NETEST: Estimating a Terrorist Network's Structure

Matthew J. Dombroski
Carnegie Mellon University
mdombros@andrew.cmu.edu

Kathleen M. Carley
Carnegie Mellon University
kathleen.carley@cmu.edu

Abstract

Since the events of September 11, 2001, the United States has found itself engaged in an unconventional and asymmetric form of warfare against elusive terrorist organizations. Defense and investigative organizations require innovative solutions that will assist them in determining the membership and structure of these organizations. Data on covert organizations are often in the form of disparate and incomplete inferences of memberships and connections between members. NETEST is a tool that combines multi-agent technology with hierarchical Bayesian inference models and biased net models to produce accurate posterior representations of a network. Bayesian inference models produce representations of a network's structure and informant accuracy by combining prior network and accuracy data with informant perceptions of a network. Biased net theory examines and captures the biases that may exist in a specific network or set of networks. Using NETEST, an investigator has the power to estimate a network's size, determine its membership and structure, determine areas of the network where data is missing, perform cost/benefit analysis of additional information, assess group level capabilities embedded in the network, and pose "what if" scenarios to destabilize a network and predict its evolution over time.

Key Words: Covert Networks, Terrorist Organizations, Bayesian Inference Models, Biased Net Theory, Biased Networks, Multi-agent Systems, Network Estimation, Social Network Analysis

This research has been supported, in part, by National Science Foundation under the IGERT program for training and research in CASOS and the ONR N000140210973. Additional support was also provided by the Center for Computational Analysis of Social and Organizational Systems and the Department of Engineering and Public Policy at Carnegie Mellon. The views and conclusions contained in this document are those of the author and should not be interpreted as representing official policies, either expressed or implied, of the Office of Naval Research, the National Science Foundation, or the U.S. government.

Citation: Dombroski, Matthew and Kathleen M. Carley, 2002, "NETEST: Estimating a Terrorist Network's Structure," CASOS conference proceedings, Pittsburgh, PA.

NETEST: Estimating a Terrorist Network's Structure

Matthew J. Dombroski
Kathleen M. Carley

The events of September 11, 2001 have illustrated the need for defense and investigative organizations to prepare and innovate for asymmetric and unconventional warfare in the 21st century. Our military and intelligence communities are unprepared to counter the elusive and deadly threat posed by terrorist and other covert organizations in this new century [PAM 525-5, 1994]. The inter-state model of warfare has been replaced by intra-state warfare, which largely consists of guerilla and terrorist forms of warfare [Smith, Corbin, and Hellman, 2001]. Military and intelligence organizations in the United States are adept at utilizing training, tools, and weapons designed to counter conventional threats to national security. However, the enemies of the United States have evolved into network forms of organization that do not obey traditional borders between states and employ unconventional techniques. A networked form of organization promotes the terrorists' covert nature and decentralizes the terrorist association, allowing parts of the organization to effectively operate almost autonomously from the rest of the organization. These deadly forms of organization employ swarming techniques that empower individuals and groups of individuals to remain "sleeping" until the time is right for a well-coordinated and powerful attack [Ronfeldt and Arquilla, 2001].

Terrorist Organizations and the Ability to Track Them

Modern terrorist organizations have learned that they can effectively counter much larger and conventional enemies using dispersed and networked forms of warfare, striking when their target is least likely to expect it. Large terrorist organizations, such as Osama bin Laden's al Qaeda and Hamas, operate in small, dispersed cells that can deploy anytime and anywhere [Ronfeldt and Arquilla, 2001]. Dispersed forms of organization allow these terrorist networks to operate elusively and secretly. There are several factors that allow a terrorist organization to remain covert. Two important factors are:

- Members sharing extremely strong religious views and ideologies that allow them to form extremely strong bonds between one another within a cell
- Members are hidden from the rest of the organization, and likely do not know much about the organization's structure, except for the cell to which the members belong.

Even if a member of the organization is detained, these factors hamper investigators' attempts to break the organization down. Investigators are finding that the tools that they have available are not adequate for the task at hand, which is to break down these organizations and permanently remove their capabilities to spread fear, crime, and terror.

Often the only data that is available on terrorist networks is generated from investigations from isolated events. Investigations provide a partial picture of the network, relating the individuals involved with higher levels in the organization through a money trail, resource trail, or communication ties. Investigators would benefit greatly from tools and techniques that combine these partial and disparate data sources into a larger picture providing valuable insight on the size, membership, structure, and collective capabilities of the network and its cellular components. Such tools will provide powerful insight to investigators, displaying explicitly where data is known, and where it is not known. Such tools and techniques will allow investigators to determine where investigations should proceed and it will allow investigators and the military to systematically attack dangerous terrorist organizations, thereby effectively disrupting their activities and breaking the network apart.

Network Estimation Using a Bayesian Approach

Hierarchical Bayesian models provide simultaneous inference of informant accuracy and social structure [Butts, 2001]. Research indicates that informant data has error in it and, taken alone, are not accurate representations of social interaction. However, Bayesian inference models effectively aggregate informant data into an accurate big picture of the social network, thereby providing valuable insight on social structure and network size. There is a great deal of uncertainty that must be dealt with when building an accurate network of social interaction. Because social interaction is often subjective and difficult to accurately measure, inferences will always be vulnerable to error and uncertainty. However, the goal of hierarchical Bayesian models is to minimize and account for all uncertainty associated with drawing inferences regarding social interaction [Butts, 2001].

To effectively describe the approach, a basic assumption is that a Bernoulli graph exists of actual interactions represented as A, where A_{ij} is either 1 or 0, representing a link or no link between nodes i and j respectively. Each observation of A from each informant is a data matrix Y. A network prior (Θ) is provided to represent all prior information on what the larger network (A) looks like. Θ is updated using a Bayesian updating procedure that aggregates each observation Y with the network prior Θ.

Another basic assumption of the model is that informants make two types of errors; reporting a tie where a tie does not exist (false positive) or reporting no tie where a tie does exist (false negative). In the case where the false positive (e^+) and false negative (e^-) probabilities are fixed and known, the data generation process can be represented using the Bernoulli mixture [Butts, 2001]:

$$p(Y_{ij} \mid A_{ij}) = \begin{cases} Y_{ij}e^{+} + (1-Y_{ij})(1-e^{+}) & \text{if } A_{ij} = 0 \\ Y_{ij}(1-e^{-}) + (1-Y_{ij})e^{-} & \text{if } A_{ij} = 1 \end{cases}$$

Applying Bayes Rule, to each arc, the posterior probability for the existence of an arc is:

$$p(A_{ij} \mid Y_{ij}) = Y_{ij}\frac{\Theta_{ij}(1-e^{-})}{\Theta_{ij}(1-e^{-}) + (1-\Theta_{ij})e^{+}} + (1-Y_{ij})\frac{\Theta_{ij}e^{-}}{\Theta_{ij}e^{-} + (1-\Theta_{ij})(1-e^{+})}$$

Posteriors are updated by simply factoring in each observation of the network, and calculating the posterior using the initial posterior as the network prior for each update.

If each error parameter is fixed and known for each observation, then the network posterior can be calculated using the following procedure:

$$p(A_{ij} \mid Y_{ijk}) \sim \frac{\theta_{ij}\prod_{k=1}^{N}(Y_{ijk}(1-e_k^{-}) + (1-Y_{ijk})e_k^{-})}{\theta_{ij}\prod_{k=1}^{N}(Y_{ijk}(1-e_k^{-}) + (1-Y_{ijk})e_k^{-}) + (1-\theta_{ij})\prod_{k=1}^{N}(Y_{ijk}e_k^{+} + (1-Y_{ijk})(1-e_k^{+}))}$$

More complex models are developed when error rates are unknown and different for each observation. Priors must be assigned for the parameters of the distributions of the error probabilities.

If the error parameters can be represented using Beta distributions, then the following error priors are assigned:

e^+~Beta(α^+,β^+)
e^-~Beta(α^-,β^-)

By relaxing the assumption that error parameters are fixed and known, the posterior cannot be solved for as it was in the simple model. The error parameters are needed to solve for the social structure and the social structure is needed to solve for the error parameters. To counter this problem, posterior draws are simulated using a Gibbs sampler. Samples are taken from the full conditionals of the parameters in the model, constructing a Markov chain, which eventually converges to the joint posterior distribution. The full conditional of the social structure is given by the equation above. The full conditional for false positives is given by:

$$p(e^{+} \mid \Theta, e^{-}, Y) \sim \prod_{k=1}^{N} Beta\left(\alpha_k^{+} + \sum_{i=1}^{N}\sum_{j=1}^{N}(1-\Theta_{ij})Y_{ijk}, \beta_k^{+} + \sum_{i=1}^{N}\sum_{j=1}^{N}(1-\Theta_{ij})(1-Y_{ijk})\right)$$

The full conditional for false negatives is given by:

$$p(e^{-} \mid \Theta, e^{+}, Y) \sim \prod_{k=1}^{N} Beta\left(\alpha_k^{-} + \sum_{i=1}^{N}\sum_{j=1}^{N}\Theta_{ij}(1-Y_{ijk}), \beta_k^{-} + \sum_{i=1}^{N}\sum_{j=1}^{N}\Theta_{ij}Y_{ijk}\right)$$

Draws are taken from each of these conditional distributions to implement the Gibbs sampler. The posterior distributions of the social structure and the error parameters are then constructed, building a big picture of the network and accounting for all uncertainty in estimating the network [Butts, 2001]..

Virtual Experiment Using the Bayesian Model

Now that the model has been introduced, the accuracy of the model and its usefulness must be determined. Two experiments were conducted to understand the interaction of informants and their accuracy in the generation of the posterior.

The first experiment examines the interaction between number of informants and their accuracy in generating an accurate posterior. For this experiment randomly generated informant observations were developed for the Roethlisberger & Dickson Bank Wiring Room friendship data. The observations were then aggregated using the Bayesian model and Hamming distances and absolute errors were calculated from the actual network. In the experiment, the number of informants was varied from 5 to 10 to 15 and their prior error distributions were varied from Beta (2,20) to Beta (2, 10) to Beta (2,5). Results of this experiment indicate that the number of informants plays a very significant role in reducing the error in the posterior. The posteriors were highly accurate for 15 informants regardless of what the prior error probabilities were.

A second virtual experiment was conducted to test whether increased samples given to informants produce more accurate posteriors. In this experiment, 10 informants were simulated. The actual network was the central graph of Krackhardt's Cognitive Social Structure data (1987). In each experiment informants were given a sample of the actual network and the observations were aggregated into the posterior. Samples of the actual network given to the simulated informants included 0, 20, 90, and 210 nondiagonal dyads of the network. All other dyads in the observations were generated using a prior error distribution of Beta (2,10). Although Krackhardt's CSS data provides informant reports, these were only used to generate the actual network, not for the informants. Results indicated that the posteriors only improved significantly for large samples (210 nondiagonal nodes). Absolute errors improved, even slightly, for increasingly large samples, but Hamming distances did not always decrease with larger samples. The reason for this phenomenon is that the Bayesian model assumes that errors are uniformly distributed across the observations. The experiment puts this assumption to the test since in real data, errors are not uniformly distributed across observations. Informants report most accurately on the portions of the network that they are most familiar with and the informants will also provide data on other portions of the network that they do not have extensive knowledge on.

Incorporating Biased Net Theory

According to biased net theory, both chance and bias create ties between population elements. Structural bias ties result from the pattern of links between nodes and not on the properties of the nodes themselves. The Bayesian models may be made more accurate if terrorist biases can be identified and incorporated into the models. Several levels of abstraction exist where biases may be identified at the cellular level, the organizational level, or the overall terrorist level. For example, Hamas may contain several structural biases that are distinct from Al Qaeda. This is an example of structural biases at the organizational level. If data analysis can identify and verify that these biases exist, then powerful inferences can be made about the structure of terrorist organizations and once these biases are incorporated into the Bayesian model, priors can be updated using biased random networks or the network priors can incorporate these biases, thereby building a more accurate posterior.

Three types of structural biases are examined. The reciprocity bias, π, refers to the state where an A to B link is more likely if node B is connected to node A. The triad closure bias, σ, refers to the state where an A to B link is more likely if both A and B have a common parent node, C, where C is connected to both A and B. The triad reciprocity bias, ρ, refers to the state where an A to B link is more likely if both A and B have a common parent node, C, and there is a link from B to A [Skvoretz, 1990].

In order to quantify the likelihoods of these different biases, we have to assess the probabilities that ties are mutual, asymmetric, or null between all nodes in the network. For dyads with no parents:

$$P(M) = d^2 + d(1-d)\pi$$

$$P(A) = 2d(1-d)(1-\pi)$$

$$P(N) = (1-d)^2 + d(1-d)\pi$$

where d is the probability that a dyad occurs by chance alone. The equations become more complicated when parents are incorporated. For dyads with one parent, the equations are:

$$P(M) = \sigma^2 + \sigma\rho(1-\sigma) + d(1-\sigma)(2\sigma(1-\rho) + \rho) + d^2(1-\sigma)^2(1-\rho)$$

$$P(A) = 2\sigma(1-\sigma)(1-\rho) + 2d(1-\sigma)(1-\rho)[(1-\sigma)(1-d) - \sigma]$$

$$P(N) = (1-\sigma)^2 + \sigma\rho(1-\sigma) - d(1-\sigma)[(1-\sigma)(2-\rho) + \sigma\rho] + d^2(1-\sigma)^2(1-\rho)$$

For dyads with k parents, $\sigma'=1-(1-\sigma)^k$ and $\rho'=1-(1-\rho)^k$ and σ' and ρ' would substitute in for σ and ρ in the above equations. By analyzing existing and verifiable network data and using the above equations, the three biases and the probability of a dyad by chance can be evaluated for certain types of networks [Skvoretz, 1990].

Virtual Experiment Using Biased Net Theory

A set of terrorist network data was analyzed to determine if biases were present in the al Qaeda terrorist network. The goal of the experiment was to determine if the probabilities for the different bias types were significant, and if so, then we could conclude that al Qaeda exhibits certain predictable and identifiable biases in its structure, based on this data set. The data set consisted of 74 nodes generated by Valdis Krebs (2001). The data set was the network of ties among the 19 hijackers from September 11, 2001 and their associates.

The nodes in each set of data were analyzed using the biased net theory equations shown above, and fitted using a grid search algorithm approach. The results for the data sets were:

Parameter	Result Krebs	Description
d	0.0097	Probability that link occurs by chance alone
π	1	Probability that a reciprocity bias exists
σ	0.1398	Probability that a triad closure bias exists
ρ	1	Probability that a triad reciprocity bias exists

The Krebs data set had a chi square value of 32.07 with 16 degrees of freedom. The fit is not very good. However, the results do indicate that reciprocity biases do exist in this network and triad closure biases result more from structural biases than by chance. Future research will examine other terrorist network data to see if these biases are consistent across different networks and comparisons will be made between different organizations and the biases that are present in those organizations.

These preliminary results indicate that structural biases may exist. The results are only preliminary and continued analysis will link these structural results with compositional biases, which result from properties of the nodes, to improve the Bayesian models and their accuracy.

References

[Butts, C. 2001. “Network Inference, Error, and Informant (In)Accuracy: A Bayesian Approach,” to appear in *Social Networks*]

[Krackhardt D. 1987. “Cognitive social structures,” *Social Networks*, Vol. 9, pp. 104-134]

[Ronfeldt, D. and J. Arquilla. September 21, 2001. “Networks, Netwars, and the Fight for the Future,” First Monday, Issue 6 No. 10. online: http://firstmonday.org/issues/issue6_10/ronfeldt/index.htm.]

[Skvoretz, J. 1990. “Biased Net Theory: Approximations, Simulations, and Observations,” *Social Networks*, Vol. 12. pp. 217-238]

[Smith, D., M. Corbin, C. Hellman. 2001. “Reforging the Sword: Forces for a 21st Century Security Strategy,” Center for Defense Information, online: http://www.cdi.org/mrp/reforging-full.pdf]

[TRADOC Pam 525-5. August 1, 1994. “Force XXI Operations,” Department of the Army, Headquarters, United States Army, Training and Doctrine Command, Fort Monroe, Virginia]