Bayesian Analysis of Cognitive Social Structures with Covariates.

Johan Koskinen*

Abstract

When studying perception of interaction in the framework of Cognitive Social Structures (Krackhardt 1987) one is often interested in correlating bias on the part of the perceivers with exogenous attributes of the perceivers (e.g. Bondonio 1998; Casciaro 1998; Casciaro, Carley and Krackhardt, 1999). In this study we propose a Bayesian probit model for explaining biases in perception in terms of known covariates.

Keywords: Binary probit. Bayesian analysis of social networks. Cognitive social structures.

1 Introduction

In various situations we are interested in a relational structure among individuals described by graphs or digraphs. As an example the structure of interest could be the co-offending of criminals in a criminal network, i.e. who commits crimes with whom. If this can be directly observed our quest is at an end but more often than not it can not be directly observed. If we only had a number of "independent" sources that reported on what the structure looked like, we would hope that we could use the information in these reports to estimate what the structure looked like. It turns out that given a few assumptions about how these reports were generated we can not only estimate the hidden structure - for example the probability that criminals A and B commit crimes together - but also obtain a measure on the reliability of each source vis-à-vis the hidden structure - given data there we can say that with

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

95% confidence there is a 75% propensity that C reports that criminals A and B commit crimes together when they actually do.

In a previous paper (Koskinen, 2002) a model for Bayesian Cognitive Social Structures was investigated. Slightly modified, we used a model originally defined by Batchelder et al (1997) and modeled the prior belief in different values on the informant accuracy by conjugate distributions. In the present paper, we expand the model to include observable covariates so that the accuracy of a perceiver is a function of the attributes of the perceiver rather than that the accuracy depends on who the perceiver is.

Perhaps more important here than in the previous paper is to deal with the problem of a model that is not identifiable. The solution is to employ a proper subjective analysis of data, subjective in the sense that we need to quantify our prior assumptions and beliefs with proper prior distributions. Naturally, we are not limited to only one view, rather the opposite. By specifying several models with different prior assumptions, we get an analysis of data that reflects several different perspectives, and properly reported the reader is given sufficient information to judge for himself what seems plausible and not.

2 Model specification

Suppose 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 a set of 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 $j, k \in V$. (For an introduction to matrix algebra see e.g. Harville, 1997). Let the set of informants be represented by $\mathcal{I} = \{1, \ldots, m\}$, and the matrix $\mathbf{X}_i$, be the adjacency matrix for the relation $R_i \subseteq V^2$, which is the information about R reported by i, for $i \in \mathcal{I}$. Now assume that 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 and absence of ties in the social network. Further assume that conditional on $\mathbf{Z}$, $\mathbf{X}_1, \ldots, \mathbf{X}_m$ are independent and their elements are independent and satisfy

$$\Pr\left(X_{ijk} = x_{ijk} \middle| \mathbf{Z} = \mathbf{z}\right) = \Pr\left(X_{ijk} = x_{ijk} \middle| Z_{jk} = z_{jk}\right). \tag{1}$$

Therefore we can write the likelihood function of $\mathbf{z}$ and $\boldsymbol{\eta}$ as a product of Bernoulli probability mass functions

$$\begin{aligned} r\left(\mathbf{x} \middle| \mathbf{z}, \boldsymbol{\eta}\right) &= \prod_{i \in \mathcal{I}} \prod_{j,k \in V} \prod_{s \in \{0,1\}} \left(\eta_{ijk}\left(s\right)\right)^{\mathbf{1}\left(z_{jk}=s, x_{ijk}=s\right)} \\ &\quad \times \left(1 - \eta_{ijk}\left(s\right)\right)^{\mathbf{1}\left(z_{jk}=s, x_{ijk}\neq s\right)}, \end{aligned} \tag{2}$$

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 0 otherwise.

3 The probit model

Let us assume that for each combination of informant and pair of actors in the network, $(i, j, k) \in \mathcal{I} \times V^2$, we have a covariate $p \times 1$ vector $\mathbf{w}_{ijk}$ and two $p \times 1$ vectors of unknown coefficients β_1 and β_0. We could then let $\eta_{ijk}\left(1\right) = \Phi\left(\mathbf{w}_{ijk}^T \beta_1\right)$, and $\eta_{ijk}\left(0\right) = 1 - \Phi\left(\mathbf{w}_{ijk}^T \beta_0\right)$, where $\Phi\left(\cdot\right)$ is the standard normal cumulative distribution function (cdf)[1]. Thus the model for an observation could be written

$$\begin{aligned} P\left(X_{ijk} = x \middle| Z_{jk} = z, \mathbf{w}_{ijk}, \beta_1, \beta_0\right) &= \Phi\left(\mathbf{w}_{ijk}^T \beta_1\right)^{zx} \\ &\quad \times \left(1 - \Phi\left(\mathbf{w}_{ijk}^T \beta_1\right)\right)^{z(1-x)} \\ &\quad \times \Phi\left(\mathbf{w}_{ijk}^T \beta_0\right)^{(1-z)x} \\ &\quad \times \left(1 - \Phi\left(\mathbf{w}_{ijk}^T \beta_0\right)\right)^{(1-z)(1-x)}, \end{aligned} \tag{3}$$

To facilitate estimation procedures, following an article by Albert and Chib (1993), we introduce the $n \times n$ vector of latent variables $\mathbf{Y}_i = (y_{ijk})$, that for each triple (i, j, k), are independently distributed

$$Y_{ijk} \sim N\left(\mathbf{w}_{ijk}^T \beta_1, 1\right) \text{ if } Z_{jk} = 1,$$

[1] In order to interpret these probabilities in the *theory of signal detection* paradigm, Batchelder et al (1997) used standard normal cdf's to infer *signal perceptibility* and *response bias*.

and

$$Y_{ijk} \sim N\left(\mathbf{w}_{ijk}^T \beta_0, 1\right) \text{ if } Z_{jk} = 0,$$

and let $X_{ijk} = 1$ if $Y_{ijk} > 0$ and $X_{ijk} = 0$ if $Y_{ijk} \leqslant 0$. Now collect the vectorised reported adjacency matrices and their corresponding latent variables in two $n^2 m \times 1$ vectors $\mathbf{X} = ((\text{vec}\mathbf{X}_1)^T, \ldots, (\text{vec}\mathbf{X}_m)^T)^T$, and $\mathbf{Y} = ((\text{vec}\mathbf{Y}_1)^T, \ldots, (\text{vec}\mathbf{Y}_m)^T)^T$. The vec-operator takes the columns of the argument matrix left to right and stacks them beneath each other. Let $\mathbf{W} = (\mathbf{W}_1^T, \ldots, \mathbf{W}_m^T)^T$, where the $n^2 \times p$ matrix $\mathbf{W}_i = (\mathbf{w}_{i11}, \mathbf{w}_{i21}, \ldots, \mathbf{w}_{in-1n}, \mathbf{w}_{inn})^T$. Further, let $\mathbf{Z}^*$ be an $n^2 \times n^2$ diagonal matrix with vec$\mathbf{Z}$ on the diagonal, and we have the following linear form

$$\mathbf{Y} = [\mathbf{C}\mathbf{W}\beta_1 + \mathbf{S}\mathbf{W}\beta_0] + \boldsymbol{\varepsilon}, \tag{4}$$

where $\boldsymbol{\varepsilon}$ is a standard normal vector and where

$$\mathbf{S} = \left(\mathbf{I}_{n^2 m} - \mathbf{I}_m \otimes \mathbf{Z}^*\right), \quad \mathbf{C} = \left(\mathbf{I}_m \otimes \mathbf{Z}^*\right),$$

in which $\otimes$ is the Kronecker product and $\mathbf{I}_r$ is the $r \times r$ identity matrix. Hence we can use standard results for linear models in our estimation procedures.

4 Priors and full conditional posteriors

It is easy to see that the introduction of the latent variables ,$\mathbf{Y}_i$, retain the structure of (3). Also, with suitably chosen priors, sampling from the exact posteriors of the parameters is fairly straightforward. Full details about the Gibbs sampling algorithm can for example be found in Gelfand and Smith (1990).

To implement the Gibbs sampler we need the full conditional posteriors of each of the parameters. When the parameters are a priori independent the joint conditional distribution of the parameters and latent variables given data is

$$\begin{aligned}
\pi\left(\beta_1, \beta_0, \mathbf{z}, \mathbf{y} \middle| \mathbf{x}\right) \quad \propto \quad & \pi\left(\beta_1\right) \pi\left(\beta_0\right) \pi\left(\mathbf{z}\right) \\
& \times \prod_{(i,j,k) \in \mathcal{I} \times V^2} \left\{\mathbf{1}\left(Y_{ijk} > 0\right) \mathbf{1}\left(x_{ijk} = 1\right)\right. \\
& \left. + \mathbf{1}\left(Y_{ijk} \leqslant 0\right) \mathbf{1}\left(x_{ijk} = 0\right)\right\} \phi\left(Y_{ijk} - w_{ijk}^T \beta_1\right)^{z_{jk}} \\
& \times \left\{\mathbf{1}\left(Y_{ijk} > 0\right) \mathbf{1}\left(x_{ijk} = 1\right)\right. \\
& \left. + \mathbf{1}\left(Y_{ijk} \leqslant 0\right) \mathbf{1}\left(x_{ijk} = 0\right)\right\} \phi\left(Y_{ijk} - w_{ijk}^T \beta_0\right)^{1 - z_{jk}},
\end{aligned}$$

where $\phi(\quad)$ is the $N(0,1)$ probability distribution function (pdf).

From the form of (4) we see that we can use standard results from linear models. Let the prior on β_1 be $N_p(\beta_1^*, \mathbf{B}_1^*)$ and independently on β_0 be $N_p(\beta_0^*, \mathbf{B}_0^*)$ so that we have independently for $s = 0, 1$

$$\beta_s | \, \mathbf{z}, \mathbf{y} \sim N_p\left(\widetilde{\beta}_s^*, \widetilde{\mathbf{B}}_s^*\right)$$

where $\widetilde{\beta}_1^* = \widetilde{\mathbf{B}}_1^*\left(\mathbf{B}_1^{*-1}\beta_1^* + \mathbf{W}^T\mathbf{C}\mathbf{Y}\right)$, and $\widetilde{\beta}_0^* = \widetilde{\mathbf{B}}_0^*\left(\mathbf{B}_0^{*-1}\beta_0^* + \mathbf{W}^T\mathbf{S}\mathbf{Y}\right)$. Further, $\widetilde{\mathbf{B}}_1^* = \left(\mathbf{B}_1^{*-1} + \mathbf{W}^T\mathbf{C}\mathbf{W}\right)^{-1}$, and $\widetilde{\mathbf{B}}_0^* = \left(\mathbf{B}_0^{*-1} + \mathbf{W}^T\mathbf{S}\mathbf{W}\right)^{-1}$.

For the latent variables $\mathbf{Y}$ we have that for each element if $Z_{jk} = s$

$$Y_{ijk} | \, \mathbf{x}, \beta_1, \beta_0, \mathbf{z} \sim N\left(\mathbf{w}_{ijk}^T\beta_s, 1\right),$$

truncated to the left at 0 if $x_{ijk} = 1$ and truncated to the right at 0 if $x_{ijk} = 0$, for $s = 0, 1$. We can write the full conditional posterior of each Z_{jk} as

$$\begin{aligned}
\pi\left(z_{jk} | \, \mathbf{x}, \beta_1, \beta_0, \mathbf{y}, \mathbf{z}_{-jk}\right) \quad = \quad & \left\{ \pi(1 | \, \mathbf{z}_{-jk}) \prod_i \Phi\left(\mathbf{w}_{ijk}^T\beta_1\right)^{x_{ijk}} \left(1 - \Phi\left(\mathbf{w}_{ijk}^T\beta_1\right)\right)^{1-x_{ijk}} \right. \\
& \left. + \pi(0 | \, \mathbf{z}_{-jk}) \prod_i \Phi\left(\mathbf{w}_{ijk}^T\beta_0\right)^{x_{ijk}} \left(1 - \Phi\left(\mathbf{w}_{ijk}^T\beta_0\right)\right)^{1-x_{ijk}} \right\}^{-1} \\
& \times \prod_i \Phi\left(\mathbf{w}_{ijk}^T\beta_1\right)^{z_{jk}x_{ijk}} \left(1 - \Phi\left(\mathbf{w}_{ijk}^T\beta_1\right)\right)^{z_{jk}\left(1-x_{ijk}\right)} \\
& \times \Phi\left(\mathbf{w}_{ijk}^T\beta_0\right)^{\left(1-z_{jk}\right)x_{ijk}} \left(1 - \Phi\left(\mathbf{w}_{ijk}^T\beta_0\right)\right)^{\left(1-z_{jk}\right)\left(1-x_{ijk}\right)} \\
& \times \pi\left(z_{jk} | \, \mathbf{z}_{-jk}\right),
\end{aligned}$$

where $\pi\left(z_{jk} | \, \mathbf{z}_{-jk}\right)$ is the prior on Z_{jk} given the rest of the matrix being equal to $\mathbf{z}_{-jk}$. Using a vague prior on $\mathbf{Z}$ (i.e. the probability of a particular graph is constant over all graphs with n nodes a priori) we get that independently for each j, k

$$Z_{jk} | \, x_{1jk}, \ldots, x_{mjk}, \beta_1, \beta_0 \sim Bernoulli\left(\frac{1}{1 + q_{jk}}\right)$$

where

$$q_{jk} = \prod_i \left(\frac{\Phi\left(\mathbf{w}_{ijk}^T\beta_0\right)}{\Phi\left(\mathbf{w}_{ijk}^T\beta_1\right)}\right)^{x_{ijk}} \left(\frac{1 - \Phi\left(\mathbf{w}_{ijk}^T\beta_0\right)}{1 - \Phi\left(\mathbf{w}_{ijk}^T\beta_1\right)}\right)^{1-x_{ijk}}.$$

By cycling through these conditional posteriors for a certain number of steps we know that after a certain burn-in period the joint output is a sample from the exact joint posterior of the parameters given data.

The reason why we have given the full conditional posteriors for the coefficients that involve a priori specified hyperparameters and not, as is usual, the posteriors corresponding to vague priors is to counter non-identifiyability. That the model is not identified means that several sets of values on the parameters give identical likelihoods. If one only use vague priors the posteriors will have several modes with equal height. More explicitly, for $\xi, \omega \in \mathbf{R}^p$,

$$P\left(\mathbf{X}=\mathbf{x}|\,\mathbf{Z}=\mathbf{z},\beta_1=\xi,\beta_0=\omega,\mathbf{W}\right)=P\left(\mathbf{X}=\mathbf{x}|\,\mathbf{Z}=\mathbf{z}^c,\beta_1=\omega,\beta_0=\xi,\mathbf{W}\right)$$

where $\mathbf{z}^c$ is the complement of $\mathbf{z}$. This implies that with vague priors everywhere, i.e. $\pi_{\mathbf{Z}}$, π_{β_1}, and π_{β_0} all constant,

$$\pi_{\beta_1,\beta_0,\mathbf{Z}}\left(\xi,\omega,\mathbf{z}|\,\mathbf{x},\mathbf{w}\right)=\pi_{\beta_1,\beta_0,\mathbf{Z}}\left(\omega,\xi,\mathbf{z}^c|\,\mathbf{x},\mathbf{w}\right),$$

obviously this is avoided when certain values a priori are deemed less likely than others. The posteriors should, as always, be interpreted in relation to the nature of the prior distributions.

5 Empirical example

To illustrate the procedures described in this paper we apply them on Krackhardts (1987) high-tech managers. Casciaro (1998) suggests that the accuracy of a perceiver with regards to an advice network is positively related to his or her hierarchical level. One reasonable model would then be to have one covariate, hierarchical level. Another suggestion as to how covariates are related to the accuracy in perception is given in Bondonio (1998). The accuracy of a perceiver i on the dyad (j,k) could be assumed to be positively related to the closeness of i and j, and how similar i and j are in terms of age and tenure. This would mean that we need to form a covariate out of two exogenous covariates, age and tenure, and the positions of the perceivers relative to the actors perceived. The dissimilarity could be measured by squared euclidean distances.

We have specified the following covariates $\mathbf{w}_{ijk}=(1,w_{ijk,1},w_{ijk,2})$, where $w_{ijk,1}$ is equal to one if i is a vice president of the firm or CEO, and nougth

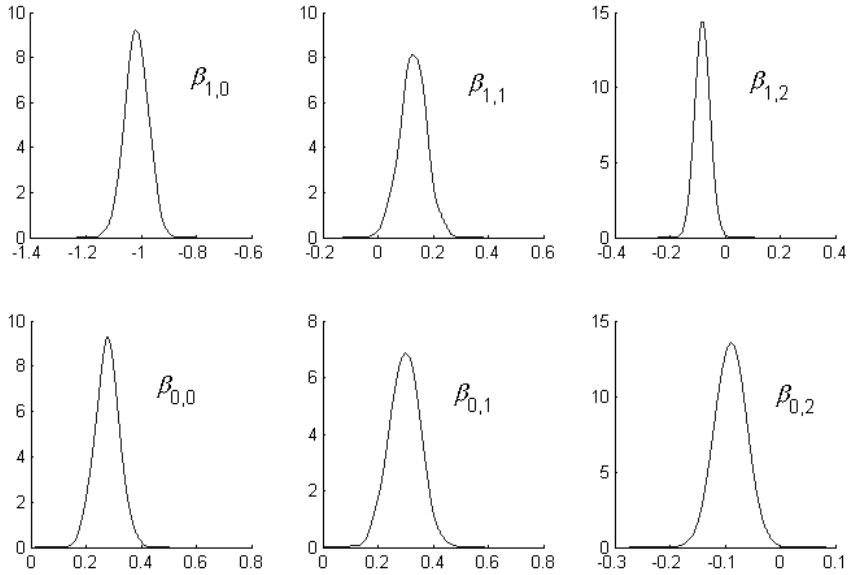

Figure 1: Posterior distributions of coefficients given data for Krackhardt's high tech managers.

otherwise; $w_{ijk,2}$ is the squared euclidean distance between i and j with respect to length of tenure and age; the first constant corresponds to the intercept.

The priors used are given in Table 1 and 2 along with the characteristics of the posteriors. The covariances between coefficients was set to 0 a prior. The posterior distributions of the parameters are given in Figure 1. The uncertainty is relatively small and the origin would not be included in any interval estimates with usual credibility levels for any of the coefficients.

Table 1. Prior and posterior expectations

	$\beta_{1,0}$	$\beta_{1,1}$	$\beta_{1,2}$	$\beta_{0,0}$	$\beta_{0,1}$	$\beta_{0,2}$
a priori	0	.5	$-.25$	0	$-.5$	.25
a posteriori	-1.0164	.1281	$-.0819$	.2768	.2976	$-.0897$

Table 2. Prior and posterior variance and covariance

	$\beta_{1,0}$	$\beta_{1,1}$	$\beta_{1,2}$	$\beta_{0,0}$	$\beta_{0,1}$	$\beta_{0,2}$
a priori	5	5	5	5	5	5
a posteriori						
$\beta_{1,0}$	.0016	−.0011	−.0006	.0004	.0000	−.0000
$\beta_{1,1}$		.0022	−.0001	−.0001	−.0001	.0000
$\beta_{1,2}$			.0005	−.0001	.0000	.0000
$\beta_{0,0}$				.0018	−.0012	−.0005
$\beta_{0,1}$					.0031	−.0002
$\beta_{0,2}$						.0006

6 Conclusions

We have here proposed a method for analysing the relation between the "accuracy" when perceiving social networks and covariates. Albert and Chib's (1993) approach in the analysis of binary data using the binary probit model, of introducing latent variables, only needs a slight modification to be applicable in the present context. This yields an easy to implement Gibbs sampling scheme involving standard statistical distributions.

The natural next step to take is to investigate the performance of the approach presented here when it comes to comparing different parameterisations and testing hypothesis. The lack of natural "reference" priors might seem a trifle inconvenient to some researchers and it would be interesting to try to apply restrictions on the parameter space such that hyper parameters do not have to be more specifically specified. An example of such an identifiability constraint is suggested for a similar model in Salabasis and Villani (2000). Regardless, interpretation of the parameters and their priors deserves more attention and it is possible that one could find an effective procedure for eduction of priors by comparative methods.

The restriction on the latent variables to have unity variance and zero covariance is restrictive but on the other hand, relaxing this assumption does not change the overall procedure considerably.

References

[1] Albert, J.H., Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88:

669-679.

[2] 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.

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

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

[5] Casciaro, Carley and Krackhardt (1999). Positive affectivity and accuracy in social network perception. *Motivation and Emotion*, 23:285-306.

[6] 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.

[7] Harville, D. A. (1997). *Matrix Algebra From a Statistician's Perspective*. New York: Springer-Verlag.

[8] Koskinen, J.H. (2002). *Bayesian Analysis of Perceived Social Networks.* Research Report No. 2, Stockholm University, Departement of Statistics.

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

[10] Salabasis, M. and Villani, M. (2000). *Panel Regression with Unobserved Classes*. SSE/EFI Working Paper Series in Economics and Finance, 353.