Social network analysis of uncertain networks

Pontus Svenson
FOI – Swedish Defence Research Agency
SE 164 90 Stockholm, Sweden
Email: ponsve@foi.se

***Abstract*—Analysing networks (primarily social, but also computer networks or networks relating events to each other) is an important part of the intelligence analysis process. Currently, computer tools for network analysis lack functionality for adequately representing and reasoning with uncertain information. In this extended abstract, we outline some approaches to adding uncertainty-handling capabilities to network analysis.**

I. Introduction

In this note on on-going work, we briefly describe the problem of representing uncertainty in social network analysis and outline some possible approaches to handling it.

Visualization and analysis of network data is an essential part of intelligence analysis. By forming networks of interesting people, organizations, places and events connected with links that indicates connection, we can create conceptual images that facilitate understanding. Network analysis is also important in mission debriefing, or after an exercise: by studying how the information has moved in the organization we can find opportunities to improve management.

Social Network Analysis (SNA) [1,3,4,5] is a set of powerful techniques for this. Together with other statistical techniques and querying tools, it can be used to identify social roles, important groups and hidden organization structures [2]. Correlation of observed data about individuals, things, places, memberships, etc. may be used to detect organized crime or terrorist cells and networks through the observation of hidden relations and co-occurrences. The methodology assumes that the ways the members of a group can and do communicate with each other affect some important properties of that group.

Intelligence analysts must be able to manipulate uncertain data and to formulate and test hypotheses about the data. Hence, the systems used for intelligence analysis must have the ability to express and reason with uncertain information. Handling uncertainty can, to some extent, be seen as the definition of fusion: we need to determine not only properties of objects and relations between multiple objects, but must also determine the probability of an object having certain characteristics and the certainty with which we can say that some objects are related.

II. Uncertainty and SNA

There exist many different methodologies for handling uncertainty, ranging from comparatively simple Bayesian analysis to more sophisticated (and complicated) analysis using fuzzy sets, Dempster-Shafer theory, or random sets.

Figuring out nested business connections across a set of known individuals or organizations is one application of SNA. Since not all people who have had contacts with a criminal are criminal themselves, there is a need for techniques which can filter out those whose contacts with known or suspected criminal individuals are either frequent or match stored patterns of suspicious behaviour. One issue thus becomes how one can automatically estimate which people among a very large community, that have been “transitively” in contact with each other, need to be investigated further and who do not. In order to do this, it is necessary to consider uncertain network data. Uncertainty in the network can be of two kinds:

- We can be uncertain whether a given node really is distinct or not, i.e., about whether two found entities should be the same or not
- We can also be uncertain about whether there really exists a link between two nodes. For instance, if person A is seen leaving the same theatre show as person B, should we assume that they are connected?

There are commercial tools to visualize networks that are currently being used by military and police intelligence analysts. There is also a great deal of sociological research done on how to analyze networks more quantitatively, by calculating various forms of measurements that are believed to be indicative of the importance (in different ways) of certain nodes and links. The currently available commercial network visualisation tools lack functionality for this.

As mentioned in the introduction, network analysis for military or police purposes have to take into account that the data that exists on the network is uncertain. Data from, for example, signals intelligence about a communications link can have an associated uncertainty. Intelligence information on resources such as a certain person seen talking to a known warlord also has quite natural uncertainties: was it really *that* person and *that* leader, and did they really talk with each other?

However, the literature from sociology and other fields on network analysis currently lack descriptions of how uncertainty could be added to social network analysis. Partly, this is because of the character of the data that has been analysed. Traditionally, network analysis was performed on data gathered by sociologists either by direct observation or taken from questionnaires. The data sets were often small, and the researcher who did the analysis possessed deep knowledge about both the data and the people it was gathered from. More recently, the vast amount of network information present in online databases and social networking applications have caused a large number of computer scientists and physicists to get interested in analysing networks. However, this research has been mostly focused on finding interesting large-scale characteristics of the networks [3], and there has thus been little need to take uncertainty into account.

In ongoing work, we are investigating how to modify standard methods from social network analysis in order to take account of such uncertainties. In order to properly handle uncertainty, it is not enough to just extend standard social network metrics from sociology to handle links with probabilities attached; we must also define proper models for how to estimate the uncertainties in the network in the first place. A very promising approach here seems to be to use the notions of possibility and necessity from possibility theory to model the uncertainty. For each node in the network, we would determine the evidence that absolutely supports the existence of the link, as well as the evidence that allows the possibility of the link. An equivalent formulation of this is to determine the belief and plausibility of Dempster-Shafer theory for the existence of the link.

Another way of representing the uncertainty in the social network is to model the network as a random set of edges. This would also enable us to analyse dynamic, uncertain networks, since we could have different random sets at different points in time. Briefly, instead of modelling a network as a set of nodes N and a set of (time-dependent) edges E(t), we would have a time-dependent random set *E*(t), whose mass function[1] m(x,t) would, for each set x and time t, give us the probability that the network had the edges in x at time t. The random set approach to stochastic network modelling would also enable us to easily define equivalence classes of networks, where the equivalence of two different networks is determined by whether the social network analysis measures used can distinguish between them or not. This would be similar to the equivalence-based sensor planning method that was developed at FOI some years ago [6].

Simulation-based methods seem to be a fruitful approach to analyze networks that contain uncertain data. Simulation is one of the traditional ways of dealing with uncertainty. In this case, we could, for example, use Monte Carlo simulations to sample graphs from a distribution which is consistent with the information on the network that we are analysing. Another option is to simulate the transport of information on the network and thereby obtain information about it. This is similar to methods already used in sociology,

Combining Monte Carlo simulation, random set modelling, and robust Bayesian analysis seems to be the most straight-forward way of handling uncertainty, as shown in the following example:

From surveillance, one gets a list of links that are present in the networks, as well as a list of links that are known not to be present (e.g., person A has not called person B's cell-phone, which is known since there is a tap on the cell-phone). By combining available knowledge about links that are known to be present and those that are known not to be present, restrictions on the possible shapes of the network can be obtained. This information can be used to construct a hypothetical set of possible networks. The set would consist of all graphs that are consistent with the known information about the network, and could be represented either as a random set or using imprecise probabilities on the edges of the graph.

This set of networks could then be used, in Monte Carlo simulations, to determine expectation values for the different structural measures that one wants to consider, or where to place new information collection resources in order to get as much new information as possible about the network. It could also be used to calculate estimates of various network properties. Different estimators would have to be used depending on the properties of the graph. It would be interesting to investigate whether the Finite-set statistics representation and analogues of the probability-hypothesis density from multi-target tracking could be used to achieve computational speedups in this process.

III. Summary and Future Work

In this note, we outlined some of the difficulties associated with doing social network analysis on uncertain networks and suggested some ways of handling them by using random set and possibility theory to model uncertainty and simulation-based methods for analysis of it. We are currently working on implementing the simulation-based analysis as well as on working out the details of a complete random set formulation of uncertain SNA.

References

[1] P. J. Carrington, J. Scott, S. Wasserman (eds), *Models and Methods in Social Network Analysis*, Cambridge University Press 2005

[2] L. Ferrara, C. Mårtenson, P. Svenson, P. Svensson, J. Hidalgo, A. Molano, A. Madsen, "Integrating data sources and network analysis tools to support the fight against organized crime", In Intelligence and Security Informatics, Proceedings of the IEEE ISI 2008 International Workshops: PAISI, PACCF and SOCO 2008, Springer-Verlag (LNCS 5075), Berlin, 2008, pp. 171-182.

[3] M. Newman, A. Barabasi, D. J. Watts (eds), *The Structure and Dynamics of Networks*, Princeton University Press 2006

[4] J. P. Scott, *Social Network Analysis: A Handbook*, Sage Publications 2000

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

[1] Note that this is not the mass function of Dempster-Shafer theory.

[6] P. Svenson and C. Mårtenson, ”SB-Plan: Simulation-Based Support For Resource Allocation And Mission Planning”, In Proceedings of the Conference on Civil and Military Readiness 2006 (CIMI 2006)