Applications of Bayesian Belief Networks in Social Network Analysis

David Koelle, Jonathan Pfautz, Michael Farry,
Zach Cox, Geoffrey Catto, and Joseph Campolongo
Charles River Analytics Inc.
625 Mount Auburn Street
Cambridge, MA 02138

Abstract

In this paper, we discuss the use of Bayesian belief networks as a tool for enhancing social network analysis. Traditional social network analysis (SNA) primarily uses graph-theoretic algorithms to compute properties of nodes in a network. However, these algorithms assume a degree of completeness and reliability of the social network data, which cannot always be assured. Applying Bayesian belief networks to social network analysis provides additional capabilities for discovering new links and identifying particular nodes in the network that cannot be achieved using more traditional methods of social network analysis. We describe these applications of Bayesian belief networks and their implementation in a SNA tool.

1. INTRODUCTION

Social network analysis (SNA) primarily focuses on applying analytic techniques to the relationships between individuals and groups, and investigating how those relationships can be used to infer additional information about the individuals and groups (Degenne & Forse, 1999). There are a number of mathematical and algorithmic approaches that can be used in SNA to infer such information, including connectedness and centrality (Wasserman & Faust, 1994).

SNA is used in a variety of domains. For example, business consultants use SNA to identify the effective relationships between workers that enable work to get done; these relationships often differ from connections seen in an organizational chart (Ehrlich & Carboni, 2005). Law enforcement personnel have used social networks to analyze terrorist networks (Krebs, 2006; Stewart, 2001) and criminal networks (Sparrow, 1991). The capture of Saddam Hussein was facilitated by social network analysis: military officials constructed a network containing Hussein's tribal and family links, allowing them to focus on individuals who had close ties to Hussein (Hougham, 2005).

In this paper, we discuss the use of Bayesian belief networks as a tool for enhancing social network analysis. The intended user of such a tool is a social network analyst who is tasked with identifying individuals of interest within the network (e.g., who is the most central person?) or inferring relationships between individuals when those relationships have not been explicitly gathered from data (e.g., two people who worked at the same small company might know of each other), while considering the impact of uncertainty inherent in the data collection process. In Section 2, we discuss the limitations of social network analysis techniques that do not incorporate notions of uncertainty or different node and link types. In Section 3, we discuss some potential applications of Bayesian belief networks as they pertain to Social Network Analysis. Finally, in Section 4, we describe the implementation of our tool that enables social network analysts to use Bayesian belief networks to enhance their SNA tasks.

2. LIMITATIONS OF SOCIAL NETWORK ANALYSIS

While traditional SNA has been used to successfully derive insights into a social network, it can be restrictive for a number of reasons. SNA assumes a well-formed social network, but real-world methods of data collection may not ensure that the resulting social network is complete and contains needed data. SNA focuses primarily on the existence of a relationship between nodes in the network, but not on attributes of that relationship or the nodes in the relationship. Furthermore, SNA does not explicitly consider the uncertainty of attributes on nodes or relationships. Finally, graph-theoretic algorithms used in SNA tend to focus on a homogenous set of entities and relationships, making it difficult to analyze networks that involve a heterogeneous set of nodes connected by a variety of link types.

2.1 ISSUES IN DATA COLLECTION

Traditional social network analysis depends on social networks that have been created with some degree of certainty. However, there are many sources of uncertainty in the data collection process. Knowledge of these sources of uncertainty, paired with the creation of analysis techniques that can utilize this uncertainty information, will improve the validity of SNA results.

The construction and analysis of social networks can be viewed as an exercise in the observation of correlations. Hypotheses about social network structure cannot be tested empirically; analysts can only observe behavior to validate networks that they have constructed (Degenne & Forse, 1999). Traditional methods of data collection usually involve interviews with various subjects to identify people who are significant to them in some way; however, nuances of traditional data collection methods, such as connotations that subjects might place on the words used in interview questions, can confound social network analysts.

Each person has a different way of perceiving their own social network, so it is difficult to obtain an objective view. For example, individuals tend to perceive themselves centrally (Kumbasar, Romney, & Batchelder, 1994), and while there are methods to reduce these ego biases in social network construction (Krackhardt, 1987), they still persist and therefore make objective analysis difficult. In addition to subjective biases, numerous studies show that individuals can identify their social networks with only a moderate level of accuracy (Bernard et al., 1990; Bernard et al., 1989), and their perception of the social network will change significantly over time (Morgan, Neal, & Carder, 1977; Coleman, Katz, & Menzel, 1957). There is also the possibility of measurement error, and the possibility that individuals may not be completely truthful in their recall (Degenne et al., 1999).

Collecting a data set that is rich enough to provide interesting conclusions requires significant effort. As a network branches out from a single individual and incorporates information about others, it becomes more detailed and interesting, but leads to an exponential buildup of identification, characterization, and reconciliation tasks. For networks based on the observation of electronic communications, this exponential build-up creates a particular analysis challenge. Face-to-face interactions represent yet another type of rich interaction data, but meaningful data sets are difficult to collect. For example, observation of a community of windsurfers on a beach in Southern California was conducted over the same 2-hour window every day for 31 days (Freeman, Freeman, & Michaelson, 1988), and while distinct patterns in communication were observed, it is probable that the observations were affected by the specific venue and time windows that were used.

Clearly, the variety of methods of data collection for social network analysis reflects the inherent difficult in capturing reliable and consistent information about human social relationships. The uncertainty of this information about the relationship limits the applicability of traditional methods of social network analysis.

2.2 HOMOGENEOUS NODE AND LINK TYPES

Social network analysis does not fully address the need to characterize different types of relationships. A social network graph can represent a variety of concepts through links, such as evaluation (A likes B, A respects B), behavioral interaction, transfers of material resources, association, affiliation, movement between places or statuses, physical connection, formal relations (such as authority), and biological relations (Wasserman & Faust, 1994).

Different links types can also indicate the strength of a particular concept. For example, one social network construction study offered subjects four choices to identify the intensity of their friendships: “No contact”, “small talk and coffee”, “exchange of favors”, and “close ties” (Heran, 1987).

While traditional graph-theoretic algorithms used for SNA may incorporate analysis of different node and link types, they tend to be homogeneous within a network (i.e., considering a single node or link type per analysis), rather than being heterogeneous within a network (i.e., multiple link and node types). Further, graph-theortic algorithms do not typically consider attributes on links or nodes (e.g., links with intensities and reliabilities; nodes with attribute sets).

3. USING BAYESIAN BELIEF NETWORKS IN SOCIAL NETWORK ANALYSIS

Bayesian belief networks can be applied to social network analysis to derive insights that are not possible using traditional SNA techniques. In the following section, we discuss three types of analyses that are enabled using Bayesian belief networks: augmenting social network algorithms with uncertainty, searching the network for nodes, and inferring new links in the network.

3.1 REASONING ABOUT UNCERTAINTY IN SOCIAL NETWORKS

In traditional SNA, graph-theoretic algorithms are useful for determining mathematically derived facts about entities in the network. For example, one common algorthm computes the “degree centrality” for a node, which is measured by adding the number of incoming links on a node, and provides some indication of how important that node might be (Wasserman et al., 1994).

However, these algorthms do not take uncertainty into account. While a node may appear to have a high value for degree centrality, the algorithm does not consider the certainty of the links, the authority from whom the link information was gathered, the recency of the link, or any other type of meta-information (i.e., qualifiers of the information) that may be known (Pfautz et al., 2006; Pfautz et al., 2005)

Bayesian belief networks can augment SNA algorithms by considering meta-information in their calculations. For example, the user of an SNA tool that incorporates uncertainty might be interested in determining the "importance" of each individual in the network. The user would create a Bayesian belief network for "Importance", which might contain one node representing the algorthmic degree centrality computation, and another node that represents the total certainty of the data used in the calculation (Figure 1). These two nodes might be parents of the "Importance" node, which the user would provide with a set of conditional probability entries.

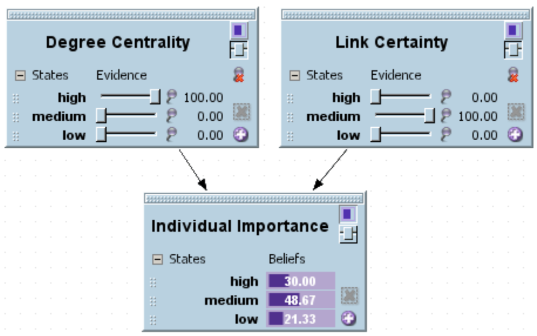

Figure 1: An "Importance" Bayesian belief network

In addition, due to the abductive reasoning capabilities of Bayesian belief networks, one could investigate questions such as, "What might be required for this individual to increase in importance?" by setting the value on an individual's "importance" node to a value, and observing what values the parent nodes would need to support that belief.

3.2 SEARCHING THE SOCIAL NETWORK

By applying a Bayesian belief network like the one above to all individuals in a social network and sorting the results, the user can find individuals of interest in a social network. This is particularly useful when the user is working with a large network (e.g., email traffic in a multinational corporation), and wants to find nodes that fit a particular set of attributes. For example, a user might be interested in individuals within the network that are likely to become a future leader in the organization. This is different from searching for simple node attributes, such as "Name" or "Age", because the notion of "Leadership Potential" is a psycho-social concept based on a combination of other attributes and relationships that cannot be handled by a simple search capability. Some of those attributes or relationships may be associated with a degree of uncertainty.

Given this example, a user might create a Bayesian belief network as illustrated in Figure 2. This uses the "Importance" Bayesian belief network discussed in Section 3.1, as well as some additional considerations: "Previous leadership experience" and "Leadership classes taken". Each of these feeds into the child node, "Likely organizational leader".

The user can now apply this Bayesian belief network to each of the individuals in the social network, and see the resulting list of individuals and values for each individual's likelihood of becoming an organizational leader. By sorting this list, the user can identify those individuals who have the highest value for leadership potential.

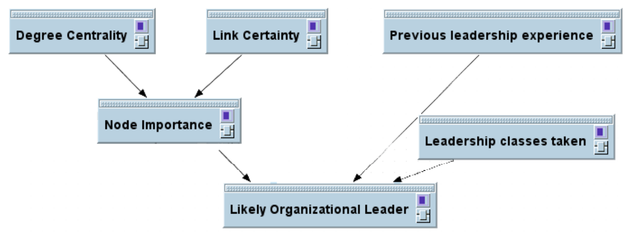

Figure 2: Bayesian network for "Likely Organizational Leader"

3.3 INFERRING LINKS

New links can be inferred from information that is already known with varying degrees of certainty. For example, well-established link data (e.g., demographic information from a census) could be used to make inferences about additional possible links (e.g., the likelihood that two people know each other based on geographic proximity and similar socioeconomic backgrounds). Analysis of group membership within a social network can indicate likely membership ties (Kubica et al., 2002). Input to a link discovery tool can be prefiltered to identify individuals who have varying levels of contact with known threats, cutting down analysis time and providing a small boost to accuracy (White & Fournelle, 2005). Probabilistic link discovery algorithms, rather than deterministic methods, have been used to assign a confidence factor to generate links and correlate them with the true group membership links (Adibi, Cohen, & Morrison, 2004). While many links found by this kind of data mining are uninteresting, rarity analysis can provide a closer look into links outside of the normal patterns that have been identified, which may be of greater interest to analysts. A mathematical definition of "rarity" for use in identifying links of interest has been developed by (Lin & Chalupsky, 2003). All of these link discovery methods can be used to generate additional data (with associated meta-information) for analysis using the above Bayesian belief network methods as well as traditional SNA algorithms. However, it is difficult to develop conclusions from inferred data that has been repeatedly aggregated (or used to infer further links), since the accuracy will necessarily decrease. It is therefore important to limit the repeated application of link inferencing methods and to properly represent and reason about the level of certainty associated with each link.

4. A SYSTEM USING BAYESIAN BELIEF NETWORKS IN SOCIAL NETWORK ANALYSIS

We have built a system that enables users to develop and apply Bayesian belief networks to reason about social networks.

The user of our system is tasked with investigating a social network that is replete with uncertainty, and must derive meaningful and useful insights into the social network and answers to questions about individuals and their relationships.

To conduct analysis on the social network, the user will first develop Bayesian belief networks that can be used to answer questions. Alternatively, a library of Bayesian belief networks may already be created by social or cultural anthropologists or other experts, who may have mechanisms for validating the Bayesian belief networks they create. Then, the user maps the Bayesian belief networks to nodes in the social network.

4.1 DEVELOPING BAYESIAN BELIEF NETWORKS

In our system, the user can create Bayesian belief networks using Charles River Analytics Inc.'s BNet.Builder© product (http://www.cra.com/bnet), as shown in Figure 3.

The nodes in the Bayesian belief network represent concepts that will be mapped to attributes of nodes or links in the social network.

In the example provided, the user is interested in discovering how likely an individual is to be an organizational leader. The user or expert has determined that this answer can be determined by a combination of the individual's previous leadership experience, whether the individual has taken leadership classes, and the individual's importance – and the individual's importance is a combination of the individual's degree centrality and link certainty.

In addition to linking the nodes together, the user or expert has entered values into the conditional probability table for the child nodes in the network.

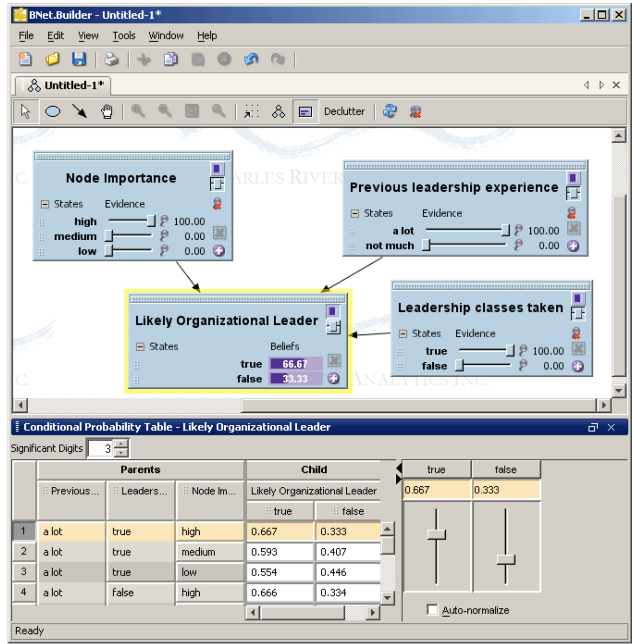

Figure 3: Creating a Bayesian belief network with BNet.Builder©

4.2 MAPPING THE BAYESIAN BELIEF NETWORK TO A SOCIAL NETWORK

The user is now ready to apply this Bayesian belief network to the social network data. To do this, the user must map each node in the Bayesian network to one of the following conditions in the social network:

- The type of node in the social network
- The value of an attribute on a node in the social network
- The comparison of an attribute on one node to a value on a connected node
- The existence of a particular link type between two nodes
- The value of an attribute on a link between two nodes
- An aggregate calculation of values on link attributes between two nodes (for example, an average of the incoming link certainties)
- The number of links between two nodes
- A value on a node or link computed using traditional graph-theoretic SNA algorithm (for example, degree centrality)
- The comparison of computed values on two connected nodes

The user may also specify the effect of a Bayesian network calculation:

- A new attribute is added to a node or link, and set with a value

- A new link is created between two nodes

The given example uses a variety of these conditions to map connections between the Bayesian network and the social network. When the Bayesian network is evaluated against each node in the social network, "Node X" will be replaced automatically by the node being evaluated.

1. Map the "Degree Centrality" BN node to the degree centrality calculation for Node X
2. Map the "Link Certainty" BN node to the average of the "certainty" attribute on each link connecting Node X to other nodes
3. Map the "Previous Leadership Experience" BN node to the value of an attribute on Node X representing previous leadership experience
4. Map the "Taken Leadership Classes" BN node to the value of a Boolean attribute on Node X representing whether the individual has taken leadership classes
5. Finally, map the end BN node, "Likely Organizational Leader", to create a new attribute on Node X representing the leadership likelihood.

Note that the "Individual Importance" BN node is not mapped to any conditions in the social network, because its values are determined from and used by other BN nodes – it is an interior node in the Bayesian network.

After making these connections, the user executes the Bayesian network on all of the individuals in the social network.

4.3 UNDERSTANDING THE RESULTS

There are a number of ways in which results from applying the Bayesian network to the social network are presented to the user.

For Bayesian networks that compute the value of an attribute on nodes or links, a table may be displayed listing each node or link, and the new attribute. The columns in this table may be sorted, so the user can easily determine who among the individuals has the highest or lowest values for a particular attribute.

In addition to the attribute value itself, our Bayesian belief network provides a degree of certainty with the answer. The social network supports this by also associating an uncertainty value with the attribute. In our example, the social network may contain one individual that seems destined to be an organizational leader, but this may be quite uncertain; there may be another individual that will not become an organizational leader, and that fact may be quite certain. The user must consider the value of the attribute with the certainty associated with that value.

If the Bayesian network results in the creation of a new link, that link is displayed differently than other links in the social network. The difference in appearance indicates that the link was determined using a computational process – i.e., it is not supported by data obtained from direct observation. Properly representing the source of the link data is important so the user can accept the link with the appropriate degree of scrutiny (i.e., allow the user to trust observed links more than inferred links).

5. CONCLUSION

The goals of the described project are to utilize Bayesian belief networks as a method for enhancing social network analysis and to demonstrate the effectiveness of this approach in an implemented system. Bayesian belief networks enable the user to perform SNA tasks that involve incomplete data, nodes with uncertain attributes, and inconclusive relationships. They also allow the user to infer new relationships between nodes that is not revealed in the original data, and to identify nodes in the network that are of particular interest because of their attributes and relationships. These capabilities make Bayesian belief networks a powerful tool in conducting social network analysis.

Acknowledgements

The work described in this paper was performed in part under Government Contract No. FA8650-04-C-6403, U.S. Army Contract No. DAAB07-02-C-L403 and U.S. Army Contract No. DAAD17-02-C-0005.

References

Adibi, J., Cohen, P., & Morrison, C. (2004). Measuring Confidence Intervals in Link Discovery: A Bootstrap Approach. In *Proceedings of ACM SIGKDD*.

Bernard, H. R., Johnsen, E. C., Killworth, P. D., McCarty, C., Shelley, G. A., & Robinson, S. (1990). Comparing Four Different Methods for Measuring Personal Social Networks. *Social Networks*, 12179-216.

Bernard, H. R., Johnsen, E. C., Killworth, P. D., & Robinson, S. (1989). Estimating the Size of an Average Personal Network and of an Event Subpopulation. In M. Kochen (Ed.), *The Small World* (pp. 159-175). Norwood, NJ: Ablex.

Coleman, J. S., Katz, E., & Menzel, H. (1957). The Diffusion of an Innovation Among Physicians. *Sociometry*, 20253-270.

Degenne, A. & Forse, M. (1999). *Introducing Social Networks*. London: Sage Publications.

Ehrlich, K. & Carboni, I. (2005). *Inside Social Network Analysis.* IBM Technical Report 05-10.

Freeman, L. C., Freeman, S. C., & Michaelson, A. G. (1988). On Human Social Intelligence. *Journal of Social and Biological Structures*, 11415-425.

Heran, F. (1987). Comment Les Francais Voisinent. *Economie et Statistique*,(195), 43-60.

Hougham, V. (2005). *Sociological Skills Used in the Capture of Saddam Hussein. http://www.asanet.org/footnotes/julyaugust05/fn3.html.*

Krackhardt, D. (1987). Cognitive Social Structures. *Social Networks,* 9109-134.

Krebs, V. E. (2006). *Uncloaking Terrorist Networks. http://www.firstmonday.org/issues/issue7_4/krebs*

Kubica, J., Moore, A., Schneider, J., & Yang, Y. (2002). Stochastic Link and Group Detection. In *Proceedings of AAAI Workshop on Link Analysis*.

Kumbasar, E., Romney, A. K., & Batchelder, W. H. (1994). Systematic Biases in Social Perception. *American Journal of Sociology,* 100477-505.

Lin, S. & Chalupsky, H. (2003). Unsupervised Link Discovery in Multi-Relational Data Via Rarity Analysis. In *Proceedings of Third International IEEE Conference on Data Mining,* (pp. 171-178).

Morgan, D. L., Neal, M. B., & Carder, P. (1977). The Stability of Core and Peripheral Networks Over Time. *Social Networks,* 199-25.

Pfautz, J., Roth, E., Bisantz, A., Fouse, A., Madden, S., & Fichtl, T. (2005). The Impact of Meta-Information on Decision-Making in Intelligence Operations. In *Proceedings of Human Factors and Ergonomics Society Annual Meeting*. Orlando, FL.

Pfautz, J., Roth, E., Bisantz, A., Llinas, J., & Fouse, A. (2006). The Role of Meta-Information in C2 Decision-Support Systems. In *Proceedings of Command and Control Research and Technology Symposium*. San Diego, CA.

Sparrow, M. (1991). The application of network analysis to criminal intelligence: An assessment of the prospects. Social Networks 13, 251-274.

Stewart, T. (2001). *Six Degrees of Mohamed Atta. http://money.cnn.com/magazines/business2*

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

White, J. V. & Fournelle, C. G. (2005). Threat Detection for Improved Link Discovery. In *Proceedings of International Conference on Intelligence Analysis*.