Social network analysis and information fusion for anti-terrorism

Pontus Svenson, FOI Ledningssystem
pontus.svenson@foi.se

Per Svensson, FOI Ledningssystem
per.svensson@foi.se

Hugo Tullberg, FOI Ledningssystem
hugo.tullberg@foi.se

0. Abstract

We discuss how methods from social network analysis could be combined with methodologies from database mediator technology and information fusion in order to give police and other civil security decision-makers the ability to achieve predictive situation awareness and enable them to plan and execute their actions in a pro-active manner, instead of just reacting to the opponent's actions. Techniques based on these ideas will be demonstrated in the EU PASR project HiTS/ISAC.

1. Introduction: the technological challenges of terrorism prevention

The threats facing society today require new methods for modelling and analysis. In fact, civil security decision makers, analysts and field operators fighting terrorism and organized crime across the European Union all need front-line integrated technologies to support their cooperative work. Our opponents are no longer organized in hierarchical structures, but instead consist of individuals and groups that are loosely organized in "dark networks". Instead of large-scale military attacks, they stage attacks or set bombs against unprotected civilians, or seek to influence crowds of legitimate demonstrators so that critical riot situations occur.

In order to construct decision support systems that take account of these new factors, new, more powerful methods and techniques from several technological domains need to be brought together and integrated.

Cross-border and cross-agency interoperability

To achieve the necessary cross-border and cross-agency interoperability, models and methods for secure sharing of information will have to be based on integrity and ownership across the information-sharing network, including *dynamically modifiable role-based access rights*, a *dynamic service-oriented system design*, and *powerful analysis tools* to support operations during stationary as well as mobile activities.

Emerging *database mediator technology* is making it possible to connect, without costly redesign and reprogramming, organizationally and geographically distributed data sources into a single homogeneous and secure virtual system.

Spatial (geographical) and temporal aspects of information are important in intelligence analysis. Currently marketed analysis systems use geographical data mostly for purposes of presentation, but in many analytical problems, spatial and temporal dimensions are needed also as important components of the analysis.

Situational awareness based on information fusion

Fundamentally uncertain intelligence information has to be interpreted, integrated, analyzed, and evaluated to provide situational awareness based on *information fusion*, in particular situational assessment and threat assessment methods.

Relevant intelligence information originates from many sources, some of which are well-established infrastructure sources, others may be secret human intelligence information sources, some are open sources like mass media or the Internet, yet others are sensors and other physical devices of many kinds, including mobile phones and payment terminals. Potentially relevant data from such sources need to be stored in databases for later proactive reanalysis, as and when sufficiently consistent indications of upcoming severely criminal behaviour occur and implicate a certain group or network of individuals.

Management of uncertainty in social network analysis

By combining modern methodology for *management of uncertain information* with concepts and methods from *social network analysis*, powerful new analysis tools and environments may be obtained for dealing with such tasks.

In the recently started HiTS/ISAC project (Gustavsson *et al.*, 2006) financed by the EU PASR programme, environments and tools will be demonstrated for solving a large class of social network interaction problems in law enforcement intelligence analysis. To a large extent, this will be done by combining and extending COTS (commercial-off-the-shelf) network algorithm software with COTS software for Bayesian network uncertainty management and database mediation.

2. The new threats

The traditional threat of large scale military invasion, where the opponent is organized in known hierarchical structures, is largely gone. The new threats come from terrorists and organized criminals, individuals and groups that are loosely organized in "dark networks".

Instead of large-scale military attacks, terrorists stage attacks or set bombs against unprotected civilians, with the attacks on the World Trade Center in New York and the bombings in Madrid and London as horrific examples. They can also influence crowds, e.g., of legitimate demonstrators, to instil riot situations. It is hard to determine whether the aftermath of the Mohammed caricatures is genuine popular anger or provoked riots.

In our modern globalized society information and rumours travel fast. As a result of this, critical situations may arise very quickly in unpredicted places. The mere threat of violence or riots may be sufficient to alter normal ways of life, which is indeed the intent of terrorism.

Organized crime or loosely connected dark networks?

It can be argued whether the notion of "organized crime" is appropriate for loosely connected networks of criminals. Nevertheless, a number of spectacular robberies in Sweden are believed to be carried out by criminals organized in dark networks.

Loosely connected networks are advantageous from a criminal's point of view since they reduce the risk of detection during planning and preparation phases. A further difficulty for law enforcement agencies is that not all actors are known in advance – the network may involve individuals without criminal records or known connections to extremist organizations.

Social network analysis

Social network analysis (Wasserman & Faust, 1994) is a family of statistical methods that support statistical investigation of the patterns of communication in groups. Social scientists use them to analyze, for example, families, organizations, corporations and internet-communities. The basis of the methodology is the assumption that the way that members of a group communicate with another affects important properties of the group.

The need to become proactive

Experience shows that the networks can be unwound and analyzed after the events. Although it provides the necessary evidence for bringing criminals to justice, it is then too late to prevent loss of life and material damage.

Social network analysis methods are clearly relevant to law enforcement intelligence work and may provide tools to discover criminal networks in their planning phase and thereby prevent terrorist acts and other large-scale crimes from being carried out. Relevant patterns to investigate include connections between actors (meetings, messages), activities of the involved actors (specialized training, purchasing of equipment) and information gathering (time tables, visiting sites).

False alarm risks

A key difficulty in social network analysis is to discriminate between legitimate and illegal activities. Like an immune system, difference from self, diversion from a normal state, is an indication of potential threat. The false alarm risk must be kept at a minimum, in order maintain trust in the security system and not to strain resources unnecessarily.

3. Social Network Analysis

Social network analysis is a part of sociology that was introduced in the 50's as an extension of sociograms that describe relations in groups. Social network analysis is a collection of mainly statistical methods to support the study of communication relations in groups, kinship relations, or the structure of behaviour, to mention a few application areas. This methodology assumes that the ways members of a group can communicate affect some important properties of that group.

Structural analysis

The emphasis in social network studies is on relations between individuals and/or groups of actors. It is sometimes referred to as *structural analysis*.

In order to study the structural properties of a group, it is necessary to model it mathematically. This is most naturally done by constructing a *graph* or *network* representing the relationships of the group. Each member of the group is mapped to a node in the graph, and edges between nodes are introduced if the corresponding members of the group communicate. Most edges link exactly two nodes; graphs where multi-edge relations are allowed are called *hypergraphs*. A hypergraph can always be embedded in an ordinary graph by introducing an extra node for each relation that involves more than two nodes. For example, several studies (Redner, 2005; Leskovec *et al*., 2005) of the citation and collaboration networks of scientists have recently been carried out. In these, the network of interest is the one where there is a link between all individuals who have co-authored a paper.

Simplifying representations

In order to avoid having to handle hypergraphs, additional nodes are introduced for each paper, and binary relations between papers and their authors are introduced. If we are studying collaboration networks, this leads to a *bipartite* graph, where there are two different kinds of nodes, and no edge can link two nodes of the same type. Often, networks of this type are transformed one step further into a network where two authors have an edge between them if they have ever co-authored a paper. Note that this transformation throws away information that could potentially be useful. An analogous example from our domain of interest might be that we need to model individuals who have met. In a bipartite graph of people and meetings, we can represent information about which particular meeting that two specific persons attended. In the transformed graph where we link people who have ever met, this information is lost. The computational speed and reduced memory requirements of the latter approach must thus be weighted against the potential loss of information resulting from choosing it.

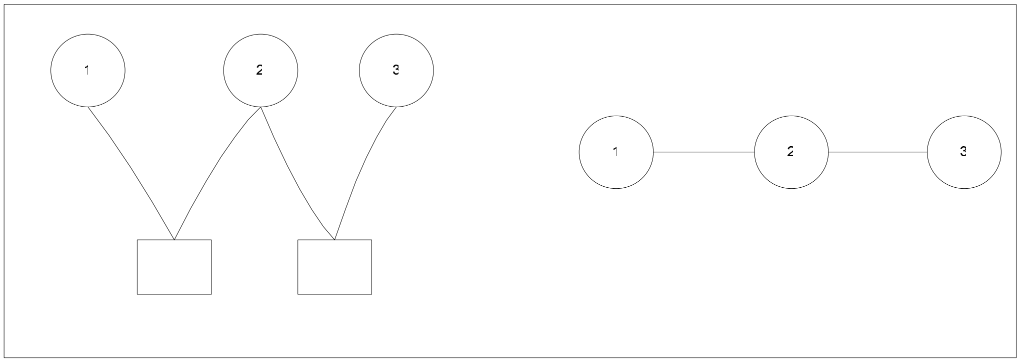

Figure 1. Example of converting a bipartite collaboration network into a network with author-author links. The left figure shows three authors (circles) who have written two papers (indicated by links to the square papers). The right figure shows the corresponding author network. Note that it is impossible to determine how many papers the authors have written from the network representation shown on the right.

For some calculations, it is convenient to work with the adjacency matrix of the network, i.e., the matrix A whose (i,j) entry is non-zero if and only if nodes i and j are connected. Edges can be either directed or undirected; a directed edge only allows information to flow in one direction.

Weights and measures

In addition to including several nodes, edges can also be extended to include a weight or probability on them. This is used to model, for example, the maximum amount of information that can flow between two nodes, or to indicate the certainty with which we know that the edge is actually present in the network.

There are several important measures that can be used to characterize a network. Perhaps the most trivial is to count the number of edges that different nodes have. This can be seen as a measure of the popularity of a node, and is one of the methods that are used by web search sites such as Google to rank search results. Relying on the number of edges alone is not always sufficient, however. Better measures are obtained by looking at the amount of information that flows through a node. Such measures are called centrality measures. The two most important centrality measures are the *betweenness centrality* and *max-flow centrality*. In betweenness centrality, the shortest paths between all nodes in the networks are first calculated. Then, each node is ranked according to the number of such shortest paths that pass through it. The hypothesis is that a node through which a large number of shortest paths pass is more important than others. Many networks, however, are highly symmetrical, which means that there will be several short paths between nodes. For such networks, *max-flow centrality* is a better measure. Here, instead of just calculating the shortest paths, all paths are considered. Given a source node **s** and a target node **t**, the method computes the total flow and ranks the nodes according to the amount of flow **f** that passes through them. A *local rank* can thus be determined by fixing source and sink nodes, while a *global rank* is obtained by averaging also over all sources and sinks. The large computational complexity of the max-flow centrality problem (Wasserman & Faust, 1994; Newman, 2005) makes is necessary to consider also approximations to it. Actors that have high centrality are often called “stars” in social network literature.

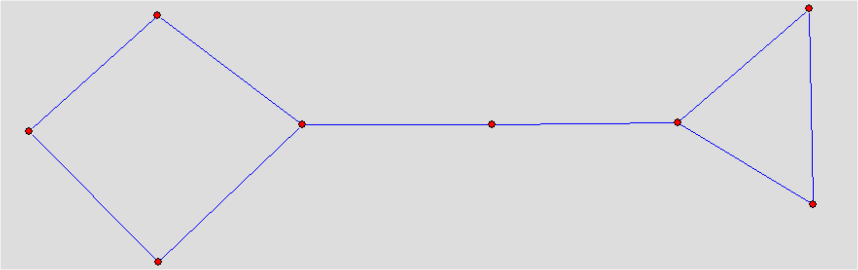

Figure 2. This figure shows a graph where the “bridge” in the middle has a high centrality.

Sociologists are often interested in actors that control the interaction between different groups. Such nodes are called “liaisons”, “bridges”, or “gatekeepers”, and can also be found by calculating the centrality measures.

Statistical analysis of large networks

Recently, many physicists and computer scientists have become interested in network analysis. This has led to an increased emphasis on studying the statistical properties of large networks, such as the internet, food webs, and even infrastructure networks (see (Svenson *et al.*, 2005) for an overview). This influx of people to the field has also led to several new approximate algorithms to compute important properties (Clauset *et al.* 2004; Newman, 2005).

Applications to crime analysis and prevention

Networks of relations between people, in some cases very large ones, will have to be set up: who knows whom, who has family relations with whom, as well as who met whom where and when, who phoned whom when. Figuring out nested business connections across the known set of individuals or organizations is a closely related issue. Since not all people who have had contacts with a criminal are criminal themselves, there is a strong need for techniques which can filter out those who have suspicious patterns of contact with a few known or suspected criminals, or with any member of a known or suspected group of criminals from a large database of contacts. People who have frequent such contacts become more or less suspect themselves, thereby potentially spreading the suspicion to even more individuals.

The papers (Carley *et al.*, 2002; Raab & Milward, 2003) provide examples of social network analysis in anti-terrorism applications and indicate both usefulness and some limitations of social network analysis as a basis for quantitative methods for situation awareness and decision-making in law enforcement applications. (Raab & Milward, 2003) discuss the organizational structure of certain drug trafficking, terrorism, and arms-trafficking networks, showing how some of them have adapted to increased pressure from states and international organizations by decentralizing into smaller units linked only by function, information, and immediate need. They also describe ways and structures of cooperation between different kinds of criminal networks, for example in financing terrorists by illegal diamond and drug trafficking. Another interesting application of social network analysis to terrorist networks is given by (Carley *et al.*, 2002). In this paper, the authors test several different methods for destabilizing terrorist networks.

4. Uncertainty management, information fusion and Bayesian Networks

In a law enforcement intelligence system, it is usually easy to define entries such as a piece of text, the name of a person, a set of coordinates, or a time. But how should the uncertainty or likelihood of each of these entities be estimated and entered into the system? Prediction and fusion of uncertain information, such as criminal agents’ likely courses of action, are examples where a degree of uncertainty is always involved. Thus, intelligence representation languages and systems need the ability to express and reason with incomplete and uncertain information. It must be possible for an intelligence analyst to manipulate uncertain data, to formulate hypotheses while taking uncertainty into account, and to test these hypotheses. Representation, management, and categorization of uncertainty in order to enable a machine to reason about potential relations are complex tasks. These are scientifically studied in the field of information fusion (www.isif.org) which provides methods for reasoning about information arising from several different uncertain sources (*Proceedings of the International Conferences on Information Fusion 1998-2005*).

Bayesian networks (BN) (Jensen, 2001) is one such uncertainty modelling and information fusion methodology used to represent and exploit uncertain causal relations between several variables. Information fusion methodology in general and BN in particular has several potential areas of application within the intelligence domain, for instance for detecting threatening insider behaviour, for probabilistic assessment of terrorist threats and for antiterrorism risk management.

By using BN methodology it is possible to deal with a large class of problems involving uncertainty in a uniform and scientifically correct manner. The BN methodology has several potential areas of application within the intelligence domain, for instance for detecting threatening behaviours by insiders (Bass, 2000), for probabilistic assessment of terrorist threats and for antiterrorism risk management.

5. Guidelines for the design of a problem-solving environment for law enforcement intelligence work

The HiTS/ISAC problem-solving environment for interoperability and situation awareness will be demonstrated and evaluated using realistic scenarios set up in cooperation with law enforcement authorities from several EU member states. In the scenario application the project will show how authorities may interoperate with information security over the network as well as illustrate how law enforcement authorities may cooperatively develop and share mission critical information across national borders.

Although database interoperability is only one of several interoperability issues that need to be addressed in a civil security intelligence system for routine operations, it is one of high technical complexity and critical importance. A modern approach for integration of heterogeneous data is to make use of mediators between data sources and those applications and software tools which use these sources (Fredriksson *et al*., 2001). Mediator systems enable automatic translation between the concepts and conventions, *schemas*, of different data sources, i.e., names and other characteristics of their data items as well as the semantic relationships between them.

In a wider context, the project strives to contribute to a deeper awareness and understanding of modern methodological opportunities among participating law enforcement authorities. These include on-demand method development based on reuse of components and subsystems in a state-of-the-art problem-solving environment. This approach should facilitate the establishing of creative, multi-disciplinary, multi-authority workgroups capable of real-time problem solving based on scientifically sound analysis methods.

Use of computerized analytical problem-solving methods needs to be based on modern open-ended and evolutionarily developing information technology. This technology should include in-depth knowledge about and access to collaborative software tools and systems, and modern analysis and inference methods. Also, integrated access to distributed legacy databases and secure, low latency and high bandwidth communications are enabling technologies of critical importance.

Organizationally, one needs to move away from "closed-room" approaches into collaborative working styles. Not only is trans-national collaboration needed between authorities in different security-related areas, such as police, coast guard, and customs services, but in order to enable effective use of modern analytical techniques and problem-solving methods there is a clear need also for cross-professional collaboration and involvement of scientifically trained analysts.

As indicated, there are many organizational as well as technological issues that need to be addressed to achieve the goals stated above. In addition, a set of hard and sometimes controversial legal issues also need to be solved. In particular, an intra-European terrorism protection legal system may need to be established which meets the following general requirements:

- graded accessibility depending on declared threat level to relevant confidential data in distributed, heterogeneous, multi-authority databases
- international, inter-authority, inter-professional cooperation based on mutual trust developed through qualified and sufficiently frequent common training sessions
- very short turn-around time for time-critical decisions in crisis situations.

6. Summary and conclusions

HiTS/ISAC is a pre-study of interoperability and situation awareness for civil security in Europe. HiTS/ISAC will demonstrate network analysis environments and tools with respect to immediate applicability and user-oriented functionality.

The HiTS/ISAC project deals with operational, methodological, and developmental issues in a demonstration scenario application. It focuses on scientifically sound analysis and secure management of distributed and usually uncertain information about events and relationships in terrorist and organized crime networks from a situation- and risk-aware decision-making perspective.

It will provide an open-ended demonstration platform on which necessary tools may be quickly built for the purpose of the scenario application. It will make plausible that this flexibility can be extended to a large number of other applications. In particular, the project will demonstrate how uncertainty management can be integrated into the analysis method development framework and used productively in investigative work. Also, it will demonstrate how legacy databases of several authorities in different countries can be effectively integrated into a problem-solving environment to provide a common user view of the combined relevant information assets of the participating authorities. Finally a road map for the way ahead in terms of technology and methodology as well as recommendations for future research will be outcomes from the project.

Key technological enablers for successful use of such methodology are distributed, mediated access to large amounts of legacy data and support for real-time collaboration among analysts.

We believe that lessons learned from this project will have the potential to profoundly influence operational processes, methods and techniques in European police intelligence work, in particular in high threat scenarios.

While the focus of the current work is on anti-terrorism, much of the methodology developed will be of comparable usefulness for military purposes in, for example, peace-support operations of the Nordic Battle group.

7. References

A. Aho, J. Hopcroft, J. Ullman (1983), *Data Structures and Algorithms*, Addison-Wesley.

T. Bass (2000). Intrusion detection systems and multisensor data fusion. *Comm. ACM* 43(4), pp. 99-105.

K. Carley, J.-S. Lee, D. Krackhardt (2002). Destabilizing Networks, *Connections* 24(3) 79-92.

A. Clauset, M. Newman, C. Moore (2004). Finding community structure in very large networks, *Physical Review* E70, 066111.

J. Fredriksson, P. Svensson, and T. Risch (2001). Mediator-Based Evolutionary Design and Development of Image Meta-Analysis Environments. *J. Intell. Information Systems*, 17:2/3, 301-322, 2001.

J. Gustavsson, J. Larsson, P. Svensson, H. Tullberg (2006). HiTS/ISAC – Highway to Security: Interoperability for Situation Awareness and Crisis Management, *this conference*.

F. V. Jensen (2001). *Bayesian networks and decision graphs*. Springer-Verlag, New York.

J. Leskovec, J. Kleinberg, C. Faloutsos (2005). Graphs over time: Densification Laws, Shrinking Diameters and Possible Explanations, *Proceedings of the 11th ACM KDD International Conference on Knowledge Discovery and Data Mining*.

M. Newman (2005). A measure of betweenness centrality based on random walks, *Social Networks* 27, 39-54.

Proceedings of the International Conferences on Information Fusion 1998-2005. International Society of Information Fusion, Mountain View, CA, USA.

J. Raab and H. B. Milward (2003). Dark networks as problems. *J. Public Administration Research and Theory*, 13(4), pp. 413-439.

S. Redner (2005). Citation Statistics from 110 Years of Physical Review, *Physics Today*, 58, 49.

P. Svenson, C. Mårtenson, C. Carling (2005). *Complex networks: Models and dynamics*, FOI-R—1766—SE.

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

Biographies

Dr Pontus Svenson is a member of the information fusion research group at the Swedish Defence Research Agency. His research interests include network analysis, data and text mining, random sets, sensor and resource management, and information operations. He has a PhD in theoretical physics from Chalmers University of Technology.

Dr Per Svensson is research director and a member of the information fusion research group at the Swedish Defence Research Agency. His research interests include network analysis, collection resource management, spatial information technology, and scientific database technology. He received a PhD in information processing in 1979 from the department of information processing, today the School of computer science and communication (CSC), KTH. Between 1996 and 2002 he was a part-time adjunct professor in scientific and statistical database technology at NADA, today CSC, KTH.

Dr Hugo Tullberg is a member of the radio system technology research group at the Swedish Defence Research Agency. His research interests include communication and information theory, error-correcting codes, graph-based systems, and iterative decoding and inference systems. He has a PhD in communication theory and systems from the University of California at San Diego.

Acknowledgements

The authors want to thank the HiTS/ISAC consortia partners, for sharing their expertise in the public safety area and for fruitful discussions during our work with this paper.