Bayesian Analysis of Perceived Social Networks

Johan Koskinen[*]

Abstract

Measurement accuracy is an inherent problem in social network analysis. The issue of actor accuracy in the reporting of their interactions with others, was raised by Bernard, Killworth and Sailer (e.g. Bernard et al.,1980) and provoked extensive debate. Krackhardt (1987) later introduced the concept of Cognitive Social Structures and several methods for aggregating different actor reports on the network into a single graph, with the aid of which for example the congruence of reports could be gauged. A statistical model for aggregating separate reports into a single consensus network, with the additional benefit of allowing estimates of actor accuracy to be obtained in the process, was proposed by Batchelder, Kumbasar and Boyd (1997). The purpose here is to investigate this approach to the problem in a Bayesian framework. The Bayesian analysis yields posterior estimates of network characteristics and of perceiver "accuracy" as well as a measure of the degree of evidence in data for different models. The procedures are illustrated by an analysis of empirical data.

Keywords: Bayesian statistical modelling. Consensus analysis. Cognitive social structures (CSS). Measurement reliability. Social network analysis.

1 Introduction

In social network analysis it is customary to model social interaction between actors by graphs. In a graph representing a social network nodes would

[*]Departement of Statistics, Stockholm University, S-106 91 Stockholm, Sweden. E-mail: johan.koskinen@stat.su.se.

typically represent actors and the edges would typically represent the status of the interaction between the actors (see for example Wasserman and Faust, 1994, for an introduction). In a study of diffusion of an innovation among physicians, for example, the nodes would be physicians and we could say that there was an arc from one physician, A, to another physicians B, if A went to B to receive information and advice (Coleman, Katz and Mentzel, 1957).

Whereas statistics has been a commonly used tool since the birth of social network analysis it is only recently that Bayesian inference has been applied to the field, for example Wong (1987), Snijders and Nowicki (1997), Nowicki and Snijders (2001), Tallberg (2002), and Frank (2001). We believe that the analysis of social networks would find Bayesian analysis beneficial since in our view this type of inference is highly suited to dealing with small data sets as well as situations where regular classical procedures are found wanting.

Consider first the network as described by the digraph $G = (V, E)$, where $V = \{1, \ldots, n\}$ is the fix set of nodes and $E \subseteq V^{(2)}$ is the set of arcs connecting distinct nodes in the graph. (In the following $V^{(2)} = \{(u, v) \in V \times V : u \neq v\}$). The matrix $\mathbf{Z} = (z_{jk})_{(j,k) \in V^{(2)}}$ is then the indicator function such that $z_{jk} = 1$ if $(j, k) \in E$ and nought otherwise.

Typically, the edge set E is taken to be the self-reports of the actors in the network. In a series of articles Bernard, and Sailer (e.g. Bernard et al. 1980, Killworth et al., 1979) questioned taking the actor reports of their own interaction at face value. Their studies made it clear that what the actors reported about their own social interaction had little to do with with whom they actually interacted. So, assuming that there is one set E which corresponds to the structure of social interaction between the actors represented by the nodes in V, be it actual behaviour or something more vaguely defined, how can we acquire information about this structure?

Say that we have for $i = 1, \ldots, m$ reports $\mathbf{X}_i = (x_{ijk})$ on $\mathbf{Z}$ where $x_{ijk} = 1$ if informant i states that $z_{jk} = 1$. Since we consider $\mathbf{Z}$ a fix but unknown parameter the slightest discrepancy between our m reports implies that the notion that $x_{ijk} = z_{jk}$ for all i, j and k is not true. Even total unanimity is not enough to convince us that $x_{ijk} = z_{jk}$ for all i, j and k since $\mathbf{X}_i = \mathbf{X}_t$ for all $i, t \in \{1, \ldots, m\}$ does not preclude that $x_{ijk} = 1 - z_{jk}$ for some j, k and all i. Thus, any inference about G based on reports by actors must be based upon some assumptions regarding the relationship between $\mathbf{Z}$ and the reports $\mathbf{X} = (\mathbf{X}_i)_{i=1}^{m}$ for us to be able to make use of our data. A probability model that describes how data is generated might give us the double benefit

of describing this relationship as well as aiding us in estimating the edge set in G.

The interest in studying perceptions of social networks is present from the very beginning of social network analysis. Moreno (1934) studied different perceptions of a classroom friendship network. The focus was on the difference between what teachers and pupils thought about the social interaction amongst pupils.

What Batchelder et al. (1997) termed *three-way social network data* refers to elicitation of the views of a network from each member of the network. The idea of collecting social network data in this three-way manner was introduced by Newcomb (1961) and has since been used by for example Krackhardt (1987), Kumbasar, Romney and Batchelder (1994), Casciaro (1998). Krackhardt (1987) proposed the use of various aggregation rules for studying this type of data, data that he termed Cognitive Social Structures (CSS). A locally aggregated structure (LAS) could for example be obtained from the actor reports by only letting arcs between actors who themselves had reported an arc belong to the LAS - i.e. we say that there is an arc from j to k iff both j and k report this to be present - (intersection rule), or alternatively, an arc from j to k is said to belong to the LAS if either j or k report this to be present (union rule). A consensus structure (CS) is a function of all the m perceivers' perception of the (j, k) status, e.g. $z_{jk} = 1$ if $\sum_i x_{ijk} > m/2$. Note that the reports in the case of the consensus structure no longer need to be supplied by members of the network but might well come from external informants. CS and both types of LAS are implemented in the software program UCINET (Borgatti et al., 1999).

Such ad hoc methods are indeed intuitive and well suited for exploratory analysis as used for example in Frank and Nowicki (1993), where longitudinal data on a network were summarised through taking the union of the edge sets from different points in time. For an extensive review of CSS see Pattison (1994). For a presentation of *cultural consensus* (Romney, Weller and Batchelder, 1986; Romney, Batchelder and Weller, 1987; Weller, 1984 and 1987) and a thorough account of the debate initiated by Bernard, Killworth, and Sailer (see e.g. Bernard et al., 1980), see Johnson (1994).

Using the data collection paradigm outlined above, Batchelder et al. (1997) specified a model based on the following conditional probabilities (the general formulation of this model is sometimes referred to as the General

Condorcet Model, see Batchelder and Romney, 1988):

$$\Pr\left\{X_{ijk} = 1 \mid Z_{jk} = z_{jk}\right\} = \left\{ \begin{array}{lll} H_{ijk} & if & z_{jk} = 1 \\ F_{ijk} & if & z_{jk} = 0 \end{array} \right. . \tag{1}$$

These probabilities are called *hit* and *false alarm* probabilities respectively. A similar parameterisation can be found in Frank (1978) but that analysis was restricted to the case when the underlying structure was an equivalence relation. Assuming that there was a fix but unknown digraph with corresponding adjacency matrix $\mathbf{Z}$, and assuming that all actors - for here the study was restricted to the case when the informants were themselves members of the network - report on each $(j, k) \in V^{(2)}$ independently of each other with probabilities (1), the likelihood function could easily be derived. In order to be able to obtain any maximum likelihood estimates (MLE) of $\mathbf{Z}$ and the probabilities (1) at all, Batchelder et al.(1997) reduced the number of parameters by assuming $H_{ijk} = H_{ij}$ and $F_{ijk} = F_{ij}$ for all $k \neq j$, i.e. the actor was assumed to be equally "competent" in judging the presence and absence of arcs from a particular actor to all others (they called this restriction out-degree homogeneity).

As mentioned above, the framework of the model was provided by, in the most current form, Batchelder et al (1997) and has some very attractive features. There is however, much to recommend a Bayesian approach.

The MLE's of $\mathbf{Z}$, H_{ij} and F_{ij} are not defined (Batchelder et al., 1997, observation 4), with e.g. the out-degree homogeneity restriction, when either of $\sum_k z_{jk}$ or $\sum_k x_{ijk}$ is 0 or $n - 1$, for an i or j, a situation which is not wholly improbable (for example in Krackhardts (1987) high-tech management team). This not only limits what data can be used, but also places heavy restrictions on the parameter space.

An additional unfortunate feature of the model based on (1) is that it is not identified since a digraph and its complement give identical likelihoods. To solve this dilemma one usually needs to implement some side constraint to the effect that $H_{ij} \geqslant F_{ij}$, for all i and j. This is a fairly strong formulation of prior belief that cannot be evaluated, i.e. its probabilistic consequences cannot be properly assessed. By quantifying our prior belief, we make assumptions such as these explicit and make it possible to compare the fit of different prior beliefs to data.

Perhaps the strongest motivation for a Bayesian approach in the light of the above is that we require less of data in order to make estimations.

Not only do we not exclude certain structures from the sample and parameter space but also allow for incomplete network data (which we will see in section 4.2.). Finally, the Bayesian approach yields probabilistic statements regarding our uncertainty about estimates.

There is no denying that introducing priors introduce new problems. These are however not primarily of a technical nature in this case rather it is the choice of priors that takes some efforts.

Krackhardt (1987) made a point of distinguishing between two conceptions of social structure, namely *i)* actual (behavioral) structure and *ii)* cognitive social structures (CSS). In other words, you could regard actor reports *i)* as biased reports on an actual structure or *ii)* as straightforward reports of the social structure as it exists in the mind of each actor. An example of the latter is Newcomb's (1961) study of perceived perfect triads. When modelling, i.e. setting priors etc., we need to think about $\mathbf{Z}$ as if *i)* is true although we believe *ii)* to be a more realistic approach. By this, we mean that the competencies of the perceivers are defined in relation to a "true" structure.

There are clear parallels to item response theory (IRT) save that the relation between the reports and the items cannot be directly observed. It would be interesting to see whether the substantial body of work on IRT could be applied here. For a comprehensive presentation of IRT see van der Linden and Hambleton (1997).

The paper is structured as follows. In section 2 the notation and likelihood function are introduced. Inference regarding model parameters is discussed in section 3 and in section 4 the details of the estimation procedures are given along with a few related results. Model selection and model inference is discussed in the fifth section. Certain aspects of the Markov chain Monte Carlo (MCMC) procedure particular to the present model framework are dealt with in section 6. The results presented in this paper are finally illustrated in section 7, using an empirical data set.

2 Model specification

Say that we are interested in a social network consisting of n actors and that we have reports on this network from m informants. The informants could also themselves be actors in this social network but they do not necessarily have to. Although we in the following assume that the social structure can

be described by a directed graph with loops very little changes when the structure is e.g. a loop less graph.

For actors $V = \{1, 2, \ldots, n\}$ and a relation $R \subseteq V^2$, let $\mathbf{Z} = (z_{jk})_{j,k \in V}$ be a matrix with elements $z_{jk} = 1$ if $(j, k) \in R$, and $z_{jk} = 0$ otherwise for $j, k \in V$. Let the matrix $\mathbf{X}_i$, for $i \in \mathcal{I} = \{1, \ldots, m\}$, be defined in the same way for relation R_i. Further assume that conditional on $\mathbf{Z}$, $\mathbf{X}_1, \ldots, \mathbf{X}_m$ are independent and their elements are independent and satisfy

$$\Pr\left(X_{ijk} = x_{ijk} \middle| \mathbf{Z} = \mathbf{z} \right) = \Pr\left(X_{ijk} = x_{ijk} \middle| Z_{jk} = z_{jk} \right). \tag{2}$$

Define for the elements X_{ijk}, in each $\mathbf{X}_i$

$$\Pr\left(X_{ijk} = s \middle| Z_{jk} = s \right) = \eta_{ijk}\left(s \right),$$

where $s = 0, 1$. The $\eta_{ijk}\left(s \right)$'s are in some respect the competencies of the informants in judging the presence or absence of ties in the social network. Now we can write the likelihood function for $\mathbf{X} = (\mathbf{X}_i)_{i \in \mathcal{I}}$ given the parameters as a product of Bernoulli probability mass functions

$$r\left(\mathbf{x} \middle| \mathbf{z}, \boldsymbol{\eta} \right) = \prod_{i \in \mathcal{I}} \prod_{j,k \in V} \eta_{ijk}\left(z_{jk} \right)^{\mathbf{1}\left\{ x_{ijk} = z_{jk} \right\}} \left(1 - \eta_{ijk}\left(z_{jk} \right) \right)^{\mathbf{1}\left\{ x_{ijk} \neq z_{jk} \right\}} \tag{3}$$

where $\boldsymbol{\eta}$ is an array containing $\eta_{ijk}\left(s \right)$, for $i \in \mathcal{I}$, $j, k \in V$ and $s = 0, 1$, and where $\mathbf{1}\left\{ A \right\} = 1$ if A is true and nought otherwise.

An obvious detriment of the formulation in (3) is that the model requires no less than $n^2\left(2m + 1 \right)$ parameters if both $\boldsymbol{\eta}$ and $\mathbf{Z}$ are considered parameters. For loop less digraphs and graphs the number of parameters are $n\left(n - 1 \right)\left(2m + 1 \right)$ and $\binom{n}{2}\left(2m + 1 \right)$ respectively. Not only would a reduction of the number of parameters facilitate estimation procedures but it is, we suggest, not a reasonable assumption that each informant should have two distinct probabilities for each (ordered) pair of actors. As mentioned earlier, Batchelder et al (1997) suggested a couple of designs meant to reduce the number of parameters. They assumed for example that an informant for each actor was equally competent in judging the ties from that actor to others. Another strategy was to assume that an informant for each actor was equally competent in judging the ties from others. Both these restrictions reduced the number of parameters from $n\left(n - 1 \right)\left(2m + 1 \right)$ to $n\left(2m + n - 1 \right)$.

Now, an obvious extension of this method of reducing the number of parameters, which Batchelder et al (1997) hinted at, is to allow for arbitrary

partitions of the perceiver-actor space. For example an informant could be assumed to be equally competent in judging the ties to and from others no matter which pair of actors is under consideration or, it could be that if the informer is himself an actor he could be considered especially competent in judging his owns ties to and from others. Should we have access to covariates, say the informants were teachers and the actors were children in a school class, these could be used to label the competencies. It could for example be assumed that all male teachers report with the same probabilities and all female teachers report with the same probabilities.

Introduce the labeling set C, $|C| = c$, that corresponds to different competency categories. Let f be the function which maps $\mathcal{I} \times V^2$ onto C. We want that $\eta_{ijk}(s)$ should only depend on what image (i, j, k) have in C. To achieve this relabel the parameters $\eta_{\nu}(s)$, for $\nu \in C$ and introduce the following counters

$$n_{\nu}(s) = \sum_{ijk:f(i,j,k)=\nu} \mathbf{1}\left\{z_{jk} = s, x_{ijk} = s\right\},$$

and

$$n_{\nu}^{*}(s) = \sum_{ijk:f(i,j,k)=\nu} \mathbf{1}\left\{z_{jk} = s, x_{ijk} \neq s\right\}.$$

Thus the likelihood can be written

$$r\left(\mathbf{x} \mid \mathbf{z}, \boldsymbol{\eta}\right) = \prod_{\nu \in C} \prod_{s \in \{0,1\}} \eta_{\nu}(s)^{n_{\nu}(s)} \left(1 - \eta_{\nu}(s)\right)^{n_{\nu}^{*}(s)}. \tag{4}$$

The use of indicator functions in the counters point in a natural way to extensions of binary relations to valued graphs.

3 Estimation

Based on observed data $\mathbf{X}$ and the generic model structure specified in (4) we want to make inference about $\boldsymbol{\eta}$ as well as $\mathbf{Z}$. Whereas the former assumes the normal role of a parameter, the latter can either be viewed as a parameter or be considered an unobserved variable. As parameter, when we estimate $\mathbf{Z}$ we estimate every element of $\mathbf{Z}$. Thus the number of parameters grows rapidly with n.

Z, treated as a latent variable, allows for a significant reduction of the parameter space provided a sufficiently rigid probability structure on V^2 is defined. The inference then made about **Z** would be, apart from estimates of its guiding parameters, a suitably chosen posterior prediction (note the analogy between these considerations and similar ones in the case of mixture models, e.g. Bock, 1996).

An additional strategy aimed at reducing the dimensions of the parameters could well be to treat $\boldsymbol{\eta}$ as an unobserved variable. Although the difference might seem small, this distinction is of some importance when interpreting the model. In the case where one treats $\boldsymbol{\eta}$ as a regular parameter and expects there to be, at least in theory, a fix quantity that captures the competency of the informant in a given region of the network, and further assume that given enough information about it we could more or less pin-point its exact value. As a variable vector $\boldsymbol{\eta}$ would vary and we would assume a different realisation of this vector for every sample we take. An extension to the model, to be treated elsewhere, is to let $\boldsymbol{\eta}$ be the response of covariate variables.

In accordance with Bayesian inference, all inference concerning the parameters is based upon the posterior distribution of the parameters conditioned on what we have actually observed, i.e. data. The uncertainty in our model stems from our parameters and this we model, a priori, with prior distributions.

In the current context it is important to remember that the posteriors contain all the relevant information regarding our parameters given data. While a distribution over the space of all digraphs might not seem all that attractive, all the typical characteristics of a network, closeness (Freeman, 1979), reachability (Doreian, 1974), influence (Katz ,1953), structural holes (Burt, 1992), etc., are easily obtained from the posterior distribution of **Z**. Since the summary measures listed above are all functions of graphs or digraphs, the posterior distributions of these measures capture our uncertainty about these parameters whose point predictions for example could be the relevant expected values. Doing this one should keep in mind that the prior represents our uncertainty, and is specified, in relation to **Z**.

4   Priors and posteriors

For inference about $\mathbf{Z}$ an uninformative or vague prior is the uniform distribution over $\{0,1\}^{n^2}$. This prior is vague in the sense that it does not favor any particular configuration of the social network. It has to be kept in mind, though, that this prior is only uninformative with respect to distinct relations but not with respect to e.g. the number of elements in the relation.

For convenience, a prior on $\boldsymbol{\eta}$ could be chosen from a conjugate family of distributions. The conjugate prior on $\boldsymbol{\eta}$ is a mixture of beta densities. If the joint prior for $\boldsymbol{\eta}$ and $\mathbf{Z}$ is the product of beta densities and some probability mass function for $\mathbf{Z}$ the marginal posterior of $\boldsymbol{\eta}$ is a mixture of beta densities with the marginal posterior of $\mathbf{Z}$ as weights.

Let $\varkappa\left(\cdot\mid\boldsymbol{\lambda}_{\nu,s}\right)$ be a beta density with parameters $\boldsymbol{\lambda}_{\nu,s}=\left(\alpha_\nu\left(s\right),\beta_\nu\left(s\right)\right)$,

$$\varkappa\left(\eta_\nu\left(s\right)\mid\boldsymbol{\lambda}_{\nu,s}\right)=\frac{\eta_\nu\left(s\right)^{\alpha_\nu(s)-1}\left(1-\eta_\nu\left(s\right)\right)^{\beta_\nu(s)}}{B\left(\alpha_\nu\left(s\right),\beta_\nu\left(s\right)\right)},$$

where $B\left(a,b\right)=\int_0^1 t^{a-1}\left(1-t\right)^{b-1}dt$, and for all $\nu,\nu'\in C$ and all $s,s'\in\{0,1\}$ let $\eta_\nu\left(s\right)$ and $\eta_{\nu'}\left(s'\right)$ be independent, and independently thereof let the prior on $\mathbf{Z}$ be $p\left(\mathbf{z}\right)$. We then have the joint prior distribution $\phi\left(\boldsymbol{\eta}\right)p\left(\mathbf{z}\right)$, where

$$\phi\left(\boldsymbol{\eta}\right)=\prod_{\nu\in C}\prod_{s\in\{0,1\}}\varkappa\left(\eta_\nu\left(s\right)\mid\boldsymbol{\lambda}_{\nu,s}\right). \tag{5}$$

Clearly the conditional distribution of $\boldsymbol{\eta}$ given data and a specific graph $\mathbf{z}$ can be written

$$\phi\left(\boldsymbol{\eta}\mid\mathbf{z},\mathbf{x}\right)=\prod_{\nu\in C}\prod_{s\in\{0,1\}}\varkappa\left(\eta_\nu\left(s\right)\mid\boldsymbol{\lambda}_{\nu,s}\left(\mathbf{z}\right)\right),$$

where $\boldsymbol{\lambda}_{\nu,s}\left(\mathbf{z}\right)=\left(\alpha_\nu\left(s\right)+n_\nu\left(s\right),\beta_\nu\left(s\right)+n_\nu^*\left(s\right)\right)$. The marginal likelihood of $\mathbf{z}$ is obtained from

$$\begin{aligned}r\left(\mathbf{x}\mid\mathbf{z}\right)&=\int_{[0,1]^{2c}}r\left(\mathbf{x}\mid\mathbf{z},\boldsymbol{\eta}\right)\phi\left(\boldsymbol{\eta}\right)d\boldsymbol{\eta}\\&=\prod_{\nu\in C}\prod_{s\in\{0,1\}}\frac{B\left(\boldsymbol{\lambda}_{\nu,s}\left(\mathbf{z}\right)\right)}{B\left(\boldsymbol{\lambda}_{\nu,s}\right)}.\end{aligned}$$

9

This gives the marginal posterior distribution

$$\pi_{\mathbf{z}} = \frac{r\left(\mathbf{x} \middle| \mathbf{z}\right) p\left(\mathbf{z}\right)}{\sum_{\mathbf{u} \in \{0,1\}^{n^2}} r\left(\mathbf{x} \middle| \mathbf{u}\right) p\left(\mathbf{u}\right)},$$

for $\mathbf{Z}$ given data and the marginal posterior

$$\phi\left(\boldsymbol{\eta} \middle| \mathbf{x}\right) = \sum_{\mathbf{z} \in \{0,1\}^{n^2}} \pi_{\mathbf{z}} \phi\left(\boldsymbol{\eta} \middle| \mathbf{z}, \mathbf{x}\right), \tag{6}$$

for $\boldsymbol{\eta}$ given data. The marginal posterior (6) is clearly a mixture of 2^{n^2} product beta densities. Note that although each component consists of densities of independent beta distributed variables the competencies are not necessarily independent a posteriori .

With the same priors the marginal posterior distribution of $\mathbf{z}$ is $\pi_{\mathbf{z}}$. Hence, if we have a function g defined on directed graphs we can easily calculate the expected value of this function given data, viz.

$$E\left(g\left(\mathbf{Z}\right) \middle| \mathbf{x}\right) = \sum_{\mathbf{z} \in \{0,1\}^{n^2}} \pi_{\mathbf{z}} g\left(\mathbf{z}\right).$$

Evaluating the summand soon quite naturally becomes a rather painstaking business as the number of terms grows rapidly with n. Whereas this is possible with small graphs or certain specifications of f, typically the evaluation has to be done through simulation, as is common in Bayesian inference.

4.1   Relationship to other methods and models for aggregation

It is interesting to note that a few of the aggregation methods in analysis of cognitive social structures can be formulated within the framework out-lined above as limiting cases of certain prior specifications of the model. These connections might serve as references when setting priors.

Assume that we disallow reflexive pairs and thus $\mathbf{Z}$ is the adjacency matrix of a directed loop less graph with vertex set $V = \{1, 2, \dots, n\}$. Assume further that we have $\mathbf{X}_i$, for $i \in V$, and let $f : V \times V^{(2)} \to C$, where

$$f\left(i, j, k\right) = \begin{cases} 1 & if \quad i \in \{j, k\} \\ 2 & otherwise \end{cases}$$

Let $\widetilde{\mathbf{z}}_{LU} = (\widetilde{z}_{jk})$ with elements $\widetilde{z}_{jk} = \max(x_{jjk}, x_{kjk})$ denote Krackhardts (1987) locally aggregated structure (LAS) from the union rule. With conjugate prior on $\boldsymbol{\eta}$, independently thereof a vague prior on $\mathbf{z}$, and $\alpha_1(1)$, $\beta_1(1)$, $\alpha_2(s)$, and $\beta_2(s)$ finite and strictly positive for $s = 0, 1$ we see that the posterior $\pi(\widetilde{\mathbf{z}}_{LU}|\mathbf{x})$ tends to unity as $\alpha_1(0)$ tends to infinity for every strictly positive and finite $\beta_1(0)$. In other words, a point mass prior on $\eta_1(0)$ centered over 1 always yields a one point posterior on $\widetilde{\mathbf{z}}_{LU}$.

In a similar way we obtain the LAS from the intersection rule, $\widetilde{\mathbf{z}}_{LI}$, with elements $\widetilde{z}_{jk} = \min(x_{jjk}, x_{kjk})$. With conjugate prior on $\boldsymbol{\eta}$, independently thereof a vague prior on $\mathbf{z}$, and $\alpha_1(0)$, $\beta_1(0)$, $\alpha_2(s)$, and $\beta_2(s)$ finite and strictly positive for $s = 0, 1$ we see that $\pi(\widetilde{\mathbf{z}}_{LI}|\mathbf{x})$ tends to unity as $\alpha_1(1)$ tends to infinity for every strictly positive and finite $\beta_1(1)$. In other words, a one point mass prior on $\eta_1(1)$ centered over 1 always yields a one point posterior on $\widetilde{\mathbf{z}}_{LI}$.

Although giving rise to degenerate priors these considerations can well be used for comparison.

Krackhardt's (1987) Consensus structure (CS) with majority rule can be seen as the MLE of the modal graph g^* of a metric model described in Banks and Constantine (1998; se Eq. 7 below). The model, in the present framework, corresponding to CS with majority rule is one where f maps $V \times V^{(2)}$ to the singleton set C and we have, for example, the following priors. Let $\eta(1)$ have density $p(\eta(1))$, say for example uniform, on the interval $(.5, 1)$, $\varphi \sim Beta(a, b)$ and $\xi \sim Bernoulli\left(\frac{1}{2}\right)$ and, with $\eta(1)$, φ and ξ independent,

$$p(\eta(0)|\eta(1)) = (1 - \varphi)\eta(1) + \varphi(\xi + .5(1 - \xi)).$$

If we let b tend to infinity and keep a finite we obtain a posterior for $\mathbf{Z}$ that is unity for the CS and nought otherwise. The connection of the present model to CS can seem tenuous as CS has a probabilistic interpretation which stems from a model where $\eta(1) = \eta(0)$.

Let $\mathbf{Z}$ be the symmetric binary adjacency matrix of a loopless graph with vertex set $V = \{1, 2, \ldots, n\}$ and assume further that we have observed adjacency matrices $\mathbf{X}_i$, on the same vertex set for $i \in \mathcal{I} = \{1, \ldots, m\}$. Say that f maps every element in $\mathcal{I} \times \binom{V}{2}$ to the singleton set C, and we set $\eta(1) = \eta(0) = \eta$, then $\{\mathbf{X}_i\}$ can be seen as the collection of m iid random variables from the distribution

$$p(\mathbf{x}_i|\mathbf{z}) = \left(1 + e^{-\tau}\right)^{\binom{n}{2}} e^{-\tau d(\mathbf{z}, \mathbf{x}_i)}, \tag{7}$$

on $\mathcal{G} = \{0, 1\}^{\binom{n}{2}}$, where τ is the log odds, $\eta / (1 - \eta)$, and $d(\mathbf{z}, \mathbf{x})$ is the symmetric class difference between $\mathbf{z}$ and $\mathbf{x}$. The parameter τ can be seen as a measure of the dispersion around the "central" element $\mathbf{z}$, and it is interesting to note that (7) has uniform measure on $\mathcal{G}$ iff $\eta = .5$. The natural choice, perhaps, for a joint prior on $\mathbf{z}$ and τ would be uniform on the first and exponential on the latter. For a general presentation of the Gibbs family of distributions over $\mathcal{G}$ and a Bayesian approach to (7) see Banks and Constantine (1998).

4.2   Partial network data

Consider the case when we have partial network data, that is to say at least one informant has not given a report for the entire network. While f is defined on $\mathcal{I} \times V^{(2)}$, for a directed loop less graph, we might have observations $\mathbf{X} = (x_{ijk})_{ijk \in S}$ where $S \subset \mathcal{I} \times V^{(2)}$. Inference regarding $\boldsymbol{\eta}_\nu$ would be unproblematic should the intersection of the inverse image of ν and S be empty for some $\nu \in C$, seeing as the prior is our best guess before we obtain data. At least this holds true if the conjugate prior $\phi(\boldsymbol{\eta})$ is used. The parameters are also a priori independent so that the prior of a $\boldsymbol{\eta}_\nu$ for a $\nu \in C$ for which there are no observations does not influence the analysis. If we let for a fix ν, $A_\nu = \{a \in \mathcal{I} \times V^{(2)} : f(a) = \nu\}$ and $A_\nu \cap S$ is non-empty but has a cardinality that is less than $|A_\nu|$ the analysis can be carried out nonetheless as long as $(x_{ijk})_{ijk \in A_\nu \cap S}$ can be seen as independent observations.

What impact partial information might have on the estimation of $\mathbf{Z}$ is not entirely self evident. Again, the model is still well specified (define $n_\nu(s)$, and $n_\nu^*(s)$ to null if there are no observations $x_{ijk} : f(i, j, k) = \nu$), only that we make inference conditioned on an event in a subspace of the sample space, $(X_a)_{a \in S}$, rather than an event in the whole sample space, $(X_a)_{a \in \mathcal{I} \times V^{(2)}}$. Moreover, should $A_\nu \cap S \in \{\emptyset, A_\nu\}$ for all $\nu \in C$ we have that

$$\pi[\mathbf{z}| (x_{ijk})_{ijk \in S}] \propto \pi[(z_{jk})_{jk \in S'} | (x_{ijk})_{ijk \in S}],$$

where S' is the set of all elements (j, k) in $V^{(2)}$ such that $(i, j, k) \in S$ for some $i \in \mathcal{I}$.

5   Model selection

Keeping with the framework outlined above it is clear that for most studies we could think of many different ways to specify the partition of the perceiver-actor space. It would seem like a very limited scope of investigation indeed if we had to restrict ourselves to choosing only one partition of the perceiver-actor space and with it only one specification of priors. The answer to this dilemma is that we allow for analysis from a variety of perspectives - what would our final analysis of data be if *this* was the case and what would it be if *that* were the case. In other words, we want to analyse data from many different angles simultaneously.

Henceforth, for the perceiver set $\mathcal{I}$ and the actor set V a model M is a specification of a function f and priors $\phi(\boldsymbol{\eta})$ and $p(\mathbf{z})$, and we let $\mathcal{M}_0$ be the class of all such models. We are not interested in using all models in $\mathcal{M}_0$, for obvious reasons, but rather we select a subset of models $\mathcal{M} \subset \mathcal{M}_0$ that we find plausibly describes the data generating process. To each of these models we associate a probability that represents our prior belief in the different models, $p(M_i)$, $M_i \in \mathcal{M}$. Assume that $\mathcal{M}$ consists of D distinct models and that we denote priors on the competencies and the true structure conditional on the model M_i, $\phi_i(\boldsymbol{\eta})$ and $p_i(\mathbf{z})$ respectively, for $i = 1, \dots, D$. The latter two, $\boldsymbol{\eta}$ and $\mathbf{z}$, are independent conditional on M_i.

The distribution of data given a model M_i is

$$r(\mathbf{x}|M_i) = \sum_{\mathbf{z} \in \{0,1\}^{n^2}} \int_{[0,1]^{2c}} r_i(\mathbf{x}|\mathbf{z}, \boldsymbol{\eta}) \, p_i(\mathbf{z}) \, \phi_i(\boldsymbol{\eta}) \, d\boldsymbol{\eta},$$

where f is specified according to M_i, and from this the posterior probability of the model conditional on data can be obtained from straightforward probability calculus as in the previous cases

$$p(M_i|\mathbf{x}) = \frac{r(\mathbf{x}|M_i) \, p(M_i)}{\sum_{i=1}^{D} r(\mathbf{x}|M_i) \, p(M_i)}.$$

A quantity that is often used in Bayesian statistics for model selection is the Bayes factor between two competing models

$$BF_{ij} = \frac{r(\mathbf{x}|M_i)}{r(\mathbf{x}|M_j)}, \quad i, j = 1, \dots, D$$

13

where $r\left(\mathbf{x}|\,M_i\right)$ is evaluated in the sample. The Bayes factor has the property that for models M_i, M_j and M_k if $BF_{ij} > 1$ and $BF_{jk} > 1$ then $BF_{ik} > 1$, so that if we in the light cast by data prefer M_i to M_j and prefer M_j to M_k we also prefer M_i to M_k. Note also that the Bayes factor does not depend on the prior belief in the different models (Kass and Raftery, 1995). While this might be an advantage in some situations this also means that we don't get any measure on the uncertainty in our model selection procedure. That is to say, we do not get any probabilistic, statements only ad-hoc rules of thumb (see e.g. ibid. p.777) that say how strong evidence we have.

We think that it is useful to report the posterior distributions of different, preferably all, models as computing them does not entail any burdensome calculations should one have the marginal likelihoods $r\left(\mathbf{x}|\,M_i\right)$. When computing the posteriors of models it is important to keep in mind that these quantities are reported conditional on the assumption that the true model is contained in $\mathcal{M}$ - i.e. models outside $\mathcal{M}$ have null probability. Because of this we might want to extend the class of models that have a strictly positive probability as much as possible. To determine the prior probability for different models might not be all that hard but if as in this case, dimensions vary between models we have to specify the priors for the parameters for each and every model. A first attempt to simplify things could be to let the probability of a given model be uniform given the dimensions of the competency vector, and have a discrete distribution to model the prior probability of the dimensions. In addition, one could let all the prior competencies be uniformly distributed. This way one only has to specify this last distribution a priori. The model (4), however is not identified with uniform priors everywhere meaning that we always have to specify proper and "informative" priors. One alternative, with $U\left(a,b\right)$ being the uniform distribution on the interval $\left(a,b\right)$, is letting

$$\left(\eta_\nu\left(1\right)|\,\eta_\nu\left(0\right)\right) \sim U\left(\eta_\nu\left(0\right),1\right) \tag{8}$$

and

$$\eta_\nu\left(0\right) \sim U\left(0,1\right) \quad \text{for all } \nu \in C. \tag{9}$$

This is however a fairly strong assumption and not "uninformative" at all but it does provide a way of automating the setting of priors. To complete this discrete model extension one could let $\left(c-1\right) \sim Bin\left(p,n^2m\right)$ a priori and the prior on individual functions given the number of classes be $p\left(f|\,c\right) =$

$S\left(n^2m,c\right)^{-1}$, where S is the Sterling number, i.e. the function that classify elements in $\mathcal{I} \times V^2$ is uniform over all functions with that number of classes. Our view is that the prior (8) and (9) on $\boldsymbol{\eta}$ is far to restrictive.

We do not see any simple solution to this problem at the moment. Rather we recommend choosing $\mathcal{M}$ so that the evaluation of the models is manageable. A little variety can be added through testing a number of hyper parameters for a conveniently chosen set of functions f.

If f differs for different models, the posterior distribution of the competencies has to be conditional on the models as well as data. The posterior distribution of $\mathbf{Z}$, on the other hand, can always be marginalised so that it is conditioned only on data.

When comparing posterior distributions given (data and) different models we can compare the posterior distributions of functions of $\mathbf{Z}$ as well as features of the posterior of $\mathbf{Z}$ itself. The usual Bayes estimate of a parameter, the posterior expected value, can be defined as a matrix with the marginal posterior probabilities of edges as elements, $p_{jk} = \sum_{\mathbf{z}} z_{jk} p\left(\mathbf{z}\mid \mathbf{x}\right)$. Similarly, the posterior covariance matrix is the $n^2 \times n^2$ matrix of covariances $Cov_{\mathbf{z}\mid \mathbf{x}}\left(z_{jk}, z_{lm}\right) = \sum_{\mathbf{z}} z_{jk} z_{lm} p\left(\mathbf{z}\mid \mathbf{x}\right) - p_{jk} p_{lm}$.

To describe the posterior uncertainty about $\mathbf{Z}$ a useful measure is the entropy defined as

$$H\left(\mathbf{z}\mid \mathbf{x}\right) = E_{\mathbf{z}\mid \mathbf{x}}\left(\log \frac{1}{p\left(\mathbf{z}\mid \mathbf{x}\right)}\right).$$

6 Markov chain Monte Carlo sampling

Since the normalizing constants of our posteriors both include a sum taken over $\{0,1\}^{n^2}$ exact evaluation of the summands is feasible only when n is relatively small. For more realistic applications of the model one can either resort to some sort of approximation of the posteriors or employ a Markov chain Monte Carlo (MCMC) sampling scheme to get a sample from the exact posterior distributions. The first approach will not be examined further here, rather we prefer the latter.

When one does not have access to the analytic expression for a joint posterior distribution but, as in this case, has a relatively simple expression for the full conditional posteriors of each of the parameters of interest, one typically uses a Gibbs sampling scheme (see e.g. Gelfand and Smith, 1990). In

our case however the expression (6) is a mixture of known pdf's with the posterior of $\mathbf{Z}$ as weights so that once we have obtained the posterior, or rather a sample from the posterior of $\mathbf{Z}$ given data, it is relatively straightforward to calculate the posterior of the competencies. Hence, we need only sample from the posterior distribution of $\mathbf{Z}$ given data.

The Metropolis-Hastings sampling scheme (Metropolis et al.,1953) is used here. The acceptance probabilities are in this case the ratio between the kernels according to the posterior distribution between the current state and the proposed state. Given, say, a directed graph G with adjacency matrix $\mathbf{Z}$, the proposed change, $\widetilde{\mathbf{Z}}$, is the adjacency matrix of a directed graph $\widetilde{G}$ chosen uniformly at random from the neighborhood of G, defined as all directed graphs such that the symmetric class difference between their edge sets and the edge set of G is exactly one.

As always, convergence to the stationary distribution, i.e. the posterior distribution of $\mathbf{Z}$, is virtually impossible to check. This is made even more difficult in the present application since there is no easy way of monitoring the walk on the state space because of its high dimensionality. We recommend in this case that one monitors for example the number of edges for each state as well as marginal likelihoods and the acceptance rates. For these quantities moving averages can be calculated and from plots of them one can get at least an idea as to when the process seems to have settled down.

Reporting the posterior of $\mathbf{Z}$ is best done with a summary measure of the distribution or the posterior distribution of some low dimensional function of $\mathbf{Z}$. All this can be obtained from the sampling scheme described above. An alternative or complement is to report the modal point of the posterior distribution. Finding the modal point is generally not feasible analytically wherefor a numerical optimization routine has to be employed. We have found simulated annealing to be a perfectly satisfactory algorithm. Details and an introduction to the procedure can be found in e.g. van Laarhoven and Aarts (1988) or Fishman (1999).

A technical detail we have to address is the issue of whether to take a single run or have several restarts. The former alternative always gives rise to dependence between observations but this is typically considered to be of minor importance. In this case however, when a Metropolis-Hastings algorithm is used with the said neighbourhood generator, the dependence between consecutive states can be considerable since only one position in the edge configuration is changed at a time. This does not imply that one has to restart the algorithm after each produced value, rather it seems reasonable

to adjust the interval i.e. the number of steps between each value that is taken as out put. It could still be considerably smaller that the burn in time.

So, how do we choose the number of graphs to be discarded between each graph that is taken as an output? Say that we from the MCMC run choose every k^{th} value. A heuristic argument for choosing k is that we should choose $k \geqslant n^2$ since $\max_{G,G' \in \mathcal{G}} \{\min_k \{k \in \mathbf{N} : \Pr(\mathbf{Z}_k = G' | \mathbf{Z}_0 = G) > 0\}\} = n^2$. However, because of the formulation of the model we expect the posterior to be concentrated around a certain point as long as we have specified "informative" priors. Hence if we at step t have a graph with high posterior probability the probabilities of graphs are expected to decrease with the distance from the present graph. We think that it suffices to choose $k = cn$ for some strictly positive c, positively a function of m, and check for multiple high density regions in the parameter space by having a few different runs with different staring points.

The posterior probability in different points of $\{0,1\}^{n^2}$ is not very useful. Because probability mass has to be allocated to every element of $\{0,1\}^{n^2}$, and since in our MCMC-output $\widehat{\pi}_{\mathbf{z}} \in \{0, 1/N, 2/N, \dots, 1\}$, estimates of the posterior probability of any one element are always flawed. The sampling scheme is still simulation consistent and properties of the posterior distribution can be expected to be reasonably well mirrored in the MCMC-sample. These properties might for example be the summary measures mentioned in Section 3, posterior entropy and certain metrics on $\{0,1\}^{n^2}$. The rationale behind the latter is that the model itself can be seen as a metric model albeit with several dimensions.

When calculating the marginal likelihood $r_i(\mathbf{x})$ an intuitive approach is sampling simultaneously from $r_i(\mathbf{x}, \mathbf{z})$ and obtaining a sample and using the relative frequency of occurrences of the sample. This is not something we recommend since the number of visits to $\mathbf{x}$ in the sample space is likely to be small as a consequence of the cardinality of the sample space.

One solution to the problem when analytical expressions are not available and when sampling from the prior distributions is not feasible or in another way not practical, is using an importance sampling scheme (Raftery, 1996). Suppressing the notational dependency on the model and writing $r(\mathbf{x})$ for a generic $r(\mathbf{x}|M)$, a simulation consistent estimate of the marginal likelihood of data is given by the harmonic mean of $r(\mathbf{x}|\mathbf{z})$ taken over the sample, $\{\mathbf{z}^{(a)}\}$, the output from the Metropolis-Hastings sampling scheme. The likelihood $r(\mathbf{x}|\mathbf{z}^{(a)})$ will most probably be very small for some $\mathbf{z}^{(a)}$ so that the

estimate will be dominated by these and the estimator therefore tends to be unstable.

An alternative is to use the fact that we have a good estimate of $\phi\left(\boldsymbol{\eta} \mid \mathbf{x}\right)$. The marginal likelihood can be calculated using Bayes theorem

$$r\left(\mathbf{x}\right) = \frac{r\left(\mathbf{x} \mid \boldsymbol{\eta}^*\right) \phi\left(\boldsymbol{\eta}^*\right)}{\phi\left(\boldsymbol{\eta}^* \mid \mathbf{x}\right)} \tag{10}$$

for an arbitrary $\boldsymbol{\eta}^* \in (0,1)^{2c}$. However,

$$r\left(\mathbf{x} \mid \boldsymbol{\eta}\right) = \sum_{\mathbf{z} \in \{0,1\}^{n^2}} r\left(\mathbf{x} \mid \mathbf{z}, \boldsymbol{\eta}\right) p\left(\mathbf{z}\right),$$

is hardly tractable. An identity equivalent to (10) of course holds for $\pi_{\mathbf{z}}$ but as noted earlier, $\widehat{\pi}_{\mathbf{z}}$ is not a good estimate for individual $\mathbf{z}$'s. As a parenthesis though, we can mention that

$$r\left(\mathbf{x} \mid g\left(\mathbf{z}\right) = Q\right) = \sum_{\mathbf{z}:g(\mathbf{z})=Q} r\left(\mathbf{x} \mid \mathbf{z}\right),$$

would be a much more stable quantity to work with than $r\left(\mathbf{x} \mid \mathbf{z}\right)$. Especially so, if the sum is not too big and if we can easily compute g for $\left\{\mathbf{z}^{(a)}\right\}$. We will not elaborate on this any further here but choose instead to carry on with the initial idea.

Set $\boldsymbol{\eta}^* = \left(\frac{1}{2}, \ldots, \frac{1}{2}\right)$, then using the identity

$$\sum_{\nu \in C} \sum_{s \in \{0,1\}} \left(n_\nu\left(s\right) + n_\nu^*\left(s\right)\right) = n^2 m,$$

which does not depend on data or $\mathbf{Z}$, we can write

$$\begin{aligned}
r\left(\mathbf{x} \mid \mathbf{z}, \boldsymbol{\eta}^*\right) &= \left(\frac{1}{2}\right)^{\sum_{\nu \in C} \sum_{s \in \{0,1\}} \left(n_\nu(s) + n_\nu^*(s)\right)} \\
&= \left(\frac{1}{2}\right)^{n^2 m},
\end{aligned}$$

and hence

$$\begin{aligned}
r\left(\mathbf{x} \mid \boldsymbol{\eta}^*\right) &= \sum_{\mathbf{z} \in \{0,1\}^{n^2}} r\left(\mathbf{x} \mid \mathbf{z}, \boldsymbol{\eta}^*\right) p\left(\mathbf{z}\right) \\
&= 2^{-mn^2}.
\end{aligned}$$

18

Now

$$\log r\left(\mathbf{x}\right) = \log \frac{2^{-m}\phi\left(\frac{1}{2}, \ldots, \frac{1}{2}\right)}{\phi\left(\frac{1}{2}, \ldots, \frac{1}{2} \middle| \mathbf{x}\right)},$$

where the denominator can be estimated from the Metropolis-Hastings sample by

$$\log \sum_{a=1}^{N} \frac{1}{N} \prod_{\nu \in C} \prod_{s \in \{0,1\}} \frac{\left(\frac{1}{2}\right)^{\alpha_\nu(s) + \beta_\nu(s) + n_\nu^{(a)}(s) + n_\nu^{*(a)}(s) - 2}}{B\left(\alpha_\nu\left(s\right) + n_\nu^{(a)}\left(s\right), \beta_\nu\left(s\right) + n_\nu^{*(a)}\left(s\right)\right)}, \qquad (11)$$

taken over $\left\{\mathbf{z}^{(a)}\right\}$.

While this is a convenient formula in that we do not have to make additional simulations, we could run into similar numerical problems as when we take the harmonic mean of the marginal likelihood of $\mathbf{z}$. Numerical problems in this case arise as a consequence of the fact that $\boldsymbol{\eta}^* = \left(\frac{1}{2}, \ldots, \frac{1}{2}\right)$ is typically not a high density region in the posterior (nor in the prior in most instances). If the posteriors are concentrated away from $\boldsymbol{\eta}^*$, the estimate of $r\left(\mathbf{x}\right)$ based on (11) could be unstable.

Once the marginal likelihood $r\left(\mathbf{x}\right)$ is calculated the entropy of the posterior distribution of $\mathbf{Z}$ given data is obtained through

$$H\left(\mathbf{z} \middle| \mathbf{x}\right) = \log r\left(\mathbf{x}\right) - E_{\mathbf{z}|\mathbf{x}}\left(\log r\left(\mathbf{x} \middle| \mathbf{z}\right) p\left(\mathbf{z}\right)\right).$$

As in previous cases, the expected value of a function of $\mathbf{Z}$ with respect to the posterior distribution can be simulation consistently estimated through taking the average of the function in the MCMC output. When using the Metropolis-Hastings algorithm and the probability of accepting a move to $\widetilde{\mathbf{z}}$ when in $\mathbf{z}$ is

$$\exp\left\{\min\left(0, \log r\left(\mathbf{x} \middle| \widetilde{\mathbf{z}}\right) - \log r\left(\mathbf{x} \middle| \mathbf{z}\right)\right)\right\},$$

the needed terms for the expectancy for the entropy are calculated from the M-H sampler steps.

7 Empirical example

When assessing accuracy defined as the tendency of an informant to accurately represent a "true" structure of e.g. interaction, it is generally difficult

to distinguish the actual accuracy from the accuracy that is an artefact of the aggregation method that is used to construct the aggregate structure. When using for example LAS with the intersection rule an actor, if he or she is part of the dyad perceived, can never be wrong when stating that a tie does not exist. Using CS with majority rule as the criteria for the "true" structure we saw above that this in our interpretation corresponds to a model where every informant is equally accurate and this accuracy does not depend on the value in the true structure and further restricted to be between $\frac{1}{2}$ and 1. When looking at matches between respondents we get a measure of who is most accurate in relation to what everyone else thinks but here the link to the actual structure is vague. It is our hope here that by quantifying our prior belief in different parameter values and models the results will be easier to interpret since the relation between data, competencies and the true structure is completely described by the generic model (4) and the only uncertainty there is concerns our uncertainty about the values of the parameters and what exact model is right.

7.1 Model specifications

To illustrate the Bayesian model for cognitive social structures we apply it to Krackhardt's high tech managers (1987). A graph that describes the organisational structure is found in appendix A and the consensus structure (with majority rule) is depicted in Figure 1. To help us formulate different specifications of the model we consulted the literature on different hypotheses regarding accuracy in perceiving social interaction. Here we focus on model specifications where the accuracy, the competencies, and the hidden structure $\mathbf{Z}$ are a priori independent. As a consequence of this we disregard hypotheses such as the (in simplified versions): A perceiving actor has more knowledge of who their own friends and their friends' friends are than their non-friends' friends are (Batchelder et al, 1997:48, Knowledge Hypothesis); Accuracy is positively related to frequency of interaction (Romney and Weller, 1984, Alternative Hypothesis 1 and 2).

Kumbasar, Romney and Batchelder (1994) in an introductory discussion elaborate on theories and results that point to the tendencies of actors to inflate their own role in a social network and overrate their estimates on their own personal, positive traits. This would suggest a model specification
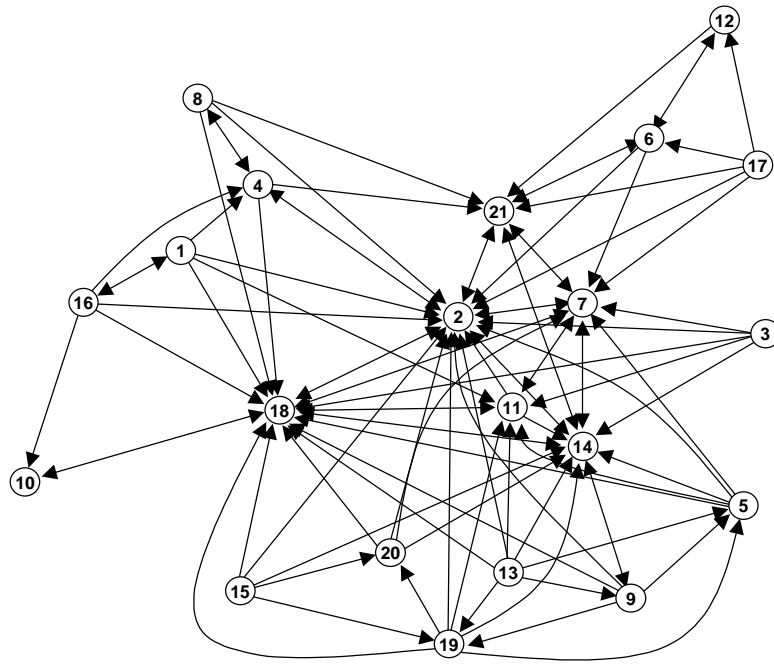
Figure 1: Sociogram of the consensus structure of the advice network of the 21 high-tech managers.

where we partition

$$f(i, j, k) = \begin{cases} \nu_{i1} & if \ i \in \{j, k\} \\ \nu_{i2} & otherwise \end{cases}, \text{ for } M_1$$

and where we think that $\eta_{\nu_{i1}}(0)$ generally is lower than $\eta_{\nu_{i2}}(0)$, possibly even an a prior expected value on the first that was below .5? This model is related to the Bias Hypothesis of Batchelder et al (1997), which predicts that an actor tends to exaggerate his or her own role in the network but again it also predicts a similar mechanism to work for this informants friends relations and this is related to the actual structure.

Another interpretation of the hypothesis above is to distinguish between cases when an actor perceives dyads when he is himself the potential sender from cases when he is the potential receiver. The prestige bias in this case could be a mechanism that only concerns the giving of advice. In other words, the perceiving actor tends to exaggerate the number of people that come to this actor to get advice. This could serve as our second model

$$f(i, j, k) = \begin{cases} \nu_{i1} & if \ i = j \\ \nu_{i2} & if \ i = k \\ \nu_{i3} & if \ i \neq \{j, k\} \end{cases}, \text{ for } M_2.$$

Casciaro (1998) hypothesises that the accuracy of a perceiver with regard to an advice network is positively related to his or her hierarchial level. This hypothesis could serve as the basis for a third and a fourth model

$$f(i, j, k) = \begin{cases} \nu_1 & if \ i \text{ belongs to the top of the hierarchy} \\ \nu_2 & otherwise \end{cases}, \text{ for } M_3,$$

and

$$f(i, j, k) = \nu_{i1}, \text{ for } M_4.$$

In the first case all actors belonging to the same hierarchial level in an organization are assumed to have the same basic competency for judging who exchanges information with whom whereas in the fourth model these competencies are assumed to be individual. In the fourth model the hypothesis is reflected in the choice of priors rather that in the partition of the actor space. A priori we believe individuals in the top of the organisation to be more competent in judging the advice network than individuals lower down in the hierarchy.

In Bondonio (1998, Hypothesis III) the accuracy of a perceiver i on the dyad (j, k) is assumed to be positively related to the closeness of i and j (in **Z**), and how similar i and j are in terms of age and tenure. The first part of this hypothesis is not investigated further here for reasons given above. The second part can be formulated in the current framework. The accuracy in this hypothesis was conceived as the accuracy of i relative to dyads (j, k) for all k, for each k, where the true structure was the locally aggregated structure from the intersection rule. To mimic this situation we can a priori consider a perceiver i to be fairly accurate when reporting on non-existing ties from himself to others. An actor i reporting on (j, k) where $i \neq j$, we assume a priori is more accurate if i and j are similar (with respect to age and tenure) than if they are not similar. For the high-tech manager data we have, there are three fairly distinct groups with respect to age and length of tenure

- A_1: Managers whose tenure exceeds 15 years,

- A_2: Managers whose tenure does not exceed 15 years and whose age is less than 45 years,

- A_3: Managers whose tenure does not exceed 15 years and whose age exceeds 45 years.

We can now formulate models number 5 and 6,

$$
f(i, j, k) = \begin{cases} \nu_i & i = j \\ \nu_a & i \in A_\delta, \ j \in A_{\delta'}, \ A_\delta = A_{\delta'} \\ \nu_b & i \in A_\delta, \ j \in A_{\delta'}, \ A_\delta \neq A_{\delta'} \end{cases} \text{, for } M_5.
$$

Model 5, M_5, forces the competencies to be equal for all actors in a certain category and as in the previous specifications we can modify this by letting

$$
f(i, j, k) = \begin{cases} \nu_{i1} & i = j \\ \nu_{i2} & i \in A_\delta, \ j \in A_{\delta'}, \ A_\delta = A_{\delta'} \\ \nu_{i3} & i \in A_\delta, \ j \in A_{\delta'}, \ A_\delta \neq A_{\delta'} \end{cases} \text{, for } M_6.
$$

It now remains to find suitable conjugate priors corresponding to each model. First of all, when the underlying hypothesis behind a model specification does not offer any guidance as to what our prior belief is for a certain $\eta_\nu(s)$ we can assign to this particular parameter a uniform prior. We can not however, as mentioned before, use uniform priors for all parameters. The following section is only meant to illustrate a process of eliciting prior belief.

7.2 Specification of hyper parameters

Thinking in terms of the first model we know that under this hypothesis we have an idea about how accurate a perceiver is as a function of what dyad is perceived in relation to the perceiver. If there is actually a tie to or from the perceiver we are sure that this perceiver will pick up on this - perhaps not with a probability of one but certainly with a probability between 0.5 and 1. Perhaps we can go as far as to say that a priori we are 95% confident that the perceiver perceiving present ties to and from himself is on average right, i.e. we want 95% of the mass of the prior to be between 0.5 and 1. If we think about it even harder we realise that this does not necessarily correspond to a 95% box over $(.5, 1)$, rather we think that values halfway in-between 0.5 and 1 are more likely than values extremely close to 0.5 and 1. Then, if we think that values between 0.5 and 0.75 are not quite as likely as values between 0.75 and 1 we want a little skewness towards one. Let us decide on an expected value a priori. Let us say that we a prior expect the perceiver (perceiving present ties to and from himself) to accurately report present ties 3 times out of 4. With $\alpha_\nu(s) = 6$ and $\beta_\nu(s) = 2$ the expected value of $\eta_\nu(s)$ is 3/4 and the probability that it lies in the interval $(.4727, .9867)$ is .95 (this is the 95% interval which a priori has the shortest length given these hyper parameters). With no additional information on the individuals we let $\eta_{\nu_{i1}}(1) \sim Beta(6, 2)$ a priori for all i for the first model.

According to M_1 an actor tends to exaggerate his or hers own importance. This translates into assuming that the actor reports that people seek his or hers advice even if they in fact don't. Using a similar line of reasoning as when choosing the prior for $\eta_{\nu_{i1}}(1)$ we could arrive at the prior $Beta(3, 5)$. This means that roughly once out of three the actor correctly reports that someone does not come to him or her to seek advice and that we are 95% sure that $\eta_{\nu_{i1}}(0)$ is between .081 and .685 and in which case 77% of the probability mass is to the left of .5.

For the perception of others relation we perhaps don't want it to be completely haphazard nor do we think that it is extremely high or extremely low. If we keep to the idea that this should be higher that for the perception of own (non-existent) ties we can choose a prior that has most of its mass to the right of 0.5. Not wanting to force ourselves to give specific priors for these cases we can always choose uniform priors which means that we do not favor any one interval of a given length. A uniform prior for both $\eta_{\nu_{i2}}(0)$ and $\eta_{\nu_{i2}}(1)$ is still in accordance with the motivation behind the model since this

means that the actor when perceiving others' ties is still on average more accurate than when an involved actor perceives non existing ties. Perhaps one could question why an actor shouldn't be accurate when judging absent ties as one would wonder why someone would "add" ties that "are not there". Missing a present tie between "distant" actors, on the other hand, is simply a case of non-detection. To capture this distinction we could always choose a fairly vague prior such as $Beta\,(3,2)$. This prior penalizes values at the extremes, i.e. values very close to 0 and 1 are deemed unlikely but not equally so. The center of mass is slightly off-balance with slightly more mass to the right of .5.

For M_2 we can use the same set of priors for ν_{i2} as we had for ν_{i1} in M_1 as we think that giving advice is considered more prestigious than receiving advice and hence the former is more affected by bias. For want of better we choose a uniform prior for ν_{i1} in M_2. We keep the same set of hyper parameters for ν_{i3} as for ν_{i2} in M_1.

In M_3 we want the priors to reflect the rather strong assumption of the underlying hypothesis. Using the rationale behind the prior for $\eta_{\nu_{i1}}(1)$ in M_1 we let $\eta_{\nu_1}(1)$ and $\eta_{\nu_1}(0)$ both be $Beta\,(6,2)$ for M_3. For actors in the lower part of the hierarchy judging the absence or presence of ties we have only specified that their competency should be lower but not by how much. Although we could choose to penalize extreme values we think it better to leave their priors uniform except for $\eta_{\nu_2}(0)$. Again we think it unlikely that actors should report ties where there are no ties unless it is motivated by for example prestige bias and again we opt for the $Beta\,(3,2)$ prior. For M_4 we choose basically the same priors as for M_3 only that we here allow actors to have competencies particular to themselves.

For the three types of regions in models 5 and 6 we choose a $Beta\,(6,2)$ for $\eta_{\nu_i}(1)$, $\eta_{\nu_i}(0)$, $\eta_{\nu_{i1}}(1)$ and $\eta_{\nu_{i1}}(0)$; $Beta\,(5,3)$ for $\eta_{\nu_a}(1)$, $\eta_{\nu_a}(0)$, $\eta_{\nu_{i2}}(1)$ and $\eta_{\nu_{i2}}(0)$; and uniform for the rest.

With no prior information regarding the actual relational structure we use a prior on $\mathbf{Z}$ that does not favor any particular graph. If we the prior probability be constant we achieve this though the amount of information can be discussed (c.p. the distribution of a function of $\mathbf{Z}$ according to this prior as seen in e.g. Figure B.2.)

25

7.3 Results

The results of the analysis can be found in Tables 1 and 2, and in the graphs in the appendices B-G[1]. As this is only meant to illustrate the procedures, details are left out and we leave to the reader to further interpret the results provided. The numerical specifics will be provided by the author upon request.

Without going into too much detail we give a brief interpretation of the posteriors for model 1 in Figure B.1. Actors 3-5, 9, 10, 18 and 21 seem to have competencies in accordance with what was hypothesised. They are accurate when perceiving existing ties to and from themselves but have a tendency to report ties to and from themselves when there are no such ties (even though these, latter, parameters are not decidedly separated from .5 - especially receivers 4, 5 and 9). When perceiving dyads in which they are not themselves involved, they are approximately 70 to 80% accurate. Perceivers 1, 15 and 20 also follow a similar pattern but $\eta_{12}(0)$ is low as are $\eta_{15,2}(1)$, $\eta_{20,2}(1)$. $\eta_{i,1}(0)$ is a lot higher than expected under the hypothesis for all other perceivers i. Note that the uncertainty about the parameters is generally greater for $\eta_{i,1}(s)$ than for $\eta_{i,2}(s)$. This is a natural consequence of the fact that we have more observations corresponding to the latter perceiver-actor category, i.e. perception of others.

Point estimates of the probabilities in M_2 are given in Table C.1. Worth mentioning is that when perceiving others' interaction everyone, but actor one, it is very unlikely that they falsely report a tie to exist $(\eta_{\nu_{i3}}(0))$. When it comes to detecting ties *to* themselves $(\eta_{\nu_{i2}}(1))$ they are also on the whole accurate although the uncertainty about these estimates are high for certain individuals. In addition these seems to be a positive association between $\eta_{\nu_{i1}}(1)$ and $\eta_{\nu_{i3}}(1)$, and $\eta_{\nu_{i1}}(1)$ seems to be negatively associated to $\eta_{\nu_{i1}}(0)$.

With the above priors there is, according to Table 1, overwhelming evidence for model 2. For example $BF_{21} = 1.3 \times 10^{10}$, and $p(M_2|\mathbf{x}) \approx 1$. The model which data gives the least support is model 3. (The rations between marginal likelihoods and the Bayes factors are equivalent with uniform priors on models.) None of the specified models give a "significant" reduction of our uncertainty about $\mathbf{Z}$, judging by the posterior entropies. The maximum entropy over $\mathbf{Z}$ is given by the uniform distribution over all $(2^{21 \times 20}$ di-$)$ graphs

---

[1]For each figure, three chains were run where every eleventh state was taken as a sample point. Of the approximately 7000 sample points from each chain, the first 500 were discarded as burn in.

and is equal to 291 and the posterior entropies lie in the range 249 to 260 [2]. This is not to say that there is no structure in data. It is informative to study Figures B.2., C.2., D.2., E.2., F.3., and G.2. These figures give the prior and posterior distributions of $\mathbf{Z}$ as represented by their respective degree distributions. According to these projections of the prior and posterior distributions of $\mathbf{Z}$, the priors and posteriors are by no means identical.

Table 1.
Marginal likelihoods and posterior entropies of $\mathbf{z}$

	M_1	M_2	M_3	M_4	M_5	M_6
$\log\left(r_i\left(\mathbf{x}\right)\right)$	-3952	-3928	-4506	-4021	-4453	-3999
$H_i\left(\mathbf{z}\mid\mathbf{x}\right)$	254.8	252.7	257.5	259.6	258.8	249.9

From Table 2., we see that the modal graphs in the posterior distributions given different models are all fairly close to each other and that the first two models have identical modes. As expected the two ad-hoc methods CS and LAS from intersection rule agree most with the first two models. Perhaps it is not surprising that CS lies closer to the model posteriors than the two types of LAS since the first is a function of "all" data (c.p. Eq. 7) whereas the latter two only use certain observations. The sociogram of the modal graph according to M_2, which could serve as our point prediction of $\mathbf{Z}$, is given in Figure 2.

---

[2]Figures in the table are based on MCMC samples. For each model 3 (M_6 only one chain each and for M_1 two) times 500000 iterations were performed, each time started in their respective modal graphs. Every 7th value was taken as a sample point and the first 500 sample points for each of the three runs were discarded as burn-in values.
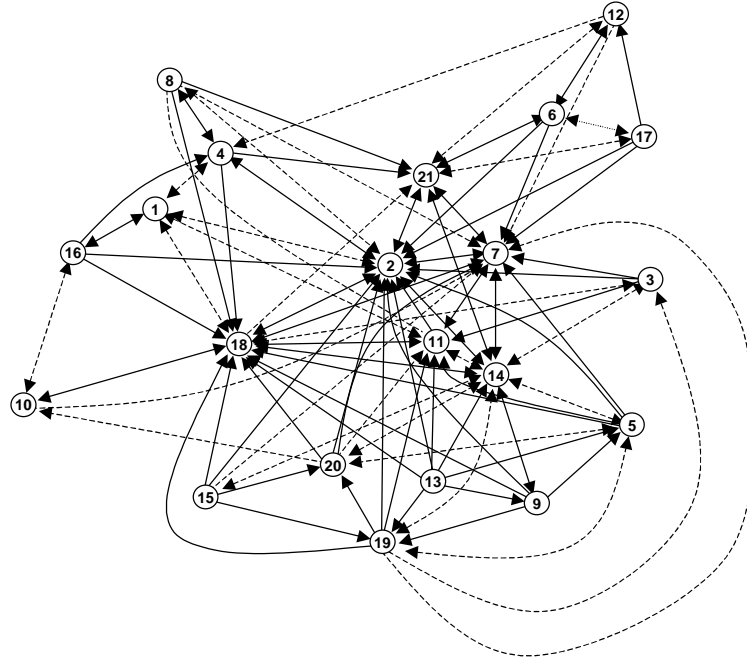
Figure 2: Sociogram of the modal graph according to M_2. Dashed lines indicate added arcs as compared to the consensus structure (no arcs were removed compared to the consensus structure).

Table 2.
Symmetric class difference between arc sets of modal
graphs in the posteriors for different models and
aggregation methods[1]

	M_1	M_2	M_3	M_4	M_5	M_6	CS	$LASI$
M_1	0							
M_2	0	0						
M_3	23	23	0					
M_4	23	23	4	0				
M_5	33	33	10	12	0			
M_6	25	25	10	10	16	0		
CS	30	30	53	53	63	55	0	
$LASI$	82	82	85	83	91	83	76	0
$LASU$	165	165	146	144	140	144	183	147

[1]Maximum difference is 420

8   Conclusions and future directions

When applying network theories to real life problems such as various sampling schemes it is essential that we know something about what mechanisms are involved when responses are generated. We could take reports about the network at face value and hope that various forms of bias cancel each other out to produce an overall bias of nought, or we could try to control for sources of bias to obtain more stable estimates. This procedure could e.g. be relevant in adaptive sampling. When it comes to prediction it is obvious that a model where covariates are more closely tied to competencies would be more natural - a logit or probit model could for example be fitted to data (c.p. Koskinen, 2002).

It would be desirable to have more in-built dependencies between reports as well as dependencies between the actual structure and the competencies. The need for the latter became clear when reference was made to prior hypothesis regarding CSS.

Although the field of study, i.e. CSS, is interesting in its own right, the data collecting paradigm might seem to demand rather a lot of the research subjects.

We call for further exploration of the influence of priors. Ideally, one ought to examine the frequency of "right" choices in model selection using data generated from a known model. This would also be beneficial to the assessment of sensitivity of model selection to different specifications of the hyper parameters. That there is no simple way to automate the setting of hyper parameters is a definite drawback, especially so since the posterior model probability can become a little hard to interpret.

In this paper $\mathbf{Z}$ and $\mathbf{X}_i$ has been binary adjacency matrices but as mentioned earlier these matrices could be more general structures. If the elements of $\mathbf{Z}$ take values in $\mathcal{F}$ and the elements of $\mathbf{X}_i$ take values in $\mathcal{X}$ and we define the indicators in (3) appropriately, the model is still essentially the same and has the same conjugate priors. For, say, digraphs we could ascribe an arbitrary label, q_{jk}, $j < k$, to each $\{j, k\} \in \binom{V}{2}$, and let $\mathcal{F} = \{(0,0), (1,0), (0,1), (1,1)\}$, so that for example $Z_{q_{jk}} = (1,0)$, if j is adjacent to k but k is not adjacent to j. Using this formulation of the model one would focus more on the dyads than if one merely considered each arc separately. Doing this one would have to put up with the fact the probability of $X_{iq} = x_{iq}$ given $Z_q = (0,0)$, and $\eta_{iq}(0,0)$ would be equal for all $x_{iq} \in \mathcal{F} \backslash \{(0,0)\}$.

One obvious extension of the model is to apply it to multiple relational networks (see e.g. Fienberg, Meyer, and Wasserman, 1985). For valued relations R_i, $i \in \mathcal{I}$, we could have reports $\mathbf{X}_i = (X_{iq})$, with X_{iq} taking values in $\mathcal{X}_i$. As an example we could have the relations esteem, disesteem, liking, disliking, positive influence, negative influence, praise and blame, as in the Sampson Monastery data (Sampson, 1969, Breiger, Boorman and Arabie, 1975). Our assumption could then be that there is an underlying structure $\mathbf{Z}$ which all these relations reflect in a stochastic way.

The mixing of the Metropolis-Hastings algorithm is a slight concern. It turns out that the algorithm is fairly sensitive to the choice of initial distribution. A further concern is the dependence between states in the MCMC sample. Although the remedy could be using a different proposal distribution the author has not found a proposal distribution which does not seriously slow down the mixing.

The implementation of the Bayesian analysis of cognitive social structures is quite computer intensive. In the absence of exact analytical expressions approximate solutions to the estimation procedure would be worth striving for.

References

[1] Banks, D., Constantine, G.M. (1998). Metric models for random graphs. *Journal of Classification*, 11:121-149.

[2] Batchelder, W.H., Romney, A.K. (1988). Test theory without an answer key. *Psychometrica*, 53:193-224.

[3] Batchelder, W.H., Kumbasar, E. and Boyd, J.P. (1997). Consensus analysis of three-way social network data. *Journal of Mathematical Sociology*, 22:29-58.

[4] Bernard, H.R., Killworth, P.D. and Sailer, L. (1980). Information accuracy in social network data IV: A comparison of clique-level structure in behavioral and cognitive network data. *Social Networks*, 2:191-218.

[5] Bock, H.H. (1996). Probability models and hypotheses testing in partitioning cluster analysis, pp. 377-453 in Arabie, P., Hubert, L.J. and De Soete, G. (eds.). *Clustering and Classification*. River Edge, NJ: World Scientific Publishing.

[6] Bondonio, D. (1998). Predictors of accuracy in perceiving informal social networks. *Social Networks*, 20:301-330.

[7] Borgatti, S.P., Everett M.G., and Freeman, L.C. (1999). *Ucinet 5 for Windows: Software for Social Network Analysis*. Natick: Analytic Technologies.

[8] Breiger R., Boorman S. and Arabie P. (1975). An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling. *Journal of Mathematical Psychology*, 12:328-383.

[9] Burt, R.S. (1992). *Structural Holes: The social structure of competition*. Cambridge: Harvard University Press.

[10] Casciaro, T. (1998). Seeing things clearly: social structure, personality, and accuracy in social netwrok perception. *Social Networks*, 20:331-351.

[11] Coleman, J., Katz, E., and Mentzel, H. (1957). The diffusion of an innovation among physicians. *Sociometry*, 20:253-270.

[12] Doreian P (1974). On the connectivity of Social Networks. *Journal of Mathematical Sociology*, 3:245-258.

[13] Fienberg, S. S., Meyer, M. M., and Wasserman, S. (1985). Statistical analysis of multiple sociometric relations. *Journal of the American Statistical Association.* 80, 51-67

[14] Fishman, G.S. (1999). *Monte Carlo: Concepts, Algorithms, and Applications.* New York: Springer-Verlag.

[15] Frank, O. (1978). Inferences concerning cluster structure, pp. 259-265 in Corsten, L.C.A. Hermans, J. (eds.). *Proceedings in Computational Statistics.* Wien: Physica Verlag.

[16] Frank, O. (2001).Using centrality modeling in network surveys. (Forthcoming in special issue of Social Networks).

[17] Frank, O. and Nowicki, K. (1993). Exploratory statistical analysis of networks. *Annals of Discrete Mathematics*, 55:349-366.

[18] Freeman L C (1979). Centrality in Social Networks: Conceptual clarification. *Social Networks* 1:215-239.

[19] Gelfand, A.E., and Smith, A. F.M. (1990). Sampling-based approaches to calculating marginal densities. *Journal Of The American Statistical Association*, 85:398-409.

[20] Johnson, J. C. (1994). Anthropological contributions to the study of social networks: a review, pp. 113-151 in Wasserman, S., Galaskiewicz, J. (eds.). Advances in Social Network Analysis: Research in the Social and Behavioral Sciences. Thousand Oaks: Sage cop.

[21] Kass, R. E., Raftery, A. E (1995). Bayes factors. Journal of the American Statistical Association, 90:773-795.

[22] Katz L (1953). A new status index derived from sociometric data analysis. *Psychometrika*, 18:34-43.

[23] Killworth, P.D., and Bernard, H.R., (1979). Information accuracy in social network data III, or: A comparison of triadic structure in behavioural and cognitive data. *Social Networks*, 2:19-46.

[24] Koskinen, J. H. (2002). *Bayesian Analysis of Cognitive Social Structures with Covariates*. Research Report, No. 3, Stockholm University, Department of Statistics.

[25] Krackhardt, D. (1987). Cognitive social structures. *Social Networks*, 9:109-134.

[26] Kumbasar, E, Romney, A.K., and Batchelder, W.H. (1994). Systematic biases in social perception. *American Journal of Sociology. 100:477-505.*

[27] Laarhoven, P.J.M. van, Aarts, E.H.L. (1987). *Simulated Annealing: Theory and Applications*. Dordrecht: Reidel.

[28] Linden, Wim J. van der, Hambleton, Ronald K. (eds.) (1997). *Handbook of Modern Item Response Theory*. New York: Springer-Verlag

[29] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953). Equations of state calculations by fast computing machine. *J. Chem. Phys.* 21:1087-1091.

[30] Moreno, J.L. (1934).*Who Shall Survive? Foundations of Sociometry, Group Psychotherapy and Sociodrama*. Washington, D.D.: Nervous and Mental Disease Publishing Co. Reprinted in 1978 (Third Edition) by Becon House, inc., Beacon, NY.

[31] Newcomb, T.M. (1961). *The Acquaintance Process*. New York: Holt, Rinehart and Winston.

[32] Nowicki, K., and Snijders, T.A.B. (2001). Estimation and prediction for stochastic blockstructures. (to appear in JASA).

[33] Pattison, P. (1994). Social cognition in context - some applications of social network analysis, pp. 79-109 in Wasserman, S., Galaskiewicz, J. (eds.). *Advances in Social Network Analysis: Research in the Social and Behavioral Sciences*. Thousand Oaks: Sage cop.

[34] Raftery, A. E. (1996). Hypothesis testing and model selection, pp.163-187, in Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (eds.). *Markov Chain Monte Carlo in Practice*. London : Chapman & Hall.

[35] Romney, A.K., Weller, S.C. (1984). Predicting informant accuracy from patterns of recall among individuals. *Social Networks*, 6:59-77.

[36] Romney, A.K., Weller, S.C. and Batchelder, W.H. (1986). Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist*, 88:313-338.

[37] Romney, A.K., Batchelder, W.H. and Weller, S.C. (1987). Recent applications of cultural consensus theory. *American Behavioral Scientist*, 31:163-177.

[38] Sampson, S. (1969). *Crisis in a Cloister*. Unpublished doctoral dissertation, Cornell University.

[39] Snijders, T.A.B., and Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification*, 14:75-100.

[40] Tallberg, C. (2002). *Modeling Stochastic Blockstructures with Covariates*. Research Report, Department of Statistics, Stockholm University.

[41] Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. New York and Cambridge: Cambridge University Press.

[42] Weller, S.C. (1984). Consistency and consensus among informants: disease concept in rural Mexican town. *American Anthropologist*, 86:966-975.

[43] Weller, S.C. (1987). Shared knowledge, intracultural variation and knowledge aggregation. *American Behavioral Scientist*, 31:178-193.

[44] Wong, G.Y. (1987). Bayesian models for directed graphs. *Journal of the American Statistical Association*, 82: 140-148.

A   The High-tech Manager Team

Figure A. 1. is an illustration of the organisational structure of Krackhardt's (1987) high-tech manager team. The 21 nodes are the members of the manager team and the edges represent departmental affiliation between hierarchical levels. The chief executive officer (CEO) has an edge to the vice president (VP) of each department. The VP is adjacent to each manager of his department. The nodes are positioned according to the age and length of tenure

of the individuals represented by the nodes. The axis of tenure has been perturbed to make the picture more visibly tractable.
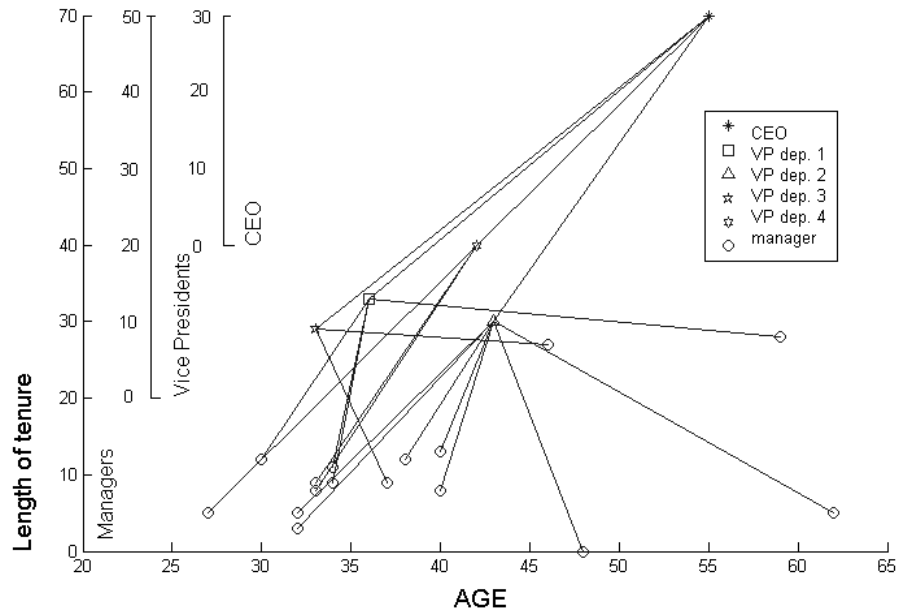
Figure A. 1. The organisational structure of Krackhardt's (1987) high-tech manager team.
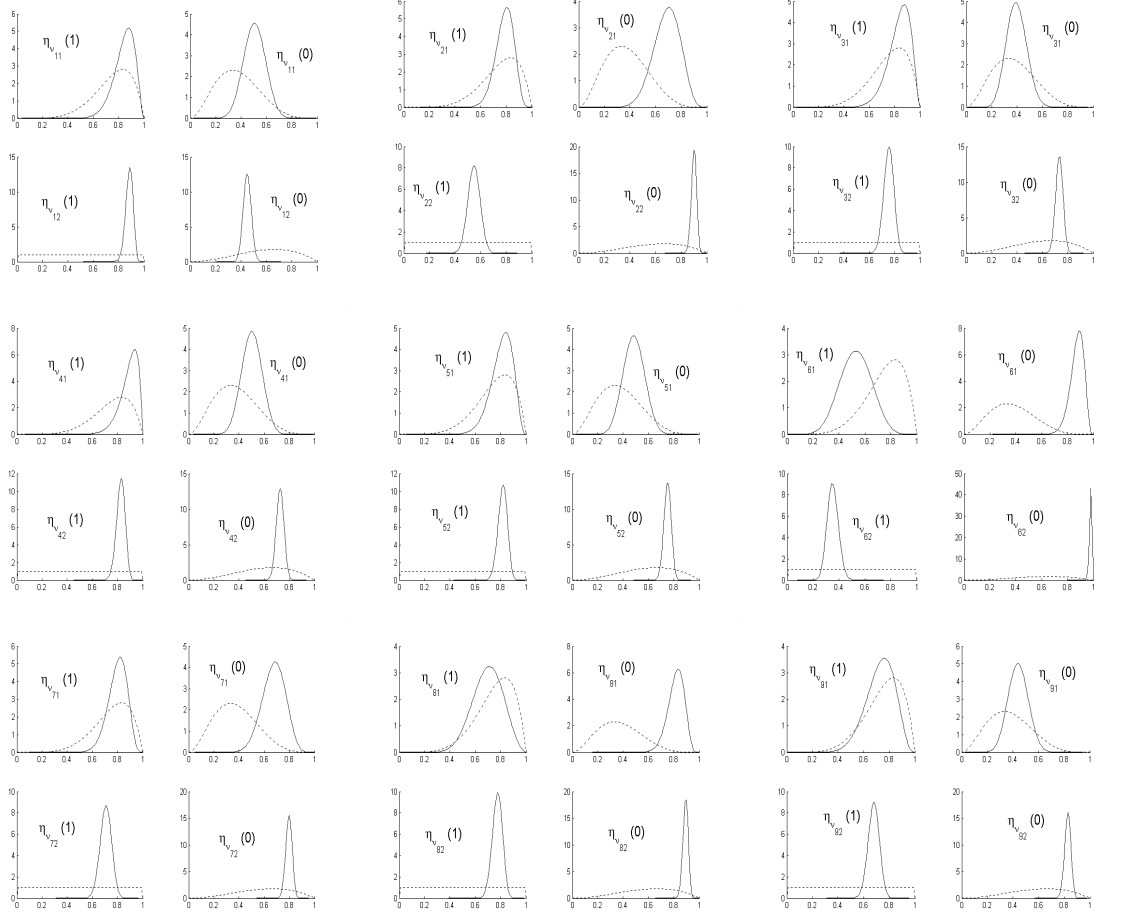
B   Results from M_1

Figure B.1. Priors (dashed lines) and posteriors (solid lines) of the competencies conditional on M_1. For each group of four panels, for perceivers i 1 through 9, the top two panels represent the analysis of $\eta_{\nu_{i1}}(1)$ and $\eta_{\nu_{i1}}(0)$ repectively. The bottom two correspond to $\eta_{\nu_{i2}}(1)$ and $\eta_{\nu_{i2}}(0)$.
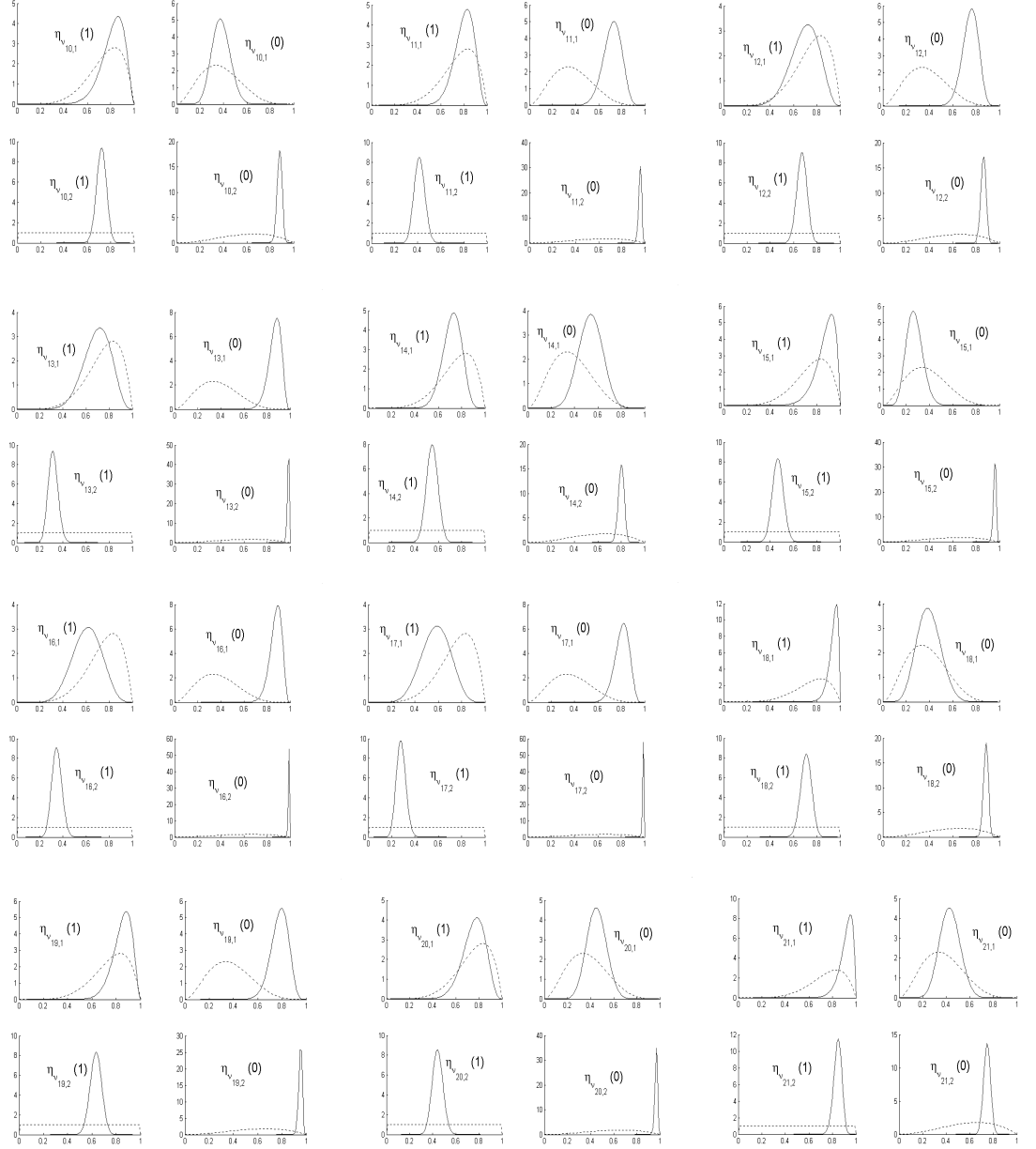
Figure B.1. continued for perceivers i 10 through 21.
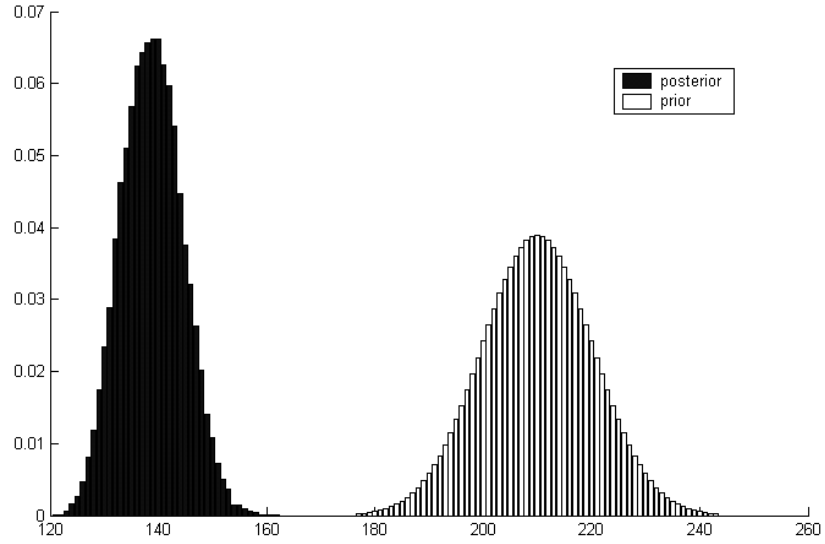
Figure B.2. The prior and posterior distribution of the total number of arcs in $\mathbf{Z}$ when the prior on $\mathbf{Z}$ is uniform over all labeled digraphs with 21 vertices. The posterior of the total number of arcs in $\mathbf{Z}$ is derived from the posterior distribution of $\mathbf{Z}$ conditional on model 1.

C Results from M_2

Table C. 1. Estimates for competencies according to model 2.
The point estimates are posterior expected values (std)

| i | $\eta_\nu(1)$ | | | | | | $\eta_\nu(0)$ | | | | | |
	ν_{i1}		ν_{i2}		ν_{i3}		ν_{i1}		ν_{i2}		ν_{i3}	
1	.71	(.158)	.87	(.088)	.88	(.028)	.86	(.087)	.24	(.093)	.45	(.031)
2	.22	(.145)	.93	(.051)	.54	(.046)	.89	(.085)	.33	(.156)	.90	(.018)
3	.87	(.116)	.75	(.123)	.75	(.027)	.38	(.122)	.42	(.102)	.74	(.028)
4	.86	(.130)	.85	(.100)	.82	(.033)	.53	(.122)	.48	(.104)	.73	(.028)
5	.89	(.101)	.70	(.125)	.81	(.034)	.41	(.129)	.53	(.105)	.76	(.027)
6	.27	(.168)	.65	(.134)	.34	(.041)	.94	(.058)	.79	(.084)	.98	(.009)
7	.86	(.126)	.75	(.090)	.70	(.041)	.79	(.100)	.49	(.142)	.80	(.025)
8	.86	(.118)	.51	(.134)	.76	(.037)	.83	(.095)	.77	(.089)	.89	(.020)
9	.55	(.177)	.80	(.127)	.67	(.040)	.34	(.118)	.50	(.098)	.83	(.024)
10	.69	(.183)	.83	(.108)	.72	(.039)	.35	(.114)	.42	(.101)	.89	(.020)
11	.54	(.183)	.86	(.080)	.41	(.043)	.94	(.059)	.48	(.120)	.96	(.012)
12	.41	(.184)	.82	(.116)	.66	(.041)	.94	(.059)	.60	(.098)	.86	(.022)
13	.63	(.147)	.72	(.155)	.30	(.040)	.91	(.080)	.82	(.073)	.98	(.009)
14	.38	(.134)	.90	(.065)	.53	(.046)	.91	(.089)	.26	(.113)	.80	(.025)
15	.88	(.111)	.79	(.134)	.45	(.043)	.06	(.063)	.41	(.096)	.96	(.012)
16	.65	(.171)	.53	(.148)	.34	(.041)	.94	(.060)	.80	(.081)	.99	(.006)
17	.52	(.180)	.61	(.152)	.27	(.038)	.82	(.096)	.78	(.082)	.99	(.007)
18	.91	(.084)	.92	(.057)	.70	(.042)	.32	(.133)	.50	(.144)	.89	(.019)
19	.90	(.092)	.73	(.120)	.62	(.042)	.74	(.118)	.78	(.088)	.95	(.013)
20	.57	(.152)	.84	(.102)	.43	(.043)	.40	(.134)	.48	(.104)	.97	(.011)
21	.87	(.117)	.89	(.076)	.85	(.032)	.63	(.121)	.27	(.103)	.76	(.027)
prior	.5	(.289)	.75	(.144)	.5	(.289)	.5	(.289)	.38	(.161)	.6	(.200)

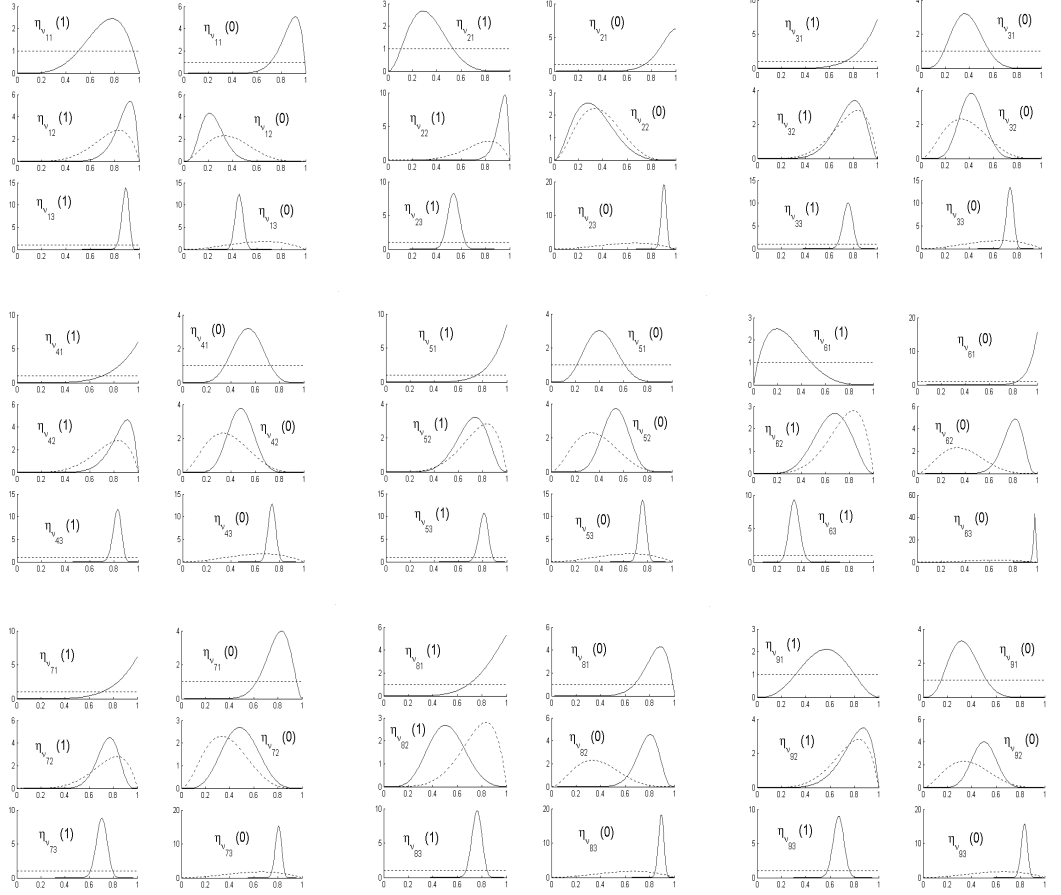Figure C.1. Priors (dashed lines) and posteriors (solid lines) of the competencies conditional on M_2. For each group of six panels,for perceivers i 1 through 9, the top two panels represent the analysis of $\eta_{\nu_{i1}}(1)$ and $\eta_{\nu_{i1}}(0)$ repectively. The intermediate two panels correspond to $\eta_{\nu_{i2}}(1)$ and $\eta_{\nu_{i2}}(0)$, and the bottom two correspond to $\eta_{\nu_{i3}}(1)$ and $\eta_{\nu_{i3}}(0)$.
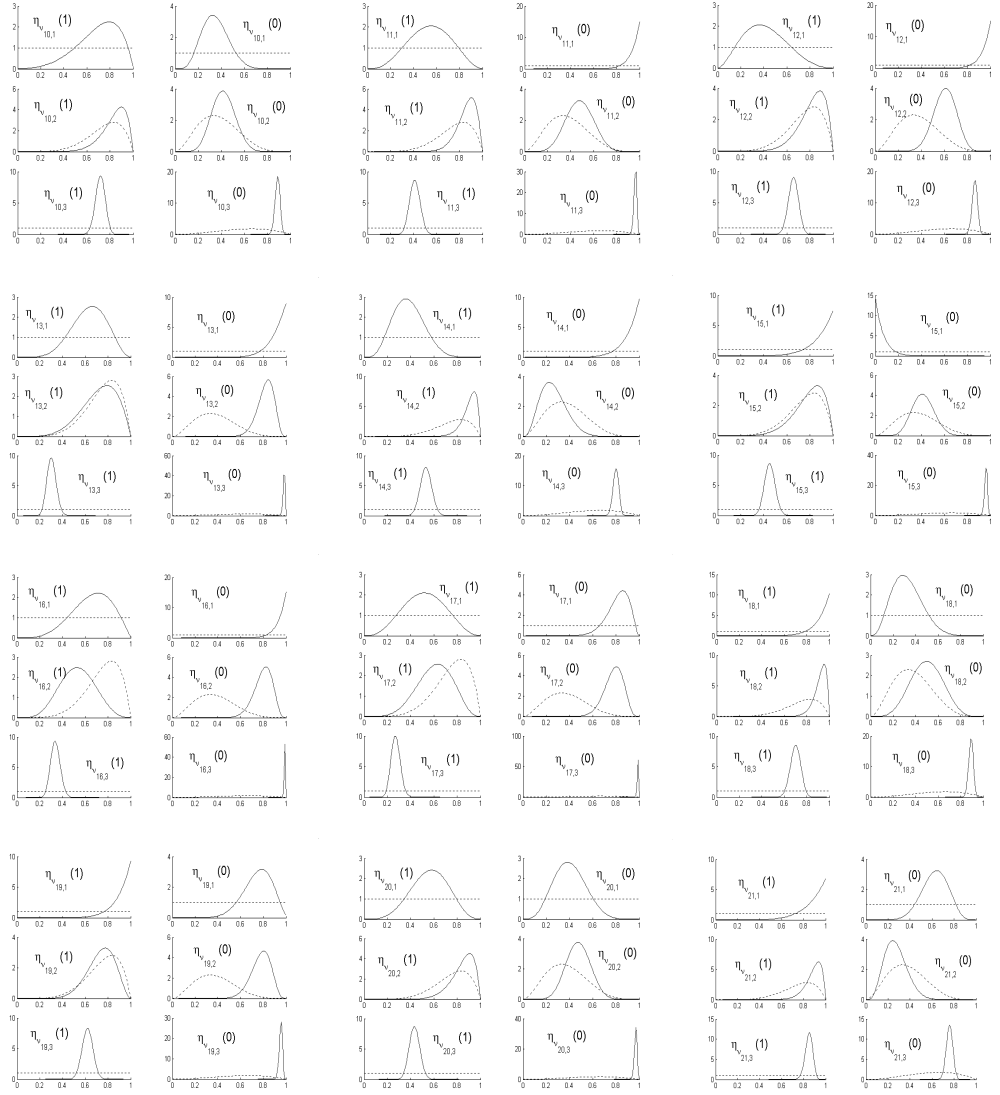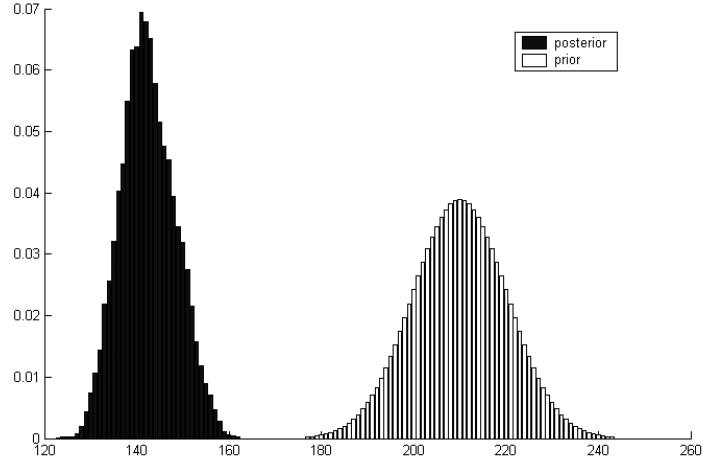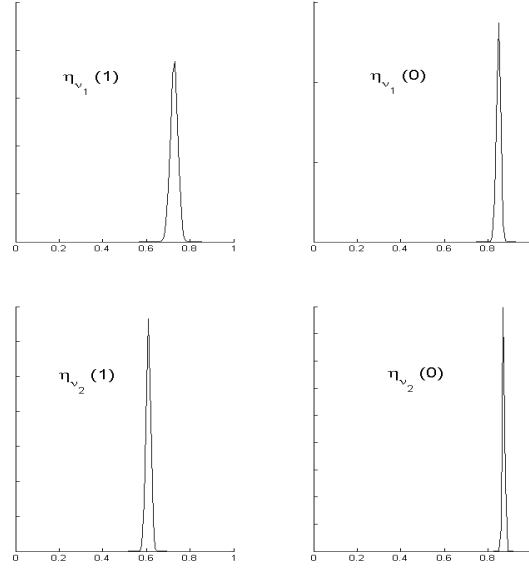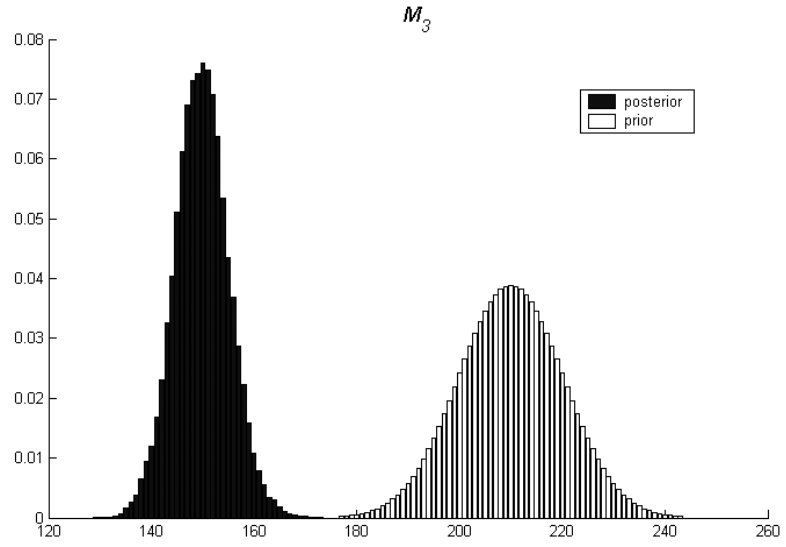
Figure C.1. continued for perceivers i 10 through 21.

Figure C.2. Prior and posterior distribution of the total number of arcs in **Z** when the prior on **Z** is uniform over all labeled digraphs with 21 vertices. The posterior of the total number of arcs in **Z** is derived from the posterior distribution of **Z** conditional on model 2.

D Results from M_3

Figure D.1. Posteriors of the competencies conditional on M_3. The left
upper panel corresponds to the probability of "correctly" reporting
a present tie for a perceiver that belongs to the top of the organisational
hierarchy. The uppermost right panel corresponds to the probability of
"correctly" reporting the non-existence of an arc for a perceiver that belongs
to the top of the organisational hierarchy. The bottom panels are the
corresponding probabilities for a perceiver that belongs to the lower
half of the organisational hierarchy.

Figure D.2. Prior and posterior distribution of the total number of arcs in **Z** when the prior on **Z** is uniform over all labeled digraphs with 21 vertices. The posterior of the total number of arcs in **Z** is derived from the posterior distribution of **Z** conditional on model 3.
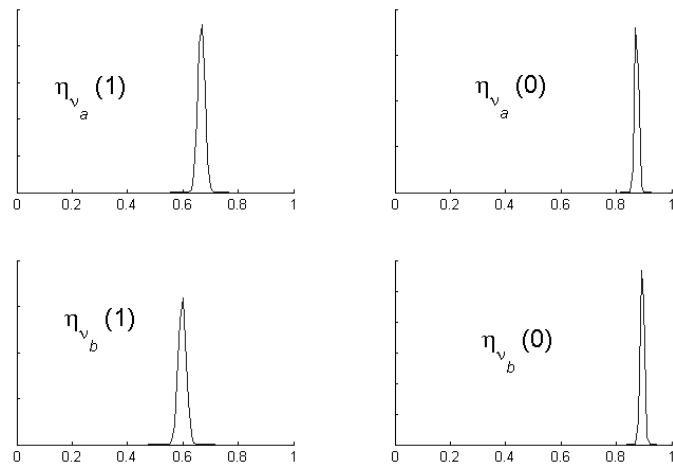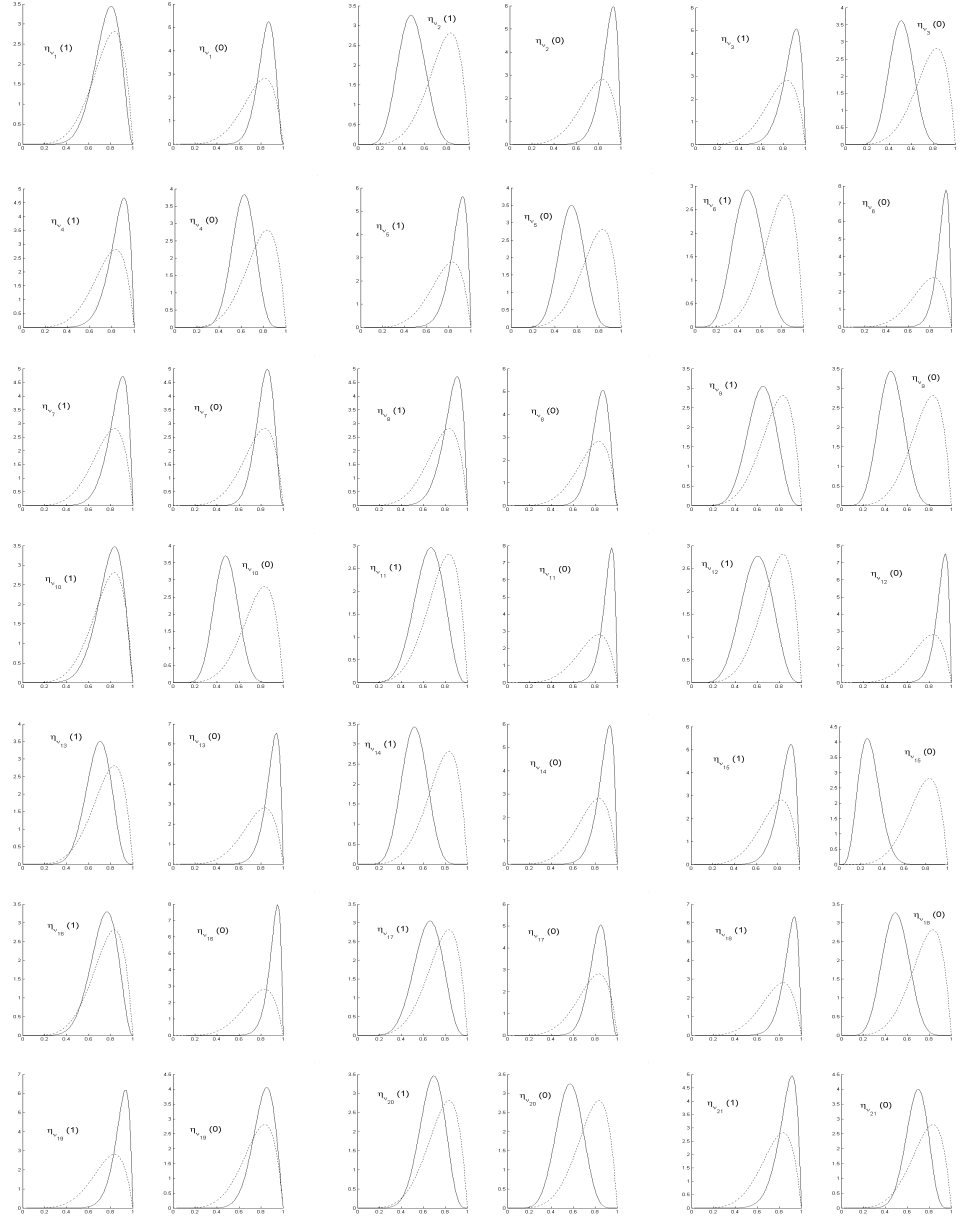
E Results from M_4
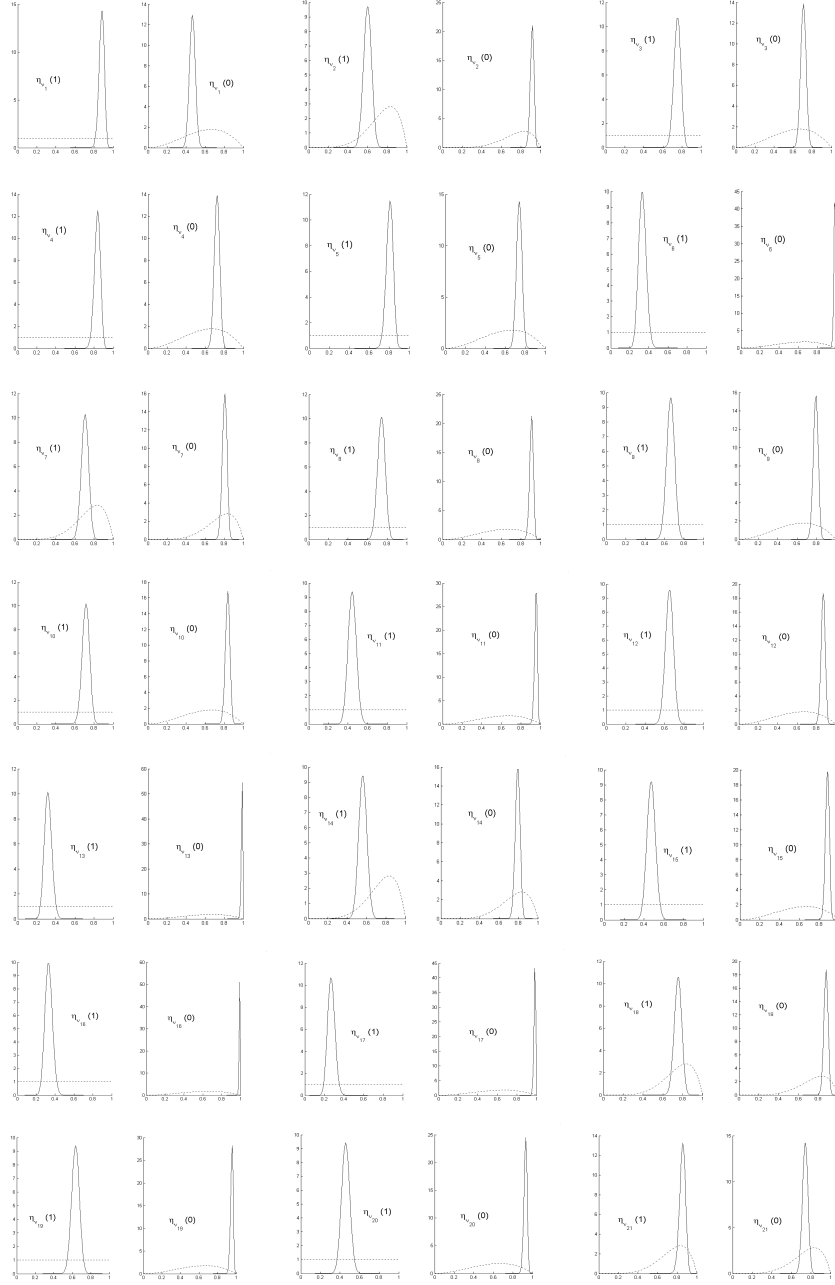
Figure E.1. Prior (dashed lines) and posterior (solid lines) distributions of $\eta_{\nu_i}(1)$ and $\eta_{\nu_i}(0)$ for perceivers i 1 through 21.
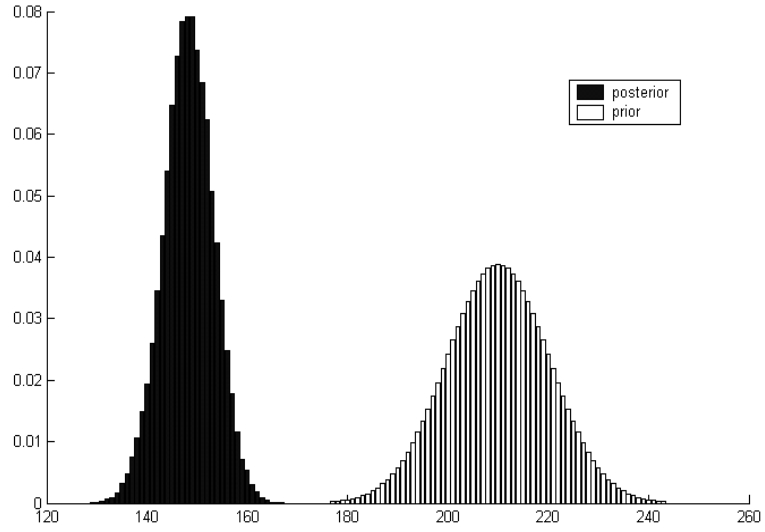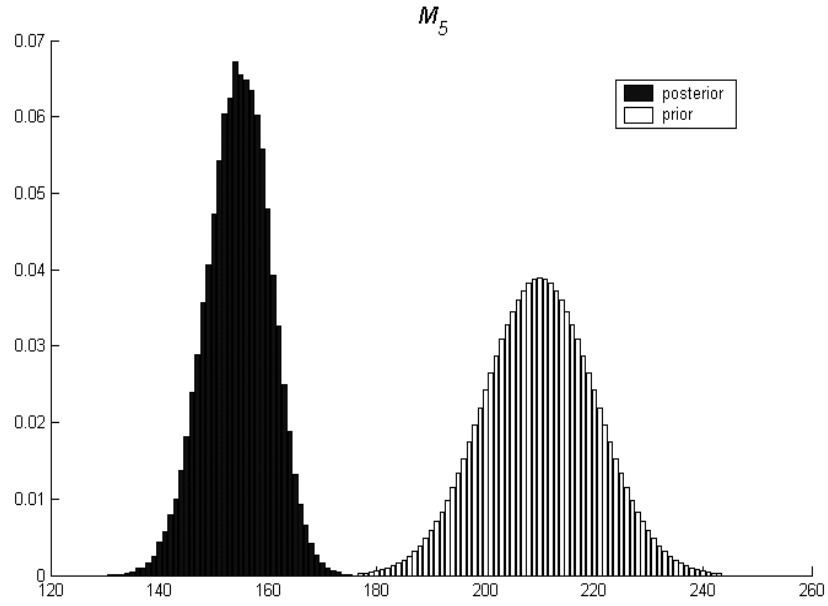
Figure E.2. Prior and posterior distribution of the total number of arcs in **Z** when the prior on **Z** is uniform over all labeled digraphs with 21 vertices. The posterior of the total number of arcs in **Z** is derived from the posterior distribution of **Z** conditional on model 4.

F   Results from M_5

Figure F.1. Posterior distributions of the competencies for categories ν_a and ν_b respectively conditional on model 5.

Figure F.2. Prior (dashed lines) and posterior (solid lines) distributions of the competencies, $\eta_{\nu_{i1}}(1)$ and $\eta_{\nu_{i1}}(0)$, for perceivers i 1 through 21 conditional on model 5.

Figure F.3. The prior and posterior distribution of the total number of arcs. in **Z** when the prior on **Z** is uniform over all labeled digraphs with 21 vertices. The posterior of the total number of arcs in **Z** is derived from the posterior distribution of **Z** conditional on model 5.

G   Results from M_6
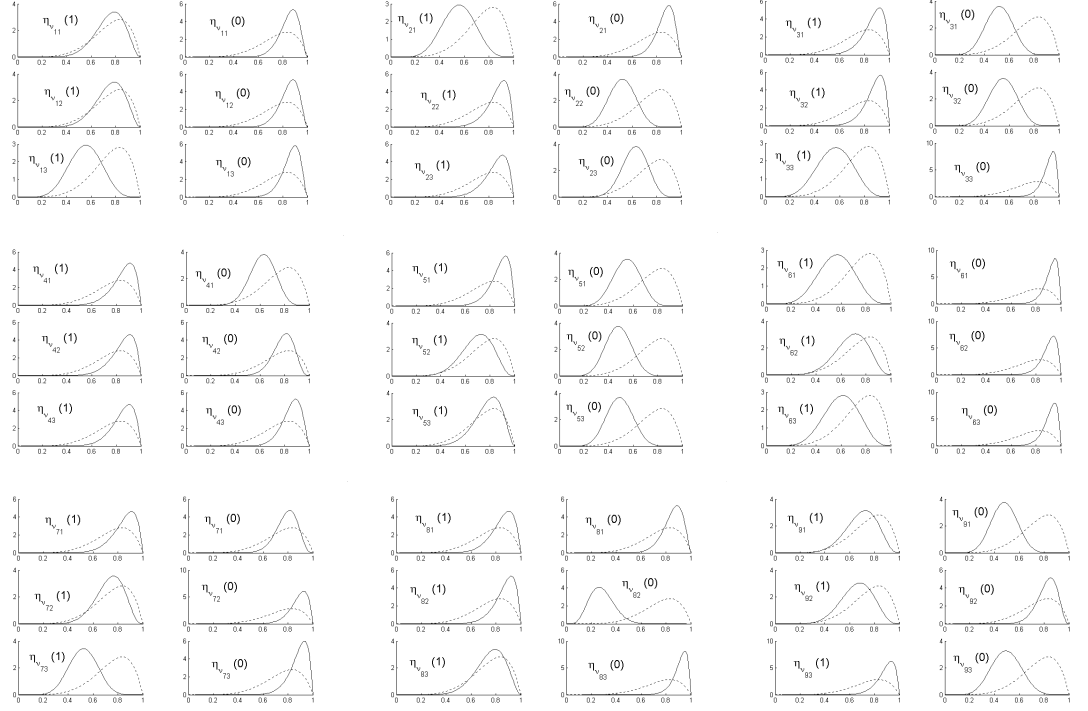
Figure G.1. Priors (dashed lines) and posteriors (solid lines) of the competencies conditional on M_6. For each group of six panels, for perceivers i 1 through 9, the top two panels represents the analysis of $\eta_{\nu_{i1}}(1)$ and $\eta_{\nu_{i1}}(0)$ respectively. The intermediate two panels correspond to $\eta_{\nu_{i2}}(1)$ and $\eta_{\nu_{i2}}(0)$, and the bottom two correspond to $\eta_{\nu_{i3}}(1)$ and $\eta_{\nu_{i3}}(0)$.
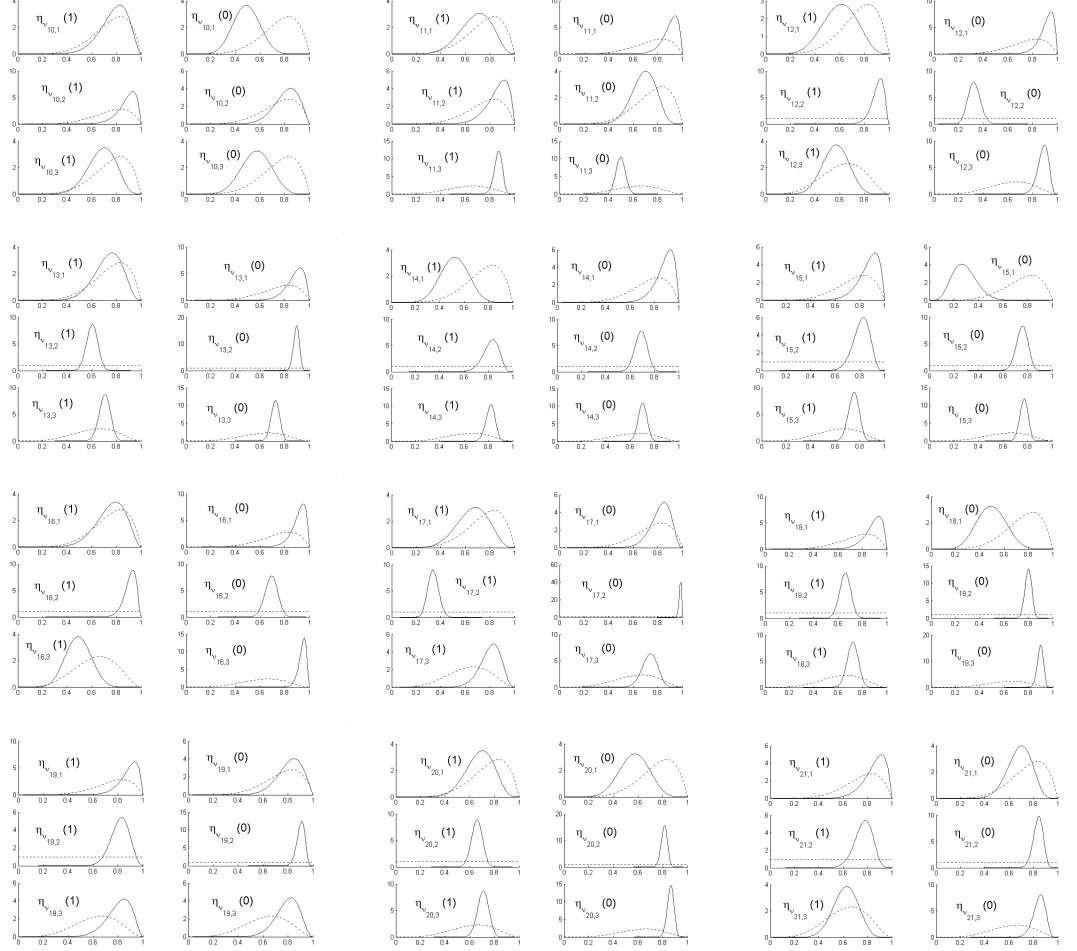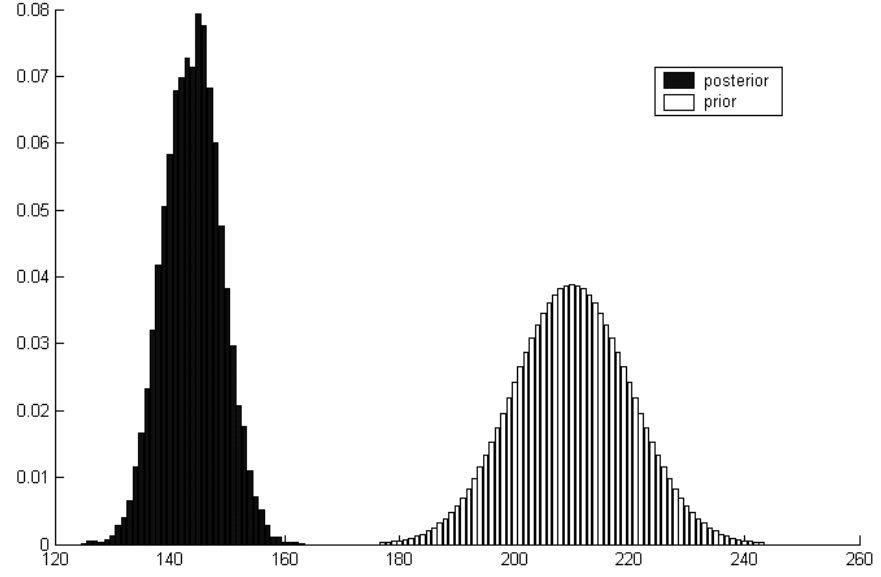
Figure G 1. continued for perceivers i 10 through 21

Figure G.2. The prior and posterior distribution of the total number of arcs. in **Z** when the prior on **Z** is uniform over all labeled digraphs with 21 vertices. The posterior of the total number of arcs in **Z** is derived from the posterior distribution of **Z** conditional on model 6.