Dynamic elicited priors for updating covert networks

JEFF GILL and JOHN R. FREEMAN

Network Science / Volume 1 / Issue 01 / April 2013, pp 68 - 94
DOI: 10.1017/nws.2012.6, Published online: 15 April 2013

Network Science **1** (1): 68–94, 2013.
doi:10.1017/nws.2012.6

Dynamic elicited priors for updating covert networks

JEFF GILL

Department of Political Science, Department of Biostatistics, and Department of Surgery (Public Health Sciences), Washington University, One Brookings Drive, Seigle Hall, St. Louis, MO 63130, USA
(*e-mail:* jgill@wustl.edu)

JOHN R. FREEMAN

Department of Political Science, University of Minnesota, 1414 Social Sciences Bldg, 267 19th Avenue South, Minneapolis, MN 55455, USA

Abstract

The study of covert networks is plagued by the fact that individuals conceal their attributes and associations. To address this problem, we develop a technology for eliciting this information from qualitative subject-matter experts to inform statistical social network analysis. We show how the information from the subjective probability distributions can be used as input to Bayesian hierarchical models for network data. In the spirit of "proof of concept," the results of a test of the technology are reported. Our findings show that human subjects can use the elicitation tool effectively, supplying attribute and edge information to update a network indicative of a covert one.

Keywords: *covert networks, Bayesian elicitation, link prediction*

1 Objectives

The safety of millions of people depends on the understanding of the workings of covert networks, especially of terrorist networks. To protect people, governments and nongovernmental organizations invest enormous amounts of time and energy to detect covert networks and to thwart terrorist and other kinds of attacks. The most fruitful contribution of currently published work is the recognition that terrorist organizations are not unitary actors, and that they are highly decentralized networks. However, there are two major deficiencies in academic studies of terrorism: all available data describe only publicly visible events, and missingness in network nodes and edges is generally treated as a regular statistical missing data problem.

Despite the recognition that terrorists should not be treated as unitary actors (Chai, 1993; Crenshaw, 1981), the study of terrorist organizations as networks is much less developed. Social network analysts have discovered that covert organizations tend to be cellular and distributed rather than hierarchical (Carley, 2004; Krebs, 2002; Rothenberg, 2002). Such networks are also dynamic and adaptive. Carley (2003, 2006) has developed methods to track the evolution of terrorist networks over time from semi-automatic parsing of signal traffic (see also Tsvetovat & Carley, 2005, 2006; Moon & Carly, 2007).

The analysis of terrorist networks presents unique challenges distinct from other social network research (Sparrow, 1991). Much work of this work investigates networks of individuals in which the data are assumed to be essentially complete, yet networks of criminals are rarely fully recorded. Consider, for example, an observed nefarious actor who is going about his or her day interacting with a variety of people. We can assume that someone engaged in terrorist activities or planning to engage in terrorist activities is concerned about being observed. Therefore, the regular interactions that are observed are not likely to be the interactions of real interest. In fact, it is likely that the interactions with the individuals of interest are the most difficult to observe. Therefore, näive network analysis would place the highest emphasis on the least important edges and the lowest emphasis on the most important edges. Thus, the key problems we face include (1) *actual size*, criminal intelligence databases can be quite large, up to thousands of individual nodes; (2) *uncertain boundaries*, the delineation of network borders is often ambiguous; (3) *dynamics*, criminal networks continuously change to avoid observation; and (4) *incompleteness*, missing edges and notes in the data will be "anything but random" (Sparrow, 1991). In this work, we concentrate on the last challenge, developing a new approach that leverages expertise that resides in subject-matter experts to reduce missingness over time and therefore steadily improve our knowledge about particular terrorist enterprises.

1.1 Missing data in network analysis

Networked terrorist organizations are fundamentally distinct types since the actors are "trading efficiency for secrecy" (Fellman & Wright, 2003; van Meter, 2002). The key problem is that unlike many other networks we study, for strategic reasons, the humans in covert networks conceal not just their identities but also their interactions with other network members. Hence, data for covert networks are often missing a large proportion of the potentially observed values (Krebs, 2002; Tsvetovat & Carley, 2006), especially on the boundaries. Furthermore, the edges of interest typically have multiple qualitative aspects and are likely to change over time, perhaps without direct observation. Although the connections between individuals change character temporally, such changes typically cannot be measured directly; therefore, they require some form of indirect estimation.

The missingness (network sampling problem) has serious *underestimation* consequences for traditional social network analysis. First, the strength of the relationship between any two nodes is underestimated for both missing edge and missing node information (Burt, 1987). Node degrees (number of edges incident to the node) are also consistently underestimated (Costenbader & Valente, 2003). The scope (measured in various ways such as the diameter or geodesic distance) of the network is underestimated (Kossinets, 2006). Relatedly, standard centrality measures are unstable and inconsistent (Costenbader & Valente, 2003).

Major advancements have been made in statistics for overcoming missingness in rectangular data sets, although there are few of these appropriate for social network data. In conventional problems, multiple imputation is the leading technique for dealing with the general missing data problem (Rubin, 1976; Little & Rubin, 1983, 2002). In social network analysis, studies of the robustness of measures like centrality to missing data, and of the impact of boundary misspecification, survey nonresponse, and censoring have been conducted (Borgatti et al., 2006; Kossinets, 2006) and evaluated (Huismann & Steglich, 2008). The problem here is that these advances usually are based on the assumption that the missing data are missing completely at random (MCAR) or missing at random (MAR). This assumption may be reasonable for some standard social networks of interest. *But terrorist and other kinds of covert network data are missing by design; these kinds of data are better characterized as non-ignorable, which violates an assumption required by multiple imputation.*

Statisticians and social scientists often stress that non-ignorable missing data require sufficiently strong assumptions and case-specific approaches that readers are likely to find the results unrealistic (Schafer, 1997). In some cases, these assumptions leverage network statistics to attempt to treat non-ignorable missing data (the general rule in network missing data) as ignorable missing data (Handcock & Gile, 2010). Ward et al. (2003) use unique features of international relations data to leverage a latent space construct to judge whether or not an absent link may be missing at random. Also, many social network analysts frame the problem in terms of *survey* network analysis. Their key assumption is that, although outward ties and related information may be missing for nonrespondents, there are a group of respondents who provide revealing information about inward ties to these nonrespondents. This available information then can be treated as exogenous in such formalisms as exponential random graph models (Robins et al., 2004). In theoretically important cases like the Irish Troubles (1969–1998) such data with incoming and outgoing ties may not even exist.[1] This means that conventional tools like single and multiple imputation, hot-decking, weighted methods, and maximum likelihood estimation with the expectation-maximization (EM) algorithm fare poorly in this context (see Little & Rubin, 2002, for detailed comparisons of these methods).

The idea of elicitation to inform missing data modeling was introduced by Savage (1954) and it now is well established in statistics. Evidence that humans are capable of producing meaningful estimates of uncertainty and of making (Bayesian) predictions for different kinds of variables continues to be found (Griffiths & Tenenbaum, 2006).[2] Priors that are elicited from subject-matter experts have a range of different characterizations depending on the literature addressed. Kass & Greenhouse (1989), for instance, coined the phrase "community of priors" to describe the range of confidence or belief that equally qualified experts can have about the same effect. Furthermore, these priors do not need to be independent as if such experts did not have commonalities such as their employers, educational background, analytical orientation, historical perspective, and so on.

[1] Still another body of work models network ties using bilinear random effects models (Hoff & Ward, 2004). In this case, missing data for units of analysis which no longer exist (Czechoslovakia, for instance) are imputed with matrix regression on known covariates. The missing data problem then is the absence of covariates (attributes).

[2] Kadane & Wolfson (1998) make the distinction between elicitation for prediction of outcome variables (for imagined values of the explanatory variables) and elicitation of "structural" parameters. Examples of the former include Garthwaite & Dickey (1988, 1992) and Kadane et al. (1980) whose work points to the need for particular controlled experiments for industrial chemical research.

2 Building a data collection system based on prior elicitation

To solve the problem of missing data in covert networks, we develop a Bayesian technology for eliciting information from experts assessing terrorist attributes and the structure of terrorist networks. Our technology uses a new form of visualization to elicit subjective probability distributions (SPDs). We show how the information in these SPDs can be used as input in Bayesian hierarchical models of social networks. This builds on the successful Bayesian elicitation work of O'Hagan (1998), who elicits prior expert information for hydrogeological analysis; Kadane & Wolfson (1998), who elicit prior expert information from an applied statistician working with childhood mental health centers and a macro-economist forecasting gross national product (GNP); Bedrick et al. (1997), who elicit prior expert information from an emergency room physician; Freedman & Spiegelhalter (1983), Carlin et al. (1993), and Jennison & Turnbull (1990) who all elicit prior expert information from researchers currently running clinical trials; and Garthwaite & Dickey (1992), who elicit prior information from an industrial chemist. Other revealing works include Kadane & Winkler (1988) and Savage (1971), both of which analyze the linkage between probability assessments and underlying utility.

2.1 The technology: Basic features

Consider attribute "strength" to be a value between zero and one. Therefore, zero means that there is no evidence of an attribute whatsoever and one means that we know for certain that there is a qualitatively strong relationship. A useful prior from an analyst, therefore, would be a number between zero and one that describes his or her probabilistic assessment from nonstatistical research that there exists an attribute. We first describe the multistep process intuitively for getting such an assessment:[3]

▹ **Analyst stage**

1. An analyst at a supported location logs onto the system and picks a particular edge to assess (subject A to subject B), or just an individual for whom there is new information to input.
2. Next, the analyst picks an attribute to assess. This could be sex, age, country of residence, country of birth, religion, etc.[4]
3. For the selected attribute or edge the analyst is asked a simple question such as:

   "On a scale of zero to one-hundred, what is your best estimate of the strength of this attribute (edge)?"
4. The follow-up question addresses certainty:

   "Provide the lowest and highest reasonable values that you can imagine around your estimate for the individual or relationship in question."
5. The analyst is then shown graphically on the terminal a probability distribution that results from these statements. She then uses a "slide" to modify it in terms of central location and width (until satisfied that the distribution on the screen accurately represents her beliefs about the probability that the attribute exists for the actor in question).

[3] We assume here that a group of experts already has been hired and that this group is under the direction of a manager. We also assume that the manager faces no budget constraints. The problem of hiring a pool of experts subject to fixed costs—of choosing a representative subset of "the crowd"—already has been solved. On this problem see, for instance, Ertekin et al. (2012).

[4] If the analyst has picked a relationship instead of an individual, then the elicitation will be on the strength of this relationship.

Of course the analyst is likely to assess a variety of attributes and edges for more than one actor in the same sitting. Therefore

▹ **Aggregation stage**
1. Each participating analyst is assigned a weight that reflects his or her expertise and demonstrated reliability.
2. The aggregate assessment for an attribute is updated using Bayes law (with weighting), thus updating the current state of thinking without ignoring previous assessments.

▹ **Production stage**
1. Weights are now assigned across elicitees, weights that reflect those elicitees' relative knowledge and reliability. These weights can differ by sub-network, region, etc.
2. A multidimensional prior structure is used to create link predictor information for a Bayesian social network model for latent variables.

The end result is an aggregated set of inputs that produce a single prior distribution for statistical network analysis.

2.2 Visualization

Over many thousands of years humans developed the capability to make judgments about uncertainty on the basis of ocular rather than numerical information. The relevant cognitive mechanisms are based on visual rather than numerical inputs (Gigerenzer & Hoffrage, 1995; Cosmides & Tooby, 1996). For this reason, cancer researchers, engineers, and other scholars are designing visualization schemes for the communication of risk (Lipkus & Holland, 1999; Ibrekk & Morgan, 1987). Chaloner et al. (1993) present strong evidence of the superiority of visual tools for eliciting information from substantive experts as opposed to statistically trained participants. Here, we present a `JAVA`-based, platform-independent software technology for elicitation based on research of the relevant literature: expert elicitation, psychological processing of queries, computer interface design, and Bayesian processing of prior information. Our freely available software package, `Elicit`, has been shown to be effective in the experiment described here.

The novelty of our visualization lies in its simplicity and ease of manipulation. Subjects see a single distributional object on the screen. They can freely change the shape of this object by moving two sliders, a task with which many humans now are familiar. The shape of the object is intuitive and its form responds immediately to subjects' movements of the sliders. The interpretive challenge is condensed down to a single issue: training the elicitees to think about uncertainty in distributional terms. Therefore, using this tool, we provide the appropriate cognitive mechanism needed for elicitation in a way and extent that conventional elicitation techniques do not.

2.3 Automation and training

From the beginning, elicitation methods have been interactive. Traditionally, researchers devote a great deal of time coaching subjects in how to complete elicitation exercises. Sometimes entire days were spent training subjects to produce the needed SPDs (O'Hagan, 1998; Winkler, 1967; Spetzler & Staël von Holstein, 1975). This very interpersonal, and therefore uneven, process creates doubt as to whether training treatments are identical in a statistical sense. Therefore, the number of subjects given the training was then quite small, often on the order of 20–25. More important, sometimes little or no information about the content of this teaching and training follows the project into publication. Performance of some subjects might be, in part, a consequence of some unknown feature of a particular training process. When the goal was to produce accurate assessments of uncertainty, variations in (unsystematic) training treatments could confound inferences about the effectiveness of certain protocols.

The problem of training unevenness led to computerized elicitation, an effort that paralleled the development of computerized protocols in other fields such as medicine (Goldman et al. 1988). The advantages of automation are threefold. First, it allows the researcher to train subjects systematically. Each subject is given exactly the same instructions. Or, subjects can be randomly assigned to different training routines. In these ways, researchers can control the training treatment. Inferences about performance are not confounded by differences in the way subjects were coached through the exercise. Second, automation allows researchers to monitor subjects. Mistakes or confusion in completing the elicitation exercise can be illuminated and recorded. The time it takes for each subject to complete each task can be recorded as well. Finally, automation allows researchers to enlarge the subject pool greatly. Many subjects can perform the elicitation exercise at the same time. Our approach makes full use of such automation for training and elicitation.

2.4 Evaluation

Generally speaking, elicitation is evaluated on normative and substantive grounds (Schaefer & Borcherding, 1973). The former amounts to "conformance to the axioms of probability theory." For example, researchers seek to determine if the probabilities subjects assign to the possibilities that a variable assumes a range of values sum to one and that alternative methods of elicitation produce the same SPD. The objective accuracy of the elicitations is the object of study in substantive evaluations. An example is the extent to which actual observations fall into the middle quartiles of the subject's SPD or in the bottom and top 1% of the SPD. These are called the interquartile range (IR) and interquartile score (SR) criteria, respectively. Although scholars often emphasize that elicitations are subjective judgments and not objective assessments of uncertainty, they themselves employ both criteria in their evaluation of subjects' performance.[5] These evaluation studies show that humans can be biased in several ways: too much emphasis on recent experiences or on single pieces of information relative to the given elicitation, overestimating the probability of rare events, underestimating the probability of very frequent events, poor assessments of events in the tails of actual probability distributions, anchoring, hindsight bias, and displacement bias by groups of subjects (Schaefer & Borcherding, 1973; Tversky & Kahneman, 1982; Garthwaite et al., 2005). We know, however, that these effects are much better controlled through computer-based and interactive elicitation tools

[5] Schaefer & Borcherding (1973) emphasize the normative criterion, more specifically, teaching their subjects how to use a direct and indirect method to produce the same SPD. However, they end up ranking subjects on the IR and SR criteria and also in illuminating objective biases in the subjects' elicitations (displacements). For a review of studies using these measures, see Lichenstein et al. (1982, 323ff).

like those developed here. Furthermore, a wealth of studies since the seminal alarm of Tversky & Kahneman (1974) in *Science* has demonstrated that careful framing of probability questions and the use of computational tools leads to reliable and consistent elicitations (Kynn 2007).

2.5 The beta distribution as an elicitation tool

When the analyst begins the elicitation exercise, he or she sees a unimodal form of the beta distribution. The beta distribution is used because of its natural simplicity and flexibility (O'Hagan, 1998). The support of the beta distribution is restricted to $[0:1]$, making it ideal for modeling existence claims in a probability context. Values in-between zero and one reflect differing estimated levels that are easily interpreted due to the anchoring of the endpoints. In addition, the probability density form of the beta distribution can represent a wide range of beliefs over a finite support ranging from a high degree of certainty (leptokurtic, peaked density), to relative certainty (perhaps a skewed density), to complete uncertainty (a uniform distribution). O'Hagan (1998) used a more informal, method of bounds, approach to construct a histogram of his hydrological engineering experts' collective belief, fit a beta distribution to this histogram, and then had these elicitees adjust this beta distribution to better represent their beliefs. Gavasakar (1988) compares the use of the beta distribution to the more typical early approach of quantile matching and finds that the beta distribution is more robust to elicitation errors. Meyer & Booker (2001) feature the beta distribution as their standard description of prior uncertainty from experts in their book-length prescriptive study. More recently, Gill & Walker (2005) demonstrate that the beta distribution is central to the Bayesian approach to prior elicitation in their study of judicial attitudes in Nicaragua.

The unimodality assumption for elicited priors also is common in the literature. Researchers have found that humans usually have little trouble providing a single most likely value of a variable. Also, humans naturally produce probability densities that are unimodal (Winkler, 1967).[6] Closely related methods like predictive modal elicitation (Chaloner & Duncan, 1983, 1987) make this assumption in inferring the prior of a beta-binomial discrete, predictive distribution. Chaloner et al. (1993) use unimodality in eliciting parameters of more complex probability densities in a clinical trial.[7]

The mathematical properties of the unimodal beta distribution make it an ideal elicitation tool for our covert network analysis. We require a graphic, descriptive means to put distributional decisions in front of subject-matter experts. The fixed and finite scale (zero to one) ensures comparability across elicitations. The unimodality removes a large number of psychologically difficult decisions. The comparable sense of variability or dispersion provides an intuitive measure of uncertainty,

[6] Winkler (1967) found that his subjects had more difficulty constructing the cumulative density functions (CDFs) for their densities. These CDFs were not unimodal. However, he attributed this to his subjects' lack of understanding of the idea of a distribution function. On this problem, see also Alpert & Raiffa (1982), especially pp. 302 and 303.

[7] Chaloner et al. (1993) use "dynamic screen displays" like ours to elicit coefficients in a proportional hazard model in a randomized trial comparing prophylaxes for toxophasmosis in a population of HIV-positive individuals.

given some training. Thus, when we want to inform networks with potentially large amounts of missingness, where some of this is structural in the form of missing edges and nodes, we are burdening our nonstatistical experts with the least amount of statistically technical work. In fact, our design eliminates the need for a discussion of infinite support for the density, an explanation of location-scale parameterizations, discussion of quantile decisions, interpretation of arbitrary scale values, and comparisons between elicitations on perceived different scales.

2.6 Construction of the prior distribution

The aggregated multistep elicited prior is built with the *general* form of the beta probability density function

$$f(y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{(y-a)^{\alpha-1}(b-y)^{\beta-1}}{(b-a)^{\alpha+\beta-1}},$$

where $a < y < b$ and $\alpha, \beta > 0$. This form generalizes the beta distribution by making the support of the random variable any region of the positive real line rather than just $[0:1]$. To be absolutely clear, this is not the *multivariate generalized beta distribution* (Libby & Novick, 1982) that introduces a scale parameter not included in the basic form. This is also not the *Dirichlet* distribution, which is the multidimensional version of the beta distribution. If we wanted to generalize our process further to accommodate simultaneous multichotomous assessments, rather than the more basic dichotomous assessments described here, we *would* use a Dirichlet distribution (in a similarly generalized form) but this is not addressed at the current time. The general form of the beta density operates on a support bounded by $[a:b]$ which are user-specified limits greater than zero, and it easily reduces to the standard form with a change of variable calculation, $X = (Y-a)/(b-a)$, so that $0 < X < 1$, but α and β are unchanged. It is easy to go back and forth between X and Y since the inverse expression is also a linear form: $Y = (b-a)X + a$.

This is a useful distribution for our purposes of elicitation because we can now talk to qualitative experts on the $[0:100]$ scale *and* let them put their own reasonable limits on the support by stipulating a and b parameters. Yet, this does not prevent us from performing analysis with the standard beta distribution, which, again, is flexible in form and has the bounds we require. The computational complexity is entirely written into the software. It therefore is not a burden for elicitation supervisors or for downstream analysts.

If X is a standard beta random variable and Y is the same random variable expressed on our preferred scale, then we will make use of the following properties of the beta distribution:

$$\text{mean:}\quad \mu_x = \frac{\alpha}{\alpha+\beta} = \frac{\mu_y - a}{b-a}$$

$$\text{variance:}\quad \sigma_x^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{\sigma_y^2}{(b-a)^2},$$

where the last expression gives us the reverse: $\sigma_y^2 = (b-a)^2\sigma_x^2$. Solving with the equations above gives us the following two important relations (Law and Kelton 1982, 205):

$$\alpha = \left[\frac{\mu_x(1-\mu_x)}{\sigma_x^2} - 1\right]\mu_x \qquad \beta = \left[\frac{\mu_x(1-\mu_x)}{\sigma_x^2} - 1\right](1-\mu_x).$$

This calculation is vital because it gives us the values of the only two parameters of interest for the elicitation. Therefore, if the user gives us the mean and variance of their assessment, we know everything about the beta distribution as a prior for input into a Bayesian updating process. Unfortunately, although researchers have found that it is easy to ask for *mean* assessments, it is considerably more difficult to elicit *variances* from subject-matter experts (O'Hagan, 1998). Therefore, we will use other methods to obtain variance estimates.

After asking for a mean assessment or "best guess" on the 0–100 scale for the strength of a node attribute or the strength of an edge between two nodes, we internally make the restrictions $\alpha > 1$ and $\beta > 1$, guaranteeing a unimodal beta distribution. From this, we know that $4\sigma_y \approx b - a$. We also know that

$$\sigma_y^2 = (b-a)^2\sigma_x^2 = (b-a)^2\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Substituting for σ_y^2 produces $\sigma_x \approx 1/4$, which sounds arbitrary. Using this estimate of σ_x as a starting point, we obtain estimates of α and β for each elicitee on each attribute or edge without taxing these analysts *directly* for a variance measure. There are several good reasons why this seemingly arbitrary value for σ_x is reasonable. First, there is a body of literature on expert input into a Program Evaluation and Review Technique (PERT) analysis for industrial and engineering project planning that advocates the value $\sigma_x = 1/6$ when the beta mode is between 0.13 and 0.87; this value of $\sigma_x = 1/6$ is somewhat smaller and therefore less conservative than ours. It comes from asking the planning expert to provide a modal value, which is related to the beta mean by $\mu_x = (4m_x = 1)/6$. If the beta distribution is unimodal and symmetric, then it resembles a normal distribution and the mode and mean are coincident. Hence, it is assumed in PERT analysis that $\sigma_x = 1/6$ (Farnum & Stanton, 1987; Lu, 2002). Also, if $\alpha = 2$ and $\beta = 2$, then $\sigma_x \approx 1/4$ anyway. Third, as noted above, the value of 1/4 is just a starting point for our software "slide" that the analyst adjusts, as described below.

We subsequently display a beta distribution with the elicitee's mean and a variance of 1/4 using a manipulable graphic as shown in Figure 1. This manipulation is done with two slides: one that the analyst uses to change $\mu_x = \alpha/(\alpha+\beta)$ and another to change $\sigma_x^2 = \alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$. The results of this process are new values of μ_x and σ_x^2, the values with which the elicitee is presumably more comfortable. This also is shown in Figure 1, where the edge elicitation updates country of origin (CoO) between two actors (nodes of interest in a real analysis). The analyst can make the adjustments desired and hit the button (designated by the edge name) on the lower left of the screen to submit her elicitation. The screen also provides the numerical values for the mean and variance that are implied by the slide positions, but not so prominently as to distract the elicitee from the graphic task. We now provide a confirmatory mechanism on a subsequent screen to provide additional internal validity for this elicitation. Figure 2 shows three prior distributions, only one of which is the beta originally specified by the elicitee.

2.7 Bayesian aggregation of elicited priors

Aggregation of estimates or beliefs has led to a vast body of literature in psychology, statistics, decision sciences, operational research, and other fields. West (1984),

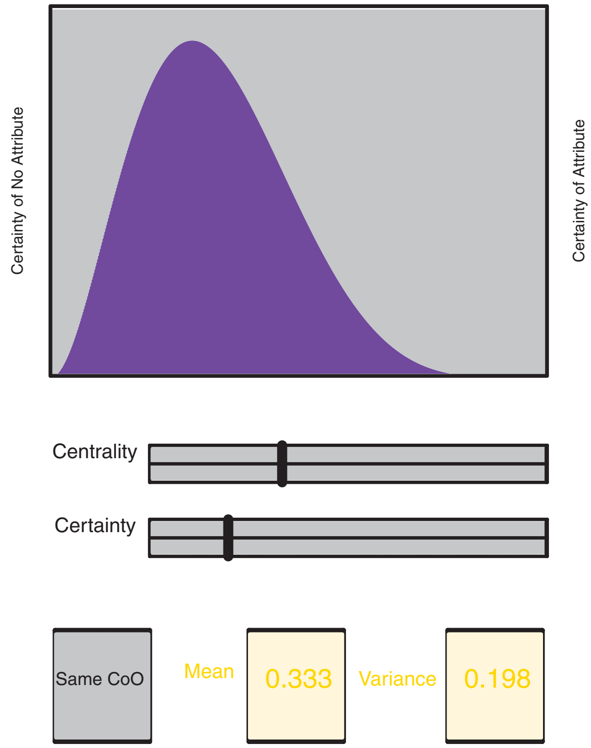

Fig. 1. Elicitation screen.

and later Risse (2003), presents a detailed theoretical formulation of Bayesian aggregation when the group is allowed to interact, finding that group consensus can be difficult when the group members communicates. Clemen and Winkler (1999) survey the literature and find that the Bayesian focus on probability statements leads to aggregations, "combined probability distributions," that can effectively represent the current state of expert opinion. More recently, and more importantly here, Johnson (2010) makes the point that "Aggregation error is not an error." This is because posterior differences through aggregation are a natural feature of properly specified models that result from the Bayesian updating mechanism on distributions. The literature on expert elicitation for economic forecasts is vast, but appears to have converged on Bayesian updating as the dominant paradigm. Our aggregation approach falls squarely in this tradition and simply exploits convenient properties of the beta distribution.

Over time, for a given node or edge, possibly many (M) parameter elicitations are collected producing $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_M]$ and $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_M]$. These can be organized in vectors as meta-terms: $[\boldsymbol{\alpha}_{i\ell k}, \boldsymbol{\beta}_{i\ell k}]$, $\quad i = 1\!:\!n$, $\ell = 1\!:\!L$, $k = 1\!:\!K$, for n elicitees, $L = (n^2 - n)/2$ possible relationships (edges), and K attributes on each case (node).[8] Consider x as an arbitrary edge to be analyzed between arbitrary nodes i and j, suppressing other conditionalities here for notational clarity. The process starts

[8] This setup assumes for now that edges are not directed, but this can easily be changed with an additional subscript.

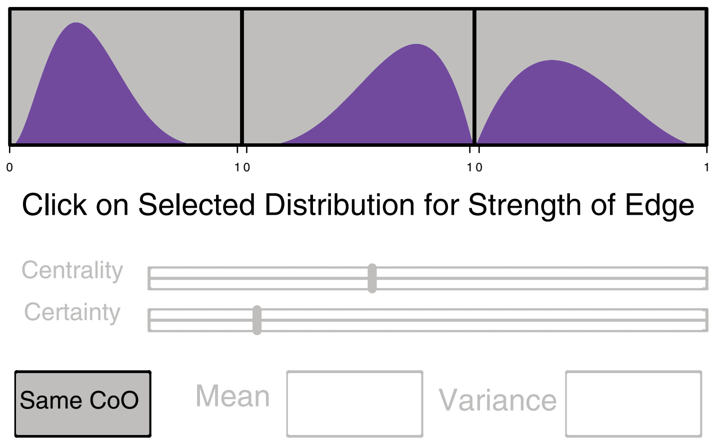

Fig. 2. Comparison screen.

with an initial probability statement on x: $p_1(x) \propto x^{\alpha_1 - 1}(1-x)^{\beta_1 - 1}$ (the kernel of a beta distribution), which can be left as deliberately vague as desired ($\alpha_1 = 1$, $\beta_1 = 1$ provides a uniform distribution). The updated distribution from the first analyst's input comes from applying Bayes' law using this new kernel, $p_2(x) \propto x^{\alpha_2 - 1}(1-x)^{\beta_2 - 1}$, producing the first posterior distribution $\pi_1(x) \propto p_1(x)p_2(x) = x^{\alpha_1+\alpha_2-2}(1-x)^{\beta_1+\beta_2-2}$, which is also a beta distribution form due to conjugacy. Repeating this multiplicative process for $m = 1, \ldots, M$ inputs from the elicitations automatically gives the Mth update:

$$\pi_M(x) \propto x^{\sum_{m=1}^{M} \alpha_m - M}(1-x)^{\sum_{m=1}^{M} \beta_m - M},$$

which is to say that x after update M is distributed as

$$x|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\omega} \sim \mathscr{BE}\left(\sum_{m=1}^{M} \alpha_m \omega_m - M + 1, \sum_{m=1}^{M} \beta_m \omega_m - M + 1\right), \qquad (1)$$

where relative weights, ω_m, are included to reflect differing levels of elicitee knowledge, reliability, and possible overconfidence. In this manner, elicitations are continuous and ongoing, improving social network information as they accumulate producing a process that managers can control.

Furthermore, unknown *edges* without direct elicitation updates can be estimated from the relative strength of corresponding *node* information. Suppose there exists single attribute information for two selected nodes i and j, labeled $\mathbf{y}_i$ and $\mathbf{y}_j$. This is in the form of the most updated (Mth) corresponding beta distribution parameters: α_i, β_i, and α_j, β_j (suppressing the $m = 1, \ldots, M$ notation here). Then a newly informed *edge* prior $\mathbf{x}_{ij}$ is calculated by "differencing" beta distributions according to $\alpha_{\mathbf{x}_{ij}} = k_\alpha \min(\alpha_i, \alpha_j)$ and $\beta_{\mathbf{x}_{ij}} = k_\beta \max(\beta_i, \beta_j)$. If the α or the β parameter pairs differ by a large amount, then the relationship attribute tends toward a beta distribution that reflects a low relationship probability. Conversely, if there is substantial agreement in parameter values, the minimum and the maximum will be very close together and aggregation will change the prior very little. Here, k_α and k_β are weighting parameters that reflect management uncertainty in the node-to-edge

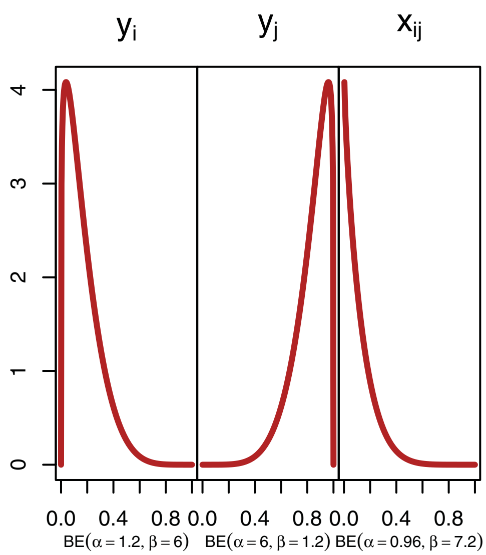

Fig. 3. Individual to relationship aggregation.

process just described, which can be defaulted to one in the absence of strong prior information. Using this mechanism, we always provide an edge assessment, $\mathbf{x}_{ij}$, as input to the statistical network model, where expert elicitation will substantially improve the quality in the presence of relevant edge observations. As an illustration of this process, suppose we have beta priors for some single attribute of nodes $\mathbf{x}_i$ and x_j that are the most updated versions in the system (after M expert elicitations): $\mathbf{y}_i \sim \mathscr{BE}(1.2, 6)$ and $\mathbf{y}_j \sim \mathscr{BE}(6, 1.2)$. These reflect divergent assessments about the strength of this attribute for case i versus case j (shown in the first two panels of Figure 3), so we would expect them to provide little or no new information on the corresponding edge x_{ij}. Suppose further that management assigns the weights $\omega_i = 0.8$ and $\omega_j = 1.2$ to the current state of these nodes, obviously reflecting differing views of estimation reliability.[9] Therefore, now the aggregated prior is a beta distribution with $\alpha = 0.96$ and $\beta = 7.2$ reflecting substantial edge skepticism, as shown in the rightmost panel of Figure 3. In considering the ω_m weights used in Equation (1), we follow Tetlock (2006), who notes that expertise is not uniform across categories of elicitation. Therefore, for instance, we might have analysts who are experts on Southeast Asian terrorism making judgments in the same sample group as people who are experts in Russian terrorism. For judging relationships in the Middle East, both may be relatively inexperienced in the Tetlock sense.

[9] The choice of weights might be informed by separate, psychologically motivated statistical analyses of expert judgments. For example, the choice of ω might be based on the "cultural groupings," response biases, and log-odds weights produced by cultural consensus theoretic methods (Batchelder, 2009). The nonlinear weights generated by Bayesian approaches to aggregating magnitude judgments also might prove useful for this purpose (Merkle & Steyvers, 2011; see also Karabatsos & Batchelder, 2003). However, note that, unlike in the former, we do not make any strong (in-)out-degree homogeneity assumptions about the accuracy of experts' responses and, in contrast to the latter, we do not assume that experts have systematic biases across all edges (nodes). Finally, we obtain "uncertainty distributions" direct from our experts, and for each edge (attribute) our aggregated distribution is based on those direct elicitations not on estimation of any statistical model.

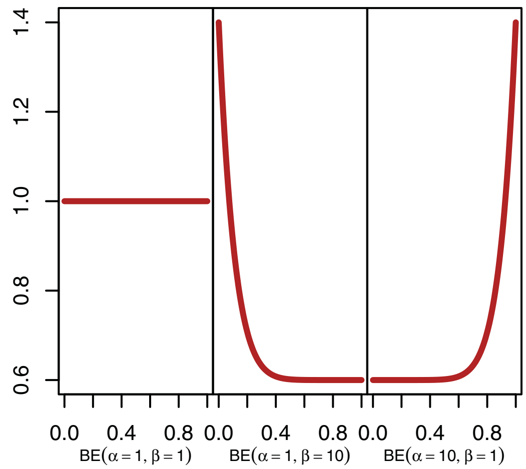

Fig. 4. Alternative beta distributions.

However, although there may be strictly a "jurisdictional" issue, the expert on Southeast Asian terrorism, in the course of her routine analysis, may have just observed data on an important link in Chechnya or Pakistan. We certainly do not want to a priori discount these observations down to zero, nor do we want to assume equivalent expertise. Therefore, assigning relative weights ends up becoming a classic management task. We suggest starting with relatively small differences and fine-tuning over time. Obvious initial criteria include level of specialization, education, experience, particularistic background, language and travel experience, as well as familiarity with the elicitation process. Importantly, analysts should neither know their individual weight nor be compensated/rewarded for it. Doing otherwise has the potential to lead to many negative bureaucratic/managerial situations.

2.8 Evaluation of elicited prior distributions

We can also evaluate elicitations and compare quality with standard statistical measures such as the mean square error. As an example, suppose that two analysts are evaluating two attributes. Analyst A gives a prior of $\mathscr{BE}(1,1)$ for both attributes and analyst B gives the prior of $\mathscr{BE}(1,10)$ for the first attribute (with true value 0) and the prior of $\mathscr{BE}(10,1)$ for the second attribute (with true value 1). These elicited distributions are shown graphically in Figure 4. Naively, we could note that the averaged forecasts of the actor having both attributes are the same. Instead, let us calculate the mean square error for both sets of elicitations. For analyst A, the mean square error (MSE) calculation for one attribute is $\text{MSE}(A) = \text{Var}_\theta(W_A) + (E_\theta(W_A) - \theta)^2 = \frac{1}{12} + \left(\frac{1}{2} - 0\right)^2 = \frac{1}{3}$. Thus, by symmetry the other MSE is also one-third, making analyst A's total MSE equal to 2/3. Analyst B's MSE for the lower side is $\text{MSE}(B) = \text{Var}_\theta(W_B) + (E_\theta(W_B) - \theta)^2 = \frac{10}{1452} + \left(\frac{1}{11} - 0\right)^2 = 0.0\overline{15}$. The other MSE is 0.006, so analyst B's total MSE is only 0.021, which is a remarkable difference. Note, however, that analyst B is still admitting a reasonable amount of uncertainty. In MSE terms, being highly uncertain about the presence of attributes connotes a lower level of performance in a way that is consistent with the argument of Tetlock (2006) and other students of expert judgment.

Evaluation is also important to identify the case in which a subject either does not understand the instructions or is deliberately misreporting the shapes of prior distributions (see Prelec, 2004, for an interesting discussion of such effects). In very small elicitation numbers, this could be disastrous as there would not be other elicitations to average these back in the correct direction and there would not be a host of alternative elicitations of a very different nature for managers to use as a comparison. Conversely, in large elicitation numbers these poor sample points would be averaged-down and could be deleted by wary managers. Since the system is intended to be dynamic and ongoing, regular diagnostics can identify this problem in all but the extreme case of single elicitations on a specific node or edge.

2.9 Interpretation of elicitations

Finally, it is important to review the *interpretation* of uncertainly required here. There is a large body of literature on eliciting *point* estimates and the problems that ensue in terms of measures of uncertainty or variance (Bedrick et al., 1997; Cooke, 1991; Cosmides & Tooby, 1996; Hogarth, 1975; Leamer, 1992; Meyer & Booker, 2001; Murphy & Winkler, 1974; Tversky & Kahneman, 1983). The original source of skepticism is Alpert & Raiffa (1982), and scholars have worked very hard to move the primitive tools that are criticized in that work. The general consensus now is that elicitees need to be pushed into thinking more in distributional terms from the start. They should get visual feedback on their elicitations and be allowed to adjust them (Bunn, 1979; Carlin et al., 1995; Chaloner et al., 1993; Chaloner & Duncan, 1983, 1987; Kadane et al., 1980; Gill & Walker, 2005; Kadane & Winker, 1988; Kadane & Wolfson, 1998). Therefore, our software shows each elicitee a graphical display of their stipulated beta distribution and allows them to manipulate it until they are confident that their beliefs have been accurately represented (rather than by specifying new parameter values).[10]

Elicitation output in distributional form solves an important interpretation problem. Tetlock (2006) and others suggest penalizing "guesses" that range around [0.4 : 0.6] because these guesses fail to "discriminate" in a useful manner. Consider two commonly provided attributes: sex and religion. Let the numbers 0 and 1 denote the condition of being female and male, respectively. For religion, 0 denotes non-Hindu and the number 1 the attribute of Hindu. Suppose analyst A says that for a particular node, the probability that one of the actors is male is 0.5 and the probability that this same actor is a Hindu also is 0.5. In contrast, analyst B is certain that the actor in question is not a man, and she also certain that the actor is a Hindu. To be more specific, analyst B says the probability that the actor is male is zero and the probability that the actor is a Hindu is one. The actor turns out to be a female Hindu. If we use Tetlock's performance measure (predicted probability − actual attribute)^2 where 0 is the best performance, it is clear that analyst A does not do as well as analyst B. Analyst A scores $(0.5-0)^2+(0.5-1)^2 = 0.5$, whereas analyst B scores $(0-0)^2+(1-1)^2 = 0$. Analyst B does better because

[10] Of course, provision must be made for training analysts in the use of the software and for making sure that all of them in fact understand the visual representation we present to them. In this training, a variety of reliability checks will be made. See Gill & Walker (2005), especially p. 855f.

even though she also predicts that the probability of the actor having one of the two traits is 0.5, she is able to discriminate perfectly the two attributes of sex and religion. Analyst A was highly uncertain and probably guessed that the actor in question has one of the two attributes. This contrast in performance is important because it illustrates how an uncertainty estimate and a location estimate can be confounded. If both analysts are forced to give distributional statements, then analyst A is likely to give something like a uniform over $[0:1]$ (a beta distribution with both parameters equal to one), and analyst B would give beta distributional forms that are highly skewed with either side showing the opposite assessments. Now measures of being "wrong" differ considerably in exactly the way we want.

3 The experiment

Our experiment was designed to show that from a *substantive* point of view, our technology is an effective tool for the elicitation of information about covert networks. Thus, this section is an empirical validation of the methodology described in Section 2.[11] What we find is that subjects are comfortable using the on-screen technology to make assessments of a partially observed relationship, and they generally are able to improve the quality of that assessment.

3.1 Experimental subjects

The experiment was conducted over the course of three weeks in the Social and Behavioral Sciences Laboratory at the University of Minnesota with the approval of the Institutional Review Board of the University. As in many elicitation experiments, our subject pool consisted of students: 57 undergraduates recruited from three large courses (20% of them enrolled in a social networks course) and 6 graduate students from the Political Science department (as checks on the effect of the undergraduate cohort). All elicitees who completed the single experimental session were compensated with their choice of a $30 iTunes card or a university bookstore gift certificate.

This group had the following characteristics: a mean age of 21, 41% female, 92% full-time students, all but 1 were U.S. citizens, 21% reported that they watched more than 8 hours of television per week. The emphasis in this experiment was on these subjects' abilities to use our visualization tool to make accurate assessment of attributes and of actor relationships in a social network through video clips, a network that shares important features of covert networks. Each subject participated in the experiment one time.

For our experiment, the subject is told that the experiment has two parts and how long it will take to complete it. The elicitee is given instructions about using keys to advance screens and what is locked (access to the Internet while experiment is underway, for instance). The subject was also asked for demographic information. The primary goal of the tutorial is to convey what the "wave" (beta distribution) probability picture represents and how each subject creates it, adjusts it, and then

[11] Background information, training materials, and timing analysis are provided on the dedicated Web page http://jgill.wustl.edu/elicitation.experiment.html.

confirms their visualization of centrality and uncertainty. There was no post-training test as pretesting indicated that the subjects understand the mechanical part of the process and the probabilistic interpretation.

3.2 Edge elicitation experiment

In the first part of the experiment, subjects were asked general questions about the character of the University's student body and about U.S. economics and demographics. These background questions were not network related. They served to further acclimate the subjects to the elicitation technology and to determine if they could use the technology to answer factual questions of varying degrees of difficulty. For brevity, we will skip here the results of this first part of the experiment. The second part assessed the ability of subjects to estimate the strength of relationships between actors in a covert-type network. Two brief DVD clips from the BBC soap opera *EastEnders* showing interacting characters were incorporated into the software system. At their core, soap operas are nothing more than a large social network where actors emerge and disappear over time. Relationships also vary dramatically by strength and contact. The choice of a British soap opera was designed to challenge elicitees. First, the characters' accents are of a heavy Cockney-esque East London variety requiring elicitees to pay close attention. Second, we suspect that very few U.S. students have had exposure to EastEnders, which appeared only on British television, thus adding an experimental control. Our choice of the EastEnders clip reflects challenges that researchers may face in practical settings where members of covert networks may have widely varying languages or speak different dialects. The idea is that, if after viewing these short clips, undergraduate students can use our technology to accurately produce information about this "foreign" network, then professional subject-matter experts can use our technology to characterize actual networks.[12]

We first show a clip that implies, but does not state, that two actors are sisters or close friends who socialize together in the local network. The question is whether their discussion of going out that evening is about them going out together to the local pub *as peers*. It is possible by paying very close attention for elicitees to cast doubt on that event, even though overtly it seems likely. At this point, we ask for an edge elicitation on the strength of the relationship with regard to going out that night together using the beta distribution tool. Subsequently, elicitees viewed another clip of roughly equal length where it is further clarified that the two actors are likely to be mother and daughter. Respondents were then provided a second opportunity to assess the probability that the two *sisters* will go out (to the pub) together, which is part of the storyline.

[12] Of course, we prefer eventually to have subjects watch a surveillance video supplied by a law enforcement or intelligence agency. However, again, this experiment was intended to be an initial proof of concept exercise. Admittedly, soap operas are different from surveillance videos; soap operas have a narrative that is intentionally revealed to the viewer over time. But we chose EastEnders because our subjects had *not* watched the soap opera over any length of time. Hence, they are not influenced by any serial storyline in the respective narrative. Our subjects only watched brief clips effectively out of context of that storyline. Moreover, they are nonspecialists, not intelligence analysts. Therefore, they present a stringent test of our technology.

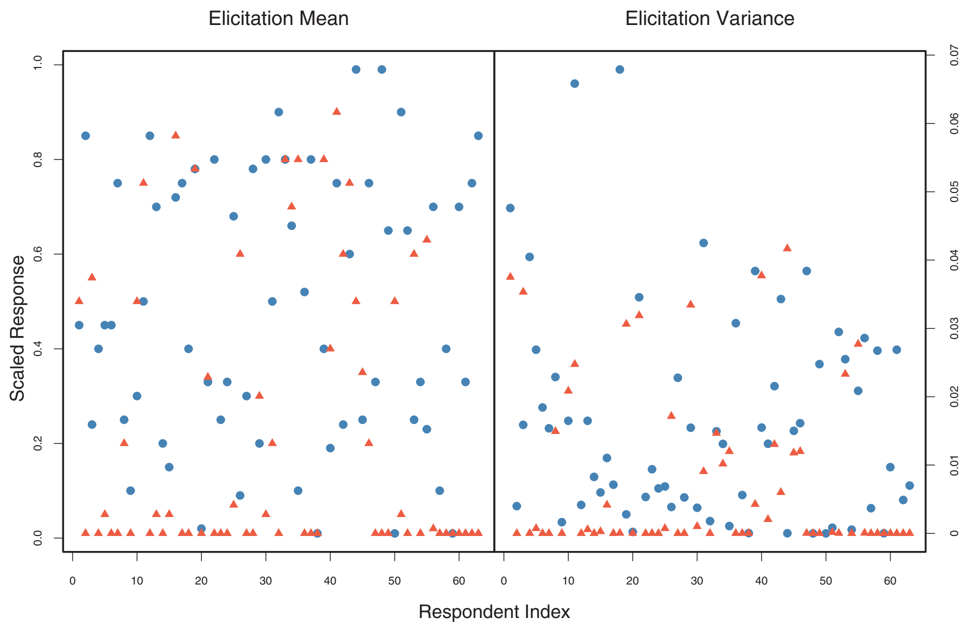

Fig. 5. Changes across two edge elicitations.

Elicitees typically lowered their probability of socializing *as sisters* on the second round: the left panel of Figure 5 shows gray circles for beta distribution means on the first round and black triangles for the second round. In fact, many lowered their variances as well, as shown in the right panel of Figure 5, with the same designations as the left panel (our software precludes means and variances below 0.01 in order to preserve distributions rather than produce point masses). Only 1 of the 57 valid responses had no adjustment of the beta distribution after seeing the second clip. We infer from this that (1) additional information provides an impetus to update elicitations and (2) elicitees are sufficiently comfortable with the automated elicitation system that they will perform this update.

3.3 Statistical modeling

The question that remains is how to best fit the elicited priors into a general Bayesian modeling framework for hidden social network data. Our approach builds on Hoff-style models (2005, 2009; see also Hoff et al., 2002).[13] Along with this model come the two important assumptions: (1) dependence between units is modeled as exogenously defined random effects and (2) each edge is conditionally independent given the node characteristics.

Define first the $n \times n$ symmetric matrix $\mathbf{Y}$ giving a mapping of links between n individuals. Here, $y_{ij} = 1$ indicates a known link between node i and node j, and $y_{ij} = 0$ indicates the absence of evidence for a link. Now define the $n \times n \times K$ array $\mathbf{X}$ where for each $n \times n$ relationship between individual i and individual j there is a

[13] This section illustrates how our technology can be applied in edge prediction of the type advanced by Hoff and his associates. We recognize that there are other approaches to edge predictions such as random forests.

K-length vector of covariate information containing attributes for i, attributes for j, and natural relationship attributes between i and j. These values could be known, in which case there are specific values in the array, or they could be unknown but possess elicited priors, in which case the array value is placeholder for the elicited distribution. A model that relates $\mathbf{X}$ and $\mathbf{Y}$ is the random effects logistic regression specification:

$$p(\mathbf{Y}|\boldsymbol{\theta}_{ij}) = \prod_{i \neq j} \frac{\exp(\boldsymbol{\theta}_{ij})}{1 + \exp(\boldsymbol{\theta}_{ij})} \qquad \boldsymbol{\theta}_{ij} = \boldsymbol{\beta}'\mathbf{x}_{ij} + z_{ij} \qquad z_{ij} = \mathbf{u}_i'\gamma\mathbf{v}_j + \epsilon_{ij}$$

where $\boldsymbol{\beta}$ is a K-length vector of coefficients to estimate, and z_{ij} is a random effects term to account for dependencies between attribute relationships. The random effects term is broken up into a $\mathbf{u}_i'$ vector of sender-specific latent factors, a $\mathbf{v}_j$ vector of receiver-specific latent factors, a γ diagonal matrix of unknown coefficients, plus a ϵ_{ij} scalar error specific to the edge. Thus, we have in this model log-odds$(y_{ij} = 1) = \boldsymbol{\theta}_{ij}$ with parameters of interest $\boldsymbol{\beta}$ and γ, giving the relative importance of covariates or latent factors, respectively. Hoff (2009) shows how posterior estimates for these unknown parameters can be obtained with Gibbs sampling and applies the process to conflict data.

Our departure from this setup comes from a different interpretation of the $\mathbf{x}$ values (which come from individual $\mathbf{x}_i$ information and edge information, $\mathbf{x}_{ij}$). These values are the attributes (suggested as *affinity*, *ideology*, *geography*, *sub-group identity*, *previous service together*, *family*) that shed light on the probability of the existence of a viable edge in the covert network ($y_{ij} = 1$). In typical social network analysis, these $\mathbf{x}_{ij}$ values are known. There are no missing data. In our context, where actors reveal partial information but hide as much as possible, some $\mathbf{x}_{ij}$ are not known. These missing data therefore need to be treated probabilistically and assigned a prior for Bayesian inference. As we explained in the previous section, our approach elicits these data *directly* from experts.

Operationally, we modify Hoff's (2009, p. 266) algorithm in the following way. Define the unknown quantities in the link prediction model to include elicited information on attributes:

- ▹ $\Theta = \boldsymbol{\theta}_{i,j}$, the set of predictors;
- ▹ $\mathbf{x}^*$ represents the $\mathbf{x}$ values that are not known with certainty and given (aggregated) beta priors from our elicitation procedure, possibly with weights from Equation (1);
- ▹ $\boldsymbol{\beta}$, the vector of regression coefficients;
- ▹ $\mathbf{U}$ and $\mathbf{V}$ are $n \times K$ matrices with orthonormal columns, $\mathbf{U} = (\mathbf{U}_{[,1]}, \ldots, \mathbf{U}_{[,K]})$ and $\mathbf{V} = (\mathbf{V}_{[,1]}, \ldots, \mathbf{V}_{[,k]})$;
- ▹ $\mathbf{D} = \text{diag}(d_1 \ldots d_K)$, an $n \times n$ diagonal matrix.

Then, given prior distributions on the model parameters, as in Hoff (2009) and our elicited priors, we can obtain their posterior distribution via Bayes rule but now with:

$$p(\phi|\mathbf{Y}) = p(\Theta, \boldsymbol{\beta}, \mathbf{x}^*, \mathbf{U}, \mathbf{D}, \mathbf{V}|\mathbf{Y}) \propto p(\mathbf{Y}|\Theta, \boldsymbol{\beta}, \mathbf{x}^*, \mathbf{U}, \mathbf{D}, \mathbf{V})p(\Theta, \boldsymbol{\beta}, \mathbf{x}^*, \mathbf{U}, \mathbf{D}, \mathbf{V}).$$

Given arbitrary starting values, $\psi_0 = (\Theta_0, \boldsymbol{\beta}_0, \mathbf{x}_0^*, \mathbf{U}, \mathbf{D}, \mathbf{V})$, we can use the same kind of MCMC scheme to generate a sequence of $\psi_1, \psi_2, \ldots$. Thus, beginning the chain at assigned starting values, repeat the following steps:

1. Sample $\boldsymbol{\beta}$ from its full conditional distribution $p(\boldsymbol{\beta}|\Theta, , \mathbf{x}^*, \mathbf{U}, \mathbf{D}, \mathbf{V})$;

2. Sample $\mathbf{x}^*$ values from their full conditional distribution $p(\mathbf{x}^*|\Theta, \boldsymbol{\beta}, \mathbf{U}, \mathbf{D}, \mathbf{V})$ and inserting these into the corresponding unknown $\mathbf{X}$ cells;
3. For $k \in 1 \ldots K$:
   (a) sample $\mathbf{U}_{[,k]}$ from $p(\mathbf{U}_{[k]}|\Theta, \mathbf{X}, \mathbf{x}^*, \mathbf{U}_{[,-k]}, \mathbf{D}, \mathbf{V})$,
   (b) sample $\mathbf{V}_{[,k]}$ from $p(\mathbf{V}_{[k]}|\Theta, \boldsymbol{\beta}, \mathbf{X}, \mathbf{x}^*, \mathbf{U}, \mathbf{D}, \mathbf{V}_{[,-k]})$,
   (c) sample $\mathbf{D}_{[k,k]}$ from $p(\mathbf{D}_{[k,k]}|\Theta, \boldsymbol{\beta}, \mathbf{X}, \mathbf{x}^*, \mathbf{U}, \mathbf{D}_{[-k,-k]}, \mathbf{V})$;
4. Sample $\Theta^* = \mathbf{X}\boldsymbol{\beta} + \mathbf{UDV}' + E*$, where $E*$ is a matrix of standard noise. Replace $\boldsymbol{\theta}_{i,j}$ by $\boldsymbol{\theta}^*_{i,j}$ with probability $[p(y_{i,j}|\boldsymbol{\theta}^*_{i,j})]/[p(y_{i,j} \mid \boldsymbol{\theta}_{i,j})] \wedge 1$. Here, the matrix $\mathbf{X}$ has values from $\mathbf{x}^*$ inserted into missing data locations.

With the elicited data on attributes, the distribution of the generated samples of the ψ parameters will converge to a posterior distribution $p(\psi \mid \mathbf{Y})$ from which posterior quantities related to covert network linkages can be approximated. How do we know this? Suppose first that $p(\mathbf{x}^*|\Theta, \boldsymbol{\beta}, \mathbf{U}, \mathbf{D}, \mathbf{V})$ was the exact correct distribution for $\mathbf{X}$. Then this procedure would obviously be a fully informed Gibbs sampler with all the complete information of full conditional distributions and therefore an ergodic Markov chain guaranteed to converge to the appropriate stationary distribution. Now consider that we instead have the case that

$$p(\mathbf{x}^*|\Theta, \boldsymbol{\beta}, \mathbf{U}, \mathbf{D}, \mathbf{V}) \xrightarrow{P} p(\mathbf{x}|\Theta, \boldsymbol{\beta}, \mathbf{U}, \mathbf{D}, \mathbf{V})$$

as the number of elicitations increases. This is equivalent to the assumption above that the empirical distribution of the missing data becomes more accurate over time, starting from a uniform form (beta distribution with both parameters unity). This also means that we always have *some* distributional form to draw from, which sets up a data augmentation justification. Divide the data into two parts, $\mathbf{X}_{\text{all}} = (\mathbf{X}, \mathbf{x}^*) = (\mathbf{x}_{\text{obs}}, \mathbf{x}_{\text{mis}})$, and notate $\boldsymbol{\Psi} = (\Theta, \boldsymbol{\beta}, \mathbf{U}, \mathbf{D}, \mathbf{V})$ such that

$$p(\boldsymbol{\Psi}|\mathbf{x}_{\text{obs}}) = \int_{\mathcal{X}} p(\boldsymbol{\Psi}|\mathbf{x}_{\text{mis}}, \mathbf{x}_{\text{obs}}) p(\mathbf{x}_{\text{mis}}|\mathbf{x}_{\text{obs}}) d\mathbf{x}_{\text{mis}}.$$

Tanner & Wong (1987, p. 530) showed that in the ith step of a sampler, starting with a plausible (recent, $i-1$) draw of $\boldsymbol{\Psi}$, it is possible then to draw m multiple independent values of the missing data to create an updated version:

$$p_{[i]}(\boldsymbol{\Psi}) = m^{-1} \sum_{j=1}^{m} p(\boldsymbol{\Psi}|\mathbf{x}_{\text{mis}}, \mathbf{x}_{\text{obs},[i-1,j]}), \tag{2}$$

which converges to the correct distribution (Tanner & Wong, 1987, pp. 537–538). Liu (2001, p. 137) notes that in the context of a Gibbs sampler, in particular, the long-run nature of the algorithm means that the sum in Equation (2) is not really necessary and can be replaced with a single draw of the missing data. This is exactly our approach above, and therefore we know from a wealth of literature (e.g., Robert & Casella, 2004, chapter 5) that the algorithm proceeds toward convergence under these conditions. Therefore, addition of the elicited attribute distribution will actually improve the performance of the algorithm in comparison to the case in which all the data are known or the data are filled in using some form of imputation.

There is one further consideration. It is unlikely from an analytical point of view that the individual covariate information provides equal information on whether a positive edge exists across all K dimensions of the data. Thus, we want to weight on substantive grounds, changing the predictors to $\boldsymbol{\theta}_{ij} = \boldsymbol{\tau}\boldsymbol{\beta}'\mathbf{x}_{ij} + z_{ij}$, where $\boldsymbol{\tau}$ is

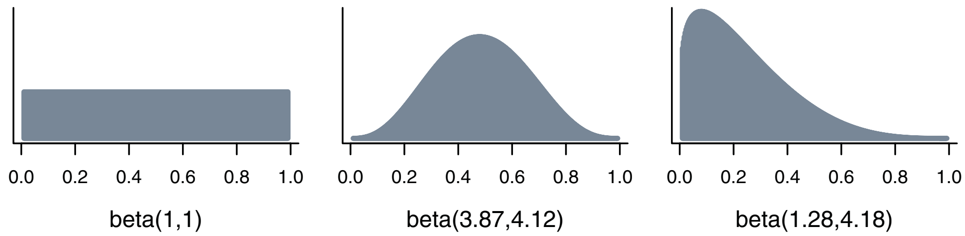

Fig. 6. Analyzed prior distributions: EastEnders example.

a K-length vector of weights summing to one. Note that this is a second level of weighting since we also weight independently on individual analyst quality.

The primary value of this model is in prediction. Although the $\mathbf{Y}$ matrix is constructed from certainties, $\hat{\mathbf{Y}}$ predictive values from

$$\hat{\mathbf{Y}} = \prod_{i \neq j} \frac{\exp(\hat{\boldsymbol{\theta}}_{ij})}{1 + \exp(\hat{\boldsymbol{\theta}}_{ij})} \tag{3}$$

take into account coefficient estimates and elicited priors. Values close to (near) 1 indicate a high probability that this node exists, conditional on data and prior information. Such predictions can be made as the network is updated, borrowing strength from other edges.

3.4 Statistical analysis example

We now return to the output from our experiment using the elicitation of information from EastEnders. We run our variant of the Hoff model above without covariates to get edge predictions from Equation (3) using three different priors: a uniform prior, the aggregate prior from the first set of elicitations, and the aggregate prior from the second set of elicitations. Aggregation across priors is performed according to the steps in Section 2.7, such that the aggregate distribution of this edge is

$$x|\boldsymbol{\alpha}, \boldsymbol{\beta} \sim \mathscr{BE}\left(\sum_{i=1}^{57} \alpha_i - 56, \sum_{i=1}^{57} \beta_i - 56\right),$$

where we weighted each of the elicitees equally, so the ω_i terms are all one and disappear from this expression. These three distributions are depicted in Figure 6. Notice the retrenchment in the third figure, indicating a collective view that Kat and Mo are less likely to go out later.

We now have prior distributions to draw the $\mathbf{x}^*$ values from in the context of the MCMC algorithm, taking values from the distributions in Figure 6. Note from the previous section that including covariates makes this draw more extensive since these would then be full expressions of the form $p(\mathbf{x}^*|\Theta, \boldsymbol{\beta}, \mathbf{U}, \mathbf{D}, \mathbf{V})$. For the MCMC algorithm described above, we performed $50,000$ iterations, throwing away the first half of the chain. All of the conventional diagnostics, run concurrently with the `R` package `superdiag`, point toward convergence (see Gill, 2008, for relevant issues), including those of Geweke (1992), where the z-scores from the difference of means tests are all under two in absolute value; Heidelberger & Welsh (1981a, 1981b), where

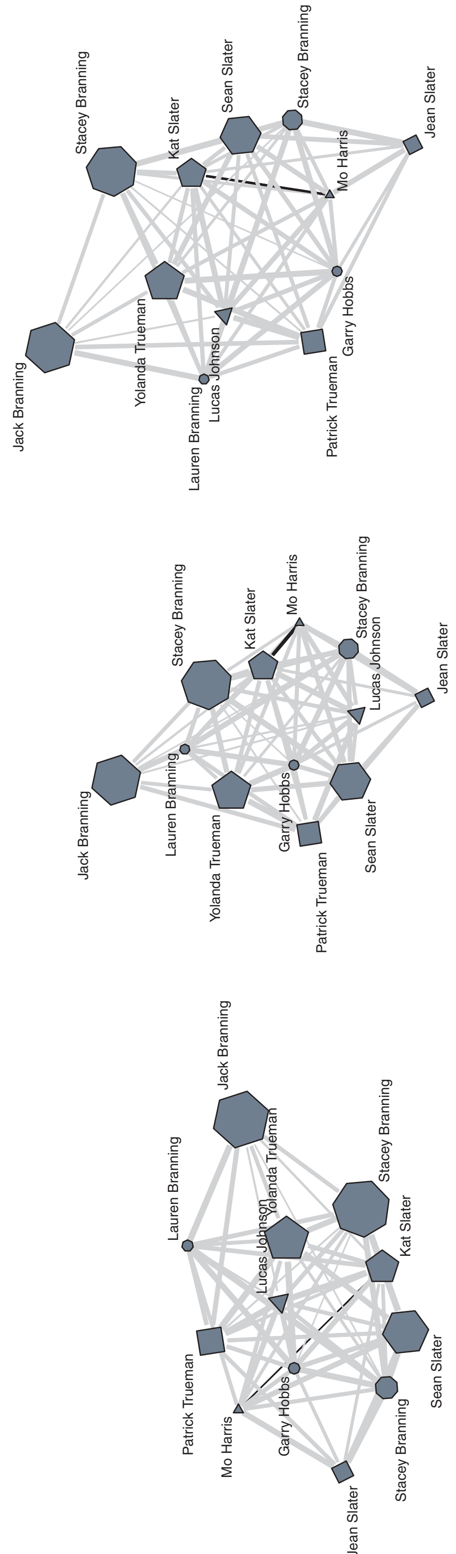

Fig. 7. Estimated edge changes between Kat Slater and Mo Harris.

the z-scores from the Cramér–von Mises test are also less than two in absolute value; and Gelman & Rubin (1992), with parallel chains returning an estimated scale reduction of 1.0131, and standard graphical analysis showing only stable patterns.

In the resulting estimated network diagrams in Figure 7 (panels 1–3), notice the thickening of the link between Kat and Mo (black line) from the uniform prior to the first informed (but slightly misguided) prior, and that this line becomes thinner as more information is revealed. The change of importance in this edge illustrates the efficacy of our elicitation process as new information is presented to observers. In the other parts of the network graphics, the remaining (gray) lines between other actors (nodes) are proportional to the number of mentions by name or appearances in the same scene during the 57-minute serialization.

4 Conclusion

This project focuses on so-called "covert networks"—social networks characterized by low visibility, low interactivity between nodes, high degrees of uncertainty about changing edges, and qualitative aspects that are likely to change over time. These are fundamentally distinct types of networks since the actors are "trading efficiency for secrecy" (Fellman & Wright, 2003). Actors in covert social networks are difficult to analyze because the targets of description are strategic, often clever, agents who work hard to prevent such description. Accordingly, Wasserman & Faust (1994) emphasized that social network analysis evaluates relationships between social entities *and* the implications of these relationships. Terrorism is a very important example of both.

The major contributions of this paper are both theoretical and practical. The integration of prior elicitation from experts and innovations in social network analysis is the key focus of this work. Our technology solves some of the practical problems with visual approaches to elicitation of SPDs. The value of graphical tools is widely recognized (Garthwaite et al., 2005, especially p. 698). Our experiment produces new insights into exactly how different visual treatments affect and improve human probability assessments. Our visualization tool for elicitation is publicly available so that other scholars can use our technology to supply data for social network analysis of difficult to measure groups.

Various disciplines have struggled with principled methods for integrating qualitative and quantitative investigation into the same modeling process. We demonstrate a means for dealing with the difficult problem of covert network analysis that can easily be generalized to other settings. The use of the additional information elicited through the system will significantly improve the performance of existing statistical methods of link prediction. We suggest a linkage into the types of models developed by Nowicki & Snijders (2001), Hoff et al. (2002), and Hoff (2005, 2009). The key idea is that unmodeled uncertainty can be replaced with qualitative assessments to produce better models of covert networks. Such applications appear in industrial management (Ford & Sterman, 1998), climatology (Morgan et al., 2001), public health (Hoffman et al., 2007), seismology (Budnitz et al., 1998), ecology (O'Leary et al., 2009), discourse analysis (Belkin et al., 1987), and other fields.

References

Alpert, M., & Raiffa, H. (1982). A progress report on the training of probability assessors. In D. Kahneman, P. Slovic, & A. Tversky (Eds.). *Judgment under uncertainty: Heuristics and biases* (pp. 294–305). Cambridge: Cambridge University Press.

Batchelder, W. H. (2009). Cultural consensus theory: Aggregating expert judgments about ties in a social network. *Social Computing and Behavioral Modeling*, **10**, 1–9.

Bedrick, E. J., Christensen, R., & Johnson, W. O. (1997). Bayesian binomial regression predicting survival at a trauma center. *The American Statistician*, **51**, 211–218.

Belkin, N. J., Brooks, H. M., & Daniels, P. J. (1987). Knowledge elicitation using discourse analysis. *International Journal of Man–Machine Studies*, **27**, 127–144.

Borgatti, S. P. (2006). Identifying sets of key players in a social network. *Computational & Mathematical Organization Theory*, **12**, 21–34.

Budnitz, R. J., Apostolakis, G., Boore, D. M., Cluff, L. S., Coppersmith, K. J., Cornell, C. A., & Morris, P. A. (1998). Use of technical expert panels: Applications to probabilistic seismic hazard analysis. *Risk Analysis*, **18**, 463–469.

Bunn, D. W. (1979). Estimation of subjective probability distributions in forecasting and decision making. *Technical Forecasting and Social Change*, **14**, 205–216.

Burt, R. S. (1987). A note on missing network data in the general social survey. *Social Networks*, **9**, 63–73.

Carley, K. (2006). A dynamic network approach to the assessment of terrorist groups and the impact of alternative courses of action. *Pre-proceedings, Visualizing Network Information* IST-063/RWS-010. Royal Danish Defence College, Copenhagen.

Carley, K. M. (2003). Dynamic network analysis. In R. Breiger, K. Carley & P. Pattison (Eds.), *Dynamic social network modeling and analysis* (pp. 1–13). Committee on Human Factors. National Research Council. Washington, DC: The National Academies Press.

Carley, K. M. (2004). Estimating vulnerabilities in large covert networks using multi-level data. *Proceedings of the NAACSOS Conference*, Pittsburgh, PA.

Carlin, B. P., Chaloner, K., Church, T., Louis, T. A., & Matts, J. P. (1993). Bayesian approaches for monitoring clinical trials with an application to toxoplasti encephalitis pro-phylaxis. *The Statistician*, **42**, 355–367.

Carlin, B. P., Chaloner, K., Louis, T. A., & Rhames, F. S. (1995). Elicitation, monitoring, and analysis for an AIDS clinical trial.In C. Gatsonis, J. S. Hodges, R. E. Kass, & N. D. Singpurwalla (Eds.), *Case studies in Bayesian statistics* (Vol. 2, pp. 48–78). New York: Springer.

Chai, S.-K. (1993). An organizational economics theory of antigovernment violence. *Comparative Politics*, **26**, 99–110.

Chaloner, K. M., Church, T., Louis, T. A., & Matts, J. P. (1993). Graphical elicitation of a prior distribution for a clinical trail. *The Statistician*, **42**, 341–353.

Chaloner, K. M., & Duncan, G. T. (1983). Assessment of a beta prior distribution: PM elicitation. *The Statistician*, **32**, 174–180.

Chaloner, K. M., & Duncan, G. T. (1987). Some properties of the Dirichlet-multinomial distribution and its use in prior elicitation. *Communications in Statistics: Theory and Methods*, **15**, 511–523.

Clemen, R. T., & Winkler, R. L. (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis*, **19**, 187–203.

Cooke, R. M. (1991). *Experts in uncertainty: Opinion and subjective probability in science*. New York: Oxford University Press.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment and uncertainty. *Cognition*, **58**, 1–73.

Costenbader, E., & Valente, T. W. (2003). The stability of centrality measures when networks are sampled. *Social Networks*, **25**, 283–307.

Crenshaw, M. (1981). The causes of terrorism. *Comparative Politics*, **13**, 379–399.

Ertekin, S., Hirsh, H., & Rudin, C. (2012). *Learning to predict the wisdom of crowds*. Paper presented at Collective Intelligence Conference 2012, Cambridge, MA.

Farnum, N. R., & Stanton, L. W. (1987). Some results concerning the estimation of beta distribution parameters in PERT. *Journal of the Operational Research Society*, **38**, 287–290.

Fellman, P.V. and Wright, R. (2003). *Modeling terrorist networks – complex systems at the mid-range*, Paper prepared for the Joint Complexity Conference, London School of Economics, 16–18 September.

Ford, D. N., & Sterman, J. D. (1998). Expert knowledge elicitation to improve formal and mental models. *System Dynamics Review*, **14**, 309–340.

Freedman, L. S., & Spiegelhalter, D. J. (1983). The assessment of subjective opinion and its use in relation to stopping rules for clinical trials. *The Statistician*, **32**, 153–160.

Garthwaite, P. H., & Dickey, J. M. (1988). Quantifying expert opinion in linear regression problems. *Journal of the Royal Statistical Society, Series B*, **50**, 462–474.

Garthwaite, P. H., & Dickey, J. M. (1992). Elicitation of prior distributions for variable-selection problems in regression. *The Annals of Statistics*, **20**, 1697–1719.

Garthwaite, P. H., Kadane, J. B., & O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, **100**, 680–700.

Gavasakar, U. (1988). A comparison of two elicitation methods for a prior distribution for a binomial parameter. *Management Science*, **34**, 784–790.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457–511.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments.In J. M. Bernardo, A. F. M. Smith, A. P. Dawid, & J. O. Berger (Eds.), *Bayesian statistics 4* (pp. 169–193). Oxford: Oxford University Press.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, **102**, 684–704.

Gill, J. (2008). Is partial-dimension convergence a problem for inferences from MCMC algorithms? *Political Analysis*, **16**, 153–178.

Gill, J., & Walker, L. D. (2005). Elicited priors for Bayesian model specifications in political science research. *The Journal of Politics*, **67**, 841–872.

Goldman, L., Cook, E. F., Brand, D. A., Lee, T. H., Rouan, G. W., Weisberg, M. C., Acampora, D., Stasiulewicz, C., Walshon, J., Terranova, G., Gottlieb, L., Kobernick, M., Goldstein-Wayne, B., Copen, D., Daley, K., Brandt A. A. Jones, D., Mellors, J., & Jakubowski, R. (1988). A computer protocol to predict myocardial infarction in emergency department patients with chest pain. *New England Journal of Medicine*, **318**, 797–803.

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science*, **17**, 767–773.

Handcock, M. S., & Gile, K. J. (2010). Modeling social networks from sampled data. *Annals of Applied Statistics*, **4**, 5–25.

Heidelberger, P., & Welch, P. D. (1981a). Adaptive spectral methods for simulation output analysis. *IBM Journal of Research and Development*, **25**, 860–876.

Heidelberger, P., & Welch, P. D. (1981b). A spectral method for confidence interval generation and run length control in simulations. *Communications of the Association for Computing Machinery*, **24**, 233–245.

Hoff, P. D. (2005). Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association*, **100**, 286–295.

Hoff, P. D. (2009). Multiplicative latent factor models for description and prediction of social networks. *Computational and Mathematical Organization Theory*, **15**, 261–272.

Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, **97**, 1090–1098.

Hoff, P. D., & Ward, M. D. (2004). Modeling dependencies in international relations networks. *Political Analysis*, **12**, 160–175.

Hoffmann, S., Fischbeck, P., Krupnick, A., & Mcwilliams, M. (2007). Using expert elicitation to link foodborne illnesses in the United States to foods. *Journal of Food Protection*, **70**, 1220–1229.

Hogarth, R. M. (1975). Cognitive processes and the assessment of subjective probability distributions. *Journal of the American Statistical Association*, **70**, 271–289.

Huisman, M., & Steglich, C. (2008). Treatment of nonresponse in longitudinal network studies. *Social Networks*, **30**, 297–308.

Ibrekk, H., & Morgan, M. G. (1987). Graphical communication of uncertain quantities to non-technical people. *Risk Analysis*, **7**, 519–529.

Jennison, C., & Turnbull, B. W. (1990). Statistical approaches to interim monitoring of medical trials: A review and commentary. *Statistical Science*, **5**, 299–317.

Johnson, V. E. (2010). Bayesian aggregation error? *International Journal of Reliability and Safety*, **4**, 359–365.

Kadane, J. B. (1980). Predictive and structural methods for eliciting prior distributions. In A. Zellner (Ed.), *Bayesian analysis in econometrics and statistics* (pp. 89–93). Amsterdam: North-Holland.

Kadane, J. B., Dickey, J. M., Winkler, R. L., Smith, W. S., & Peters, S. C. (1980). Interactive elicitation of opinion for a normal linear model. *Journal of the American Statistical Association*, **75**, 845–854.

Kadane, J. B., & Winkler, R. L. (1988). Separating probability elicitation from utilities. *Journal of the American Statistical Association*, **83**, 357–363.

Kadane, J. B., & Wolfson, L. J. (1998). Experiences in elicitation. *Journal of the Royal Statistical Society, Series D*, **47**, 3–19.

Karabatsos, G., & Batchelder, W. H. (2003). Markov chain estimation for test theory without an answer key. *Psychometrika*, **68**, 373–389.

Kass, R. E., & Greenhouse, J. B. (1989). Comments on the paper by J. H. Ware. *Statistical Science*, **4**, 310–317.

Kossinets, G. (2006). Effects of missing data in social networks. *Social Networks*, **28**, 247–268.

Krebs, V. E. (2002). Mapping terrorist cells. *Connections*, **24**, 43–52.

Kynn, M. (2007). The 'heuristics and biases' bias in expert elicitation. *Journal of the Royal Statistical Society, Series A*, **171**, 239–264.

Law, A. M., & Kelton, W. D. (1982). *Simulation modeling and* analysis. New York: McGraw-Hill.

Leamer, E. E. (1992). Bayesian elicitation diagnostics. *Econometrica*, **60**, 919–942.

Libby, D. L., & Novick, M. R. (1982). Multivariate generalized beta distributions with applications to utility assessment. *Journal of Educational Statistics*, **7**, 271–294.

Lichenstein, S., Fischoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In A. Tversky & D. Kahneman (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). New York: Cambridge University Press.

Lipkus, I. M., & Holland, J. G. (1999). The visual communication of risk. *Journal of the National Cancer Institute Monographs*, **25**, 149–163.

Little, R. J. A., & Rubin, D. B. (1983). On jointly estimating parameters and missing data by maximizing the complete-data likelihood. *The American Statistician*, **37**, 218–220.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York: John Wiley & Sons.

Liu, J. S. (2001). *Monte Carlo strategies in scientific computing*. New York: Springer.

Lu, M. (2002). Enhancing project evaluation and review technique simulation through artificial neural network-based input modeling. *Journal of Construction Engineering and Management,* **128**(5), 438–445.

Merkle, E. C., & Steyvers, M. (2011). A psychological model for aggregating judgments of magnitude. In J. Salemo, S. J. Yang, D. Nau, & S.-K. Chai (Eds.), *Social computing, behavioral–cultural modeling and prediction*, Lecture Notes in Computer Science Vol. 6589. Heidelberg: Springer Berlin.

Meyer, M. A., & Booker, J. M. (2001). *Eliciting and analyzing expert judgment: A practical guide*. Philadelphia, PA: SIAM Press.

Moon, I.-C., & Carley, K. M. (2007). Modeling and simulation of terrorist networks in social and geo-spatial dimensions. *IEEE Intelligent Systems: Special Issue on Social Computing* **22**, 40–49.

Morgan, M. G., Pitelka, L. F., & Shevliakova, E. (2001). Elicitation of expert judgments of climate change impacts on forest ecosystems. *Climatic Change*, **49**, 279–307.

Murphy, A. H., & Winkler, R. (1974). Subjective probability forecasting experiments in meteorology: Some preliminary results. *Bulletin of the American Meteorological Society*, **55**, 1206–1216.

Nowicki, K., & Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, **96**, 1077–1087.

O'Hagan, A. (1998). Eliciting expert beliefs in substantial practical applications. *The Statistician*, **47**(Part 1), 21–35.

O'Leary, R. A., Choy, S. L., Murray, J. V., Kynn, M., Denham, R., Martin, T. G., Mengersen, K. (2009). Comparison of three expert elicitation methods for logistic regression on predicting the presence of the threatened brush-tailed rock-wallaby Petrogale penicillata. *Environmetrics*, **20**, 379–398.

Prelec, D. (2004). A Bayesian truth serum for subjective data. *Science*, **306**, 462–466.

Risse, M. (2003). Bayesian group agents and two modes of aggregation. *Synthese*, **135**, 347–377.

Robert, C., & Casella, G. (2004). *Monte Carlo statistical methods* (2nd ed.). New York: Springer.

Robins, G., Pattison, P., & Woolcock, J. (2004). Missing data in networks: Exponential random graph (p*) models for networks with non-respondents. *Social Networks*, **26**, 257–283.

Rothenberg, R. (2002). From whole cloth: Making up the terrorist network. *Connections*, **23**, 36–42.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, **63**, 581–592.

Savage, L. J. (1954). *The foundations of statistics*. New York: John Wiley & Sons.

Savage, L. J. (1971). Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, **66**, 783–801.

Schaefer, R. E., & Borcherding, K. (1973). The assessment of subjective probability distributions: A training experiment. *Acta Psychologica*, **37**, 117–129.

Schafer, Joseph L. (1997). *Analysis of Incomplete Multivariate Data*. New York: CRC Press.

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

Spetzler, C. S., & Staël von Holstein, C.-A. S. (1975). Probability encoding and decision analysis. *Management Science*, **22**, 340–358.

Tanner, M. A., & Wong, W. H. (1987) The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, **82**, 528–540.

Tetlock, P. (2006). *Expert political judgment: How good is it? how can we know?* Princeton, NJ: Princeton University Press.

Tsvetovat, M., & Carley, K. M. (2005). Structural knowledge and success of anti-terrorist activity: The downside of structural equivalence. *Journal of Social Structure,* **6**.

Tsvetovat, M., & Carley, K. M. (2006). Improving effectiveness of communications sampling of covert networks. *E-Social Science Conference,* Manchester, UK.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, **185**, 1124–1131.

Tversky, A. & Kahneman, D. (1982). Judgment Under Uncertainty: Heuristics and Biases. In A. Tversky & D. Kahneman (Eds.), *Judgment Under Uncertainty: Heuristics and Biases* (pp. 587–600). New York: Cambridge University Press.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, **90**, 293–315.

van Meter, K. M. (2002). Terrorists/liberators: Researching and dealing with adversarial social networks. *Connections*, **24**, 66–78.

Ward, M. D., Hoff, P. D., & Lofdahl, C. L. (2003). Identifying international networks: Latent spaces and imputation. In R. Breiger, K. Carley, & P. Pattison (Eds.), *Dynamic social network modeling and analysis: Workshop summary and papers* (pp. 345–359). Washington, DC: The National Academies Press.

Wasserman, S., & Faust, K. (1994). *Social network analysis*. New York: Cambridge University Press.

West, M. (1984). Bayesian aggregation. *Journal of the Royal Statistical Society, Series A*, **147**, 600–607.

Winkler, R. (1967). The assessment of prior distributions in Bayesian analysis. *Journal of the American Statistical Association*, **62**, 776–800.