-109.
7. McCulloh I, Carley K (2008a) Social Network Change Detection. Technical Report, Carnegie Mellon University, School of Computer Science, Institute for Software Research, CMU-ISR-08-116.
8. McCulloh I, Carley K (2008b) Detecting Change in Human Social Behavior Simulation. Technical Report, Carnegie Mellon University, School of Computer Science, Institute for Software Research, CMU-ISR-08-135.
9. Moon I, Carley K (2007) Modeling and Simulation of Terrorist Networks in Social and Geospatial Dimensions. IEEE Intelligent Systems, Special issue on Social Computing - Sep/Oct '07, 22, 40-49.

Acknowledgements

The work by CMU on DyNetML 2.0, *ORA and Construct, was supported in part by the Air Force Research Laboratory (AFRL) CAPES Project: FA865007C6769, the Department of Defense Multi-disciplinary University Research Initiative (MURI) in cooperation with George Mason University (GMU) and Air Force Office of Scientific Research (AFOSR): 600322 and the Office of Naval Research (ONR) DNA: N00014-06-1-0104. Additional support for *ORA was provided by the center for Computational Analysis of Social and Organizational Systems (CASOS) and the Institute for Software Research at Carnegie Mellon University. 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 United States Air Force, United States Department of the Navy, U.S. Department of Defense, or the U.S. government.