ILLiad TN: 116859
Borrower: ENS

Lending String: hnt

Patron: Staff: Rochelle Economou

Journal Title: Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers

Volume: **Issue:**
Month/Year: 2003 **Pages:** 361-370
363

Article Author: Kathleen Carley

Article Title: Linking Capabilities to Needs

Imprint:

ILL Number: 116760

Call #: 302.3 W926

Location: OVRSZQ-2

Mail
Charge
Maxcost:

Shipping Address:
Engineering & Science Library
Carnegie Mellon University
Wean Hall
Pittsburgh, PA 15213

Fax:
Ariel: es4bb928.library.cmu.edu

Linking Capabilities to Needs

Kathleen M. Carley
Institute for Software Research International
Carnegie Mellon University

Abstract

Social network analysis is an established field with a wealth of theory, method, and tools for handling network or relational data. Recent advances have expanded this repertoire to include larger scale more dynamic networks under conditions of uncertainty. This moves the field closer to the needs of the defense department and intelligence agencies. However, the needs are not completely met. This paper briefly outlines where the field is, what it has to offer, and where the needs are not being met.

Acknowledgement

The research reported herein was supported by the National Science Foundation NSF IRI9633 662, the NSF IGERT in CASOS, the Office of Naval Research (ONR) Grant No. N00014-97-1-0037 and Grant No. 9620.1.1140071, Additional support was provided by the NSF IGERT 9972762 for research and training in CASOS and by the center for Computational Analysis of Social and Organizational Systems at Carnegie Mellon University (http://www.casos.ece.cmu.edu). 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 National Science Foundation or the U.S. government.

Citation

Kathleen M. Carley, forthcoming, "Linking Abilities to Needs" in the Summary of the NRC workshop on Social Network Modeling and Analysis, Ron Breiger and Kathleen M. Carley (Eds.), National Research Council.

Linking Capabilities to Needs

The increase in concerns with terrorism world wide have led the researchers, policy makers, the defense community and the intelligence community to ask – what techniques, tools, methods, findings can help defeat terrorism. Social network analysis is one area that shows promise. The recent press on "linking the dots" and the "new" science of networks has brought this area into the limelight. In attempt to determine whether is a match between capabilities and needs an NRC workshop on Social Network Modeling and Analysis was run.

The purpose of this paper is to juxtapose the work discussed and not discussed at this meeting with the needs of the defense and intelligence community to the extent that is possible at the unclassified level. This will be done first by describing what the social networks area, broadly defined, has to offer particularly with respect to data collection, analysis, special training, intelligence and policy needs, and effects based operations. This is followed by a short description of what was missed. Finally, a few concluding comments are provided on what can be done to reduce the mismatch between capabilities and needs.

What Social Networks Has to Offer

Most work in social network analysis (SNA) has focused on small, bounded networks, with 2-3 types of links (such as friendship and advice) among one type of node (such as people), at one point in time, with close to perfect information. The research presented here is striving to overcome these limitations. As such it is increasingly useful to members of the defense and intelligence community. For in those communities, data is rarely complete, often there are massive quantities of data, relevant relational information is both multi-modal and multi-link, and the networks range in size from smaller command and control teams, to international terrorist networks. Recent work is helping to reduce this gap between capabilities and needs. In particular there are four areas where network analysis, such as that described in this workshop, is particularly relevant. These are data collection, analysis, special training, intelligence and policy needs, and effects based operations. Cross cutting these areas the work on networks has been done on the micro, meso and macro levels. At each level, issues of diffusion and so information warfare arise. At each level, issues of power, influence, control, and the role of technology arise. And, at each level issues of governance arise. The upshot is that network tools, techniques and theories can be used at each of these levels and across these levels to inform policy makers.

Data Collection

Much of the recent work provides some guidance for data collection. One of the key concerns with SNA is that it is data greedy and requires intensive effort. The image is of questionnaires to small groups asking each person who interacts with whom generating full information. However, the work presented herein suggests that full information is not needed, that the data can be collected in a more automated fashion, that it is possible to estimate the whole network from parts of it, and that it is key to collect new types of relational data.

We see here new techniques for collecting data. For example, by focusing on the paths among the key actors, data can be collected on the core of the network rapidly (Klovdahl, Borgatti). We can think of this as "socially informed" search. Another difference here is the emphasis on new types of data such as the meta-matrix (Carley), role data (Johnson), internet ties (Faloutsos), and multiple types of relations such as finance plus communication (Jensen, Morris, Contractor). A third difference is the growing emphasis on collecting indicators of the shape and nature of networks rather than on collecting the exact networks. Examples here are the work on spatial and proximity effects (Butts) and temporal issues (Morris). Another implication of the work on large scale and functional networks is that general demographic data and the nature of the tasks being done place constraints on the resulting networks limiting the forms that they can take on (Jensen, Carley). These factors pave the way for expanding the area to collect the type of data useful in defense and intelligence contexts.

Nevertheless, there is much to be done. It is still the case that most data collection techniques are too personnel intensive. What is needed are additional tools for doing rapid assessment and for collecting vast quantities of data. In terms of rapid assessment the issue is what high level indicators can be quickly gathered that indicate the overall shape of the network and the key actors within it. Tools for estimating the size, shape and other key dimensions of networks from leading indicators are needed. The key scientific question is from the space of possible networks, what key indicators suggest which portion of this space the network of interest is in. A related question is what constraints due institutions, social conventions, cognitive limitations and capabilities, geography, etc. place on the shape and structure of the social networks.

As to more automated data collection several things are needed. New non-invasive tools for data capture need to be developed and deployed so that they can be easily used in organizational and social analysis. These range from automatic data capturing tools, to automated text analysis (such as KEDDS, automap, etc.), to tools for turning standard data such as web pages, etc. into network data. Further, an inventory needs to be conducted of standard currently collected data to determine which data can be used and converted into network data. Examples of such data are web pages, HR reports, personnel records, phone logs. Then shareable tools need to be developed so that researchers can take the extant data and generate a network from it. For example, since KEDDS extracts events, techniques for inferring interactions and so networks from these events are needed (Ward, Hoff and Lofdahl).

In many cases, there is a need to reason about networks and how they will change where there is little actual data. To address this problem, two possible approaches were described. On the one hand, computational techniques can be used to enumerate and analyze classes of networks with specific properties (Bienenstock and Bonacich, Wasserman). Here, statistical models can be used to characterize distributions (Hockman). Then computational techniques can be used to generate sample networks with these characteristics.

On the other hand, computational techniques can be used to grow networks from limited data and basic principles about the way networks evolve. A variety of principles for how networks evolve were discussed including changes in interaction due to changes what is known or owned, relative similarity (homophilly), relative expertise, co-work on same tasks, contributions to the public good, who else is around, changes in status or role (Carley, Contractor and Monge, Freeman, Borgatti, Friedkin). Given these principles, computational techniques can be used to both generate networks consistent with these principles and then evolve them over time. Using such techniques it then possible to address questions such as how and to whom information will diffuse (Carley), how contagious diseases will spread (Morris), how cultures will evolve (Macy), how to destroy and preserve networks (Carley, Stanley and Havlin). From a data collection standpoint, such tools can be used in many ways, including to provide guidance on new data collection efforts, to generate hypothetical data that can be used for planning when actual data is not available (as in the case of biological and chemical attacks), and to supplement real data by filling gaps or suggesting trends.

Analysis

Advances in analysis have been moving on a number of dimensions. Advances in statistical approaches for analyzing networks are expanding our ability to formally handle relational data and to utilize more realistic network distributions (Hoff, Hockman, Snijders, Wasserman). Advances in visualization techniques are enabling greater understanding of the shape of networks and the impact of various change strategies (Borgatti, Richards). Meanwhile work on large-scale networks suggests that there is little information content or variability in existing measures, or that existing tools are computationally slow when faced with large networks. Thus, leading to a quest for alternative measures, metrics that scale well, increased concern with optimized or parallelizable code, fast algorithms, and estimation techniques (Stanley, Faloutsos). Work on scale free networks is one possible avenue (Faloutsos) although questions abound on the relevance of such work to traditional social networks such as friendship. That is, are such networks good enough approximations to be useful?

All this being said there are quite a few questions left unanswered that are critical in the current context. For example, for each of the existing measures, what is its robustness, resistance and sensitivity (Bienenstock and Bonacich, Borgatti, Butts, Carley, Pattison, Snijders, Wasserman). How do different measures behave for networks drawn from different distributions? How does the distribution of nodal characteristics relate to the overall structure (Snijders)? How do the various measures correlate, and how does this change as a function of the error in information, the size and density of the network. It would benefit the area to identify a few distinct core measures that are consistent and robust indicators of the shape of the underlying network.

With the movement to new types of data there is a corresponding movement to apply old metrics in new contexts to see how they behave and to develop new metrics specifically tuned to the underlying type of data. Measures that link multiple networks together such as cognitive load (Carley) or that indicate specific temporal patterns of ties such as sequences of financial transactions (Jensen) are particularly promising. A great deal of fundamental work needs to be done in the area of metrics, algorithms and the assessment of metrics.

In many cases, computational approaches, particularly simulation is playing a key role in analysis. Such work is enabling a greater understanding of the way in which networks evolve under standard and attack conditions (Macy, Carley), the way in which influence, ideas, and diseases propagates (Friedkin, Morris). In this sense simulations can be used to put "probabilities" on possible futures – suggesting not only that the network might change but how and the likelihood of such change. Further, simulations can be used to suggest areas where data might be missing. An example of this would be to turn KeyPlayer around (Borgatti) and ask where might there be a key player that we are not observing. Simulations can also be used to suggest when is it worth gathering more information as in the DyNet model (Carley). A key avenue that will increase the value of these uses of simulation would be to place the simulation in a decision context and ask what are the cost and benefits of such additional data collection and the chance of being wrong.

Special Training, Intelligence and Policy Needs

A key need in the defense and intelligence communities is for tools to help with training, intelligence, and policy analysis. In particular tools that enable the analyst to utilize the bits and pieces of disparate information in a systematic, and if possible, quantitative fashion that moves the analysis beyond informed guess work. Here, to the extent that network tools are combined with computational modeling techniques, simulation tools can be created for doing "what if" scenario analysis and suggesting possibilities. Such tools are "thinking tools" for reasoning about events. The use of such tools would not be to predict specific events but to reduce the likelihood of being "surprised." Such tools may be particularly useful in reasoning about and planning operations related to the destabilization of networks, the impact of changes in trade or cooperative agreements on inter-national, inter-state, and inter-agency behavior, understanding inter-player coordination in situations other than war, knowledge management within and among agencies, and so on. Work on governance (Lazer), influence (Friedkin), and network stability (Carley, Stanley, Borgatti, Bienenstock and Bonacich) are key in this context.

Combining network techniques with computational techniques can thus promote Effects Based Planning (EBP). Effects based planning is, according to the Joint Forces Command Glossary "an operational planning process to conduct effects based operations within rapid decisive operations. ... EBP closely mirrors the current joint planning process, yet focuses upon the linkage of actions to effects to objectives. ... It employs virtual, near-simultaneous planning at all echelons of command."

Using dynamic network models in effects based planning will require researchers to seriously address the issue of realism. How real is real enough in these models? Are simplistic models based on central social science principles sufficient or are social network flight simulators necessary? If the latter, a core area to address is how the simulation models can meaningfully make use of abstract data, easily available data, and data at varying levels of fidelity.

Core concerns in developing such models include the following. It is imperative for validation, development of metrics, and usefulness of results that the data collected from the computational model and the data collected from the real world are the same kind of data and at the same level of granularity. Second, the models should be easy to use. In part this means greater attention to user-interfaces that are rarely developed by academicians. To facilitate this, funding agencies need to provide funds not just for the basic model but also for the development of the interface, the development of ontologies and the development of links to other programs. Finally, the computational tools need to include "meaningful" variables. To aid in this, subject matter experts should be consulted to develop a vocabulary for the tool that maps from the academicians ideas to the terms of the user.

Effects Based Operations

Network techniques can also enable certain types of effects based operations. Effects based operations, according to the Joint Forces Command Glossary, is "a process for obtaining a desired strategic outcome or 'effect' on the enemy, through the synergistic, multiplicative, and cumulative application of the full range of military and nonmilitary capabilities at the tactical, operational, and strategic levels." Many types of effects are particularly amenable to being planned for using network techniques and models – destabilization of the adversaries networks, reducing the ability of the adversary to acquire intelligence or coordinate, decreasing/increasing trust in the adversaries leaders or key operators (information warfare effects), and reduction in the adversaries "reach."

In order to use network analysis for effects based operations a number of advances are needed. First, network analysis need to move beyond characterizing the structure to understanding its dynamics and how to influence it. Further, in this area particularly, it is important to link the social, anthropological and communication approach so common in social networks with the understanding of events and actions coming out of political science. This will enable the analyst to ask, what types of networks can be handled through diplomacy, war, operations other than war and with what effects. This may be particularly important when the adversary is not a nation-state.

Is There Anything Else?

While the work covered at this meeting addressed many of the core areas in social networks, there are still areas left uncovered. It is not my intent to comprehensively review these. Rather, I will mention only a few where the tools may have direct relevance in the defense and intelligence communities.

Network analysis can be used to characterize difference across cultures and to suggest strategies for changing the underlying culture. One way in which culture can be viewed is as the distribution of knowledge, beliefs, and norms across people – i.e., the knowledge network. Work melding social networks and knowledge networks demonstrates that the underlying culture can be changed through the flow of targeted information. The basic idea is that as individuals interact through the social network, they learn, thus changing the knowledge network, and so the overall culture. Basically, the co-evolution of social structure (the social network) and culture (the knowledge network) can be managed. In this area, individuals interested in communication, organizational knowledge management, and social networks are coming together to ask questions such as – how can we identify the culture-change agent in the social network?, what is the core piece of knowledge that needs to be changed?, and how can we rapidly assess the culture? Such work provides new tools of use in the area of information warfare.

Network techniques are also being used to analyze coded texts. Work is being conducted to create automated tools for analyzing texts as networks of concepts (often called mental models) and for extracting social, knowledge and event networks from texts. Postprocessors are then storing the coded data in formats that can be analyzed using standard network analysis tools like Mage, Pajek, Ucinet. One of the key possible uses of these tools is to facilitate rapid first past coding of intelligence reports. Another use here is to contrast differences in the structural properties of the mental models of experts with novices, or to contrast the reports of different intelligence officers.

Another area of advancement is in the area of visualizing networks. Here research is on-going on how to simultaneously represent nodes and "meta-nodes" and automatically create meta-nodes. A meta-node is a node containing a set of other nodes that have some structural property in common. Simultaneous representation of multiple networks, zooming in and out in extremely large networks, and graphic representation of critical nodes are all key aspects of the new visualization techniques.

The final area I want to mention is health. New epidemiological models now take not only geographic factors but the networks in to account. This has profound implications for estimating the size of potentially exposed populations and evaluating prevention, containment and treatment policies. Such models are useful in combating bio-terrorism.

Finally, and it is important to stress this point, tools and techniques for dealing with relational data are currently being developed in a number of fields. In this workshop there was a small representation of the work in computer science and physics on graphs, networks, and relational data. In addition to this, there is on going work in other areas of computer science, information science, electrical engineering, biology and chemistry that is of value here. Major advances may well require linking network analysis to other techniques such as data-mining, data-fusion, and machine learning. The key here, is that there are an increasing number of applications as the field moves from standard network analysis to dynamic network analysis, from looking at small to large networks, from looking at static to dynamic networks, from looking at single mode (e.g. people) to multi-mode (e.g., people and knowledge) networks. Dynamic network analysis and the associated tools and techniques will be useful in addressing both tactical and strategic issues. Further, it will be useful in looking at both short and long term issues; e.g., in how to destabilize a covert network and in how to alter the underlying culture to reduce the possibility of new recruits.

Outreach – Connecting the Dots

As we have seen, many of the ideas, measures, and tools being developed hold promise relative to the needs of the defense and intelligence community. However, there is still large gap. Filling this gap is not simply a straightforward excursion moving the theoretical to the applied. Rather, it requires fundamental new science on dynamic networks under varying levels of uncertainty. Further, it requires channel through which the needs of the intelligence and defense community can be better expressed to the scientific community.

On the academic side, it would benefit the field, the government and the economy if a larger number of masters and Ph.D. students were trained in formal (statistical, mathematical and computational) network techniques. In fields where the link between science and government is tighter, roughly 50% of the students graduating go into industry or government positions rather than academics. In contrast, not only are fewer graduate students produced in the social network area (almost an order of magnitude fewer) but many fewer go into non-academic positions (almost two orders of magnitude).

Further the two key factors academics need to consider are scale and communication. That is, the network community needs to take seriously whether the data collection techniques, measures, methods, algorithms scale with the size of the network. It is imperative that the issue of scale be addressed so that the academic researcher and government analyst can utilize the ideas, theories, tools etc. regardless of the size of the network. One thing the community needs to be concerned with is collecting and sharing databases of networks of varying sizes – from the tiny, 4 nodes, to the mega, millions of nodes – that can be used to test all theories, measures, algorithms, etc.

The second key factor is communication. The way in which academics in this field ask questions are often tied to the body of theory rather than the needs of the time. It is important when doing research to ask both – how do I ask my questions among academic peers – and – how do I ask my question from an applied angle. Or in other words – "why should defense or intelligence agencies care?".

On the part of the defense and intelligence community there is also work to be done. Here, personnel should go to training courses or bring in faculty to run short courses. In network analysis there are a number of training courses – such as that at the ISPCR in Michigan, the CASOS summer institute, the training sessions at the Sunbelt Social Networks Meetings and the Academy of Management. Having even a rudimentary training would help with communication as more terms of art would be shared. Another key point is that this applied community needs to provide more detailed guidance as to the level and types of measures needed and to questions of concern. It is not sufficient to tell the academician – well what you have is not quite right. Rather, ways of discussing needs, without compromising national security, must be found. This is very possible and is done quite well in other areas, such as technology design, data mining, and engineered systems. Techniques that might help in this regard include getting academicians in the area to help write the broad area announcements, sending personnel to spend one term on a campus as a visitor, creating cleaned or hypothetical datasets that can be shared with the academic community, and of course, workshops like this where the two communities can exchange views.