

A Semiparametric Bayesian Approach to Network Modelling using Dirichlet Process Priors

Pulak Ghosh, Paramjit S. Gill, Saman Muthukumarana and Tim B. Swartz *

*Pulak Ghosh is Associate Professor, Department of Biostatistics and Winship Cancer Institute, Emory University, 1518 Clifton Road NE, Atlanta GA, 30322. Paramjit Gill is Associate Professor, Mathematics, Statistics and Physics Unit, Irving K. Barber School of Arts and Sciences, University of British Columbia Okanagan, 3333 University Way, Kelowna BC, Canada V1V1V7. Tim Swartz is Professor and Saman Muthukumarana is a PhD candidate, Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby BC, Canada V5A1S6. Gill and Swartz have been partially supported by grants from the Natural Sciences and Engineering Research Council of Canada. The authors thank the Associate Editor and a Referee for helpful comments that have improved the paper.

Abstract

This paper considers the use of Dirichlet process priors in the statistical analysis of network data. Dirichlet process priors have the advantage of avoiding the parametric specification for distributions which are rarely known and for facilitating a clustering effect which is often applicable to network nodes. The approach is highlighted on two network models and is conveniently implemented using WinBUGS software.

Keywords : Bayesian semiparametric modelling, Clustering, Dirichlet process, Network models, Social relations. WinBUGS software.

1 INTRODUCTION

The analysis of network data is an active research topic. The range of applications is vast and includes such diverse areas as the detection of fraud in the telecommunications industry (Cortes, Pregibon and Volinsky 2003), the development of adaptive sampling schemes for populations at risk of HIV/AIDS infection (Thompson 2006), the study of conflicts between nations (Ward and Hoff 2007, Hoff 2009), the quantification of social structure in elephant herds (Vance 2008) and the investigation of the cooperative structure between lawyers (Lazega and Pattison 1999).

Not only are the areas of application varied, the statistical approaches to the analysis of network data are also varied. The approaches depend on many factors including the inferential goal of the analysis whether it be description, testing or prediction, the size of the data set and the nature of the data. Data may be continuous or discrete, there may be complex dependencies amongst nodes, relationships may be directed or non-directed, data may be dynamic, multivariate, have missing values, include covariates, lack balance, etc. Network analyses have been considered under both classical and Bayesian paradigms.

Although a complete review of the network literature strikes us as a daunting task, we remark on some of the prominent approaches to the statistical analysis of network data. With continuous observations between network nodes, Warner, Kenny and Stoto (1979) introduced the social relations model whose structure considers dependencies in the measurements between nodes. In the social relations model, nodes (e.g. subjects) have dual roles as both actors and partners where measurements between nodes are dependent on both actor and partner effects. Social relations models (also referred to as round robin models) were originally studied using analysis of variance methodology. Other inferential approaches have since been explored including maximum likelihood (Wong 1982), multilevel methods (Snijders and Kenny 1999) and Bayesian methods (Hoff 2005,

Gill and Swartz 2007).

More research effort has taken place in the context of binary network data where a greater amount of mathematics and graph theory have come into play (Besag 1974, Frank and Strauss 1986). In the context of binary network data, a seminal contribution is due to Holland and Leinhardt (1981) who broke away from the often unrealistic assumption of independence between pairs of nodes and proposed the p_1 -model for directed graphs. The original p_1 -model has been expanded upon in many ways including empirical Bayes approaches (Wong 1987), fully Bayesian approaches (Gill and Swartz 2004) and the consideration of more complex dependencies (Wasserman and Pattison 1996). All of these models fall under the general framework of exponential random graph models whose various limitations have been discussed by Besag (2001) and Handcock (2003). A main feature of exponential random graph models is that the entire network is modelled. A distinct approach to the analysis of binary network data involves modelling the individual nodal relationships; these models have been generalized in various ways and are referred to as latent factor models (Hoff, Raftery and Handcock 2002, Handcock, Raftery and Tantrum 2007). Finally, a recent approach which is related to the latent factor methodology provides a greater emphasis on the socio-spatial structure typically inherent in networks (Linkletter 2007). The approach requires the existence of meaningful spatial covariates and appears well suited for prediction.

This paper investigates the suitability of Dirichlet process priors in the Bayesian analysis of network data. The Dirichlet process (Ferguson 1974) which was once a mathematical curiosity is becoming a popular applied tool (Dey, Müller and Sinha 1998). Dirichlet process priors allow the researcher to weaken prior assumptions by going from a parametric to a semiparametric framework. This is important in the analysis of network data where complex nodal relationships rarely allow a researcher the confidence in assigning paramet-

ric priors. The Dirichlet process has a secondary benefit due to the fact that its support is restricted to discrete distributions. This results in a clustering effect which is often suitable for network data where groups of individuals in a network can be thought of as arising from the same cohort. Importantly, we demonstrate how Dirichlet process priors can be easily implemented in network models using WinBUGS software (Spiegelhalter, Thomas and Best 2003). The ease in which this can be done increases the potential of the methodology for widespread usage.

In section 2, we provide an overview of the Dirichlet process with an emphasis on issues that are most relevant to the implementation of the network models that are considered in this paper. In section 3, we provide three examples which demonstrate the utility of Dirichlet process mixture models in the context of social networks. The first example in section 3 is a simulation study involving a simple but popular binary network model. We demonstrate that the inferences are what we expect under a variety of conditions. The second example concerns an enhanced binary network model that studies the working relationships between lawyers. This is a variation of the p_1 -model of Holland and Leinhardt (1981) and stratifies the lawyers according to their professional rank. The third example involves a social relations model previously studied by Gill and Swartz (2007) where the observations between nodes are measured on a continuous scale. In each of the three examples, the Dirichlet process can be easily implemented using WinBUGS software. Some concluding remarks are provided in section 4.

2 THE DIRICHLET PROCESS

In a Bayesian framework, parameters are not viewed as fixed quantities whose values are unknown to us. Rather, parameters are thought of as random quantities that arise from

probability distributions. For the sake of discussion, consider random effects $\theta_1, \dots, \theta_n$ from a parametric Bayesian model where

$$\theta_i \stackrel{\text{iid}}{\sim} G_0. \tag{1}$$

In (1), we specify the parametric distribution G_0 , and note that sometimes G_0 may depend on additional parameters. For example, G_0 may correspond to a normal distribution whose mean and variance are left unspecified. We also note that the θ 's may be scalar or vector-valued.

With a Dirichlet process (DP) prior, we instead write

$$\begin{aligned} \theta_i &\stackrel{\text{iid}}{\sim} G \\ \text{where } G &\sim \text{DP}(m, G_0). \end{aligned} \tag{2}$$

In (2), we are stating that the parameter θ arises from a distribution G but G itself arises from a distribution of distributions known as the Dirichlet process with concentration parameter $m > 0$ and mean $E(G) = G_0$. The Dirichlet process in (2) is defined (Ferguson 1974) as follows: For finite k and any measurable partition (A_1, \dots, A_k) of \mathcal{R} , the distribution of $G(A_1), \dots, G(A_k)$ is Dirichlet($mG_0(A_1), \dots, mG_0(A_k)$). It is apparent that the baseline distribution G_0 may serve as an initial guess of the distribution of θ and that the concentration parameter m determines our a priori confidence in G_0 with larger values corresponding to greater degrees of belief. Under (2), we think of a distribution G arising from the Dirichlet process followed by a parameter θ arising from G .

An illuminating and alternative definition of the Dirichlet process was given by Sethuraman (1994). His constructive definition of (2) which is also known as the stick breaking representation is given as follows: Generate a set of iid atoms $\theta_i^* \sim G_0$ and generate a set

of weights $w_i = y_i \prod_{j=1}^i (1 - y_j)$ where the y_i are iid with $y_i \sim \text{Beta}(1, m)$ for $i = 1, \dots, \infty$.

Then

$$G = \sum_{i=1}^{\infty} w_i I_{\theta_i^*} \quad (3)$$

where $I_{\theta_i^*}$ is a point mass at θ_i^* .

For our purposes, the Sethuraman (1994) construction is most useful. First, we see that the stick breaking mechanism creates smaller and smaller weights w_i . This suggests that at a certain point we can truncate the sum (3) and obtain a reasonable approximation to G (Muliere and Tardella 1998). Ishwaran and Zarepour (2002) suggest that the number of truncation points $L = n$ when the number of random effects n is small and $L = \sqrt{n}$ when large. Secondly, in WinBUGS modelling, it is required to specify the distributions of parameters. Whereas the Ferguson (1974) definition does not provide an adequate WinBUGS specification, the truncated version of (3) can be easily implemented. Finally, the stick breaking construction clearly shows that a generated G is a discrete probability distribution which implies that there is non-negligible probability that θ 's generated from the same G have the same value. As later demonstrated in the examples, it is often desirable to facilitate clustering in network modelling.

In a typical MCMC application, there are considerable programming challenges that face a user. In particular, one needs to determine a Markov chain which has the posterior as its invariant distribution. The chain also needs to be an appropriate chain that reaches practical convergence in practical computing times. This is sometimes facilitated by breaking the parameter vector into smaller components where simulation is carried out componentwise. A good introduction to MCMC methods is given by Gilks, Richardson and Spiegelhalter (1996). The appeal of WinBUGS software is that the programming demands are often significantly reduced. A WinBUGS implementation requires only the

specification of the likelihood, the prior distribution and the data. WinBUGS determines the Markov chain in the background and provides the user with MCMC output from which inferences can be obtained. When possible (i.e. with “conjugate” distributions), WinBUGS uses the Gibbs sampling algorithm as the Markov chain. In more complex situations, WinBUGS imbeds Metropolis steps with normal proposal densities. Detailed information on WinBUGS is available from the WinBUGS website www.mrc-bsu.cam.ac.uk/bugs/.

In our applications where Dirichlet process priors are used in network models, WinBUGS output allows us to readily assess clustering. Given a single iteration from the Markov chain, we simply observe which θ 's have the same value as θ_i . Over many iterations, the proportion of times that θ_i is the same as θ_j is an estimate of the posterior probability that the i th and j th subjects cluster together. An advantage of Bayesian clustering is that probabilistic statements can be made concerning clustering. We contrast this with many classical deterministic algorithms where there is no measure of clustering strength.

Although the DP is a highly technical tool, the simple introduction above is all that is required to use Dirichlet process priors in the network models considered in this paper.

3 EXAMPLES

We consider three examples that demonstrate the utility of Dirichlet process mixture models in the context of social networks.

3.1 *Example 1: A Simulation Study*

We report on a simulation study that investigates the performance of clustering using the Dirichlet process mixture in a simple binary network model. The model is a variation of logistic regression where binary responses describe the presence of ties between nodes. The simulated network data consist of an n by n matrix Y where $y_{ij} = 1$, $i \neq j$ indicates that subject i has a tie towards subject j , and $y_{ij} = 0$ denotes the absence of such a tie. Each $y_{ij} \sim \text{Bernoulli}(p_{ij})$ is assumed independent of other y 's and the independence assumption is a common criticism of the simple model. We use a logistic link for $p_{ij} = \text{Pr}(y_{ij} = 1)$ whereby

$$\begin{aligned} \log\left(\frac{p_{ij}}{1-p_{ij}}\right) &= \mu + \alpha_i + \beta_j \\ \log\left(\frac{p_{ji}}{1-p_{ji}}\right) &= \mu + \alpha_j + \beta_i. \end{aligned} \tag{4}$$

In (4), the parameters α_i and β_i quantify the strength with which subject i produces and attracts ties respectively. With the inclusion of the α and β random effects, a type of dependency is introduced among dyads which share a common subject. The parameter μ measures the overall density of ties in the network.

In order to induce clustering amongst the random effects, we divide $n = 100$ subjects into four groups of equal size. This is a substantial network as each subject has $2(99) = 198$ observations that describe its associated ties. The large size of the dataset helps demonstrate the utility of the approach. We set $\mu = 0$ and set the random effects according to $(\alpha_i, \beta_i) = (-1, -1), (1, 1), (-1, 1), (1, -1)$ for the four groups.

A Bayesian model for this network consists of the Bernoulli model description for Y , the logistic link (4), and the diffuse prior distributions $\mu \sim \text{Normal}(0, 10000)$, $(\alpha_i, \beta_i)' \stackrel{\text{iid}}{\sim} \text{Normal}_2(0, \Sigma_{\alpha\beta})$ and $\Sigma_{\alpha\beta}^{-1} \sim \text{Wishart}_2(2, I)$. To implement the Dirichlet process mixture version of the model, these priors are maintained except that the prior distribution for

(α_i, β_i) is modified according to (2) where the number of truncation points $L = 20$ and the baseline distribution G_0 is the bivariate normal. The prior for the concentration parameter is given by $m \sim \text{Uniform}(0.4, 10)$ which is similar to the choices made by Ohlssen, Sharples and Spiegelhalter (2007).

In testing the adequacy of the model, we note that all of the 100 subjects are correctly clustered into their corresponding groups. The posterior probabilities of pairs of subjects (from the same group) clustering together range from 0.77 to 0.99. For pairs of subjects from different groups, the posterior probabilities of clustering are all 0.00. WinBUGS simulations for this substantial dataset require roughly two hours of computation for 20000 iterations.

We then modify the density parameter from $\mu = 0$ to $\mu = -1$ prior to simulating the data Y . This has the effect of radically decreasing the number of ties between subjects. Again, we find perfect clustering for the 100 subjects.

As a third test of the utility of the model, we introduce some variation in the random effects (α_i, β_i) as might be expected in most networks. We generate the (α_i, β_i) from a bivariate normal distribution having zero correlation and standard deviation 0.1 in both the α and β parameters. This time, the clustering is again perfect in the sense that none of the subjects from a given group cluster with subjects outside of their own group. However, there is a little bit of sub-clustering of subjects within their own groups. In particular,

- the group with mean $(-1, -1)$ has two sub-clusters of sizes 3 and 22
- the group with mean $(1, 1)$ has two sub-clusters of sizes 5 and 20
- the group with mean $(-1, 1)$ is a single cluster
- the group with mean $(1, -1)$ has three sub-clusters of sizes 2, 4, and 19 .

Cluster membership is based on posterior probability of pairwise clustering exceeding 0.5. When the threshold level is reduced to 0.25, we again observe perfect clustering with each of the 100 subjects assigned to its original group.

3.2 *Example 2: An Enhanced Binary Network Model*

We now consider an exponential random graph model previously studied by Gill and Swartz (2004). The data is an n by n matrix $Y = (y_{ij})$ describing the relationships between n nodes where $y_{ij} = 1$ denotes a tie from node i to node j and $y_{ij} = 0$ denotes the absence of such a tie, $i \neq j$. The p_1 -model of Holland and Leinhardt (1981) states

$$\text{Prob}(Y) \propto \exp \left(\sum_{i < j} \phi y_{ij} y_{ji} + \sum_{i \neq j} (\theta + \alpha_i + \beta_j) y_{ij} \right) \quad (5)$$

where (5) implies the independence of the dyads $D_{ij} = (y_{ij}, y_{ji})$, $i < j$. The parameter ϕ measures the average degree of reciprocity or mutuality of ties in the population whereas θ measures the density of ties. The subject specific effects α_i and β_i represent the ability of subject i to extend and attract ties respectively. The Bayesian model specification then assigns prior distributions to the primary parameters of interest

$$\phi \sim \text{Normal}(\mu_\phi, \sigma_\phi^2), \quad (6)$$

$$\theta \sim \text{Normal}(\mu_\theta, \sigma_\theta^2), \quad (7)$$

$$(\alpha_i, \beta_i)' \stackrel{\text{iid}}{\sim} \text{Normal}_2(0, \Sigma_{\alpha\beta}). \quad (8)$$

To complete the Bayesian model specification, hyperpriors are assigned as follows:

$$\begin{aligned}
\mu_\phi &\sim \text{Normal}(\mu_0, \sigma_0^2), & \mu_\theta &\sim \text{Normal}(\mu_0, \sigma_0^2), \\
\sigma_\phi^{-2} &\sim \text{Gamma}(a_0, b_0), & \sigma_\theta^{-2} &\sim \text{Gamma}(a_0, b_0), \\
\Sigma_{\alpha\beta}^{-1} &\sim \text{Wishart}_2(r_0, \Sigma_0).
\end{aligned}
\tag{9}$$

The parameters subscripted with a 0 in the hyperpriors (9) are set to provide diffuse distributions. To implement the Dirichlet process mixture version of the model, priors (6) through (9) are maintained except that (8) is modified as follows:

$$\begin{aligned}
(\alpha_i, \beta_i)' &\stackrel{\text{iid}}{\sim} G \\
G &\sim \text{DP}(m, \text{Normal}_2(0, \Sigma_{\alpha\beta})) \\
m &\sim \text{Uniform}(0.4, 10.0).
\end{aligned}
\tag{10}$$

To investigate the enhanced Dirichlet process mixture model, we consider a subset of the law firm data originally studied by Lazega and Pattison (1999). The directed data matrix Y specifies whether or not advice was given between lawyers in a law firm consisting of 36 partners and 35 associates. The use of the Dirichlet process provides an approach to modelling the heterogeneity amongst the lawyers with respect to the parameters α and β . In the law firm example, one line of reasoning suggests that:

- senior lawyers are more likely to give advice but are less likely to receive advice (positive α and negative β)
- junior lawyers are more likely to receive advice but are less likely to give advice (negative α and positive β)
- intermediate lawyers are likely to provide advice to the same extent that it is sought (comparable α and β) .

The idea of partitioning the network actors into classes is related to the concept of blockmodelling. Wasserman and Faust (1994, chapters 10 and 16) describe in detail a priori and a posteriori blockmodelling. In a priori blockmodelling, exogenous attributes of actors are used for partitioning. Although this may appear sensible, there may very well be actors who do not fit the mold for a priori blockmodelling and may be thought of as a cluster of their own. For example, there may be young associates brimming with confidence who rarely ask for advice but readily offer their opinions. We prefer to let the data determine the clusters and this is possible with the proposed Dirichlet process mixture model. With a priori blockmodelling, the purpose is to describe overall propensities. However, excessive rogue cases can adversely affect model fit. Another objection to a priori blockmodelling is that often many models are fit before satisfactory covariates are determined. This suggests the problem of multiple comparisons where the final model may only include covariates that fit the dataset in question and may not provide adequate fit to the population of interest.

In a posteriori blockmodelling, estimates of the subject parameters α_i and β_i are obtained, and then standard clustering methods are applied to the estimates with the intention of grouping individuals. A posteriori blockmodelling strikes us as somewhat of an ad hoc procedure. We prefer a principled Bayesian approach where the individuals are clustered as a by-product of the DP mixture model.

Figure 1 provides a plot depicting the relationship between providing advice and receiving advice. For each of the 71 lawyers, out-degree (number of individuals to whom advice was given) is plotted against in-degree (number of individuals from whom advice was received). As expected, we observe that the younger associates generally give less advice than they receive. For example, one associate gave advice to only two colleagues yet received advice from 26 different colleagues. However, we notice that there are ex-

ceptions to the general heuristics. For example, there is a partner who gave advice to 11 colleagues yet received advice from 30 colleagues.

We fit the Bayesian DP model to the lawyer data and consider the clustering of (α_i, β_i) amongst the 71 lawyers. In a single iteration of MCMC, lawyers are clustered according to whether their (α_i, β_i) values are the same. In subsequent iterations of MCMC, the cluster membership may differ. With the MCMC output, we are able to calculate the proportion of iterations that any given pair of lawyers cluster together and this provides an estimate of the posterior pairwise probability of clustering. We contrast this feature with a posteriori blockmodelling where clustering is based on a deterministic algorithm and there is no probability measure associated with resultant clusters. In Figure 2, we provide a plot which highlights the pairwise clustering involved in the DP analysis. For every pair of lawyers, a black square represents the posterior probability of clustering using a threshold value of 0.5. An interesting observation from Figure 2 is that the grid is roughly divided into four quadrants. It appears that partners (the top left quadrant) tend to cluster together and that associates (the bottom right quadrant) tend to cluster together. In other words, partners tend to behave similarly and associates tend to behave similarly. What this suggests is that the original intuition of three groups of lawyers is not quite right, and this argues again for the DP approach. In the DP approach, the data determine the clusters. In a priori blockmodelling, one may fail to find suitable covariates to improve model fit.

3.3 Example 3: A Social Relations Model

We consider a simplification of the social relations model considered by Gill and Swartz (2007). The model involves paired continuous observations y_{ijk} and y_{jik} where y_{ijk} represents the k -th response of subject i as an actor towards subject j as a partner,

$k = 1, \dots, n_{ij}$, $i \neq j$. In y_{jik} , the roles are reversed. We let n denote the number of subjects. The model expresses the paired responses in an additive fashion

$$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$$

$$y_{jik} = \mu + \alpha_j + \beta_i + \varepsilon_{jik}$$

where μ is the overall mean, α_i is the effect of subject i as an actor, β_j is the effect of subject j as a partner and ε_{ijk} is the error term. We refer to μ , the α 's and the β 's as first-order parameters. The Bayesian model specification then assigns prior distributions

$$\mu \sim \text{Normal}(\theta_\mu, \sigma_\mu^2), \quad (11)$$

$$(\alpha_i, \beta_i)' \stackrel{\text{iid}}{\sim} \text{Normal}_2(0, \Sigma_{\alpha\beta}), \quad (12)$$

$$(\varepsilon_{ijk}, \varepsilon_{jik})' \stackrel{\text{iid}}{\sim} \text{Normal}_2(0, \Sigma_\varepsilon) \quad (13)$$

where

$$\Sigma_{\alpha\beta} = \begin{pmatrix} \sigma_\alpha^2 & \rho_{\alpha\beta}\sigma_\alpha\sigma_\beta \\ \rho_{\alpha\beta}\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix}, \quad \Sigma_\varepsilon = \sigma_\varepsilon^2 \begin{pmatrix} 1 & \rho_{\varepsilon\varepsilon} \\ \rho_{\varepsilon\varepsilon} & 1 \end{pmatrix}.$$

The parameters $\{\sigma_\alpha, \sigma_\beta, \rho_{\alpha\beta}, \sigma_\varepsilon, \rho_{\varepsilon\varepsilon}\}$ are called the variance-covariance parameters (or variance components). Note that the joint distributions (12) and (13) induce a dependence structure amongst the observations y_{ijk} . The interpretation of the variance-covariance parameters is naturally problem specific. However, for the sake of illustration, suppose that the response y_{ijk} is the k -th measurement of how much subject i likes subject j . In this case, $\rho_{\alpha\beta}$ represents the correlation between α_i and β_i , and we would typically expect a positive value. That is, an individual's positive (negative) attitude towards

others is usually reciprocated. To complete the Bayesian model specification, hyperpriors are assigned as follows:

$$\begin{aligned}
\theta_\mu &\sim \text{Normal}(\theta_0, \sigma_{\theta_0}^2), & \sigma_\mu^{-2} &\sim \text{Gamma}(a_0, b_0), \\
\Sigma_{\alpha\beta}^{-1} &\sim \text{Wishart}_2((\nu_0 R_0)^{-1}, \nu_0), & & (14) \\
\sigma_\varepsilon^{-2} &\sim \text{Gamma}(c_0, d_0), & \rho_{\varepsilon\varepsilon} &\sim \text{Uniform}(-1.0, 1.0)
\end{aligned}$$

where $X \sim \text{Gamma}(a, b)$ implies $E(X) = a/b$ and hyperparameters subscripted with a 0 are set to give diffuse prior distributions (Gill and Swartz 2007).

We now consider a modification of the above social relations model where the prior assumptions (11) through (14) are maintained except that (12) is modified according to

$$\begin{aligned}
(\alpha_i, \beta_i)' &\stackrel{\text{iid}}{\sim} G \\
G &\sim \text{DP}(m, \text{Normal}_2(0, \Sigma_{\alpha\beta})) \\
m &\sim \text{Uniform}(0.4, 10.0).
\end{aligned} \tag{15}$$

Via the DP prior, we have weakened the parametric normality assumption concerning (α_i, β_i) and have also introduced the potential for clustering individuals according to (α_i, β_i) . In the context of interpersonal attraction, this is important as one can imagine four broad classifications of individuals:

- those who like others and are also liked
- those who like others and are disliked
- those who dislike others and are liked
- those who dislike others and are also disliked .

Whereas social relations models focus on the variance components which are characteristics of the population, the social relations model using the Dirichlet process also permits the investigation of individuals.

To demonstrate the approach, we consider a study of students who lived together in a residence hall at the University of Washington (Curry and Emerson 1970). Data were collected on $n = 48$ individuals and measured on occasions $k = 1, 2, 3, 4, 5$ according to their pairwise levels of attraction. There is a missing data aspect to the problem as measurements were only taken between pairs of 8 individuals in each of six dorm groups. Markov chain Monte Carlo simulations were carried out in WinBUGS using the original normal prior and the DP prior. We allow 5000 iterations for the sampler to converge and another 10000 iterations for sampling from the posterior. Convergence is checked visually and by using several starting points.

In Figure 3, we provide a plot of the posterior means of the 48 (α_i, β_i) pairs using the DP prior. We have also included the line $y = x$ for comparison purposes. Figure 3 suggests a tendency of individuals to cluster together with points scattered about the line $y = x$ corresponding to individuals who extend friendship to a similar extent that friendship is returned. The outlier in the bottom right corner corresponds to an individual who likes others but is disliked. The two clusters of points in the top left corner correspond to individuals who may be regarded as having false personalities; they do not generally like others although they convey signals that in turn cause them to be liked. For comparison, Figure 4 provides a plot of the posterior means of the 48 (α_i, β_i) pairs using the normal prior (12). We observe that the posterior inferences for the pairs (α, β) differ considerably from those obtained using the DP mixture model. To investigate the fit of the DP prior in this example, we calculate the log pseudo marginal likelihood (LPML) proposed by Gelfand, Dey and Chang (1992) as a model selection technique. Using the LPML, the

DP prior is preferred (LPML = -5016.9) over the normal prior (LPML = -5180.9).

To investigate the effect of the prior choice involving the concentration parameter m in (15), we consider various priors. For instance, let $m \sim \text{Gamma}(2.0, 0.1)$ which is greatly different from the $\text{Uniform}(0.4, 10.0)$ prior. In comparing these two priors, we find that the posterior distributions of m differs substantially with $E(m|y) = 9.3$ under the Gamma prior and $E(m|y) = 6.6$ under the Uniform prior. However, our applied focus does not concern m . When looking at the posterior distributions of the (α_i, β_i) pairs under the two priors, we see very little difference. This is comforting and provides us with a sense of prior robustness with respect to the concentration parameter m .

4 DISCUSSION

In this paper, we have considered the use of Dirichlet process priors for network problems. The relaxation of parametric assumptions and the ability to facilitate clustering are both seen as advantages in network analyses. Furthermore, the models that we have considered are easily implemented using WinBUGS software.

It is worth asking where DP priors can be reasonably employed in network models. There are many networks where data can be modelled using a random effects specification. When some of the random effects might possibly be the same, then it is good to have methodology to accommodate and identify this type of clustering, and DP mixture modelling accomplishes this goal. For example, in various disease transmission networks, it is useful to identify individuals who have high probabilities of transmission. By clustering these individuals, patterns of behaviour may be deduced and this may be useful in disease prevention. As another example, consider the complex network structures that can be studied between states or nations. These structures may involve trade, information

flow, immigration/tourism, military cooperation, etc. Here, it may be useful to cluster the states or nations so that ideological categorizations can be inferred. For example, it may be interesting to know which eastern countries (if any) are close ideologically to western countries.

There are a number of future directions for this line of research. We are interested in using the Dirichlet process in more complex network problems with more complex dyadic dependencies. We are also interested in the treatment of longitudinal data and dynamic data networks. The development of complementary software to handle the special features of Dirichlet modelling may also be of value. Like other packages such as CODA, we imagine software written in R that processes WinBUGS output.

We emphasize the simplicity with which WinBUGS code facilitates the implementation of DP process mixture models for the network problems described in this paper. The WinBUGS program for Example 3 is available from the fourth author's website at www.stat.sfu.ca/~tim. The program consists of roughly 40 lines of code.

5 REFERENCES

- Besag, J. (1974). "Spatial interaction and the statistical analysis of lattice systems", *Journal of the Royal Statistical Society, Series B*, 36, 192-236.
- Besag, J. (2001). "Markov chain Monte Carlo for statistical inference", Working Paper No. 9, Center for Statistics and the Social Sciences, University of Washington.
- Cortes, C., Pregibon, D. and Volinsky, C. (2003). "Computational methods for dynamic graphs", *Journal of Computational and Graphical Statistics*, 12, 950-970.
- Curry, T.J. and Emerson, R.M. (1970). "Balance theory: a theory of interpersonal attraction", *Sociometry*, 33, 216-238.

- Dey, D., Müller, P. and Sinha, D., editors (1998). *Practical Nonparametric and Semiparametric Bayesian Statistics*, Lecture Notes in Statistics, Vol. 133, New York: Springer-Verlag.
- Ferguson, T.S. (1974). “Prior distributions on spaces of probability measures”, *Annals of Statistics*, 2, 615-629.
- Frank, O. and Strauss, D. (1986). “Markov graphs”, *Journal of the American Statistical Association*, 81, 832-842.
- Geisser, S. (1980). “Discussion on Sampling and Bayes’ inference in scientific modelling and robustness by G.E.P. Box”, *Journal of the Royal Statistical Society, Series A*, 143, 416-417.
- Gelfand, A.E., Dey, D.K. and Chang, H. (1992). “Model determination using predictive distributions with implementation via sampling-based methods (with discussion)”, In *Bayesian Statistics 4* (editors J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith), 147-169. Oxford: Clarendon.
- Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (eds.) (1996). *Markov Chain Monte Carlo in Practice*, Chapman & Hall: London.
- Gill, P.S. and Swartz, T.B. (2004). “Bayesian analysis of directed graphs data with applications to social networks”, *Journal of the Royal Statistical Society, Series C*, 53, 249-260.
- Gill, P.S. and Swartz, T.B. (2007). “Bayesian analysis of dyadic data”, *American Journal of Mathematical and Management Sciences: Special Volume on Modern Advances in Bayesian Theory and Applications*, 27, 73-92.
- Handcock, M.S. (2003). “Assessing degeneracy in statistical models of social networks”, Working Paper No. 39, Center for Statistics and the Social Sciences, University of Washington.
- Handcock, M.S., Raftery, A.E. and Tantrum, J.M. (2007). “Model-based clustering for social networks”, *Journal of the Royal Statistical Society, Series A*, 170, 301-354.

- Hoff, P.D. (2005). “Bilinear mixed effects models for dyadic data”, *Journal of the American Statistical Association*, 100, 286-295.
- Hoff, P.D. (2009). “Multiplicative latent factor models for description and prediction of social networks”, *Computational and Mathematical Organization Theory*, To appear.
- Hoff, P.D., Raftery, A.E. and Handcock, M.S. (2002). “Latent space approaches to social network analysis”, *Journal of the American Statistical Association*, 97, 1090-1098.
- Holland, P.W. and Leinhardt, S. (1981). “An exponential family of probability distributions for directed graphs”, *Journal of the American Statistical Association*, 76, 33-65.
- Ishwaran, H. and Zarepour, M. (2002). “Dirichlet prior sieves in finite normal mixtures”, *Statistica Sinica*, 12, 941-963.
- Lazega, E. and Pattison, P.E. (1999). “Multiplexity, generalized exchange and cooperation in organizations: a case study”, *Social Networks*, 21, 67-90.
- Linkletter, C.D. (2007). “Spatial process models for social network analysis”, *PhD Thesis, Department of Statistics and Actuarial Science, Simon Fraser University*.
- Muliere, P. and Tardella, L. (1998). “Approximating distributions of random functionals of Ferguson-Dirichlet priors”, *Canadian Journal of Statistics*, 26, 283-297.
- Ohlssen, D., Sharples, L.D. and Spiegelhalter, D.J. (2007). “Flexible random-effects models using Bayesian semi-parametric models: applications to institutional comparisons”, *Statistics in Medicine*, 26, 2088-2112.
- Sethuraman, J. (1994). “A constructive definition of Dirichlet priors”, *Statistica Sinica*, 4, 639-650.
- Snijders T.A.B. and Kenny, D.A. (1999). “The social relations model for family data: a multilevel approach”, *Personal Relationships*, 6, 471-486.

- Spiegelhalter, D., Thomas, A. and Best, N. (2003). *WinBUGS (Version 1.4) User Manual*, Cambridge: MRC Biostatistics Unit.
- Thompson, S.K. (2006). “Adaptive web sampling”, *Biometrics*, 62, 1224-1234.
- Vance, E.A. (2008). “Statistical methods for dynamic network data”, *PhD Thesis, Department of Statistical Science, Duke University*.
- Ward, M.D. and Hoff, P.D. (2007). “Persistent patterns of international commerce”, *Journal of Peace Research*, 44, 157-175.
- Warner, R.M., Kenny, D.A. and Stoto M. (1979). “A new round robin analysis of variance for social interaction data”, *Journal of Personality and Social Psychology*, 37, 1742-1757.
- Wasserman, S. and Faust, K. (1994). *Social Network Analysis*, Cambridge: Cambridge University Press.
- Wasserman, S. and Pattison, P. (1996). “Logit models and logistic regression for social networks: an introduction to Markov graphs and p*”, *Psychometrika*, 61, 401-425.
- Wong, G.Y. (1982). “Round robin analyses of variance via maximum likelihood”, *Journal of the American Statistical Association*, 77, 714-724.
- Wong, G.Y. (1987). “Bayesian models for directed graphs”, *Journal of the American Statistical Association*, 82, 140-148.

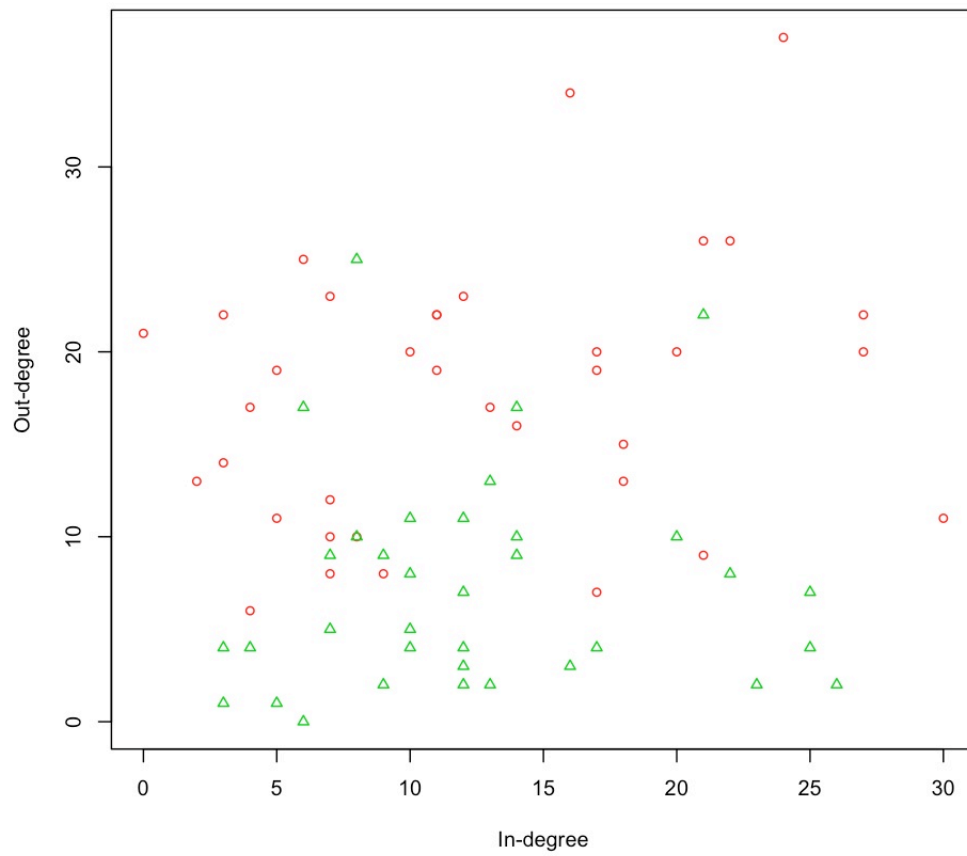


Figure 1: Plot of out-degree versus in-degree for the 71 lawyers in Example 2 where the lawyers labelled with triangles are associates and the lawyers labelled with circles are partners.

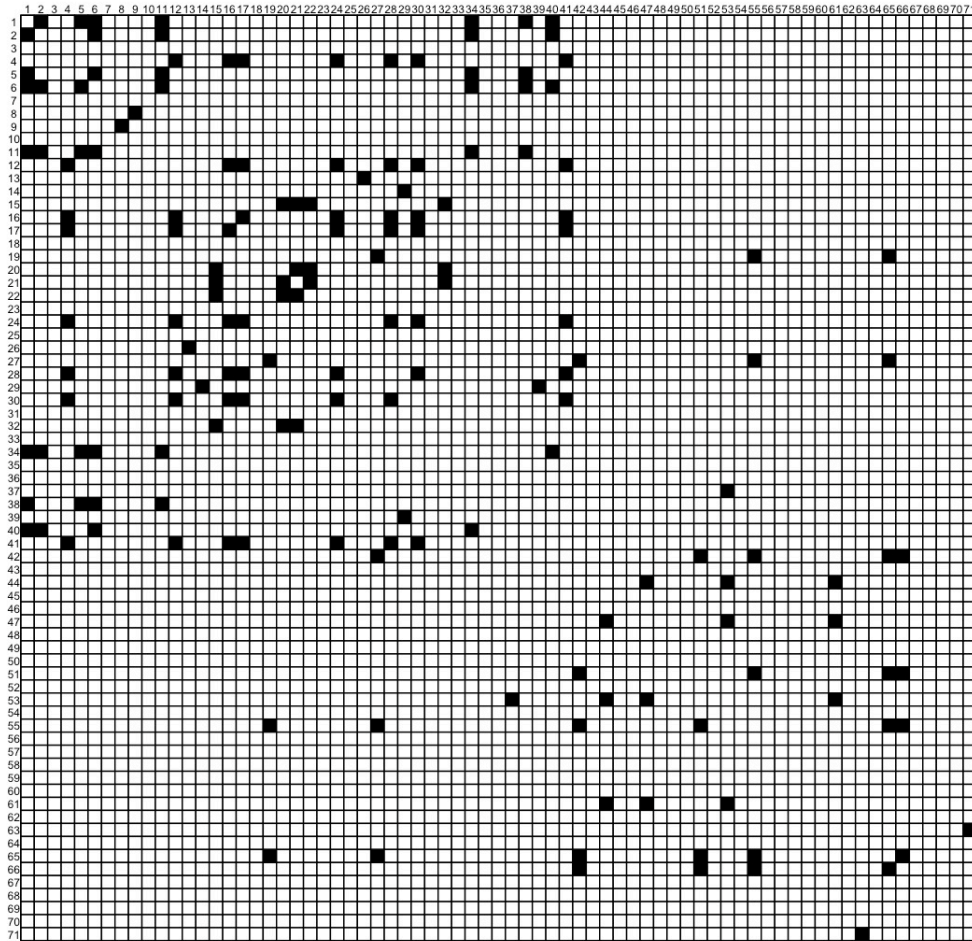


Figure 2: Plot of pairwise clustering of the 71 lawyers based on the DP model in Example 2. Black (white) squares indicate posterior probabilities of clustering greater than (less than) 0.5. Labels 1-36 correspond to partners and labels 37-71 correspond to associates.

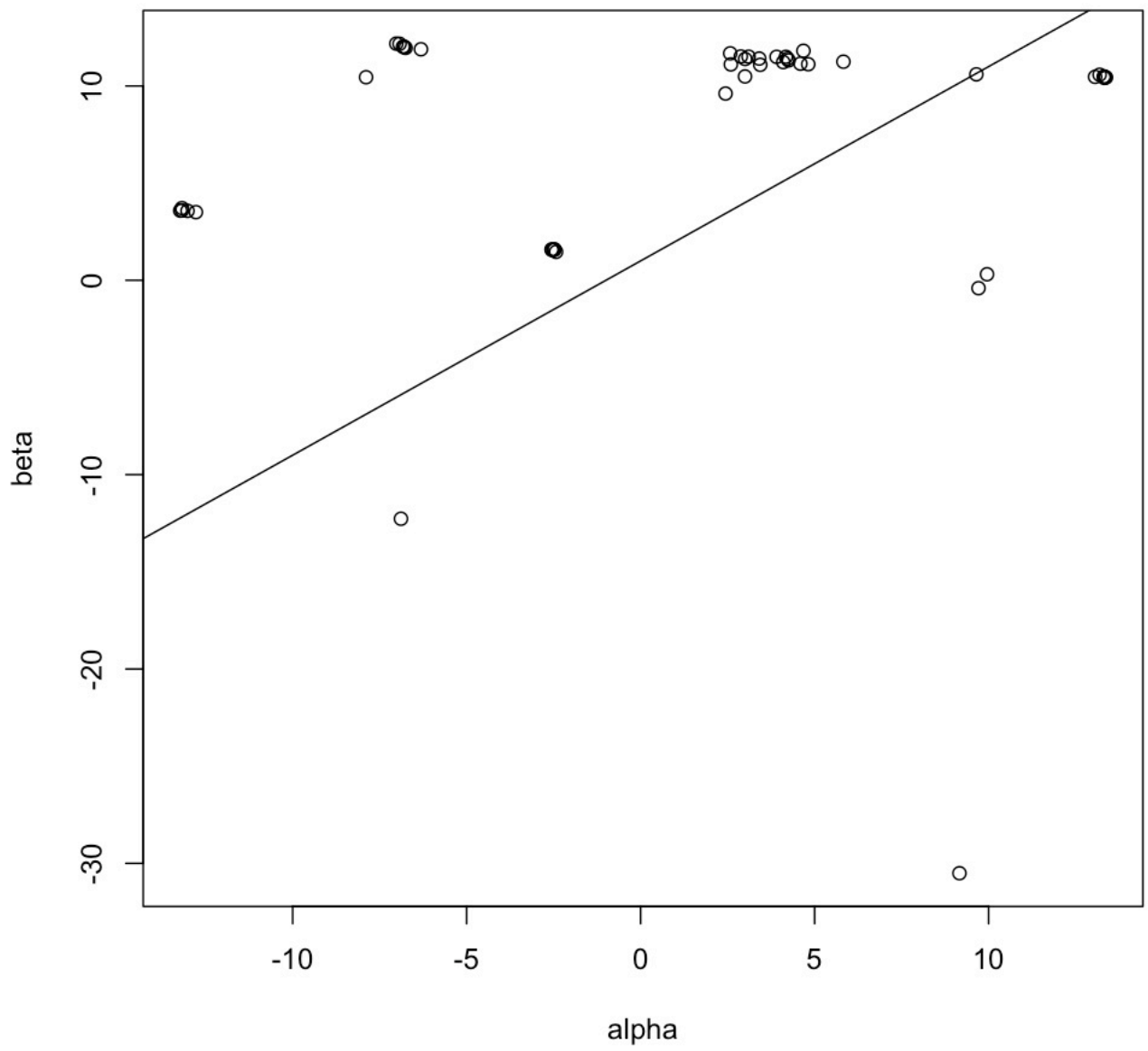


Figure 3: Posterior means of the (α_i, β_i) pairs under the DP prior in Example 3.

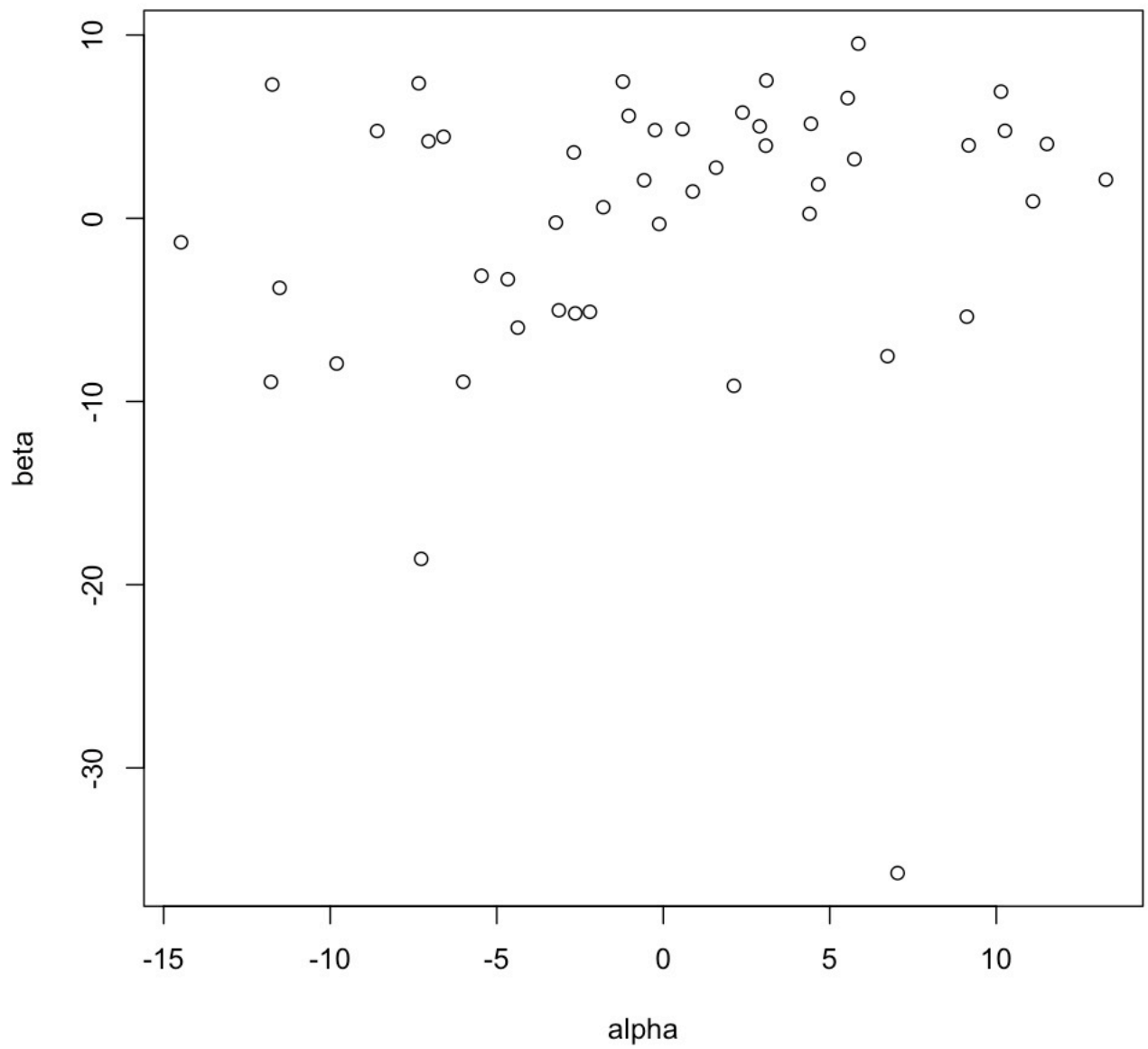


Figure 4: Posterior means of the (α_i, β_i) pairs under the normal prior in Example 3.