Supporting Information for “Interlocking directorates in Irish companies using bipartite networks: a latent space approach”

Nial Friel, Riccardo Rastelli, Jason Wyse and Adrian E. Raftery

April 12, 2016

Here we give details of the Bayesian method used to estimate the bipartite latent space social network model defined in the main text. We describe the Markov chain Monte Carlo (MCMC) algorithm used to estimate the parameters. Then, we address the issue that the likelihood is invariant to rotations, reflections or translations of all the latent positions, illustrating a procedure to solve this problem.

1 Markov chain Monte Carlo estimation

1.1 Metropolis within Gibbs updates

Markov chain Monte Carlo (MCMC) sampling was used to find the joint posterior distribution of all model parameters. To do so a standard Metropolis-within-Gibbs sampler was used by iteratively sampling each model parameter from its full conditional distribution given all the other parameters.

The director and board latent positions as well as the intercepts $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ were updated using random walk Metropolis within Gibbs updates. As an example, for a variable θ being updated, we propose a new value θ' drawn from a Gaussian density centered at θ, and accept θ' as the next value of θ in the Markov chain with probability

$$\alpha(\theta, \theta') = \min\left\{1, \frac{\pi(\theta'|\dots)}{\pi(\theta|\dots)}\right\}.$$

The proposal distribution is a D-variate Gaussian for the latent positions of boards and directors, and a univariate Gaussian for the intercept parameters. The variances of such proposal distributions are tuned so that acceptance rates are in the range 30–35%.

1.2 Updating scheme

Initialise all of the parameters to random feasible values. Then, for every iteration:

- For $i = 1, \dots, N$ update the position of director i using Metropolis-Hastings.

- For $t = 1, \dots, T$ and for $j = 1, \dots, M$ update the position of board j at time t using Metropolis-Hastings.
- For $t = 1, \dots, T$ update γ_t using Metropolis-Hastings.
- For $t = 1, \dots, T$ update β_t using Metropolis-Hastings.
- Update $\tau_w, \tau_w^0, \tau_\gamma, \tau_\gamma^0, \tau_\beta, \tau_\beta^0$ using conjugate Gamma full conditional updates.

The results in the article are based on an MCMC run of 1 million iterations with the first $500,000$ discarded as burn in, taking every 50^{th} iteration thereafter.

1.3 Posterior and full conditional distributions

The joint posterior distribution of the parameters can be written as:

$$\begin{aligned}\pi\left(\mathcal{X}, \mathcal{W}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \tau_w, \tau_w^0, \tau_\gamma, \tau_\gamma^0, \tau_\beta, \tau_\beta^0 \middle| \mathcal{Y}\right) \propto\ & P\left(\mathcal{Y}|\mathcal{X}, \mathcal{W}, \boldsymbol{\gamma}, \boldsymbol{\beta}\right) \times \\ & \pi\left(\mathcal{X}\right) \times \\ & \pi\left(\mathcal{W}\middle|\tau_w, \tau_w^0\right) \pi\left(\tau_w\right) \pi\left(\tau_w^0\right) \times \\ & \pi\left(\boldsymbol{\gamma}\middle|\tau_\gamma, \tau_\gamma^0\right) \pi\left(\tau_\gamma\right) \pi\left(\tau_\gamma^0\right) \times \\ & \pi\left(\boldsymbol{\beta}\middle|\tau_\beta, \tau_\beta^0\right) \pi\left(\tau_\beta\right) \pi\left(\tau_\beta^0\right).\end{aligned} \tag{1}$$

The full conditional distributions follow. As in the main text, $f\left(s; \mu, \nu\right)$ denotes a Gaussian density evaluated at s and with mean μ and variance ν, and $g\left(s; u, v\right)$ denotes a Gamma density evaluated at s with shape u and rate v. Also, $\mathbb{1}_{\mathcal{A}}$ denotes the indicator function for the event $\mathcal{A}$, i.e. its value is 1 if $\mathcal{A}$ is true or 0 otherwise. For board j and for every $j = 1, \dots, M$: $\mathcal{T}_j^{in}$ denotes the time index of first appearance of board j on the ISE, whereas $\mathcal{T}_j^{out}$ denotes the index of the last appearance.

For any $i = 1, \dots, N$:

$$\pi\left(\mathbf{x}_i|\dots\right) \propto \left[\prod_{d=1}^{D} f\left(x_{id}; 0, \frac{1}{\tau_x}\right)\right] \prod_{j=1}^{M} \prod_{t=\mathcal{T}_j^{in}}^{\mathcal{T}_j^{out}} \left[p_{ij}^{(t)}\right]^{\mathbb{1}_{\{i \in \mathcal{D}_t\}}}. \tag{2}$$

For any $j = 1, \dots, M$:

$$\pi\left(\mathbf{w}_j^{(1)}\middle|\dots\right) \propto \left[\prod_{d=1}^{D} f\left(w_{jd}^{(1)}; 0, \frac{1}{\tau_w^0}\right) f\left(w_{jd}^{(2)}; w_{jd}^{(1)}, \frac{1}{\tau_w}\right)\right] \left[\prod_{i \in \mathcal{D}_1} p_{ij}^{(1)}\right]^{\mathbb{1}\left\{\mathcal{T}_j^{in}=1\right\}}; \tag{3}$$

$$\pi\left(\mathbf{w}_j^{(T)}\middle|\dots\right) \propto \left[\prod_{d=1}^{D} f\left(w_{jd}^{(T)}; w_{jd}^{(T-1)}, \frac{1}{\tau_w}\right)\right] \left[\prod_{i \in \mathcal{D}_T} p_{ij}^{(T)}\right]^{\mathbb{1}\left\{\mathcal{T}_j^{out}=T\right\}}. \tag{4}$$

For any $t = 2, \dots, T-1$ and $j = 1, \dots, M$:

$$\pi\left(\mathbf{w}_j^{(t)}\middle|\dots\right) \propto \left[\prod_{d=1}^{D} f\left(w_{jd}^{(t)}; w_{jd}^{(t-1)}, \frac{1}{\tau_w}\right) f\left(w_{jd}^{(t+1)}; w_{jd}^{(t)}, \frac{1}{\tau_w}\right)\right] \left[\prod_{i \in \mathcal{D}_t} p_{ij}^{(t)}\right]^{\mathbb{1}\left\{t \geq \mathcal{T}_j^{in}\right\}\mathbb{1}\left\{t \leq \mathcal{T}_j^{out}\right\}}. \tag{5}$$

For any $t = 1, \ldots, T$:

$$\pi\left(\gamma^{(t)}|\ldots\right) \propto f\left(\gamma^{(t)}; 0, \frac{1}{\tau_\gamma^0}\right)^{\mathbb{1}_{\{t=1\}}} f\left(\gamma^{(t)}; \gamma^{(t-1)}, \frac{1}{\tau_\gamma}\right)^{\mathbb{1}_{\{t>1\}}} f\left(\gamma^{(t+1)}; \gamma^{(t)}, \frac{1}{\tau_\gamma}\right)^{\mathbb{1}_{\{t<T\}}} \prod_{j\in\mathcal{B}_t}\prod_{i\in\mathcal{D}_t} p_{ij}^{(t)}; \tag{6}$$

$$\pi\left(\beta^{(t)}|\ldots\right) \propto f\left(\beta^{(t)}; 0, \frac{1}{\tau_\beta^0}\right)^{\mathbb{1}_{\{t=1\}}} f\left(\beta^{(t)}; \beta^{(t-1)}, \frac{1}{\tau_\beta}\right)^{\mathbb{1}_{\{t>1\}}} f\left(\beta^{(t+1)}; \beta^{(t)}, \frac{1}{\tau_\beta}\right)^{\mathbb{1}_{\{t<T\}}} \prod_{j\in\mathcal{B}_t}\prod_{i\in\mathcal{D}_t} p_{ij}^{(t)}. \tag{7}$$

The full conditional distributions of the precision parameters are as follows:

$$\pi\left(\tau_w|\ldots\right) \propto g\left(\tau_w;\ a + \frac{1}{2}MD(T-1),\ b + \frac{1}{2}\sum_{t=2}^{T}\sum_{j=1}^{M}\sum_{d=1}^{D}\left[w_{jd}^{(t)} - w_{jd}^{(t-1)}\right]^2\right); \tag{8}$$

$$\pi\left(\tau_w^0|\ldots\right) \propto g\left(\tau_w^0;\ a + \frac{1}{2}MD,\ b + \frac{1}{2}\sum_{j=1}^{M}\sum_{d=1}^{D}\left[w_{jd}^{(1)}\right]^2\right); \tag{9}$$

$$\pi\left(\tau_\gamma|\ldots\right) \propto g\left(\tau_\gamma;\ a + \frac{1}{2}(T-1),\ b + \frac{1}{2}\sum_{t=2}^{T}\left[\gamma^{(t)} - \gamma^{(t-1)}\right]^2\right); \tag{10}$$

$$\pi\left(\tau_\gamma^0|\ldots\right) \propto g\left(\tau_\gamma^0;\ a + \frac{1}{2},\ b + \frac{1}{2}\left[\gamma^{(1)}\right]^2\right); \tag{11}$$

$$\pi\left(\tau_\beta|\ldots\right) \propto g\left(\tau_\beta;\ a + \frac{1}{2}(T-1),\ b + \frac{1}{2}\sum_{t=2}^{T}\left[\beta^{(t)} - \beta^{(t-1)}\right]^2\right); \tag{12}$$

$$\pi\left(\tau_\beta^0|\ldots\right) \propto g\left(\tau_\beta^0;\ a + \frac{1}{2},\ b + \frac{1}{2}\left[\beta^{(1)}\right]^2\right). \tag{13}$$

2 Likelihood invariance and post processing of latent positions

The likelihood $\mathcal{L}_\mathcal{Y}\left(\mathcal{X}, \mathcal{W}, \boldsymbol{\gamma}, \boldsymbol{\beta}\right)$, depends on the latent positions only through the distances $||\mathbf{x}_i - \mathbf{w}_j^{(t)}||$ for all i, j, t. If we apply a rotation, reflection or translation to all positions, the value of the likelihood will not be affected because of this. We therefore postprocess the posterior samples of latent positions to account for the fact that rotation or translation may have occurred during the sampling process. We use Procrustes matching to correct the samples of latent positions, by finding a rotation matrix for each posterior sample which gives the best match to a reference sample.

The first part of our postprocessing requires a reference set of latent positions, to which others will be matched. As a reference set we take the latent positions belonging to the MCMC iterate which obtained the highest value of the full log posterior density. We then create a $(N + TM) \times D$ matrix containing all of the latent positions for that particular iteration, namely

$$\widehat{\mathbf{Z}} = \left[\widehat{\mathbf{x}}_1 \ldots \widehat{\mathbf{x}}_N\ \widehat{\mathbf{w}}_1^{(1)} \ldots \widehat{\mathbf{w}}_M^{(1)}\ \ldots\ \ldots\ \widehat{\mathbf{w}}_1^{(T)} \ldots \widehat{\mathbf{w}}_M^{(T)}\right].$$

We repeat the same operation with all the other iterations obtaining a sequence of matrices $\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}, \ldots, \mathbf{Z}^{(L)}$, where L is the number of iterations. Then we apply a Procrustes transformation to each of these matrices, using $\widehat{\mathbf{Z}}$ as a reference.

Each rotation matrix can be evaluated in the following way. We denote by $\mathbf{Z}_{ref}$ the reference matrix and by $\mathbf{Z}$ the matrix to be rotated. Then both matrices are centred and

$$\mathbf{X} = \left(\mathbf{Z}_{ref}^{centred}\right)' \mathbf{Z}^{centred}$$

is evaluated, with the dash denoting the transposed matrix. Let $\mathbf{X} = \mathbf{U}\mathbf{V}'$ be the singular value decomposition of $\mathbf{X}$. Then the rotation matrix is given by $\mathbf{V}\mathbf{U}'$. Once all the matrices have been rotated, the corresponding directors' and boards' positions are recovered from $\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(L)}$. This procedure guarantees that, within each iteration, all of the latent distances (and hence the likelihood values) are preserved.