

# A Bayesian Approach to the Problem of Unknown Networks in Spatial Autoregressive Models

陳捷

Jesse Chieh Chen<sup>\*</sup>

Draft: 2024-05-29

## Abstract

We introduced a Bayesian method for estimating unknown networks in the context of **Spatial Autoregressive Model** models by introducing a network formation model, **Exponential Random Graph Model** in this paper, as a prior distribution. This method brings two main advantages: first, this method is a natural extension to common high-dimensional methods; second, this method enables the econometrician to incorporate prior knowledge or belief about the network formation process with great flexibility. Via simulation studies, we demonstrated this approach is practical in recovering unknown networks ties and higher-order characteristics of the network.

## 1 Introduction

**Spatial Autoregressive Model (SAR)** is one of the main methods econometrician employ to study network effects. This approach requires the underlying network to be known to the researcher. However, it is usually difficult to obtain network data. For example, if we are interested in peer effect within a certain class, it is easy to know the pupils' individual characteristics (age, sex, GPA, etc), but difficult to know the friendship network. Hence it is natural to extend the canonical **SAR** model with known networks to the case where the network is unknown, but is *believed* to exist by the researcher.

The estimation of unknown network has been of interest for a long time. There are interests from various disciplines that seeks to tackle the problem of unknown network from different perspectives. In biology, the interest are mainly focused on retrieving the covariance between different covariates (see Yuan and Lin (2007)) or recovering the **Directed Acyclic Graph (DAG)** structure of covariates (see Madigan and Raftery (1994)). These methods usually apply to relatively small number of nodes or covariates (usually no larger than 10) and deal with structures with relatively low complexity.

Compared to the research interest from biology, social networks to contain more nodes and more complex structures. The interest among economists about social network mainly began with Ord (1975) and later with Manski (1993). Following it, many researchers have studied the problem of estimating the effect of social interactions and network formation (see Anselin (2009) for a survey of these results). These methods focuses on known networks. Recent literature seeks to tackle the problem

---

<sup>\*</sup>National Taiwan University. Contact: <https://jessekelighine.com>.

of unknown network. The problem is two-fold: the first is about identification, the second is about estimation. On the identification problem, most of the literature utilize some assumption about sparsity of the social network adjacency matrix (e.g. Rose (2017)); Blume et al. (2015) utilizes a *partially* observed network and assumes some individual are known not to connect. Other efforts such as Bonaldi, Hortaçsu, and Kastl (2015) and Manresa (2016) further assume that there is no *endogenous effect* (Manski, 1993).

Many of these assumptions also carry through to estimation, i.e., sparsity assumptions translates directly to penalized or regularized estimations. For example, Lam and P. C. Souza (2020) and Meinshausen and Bühlmann (2006) uses penalized methods to estimate unobserved networks. One recent working paper de Paula, Rasul, and P. Souza (2023) identifies the unknown network *without* sparsity assumptions but still utilizes adaptive elastic net (Camer and Zhang, 2014) in estimation. The successful use of high-dimensional methods naturally suggests a Bayesian approach.

Bayesian inference methods are present in the literature of social networks (e.g., Hsieh and Kippersluis (2018)) along with some software support (e.g., Caimo and Friel (2014)). The classical **SAR** model with known network can be estimated with Bayesian methods. However, in the fields of unknown networks, Bayesian treatment is largely absent. Recently, Krisztin and Piribauer (2022) also proposed a Bayesian treatment to the problem of unknown networks in **SAR** models. In their paper, they considered **Erdős-Rényi (ER)** type priors. This is, to our knowledge, the first attempt in recovering unknown networks in the context of **SAR** via a Bayesian methodology.

The approach presented by Krisztin and Piribauer (2022) is the Bayesian analogue to the frequentest high-dimensional method. It is well known that high-dimensional methods closely relates to Bayesian methods, and **ER** type priors relates directly to sparsity (as they mentioned themselves) and only incorporates “local” information (dyadic). However, Bayesian methods allows us to go beyond regularizations of simple forms, e.g., Ridge or Lasso, and form regularizations with economic meaning and microfoundation. Our approach goes beyond the sparsity assumption and motivate the Bayesian approach from a network formation point of view. This allows us to also incorporates “global” and “structural” information (e.g., mutuality, triangles) in the prior while preserving “sparsity.” Thus, this paper can be viewed as an extension of Krisztin and Piribauer (2022).

The rest of this paper is structured thus: in **section 2**, the motivation and model setup are given; in **section 3**, the full Bayesian models are specified; in **section 4**, the procedure for posterior sampling is presented; in **section 5**, simulations studies are performed; in **section 6**, a discussion and conclusion is given.

## 2 Model Setup

This research focuses on **SAR**. The canonical **SAR** assumes the form

$$\underset{(N \times 1)}{\mathbf{y}} = \lambda \underset{(N \times N)}{\bar{\mathbf{W}}} \mathbf{y} + \underset{(N \times K)}{\mathbf{X}_1} \underset{(K \times 1)}{\boldsymbol{\beta}_1} + \bar{\mathbf{W}} \mathbf{X}_2 \boldsymbol{\beta}_2 + \mathbf{u} \quad (1)$$

where  $\mathbf{y}$  is the vector of outcome variables (e.g., GPA of all students in a class),  $\bar{\mathbf{W}}$  is the row-normalized adjacency matrix representation of the underlying network  $\mathbf{W}$  (e.g., friendship network of the class),<sup>1</sup>  $\mathbf{X}_1, \mathbf{X}_2$  are the exogenous variables (e.g.,

<sup>1</sup>In many studies, the network matrix is *row-normalized*. (See, e.g., LeSage and Pace (2009)) That is, rows of the network matrix in the **SAR** sums to one. This is usually done for two purposes: to make the direct effect  $\lambda$  easier to interpret (as  $N \rightarrow \infty$ ) and to ensure asymptotic properties of the estimators. (See, e.g., Lee (2003)) We also incorporate this in our **SAR** model. Hence, we denote the row-normalized adjacency matrix by  $\bar{\mathbf{W}}$ , where as the binary adjacency matrix by  $\mathbf{W}$ .

race, gender, etc. of each student). The coefficients  $\lambda$ ,  $\beta_1$ , and  $\beta_2$  have different interpretations à la Manski (1993):  $\lambda$  represents the *peer effect* or *endogenous effect*;  $\beta_1$  represents the *exogenous effect*; and  $\beta_2$  represents the *correlated effect*.

Most of the studies about SAR assumes that the network  $\mathbf{W}$  is observed or at least partially observed. This assumption is problematic in two ways: practically, it is usually difficult to obtain network data  $\mathbf{W}$ ; theoretically, a social network  $\mathbf{W}$  is usually formed endogenously. Thus, methods such as the one provided by Goldsmith-Pinkham and Imbens (2013) are used to filter out endogenous network formation effects such as homophily. However, these methods are only useful if  $\mathbf{W}$  is (partially) observed. Furthermore, we wish to specify a network formation that is more general than homophily.<sup>2</sup>

In this paper,  $\mathbf{W}$  is assumed to be completely unobserved, but some of its characteristics are known *a priori* to the researcher. For example,  $\mathbf{W}$  is known to have a relatively high number of mutual links (mutuality), or  $\mathbf{W}$  is known to exhibit the characteristics of “small world.” (See, e.g., Watts and Strogatz (1998)) In order to incorporate these *a priori* characteristics in the model, it is only natural to place a prior distribution on  $\mathbf{W}$ , interpreted as the *network formation* probability. With the prior on  $\mathbf{W}$ , the “complete” probability of observing a data set not only comes from the probability induced by (1), but also from the prior  $\pi(\mathbf{W})$ , i.e.,

$$f(\mathbf{y} | \lambda, \mathbf{W}, \beta_1, \beta_2) \cdot \pi(\mathbf{W})$$

where the likelihood part, i.e.,  $\pi(\mathbf{y} | \lambda, \mathbf{W}, \beta_1, \beta_2)$ , is induced by (1).

This model is reminiscent of maximum likelihood or quasi-maximum likelihood approaches, which is widely used in the SAR literature,<sup>3</sup> but now extended to a more general framework that incorporates the network formation process via the Bayesian framework. Here the prior has significant economic meaning: The prior placed on  $\mathbf{W}$  is not merely a statistical tool used to facilitate the usage of Bayesian methods, it encodes the *network formation process*. The prior  $\pi(\mathbf{W})$  can assume the form of any game-theoretical model of the formation of  $\mathbf{W}$  that yields a probability measure on  $\mathcal{W} = \{0, 1\}^{N \times N}$ , the space of adjacency matrices. In this paper, we focus on a particular type of prior that is tremendously flexible: **Exponential Random Graph Model (ERGM)**.

## 2.1 ERGM Prior

**ERGM** is family of model that yields a probability distribution on adjacency matrices that belongs to the exponential family. Let there be  $N$  nodes (actors) and a set of edges (relations) described by an adjacency matrix in space  $\mathcal{W} = \{0, 1\}^{N \times N}$ . The **ERGM** model specifies that the probability of observing any  $\mathbf{W} \in \mathcal{W}$  be

$$\pi(\mathbf{W} | \boldsymbol{\theta}) = \frac{1}{z(\boldsymbol{\theta})} \exp(\mathbf{s}(\mathbf{W})^\top \boldsymbol{\theta}) \quad \text{where} \quad z(\boldsymbol{\theta}) = \sum_{\tilde{\mathbf{W}} \in \mathcal{W}} \exp(\mathbf{s}(\tilde{\mathbf{W}})^\top \boldsymbol{\theta})$$

<sup>2</sup>The homophily network formation model yields the following probability mass function on the space of all possible networks  $\mathbf{W}$ :

$$\pi(\mathbf{W} | \mathbf{X}) = \prod_{i \neq j; i, j \in \{1, \dots, N\}} p^{\mathbf{1}\{X_i = X_j\}} (1 - p)^{1 - \mathbf{1}\{X_i = X_j\}}$$

where  $\mathbf{1}$  denotes the indicator function and  $p$  is a parameter of link formation if individual  $i$  and  $j$  share some characteristics. It is clear that this network formation model only depends on dyadic, hence local, information.

<sup>3</sup>For theoretical and computational treatments, see, e.g., Ord (1975), Smirnov and Anselin (2001), and Lee (2004).

where  $\mathbf{s}(\mathbf{W})$  is a known vector of sufficient statistics,  $\boldsymbol{\theta}$  is a vector corresponding natural parameters, and  $z(\boldsymbol{\theta})$  is the normalizing constant. The **ERGM** prior is chosen in this study for two main reasons:

**Flexibility** **ERGM** is flexible in the sense that the vector of sufficient statistics can be quite arbitrary, depending on the relevant characteristics of the networks. Often these sufficient statistics contains so-called *configurations* — small structures into which nodes organize themselves, e.g., number of  $k$ -stars, cliques, cyclic triads, etc. Compared to network models such as the **ER** model (Bernoulli model), the **ERGM** model enables a rich specification of probabilities that can encompass both local and global properties of the network. These configurations are studied extensively in past literatures. <sup>4</sup>

**Tractability** Although **ERGM** assumes the attractive exponential family form, sampling **ERGM** or estimation parameters  $\boldsymbol{\theta}$  given  $\mathbf{W}$  has been somewhat difficult due to the high-dimensionality of  $\mathbf{W}$  and the intractable normalizing constant  $z(\boldsymbol{\theta})$ . These problems are tackled via various Monte Carlo methods, such as tie-no-tie proposal distributions (Hunter, Handcock, et al., 2008), **Maximum Pseudolikelihood (MPLE)** to approximate likelihood ratios with local information (Strauss and Ikeda, 1990), and **Markov Chain Maximum Likelihood (MC-MLE)** to approximate likelihood ratios with numerical-integration-like techniques (Geyer and Thompson, 1992). <sup>5</sup>

In a Bayesian setting, we will need to specify a prior distribution on  $\boldsymbol{\theta}$  and sample from its posterior distribution. However, since **ERGM** contains a normalizing constant  $z(\boldsymbol{\theta})$  of which the form is not known, the posterior of  $\boldsymbol{\theta}$  will also contain  $z(\boldsymbol{\theta})$ , an unknown function of  $\boldsymbol{\theta}$ . This problem, along with the fact that posterior  $\boldsymbol{\theta}$  can only be obtains via **Markov Chain Monte Carlo (MCMC)** methods, is referred to as *doubly intractable* (Caimo and Friel, 2011). This *doubly intractable* problem can be solved using a specialized **MCMC** method: **Exchange Algorithm (EX)**. The usage of this method is proposed by Caimo and Friel (2011). More details about this method can be found in section 4.

**Interpretation** **ERGM** is primarily a network formation model with good statistical properties, not necessary with profound microfoundation. However, even **ER** type network formation, which is a special case of **ERGM**, can be infused with microfoundation. A simple extension of **ER** with different linking probability for each dyad, i.e., homophily model, can also be motivated from some microfoundation. Moreover, as demonstrated by Mele (2017), a general **ERGM** can also have microfoundation. Although it is still difficult to interpret common terms in the **ERGM** model, e.g., **2-star**, **ctriad**, etc, this paper still demonstrates that one can go beyond “local” information, i.e., beyond **edges**, in specifying a prior distribution to incorporate *a priori* knowledge in the network formation.

### 3 Distribution Specifications

In this section, we provide the full specification of the model. Following (1) with  $\mathbf{u}$  following  $N$  dimensional multinomial Gaussian distribution with mean 0 and variance  $\sigma^2 \mathbf{I}$ , the distribution of  $\mathbf{y}$  also following multinomial Gaussian distribution with mean

$$(\mathbf{I} - \lambda \bar{\mathbf{W}})^{-1}(\mathbf{X}_1 \boldsymbol{\beta}_1 + \bar{\mathbf{W}} \mathbf{X}_2 \boldsymbol{\beta}_2)$$

<sup>4</sup>See Robins et al. (2007) for a general introduction to **ERGM** in the context of social networks; See the manual of Handcock et al. (2023) for commonly used configurations.

<sup>5</sup>These methods are implemented in the **ergm** package for R (Handcock et al., 2023).

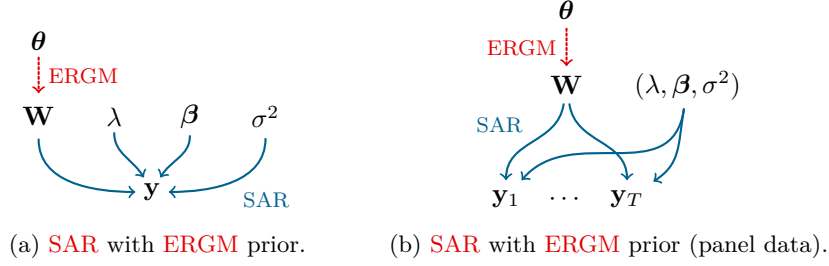


Figure 1: DAG Representations.

and variance

$$\sigma^2(\mathbf{I} - \lambda \bar{\mathbf{W}})^{-1}(\mathbf{I} - \lambda \bar{\mathbf{W}})^{-\top}.$$

The corresponding likelihood function is

$$f(\mathbf{y} | \lambda, \mathbf{W}, \beta, \sigma^2) = (2\pi)^{-N/2} (\sigma^2)^{-N/2} \det(\mathbf{M}) \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{M}\mathbf{y} - \mathbf{H}\beta\|_2^2\right) \quad (2)$$

where  $\mathbf{M} := \mathbf{I} - \lambda \bar{\mathbf{W}}$ ,  $\mathbf{H} := [\mathbf{X}_1, \bar{\mathbf{W}}\mathbf{X}_2]$ , and  $\|\mathbf{x}\|_2^2 = \mathbf{x}^\top \mathbf{x}$  denotes the square of Euclidean norm.

### 3.1 Priors: SAR

We now specify the simplest case where we assume that the parameter  $\theta$  for the ERGM probability  $\pi(\mathbf{W} | \theta)$  is known. For each parameter in the SAR model, we specify the following priors:

$$\begin{cases} \beta & \sim \mathcal{N}(\mu_\beta, \mathbf{V}_\beta), \\ \sigma^2 & \sim \mathcal{IG}(a_{\sigma^2}, b_{\sigma^2}), \\ \lambda & \sim \text{Unif}(-1, 1), \\ \pi(\mathbf{W} | \theta) & \propto \exp(\mathbf{s}(\mathbf{W})^\top \theta). \end{cases} \quad (3)$$

The specification of  $\beta$  and  $\sigma^2$  are standard conjugate priors; uniform prior for  $\lambda$  is standard in Bayesian estimations of the SAR model with known networks (LeSage and Pace, 2009); of course, we specify  $\mathbf{W}$  to follow ERGM. This set of priors are represented in a DAG in Figure 1a. This setup, apart from the prior  $\pi(\mathbf{W} | \theta)$  for the unknown network, is standard in SAR literature where Bayesian methods are employed.

### 3.2 Priors: ERGM

In practice, the researcher would not know *a priori* the value of  $\theta$ . Thus, we need to also specify a prior on the ERGM parameters. The standard choice would be Gaussian priors, same as the specifications in Caimo and Friel (2011):  $\theta \sim \mathcal{N}(\mu_\theta, \mathbf{V}_\theta)$ .

### 3.3 Panel Data

In practice, a single observation of  $\mathbf{y}$  and  $\mathbf{X}_1, \mathbf{X}_2$ 's without observing  $\mathbf{W}$  contains too little information about  $\mathbf{W}$  to be useful.<sup>6</sup> Thus, we introduce a time subscript

<sup>6</sup>Clearly, with only one observation, the parameters of interest are not identified in the frequentist sense. As mentioned in section 1, de Paula, Rasul, and P. Souza (2023) provided a frequentist

---

**Algorithm 1** Gibbs Sampling.

---

```
1: Initialize  $(\lambda, \mathbf{W}, \beta, \sigma^2, \theta)$ 
2: while true do
3:    $\beta' \leftarrow$  sample from  $\pi(\beta | \{\mathbf{y}_t\}, \lambda, \mathbf{W}, \sigma^2)$  exactly
4:    $\sigma^{2'} \leftarrow$  sample from  $\pi(\sigma^2 | \{\mathbf{y}_t\}, \lambda, \mathbf{W}, \beta')$  exactly
5:    $\lambda' \leftarrow$  sample from  $\pi(\lambda | \{\mathbf{y}_t\}, \mathbf{W}, \beta', \sigma^{2'})$  with Metropolis-Hastings Algorithm (M-H)
6:    $\mathbf{W}' \leftarrow$  sample from  $\pi(\mathbf{W} | \{\mathbf{y}_t\}, \lambda', \beta', \sigma^{2'}, \theta)$  with Hamiltonian Monte Carlo (HMC)
7:    $\theta' \leftarrow$  sample from  $\pi(\theta | \mathbf{W}')$  with EX
8:    $(\lambda, \mathbf{W}, \beta, \sigma^2, \theta) \leftarrow (\lambda', \mathbf{W}', \beta', \sigma^{2'}, \theta')$ 
9: end while
```

---

$t = 1, \dots, T$ , i.e., we repeatedly observe realizations of  $\mathbf{y}_t$ . Explicitly, we now have a subscripted version of (2) as

$$f(\mathbf{y}_t | \lambda, \mathbf{W}, \beta, \sigma^2) = (2\pi)^{-N/2} (\sigma^2)^{-N/2} \det(\mathbf{M}) \exp \left( -\frac{1}{2\sigma^2} \|\mathbf{M}\mathbf{y}_t - \mathbf{H}_t\beta\|_2^2 \right).$$

And the likelihood of observing the entire data is

$$f(\{\mathbf{y}_t\} | \lambda, \mathbf{W}, \beta, \sigma^2) = \prod_t f_t(\mathbf{y}_t | \lambda, \mathbf{W}, \beta, \sigma^2)$$

where the subscripted  $f_t$  simply denotes the dependency of  $\mathbf{X}_t$ . The **DAG** representation is shown in **Figure 1b**. The conditional posteriors are presented in **Appendix A**. In subsequent discussions, we will focus on this case of having panel data.

## 4 Sampling Procedure

To estimate our model, we employ the Gibbs sampling algorithm. In the following subsection, we describe the sampling strategy for each parameter.

### 4.1 Sample Posterior $\beta$ and $\sigma^2$

For  $\beta$  and  $\sigma^2$ , standard normal-normal and normal-inverse-gamma conjugacy results apply immediately. Their corresponding full conditional posterior densities are presented in **subsection A.2** and **subsection A.3** respectively.

### 4.2 Sample Posterior $\lambda$

The posterior of  $\lambda$  with a uniform prior assumes no known form, as shown in **subsection A.1**. Hence, **MCMC** techniques must be utilized. In LeSage and Pace (2009) and also Ritter and Tanner (1992), they recommended the usage of **Griddy-Gibbs Algorithm (G-G)** instead of **M-H** for efficiency. However, in our case, we found that the efficiency gain in using **G-G** is not significant. Furthermore, **G-G** tends to be more unstable compared to **M-H**, especially in the early phase of Gibbs sampling due to the unstableness of the likelihood function. Therefore, a simple **M-H** algorithm is adopted for sampling posterior  $\lambda$  in our case.

---

identification result of time-invariant unknown network in **SAR** with panel data. In a pure Bayesian sense, there is no problem of identification, i.e., identification is only a problem with the likelihood function. As long as a posterior distribution is defined, one can make valid Bayesian inference.

### 4.3 Sample Posterior $\mathbf{W}$

Technically, the posterior of  $\mathbf{W}$  also follows **ERGM**. However, its form is not standard, which inhibits us to use standard packages such as **ergm** in R. Thus, a custom sampler utilizing a discrete variation of **HMC** from Pakman and Paninski (2015) is adopted. More implementation details can be found in **Appendix B**.

### 4.4 Sample Posterior $\theta$

The method for sampling posterior  $\theta$ , the **EX** algorithm, is directly adopted from Caimo and Friel (2011). As mentioned in **subsection 2.1**, sampling posterior  $\theta$  is a *double intractable* problem, since the posterior density of  $\theta$  is

$$\pi(\theta | \mathbf{y}, \lambda, \mathbf{W}, \beta, \sigma^2) \propto \pi(\mathbf{W} | \theta) \pi(\theta) = \frac{1}{z(\theta)} \exp(\theta^\top \mathbf{s}(\mathbf{W})) \pi(\theta)$$

where  $z(\theta)$  is unknown to the researcher. The **EX** algorithm solves this problem by introducing an auxiliary state  $\tilde{\mathbf{W}}$ , which is sampled from **ERGM** with parameter  $\theta$ . This auxiliary state is used to approximate the unknown normalizing constant  $z(\theta)$ . Caimo and Friel (2011) further proposed the **Adaptive Directional Sampler (ADS)**, an extension of **EX**, for better acceptance rate. In our case, we found that the additional complexity of the **ADS** is not worth the marginally better acceptance rate. Thus, a basic **EX** algorithm is implemented.

\* \* \*

The full algorithm for posterior sampling is presented in **Algorithm 1**. One detail to note is the order in which each parameter is sampled, as different ordering will effect the convergence rate of the algorithm. It is preferable that one first samples from  $\beta$ , as a good starting  $\beta$  quickly in the moves the likelihood in correct direction. A bad choice would be to start with  $\lambda$  or  $\mathbf{W}$ , since bad choices of either of these would greatly effect the matrix  $(\mathbf{I} - \lambda \mathbf{W})^{-1}$ , making the likelihood unstable. Nevertheless, Gibbs sampling guarantees the convergence of the posterior samples, regardless of ordering. Further implementation details can be found at **Appendix C**.

## 5 Simulation Study

In this section, we present three simulation studies. The first study uses a artificial network generated directly from an **ERGM**; the second study uses the *Sampson's Monk* network (directed) with artificially generated data; the their study uses the *Zachary's Karate Club* network (undirected) also with artificially generated data.

### 5.1 Basic **ERGM**

In this first simulation study, we consider a data set that contains  $N = 15$  nodes and a panel length of  $T = 12$ .<sup>7</sup> The network, shown in **Figure 2**, is generated from an **ERGM** with sufficient statistics **edges** = -2 and **mutual** = 2. Two groups of simulations are run here:

Group P1:  $\mathbf{X}_1$  is an  $N \times 1$  matrix of exogenous variables generated from a normal distribution and  $\mathbf{X}_2$  equals  $\mathbf{X}_1$ .

---

<sup>7</sup>Note that with an entirely unknown network, it is impossible to recover  $\mathbf{W}$  with **Maximum Likelihood Estimator (MLE)** given this panel length. Hence, we do not intend to compare the efficiency of our results with **MLE** or frequentist methods.



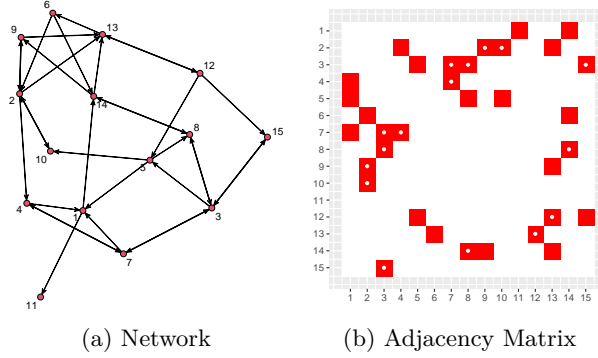


Figure 2: Simulated ERGM network for subsection 5.1.

Note: The white dots in the adjacency matrix denotes mutuality.

Group P2:  $\mathbf{X}_1$  is an  $N \times 2$  matrix of exogenous variables generated from a normal distribution.

That is, Group P1 is run *with* correlated effect while Group P2 *without*. The main focus of this first simulation study is on different prior specifications of  $\theta$  and model specification. The simulations results of the parameters are reported in Table 1 respectively; and the results of the adjacency matrices are shown in Figure 3 and Figure 4 respectively. The results are obtained by running Algorithm 1 for 3000 iterations and discarding the first 10% of iterations as burn-in.

**Posterior Inference on  $\lambda$ .** In all specifications presented in Table 1, no significant result is obtained for  $\lambda$ . This is true even for larger specification of  $\lambda$  (for  $\lambda = 0.15$  and  $\lambda = 0.2$ ).<sup>8</sup> In an SAR model (with Bayesian treatment) where the network is known, we essential make inference on the posterior distribution  $\pi(\lambda, \beta, \sigma^2 | \{\mathbf{y}_t\}, \mathbf{W})$ ; however, in our case, our inference is on the distribution  $\pi(\lambda, \mathbf{W}, \beta, \sigma^2 | \{\mathbf{y}_t\})$ . Whether or not  $\mathbf{W}$  is conditioned changes the distribution of  $\lambda$  substantially.

One the other hand, as presented in the lower half of Table 1 we obtain all significant results for  $\lambda$ . This is clearly related to the Data Generation Process (DGP). However, the exact nature of why this is unclear to the author.

**Posterior Inference on  $\mathbf{W}$ .** In Figure 3, a gradient between white and red is used to represent the posterior mean: the darker the color, the closer if the mean to 1; similarly, the lighter, the closer is the mean to 0.

Overall, the retrieval of matrix  $\mathbf{W}$  is surprisingly good in all specifications with only  $T = 12$ . As expected, the posterior mean is noisier if the prior belief is less defined. Nevertheless, in all specifications, most entries are completely white, i.e., the information from the likelihood function is sufficiently strong to rule out many improbable links.<sup>9</sup>

<sup>8</sup>The simulation results of  $\lambda = 0.15$  and  $\lambda = 0.2$  is not presented here, since the results are almost identical to Table 1, except for the posterior mean of  $\sigma^2$ . The posterior mean of  $\sigma^2$  with larger  $\lambda$  is larger than those presented in Table 1, since the unstable nature of  $(\mathbf{I} - \lambda \mathbf{W})^{-1}$  is reflected mostly in  $\sigma^2$ .

<sup>9</sup>In many applications, It might be necessary to obtain a binary adjacency matrix from the posterior samples, not just the posterior mean. From a purely decision theoretic point of view, we need to specify a specific loss function for this objective. One possible candidate is the zero-one loss, which corresponds to the Maximum A Posteriori (MAP) decision rule. However, the loss function would depend on the application in different scenarios. Hence, we simply present the posterior mean of  $\mathbf{W}$  as a convenient summary of the posterior distribution.



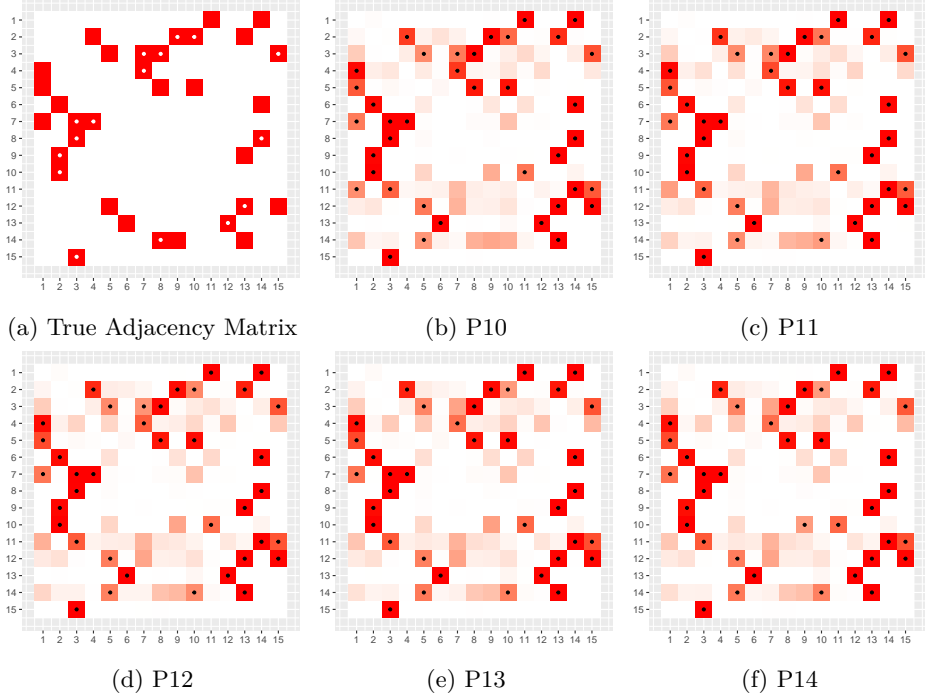


Figure 3: Posterior Adjacency Matrices.

*Note:* The black dots in the posterior adjacency matrices denotes the probability of entry being one exceeding 50%.

Compared to the results shown in Figure 3, it is clear that the posteriors in Figure 4 are a lot less defined. A heuristic explanation is that the term  $\mathbf{W}\mathbf{X}_2\beta_2$  provides a lot of information about  $\mathbf{W}$ , where as a lack of such term makes the exact entries of  $\mathbf{W}$  more ambiguous.

**Posterior Inference on  $\theta$ .** Different prior distributions of  $\theta$  leads to different result: the more specific and “correct” the prior is, the less noisy posterior mean of  $\mathbf{W}$  is.<sup>10</sup> The most important difference is whether each  $\theta$  obtains significant results. In all cases,  $\theta_{\text{edges}}$  is significant (95% credible interval), since *sparseness*

---

From a more practical perspective, one can simply categorise the posterior mean of  $\mathbf{W}$  by some threshold value and declare a certain link is retrieved if the posterior mean surpasses the value. In decision theoretic framework, we can obtain this decision rule by considering the following loss function

$$L(\hat{\mathbf{W}}, \mathbf{W}_0) = \#\{(i, j) : \hat{\mathbf{W}}_{i,j} \neq (\mathbf{W}_0)_{i,j}\},$$

that is, counting the total miss-categorized entries. It is straightforward to derive that the decision rule corresponding to this loss function should be

$$\hat{\mathbf{W}}_{i,j} = \mathbf{1} \{ \pi((\mathbf{W}_0)_{i,j} = 1 | \{\mathbf{y}_t\}) \geq 0.5 \},$$

that is, let the  $(i, j)$ -entry of the decision  $\hat{\mathbf{W}}$  be 1 if the posterior probability the entry being one is over 0.5. In the empirical illustration from Krisztin and Piribauer (2022), they adopted a similar strategy by categorizing the posterior mean of the adjacency matrix in three categories via cut-off values 0.5 and 0.75.

<sup>10</sup>In this case, there is a “correct” prior since the network is generated from a known ERGM with known parameters. Of course, there is no “correct” prior in the Bayesian framework. Hence, by “correct” we simply mean closer to the ERGM that generates this specific network.

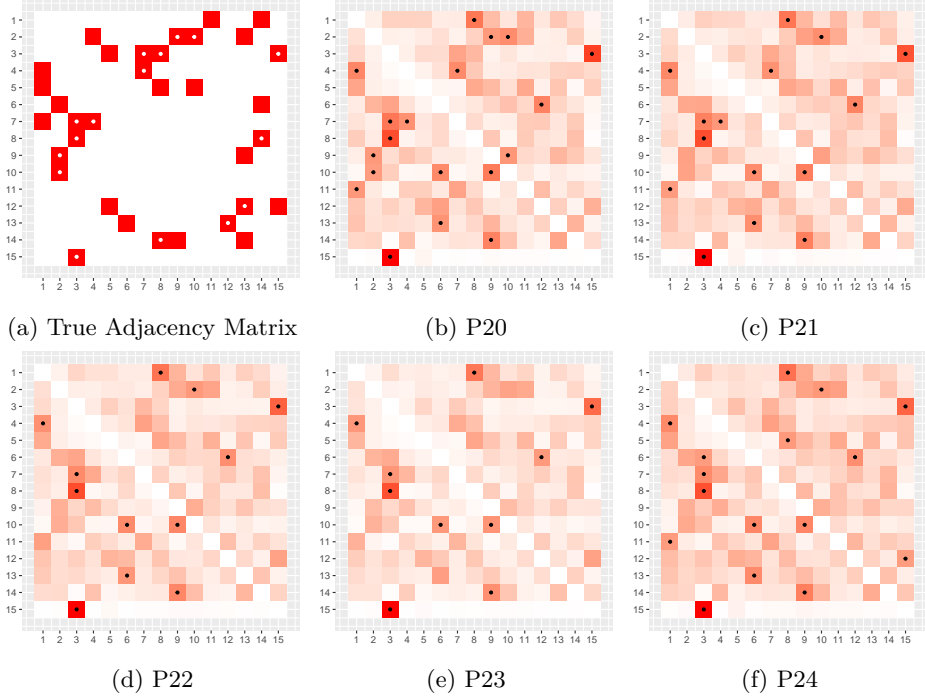


Figure 4: Posterior Adjacency Matrices.

is perhaps the most important aspect of this simulated matrix; on the other hand,  $\theta_{\text{mutual}}$  obtains significant result only in the case where the prior belief on it is very strong (first case in Table 1). Nevertheless, all prior specifications leads to roughly similar posterior inferences. This result is similar to the results reported in Krisztin and Piribauer (2022), as they also reported that the “sparsity” parameter can be effectively chosen by the model, rather than specifying it *a priori*.

\* \* \*

The results of this simulation is promising. However, as one can see in Figure 2, this simulated networks is not what we typically would observe in application. In the next two sections, we perform simulation on two well-known networks.

## 5.2 Sampson’s Monk: Directed Network

In this simulation, we consider Sampson’s monk network ( $N = 18$ ), as shown in Figure 5. A characteristic of the Sampson’s monk network is that it is a lot less sparse compared to the network used in subsection 5.1. This fact will be reflected in the simulation we see later. Similar to the setting in subsection 5.1, we investigate two groups of simulations:

Group S1:  $\mathbf{X}_1$  is an  $N \times 2$  matrix of exogenous variables with the first column generated from a normal distribution and the second column generated from a Bernoulli distribution. The second column is held constant across time to simulate characteristics that is time-invariant. Also, in this group  $\mathbf{X}_2$  is set to equal to  $\mathbf{X}_1$ . The coefficients corresponding to the exogenous effects are larger than the coefficients corresponding to the correlated effects.

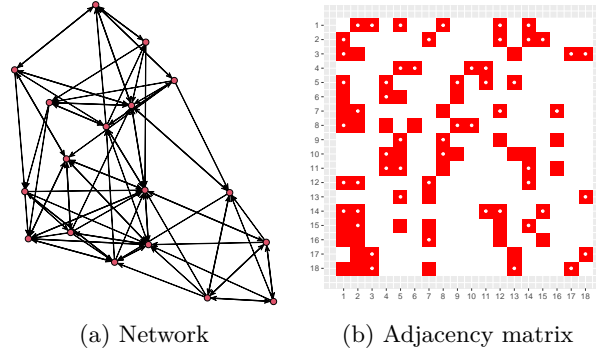


Figure 5: Sampson's Monk

Group S2: Similarly,  $\mathbf{X}_1$  and  $\mathbf{X}_2$  are  $N \times 2$  matrices of exogenous variables with the first column generated from a normal distribution and the second column generated from a Bernoulli distribution that is held constant across time. But in this group, some correlated effects are larger than some exogenous effects.

Three **ERGM** terms (**edges**, **mutual**, and **ctriad**) are incorporated in the estimation. The goal of this simulation group is to see how the relative size of the correlated and exogenous effects affect the estimation quality. We also vary the panel length observed by the econometrician in the simulations.

The results of **Group S1** and **Group S2** are presented in **Table 2** and **Figure 6**.<sup>11</sup> The main finding is intuitive and easily observed in **Figure 6**: the shorter the panel length, the less defined the posterior distributions, especially in the posterior distribution of the adjacency matrix. Also notice that the posterior mean of  $\mathbf{W}$  is somewhat less defined than the results obtained in **subsection 5.1**, this is perhaps due to the fact that the Sampson's monk network is more complex than the artificial network used in **subsection 5.1**.

Comparing **Group S1** and **Group S2**, it seems like **Group S2** performs better in terms of adjacency matrix recovery with panel length  $T = 18$  and  $T = 10$ . However, in the simulation with panel length  $T = 8$ , it seems like the posterior network is nearly degenerate, i.e., a plurality of the entries in the posterior adjacency matrix appear over 50%. The problem of degeneracy is characterized by that the sampled network adjacency matrix using **MCMC** techniques converges to a matrix with all entries being one. This problem is well-known in the literature of **ERGM**, see e.g., Handcock (2003) and Rinaldo, Fienberg, and Zhou (2009). In certain specifications with the Sampson's monk network, the sampling of posterior adjacency network tend to degenerate. However, in many cases when degeneracy occurs, we can normalize the posterior mean of the adjacency matrix and obtain somewhat reasonable results. This is perhaps due to the fact that this particular network is not very sparse and the panel length is too short, but still some information about the network is retained.

In terms of the other parameters, similar to the results in **subsection 5.1**, we cannot obtain significant result for  $\lambda$  in **Group S1** consistently. The exogenous and correlated effects  $\beta$  are estimated reasonably well, except for smaller effect sizes. This is a general problem across all econometric models, not a insufficiency of our

<sup>11</sup>The *true*  $\theta$  presented in **Table 2** is, of course, not literally *true*. It is the **ERGM** estimate of the parameters given the network.

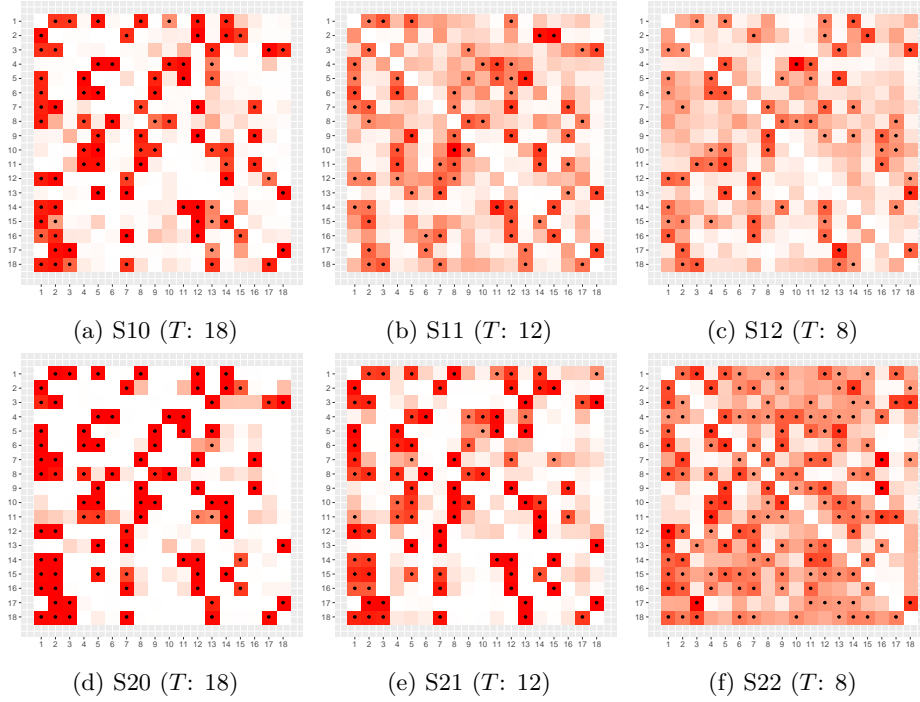


Figure 6: Posterior Adjacency Matrices.

approach. The estimations for  $\sigma^2$  worsen when panel length is reduced, which is also to be expected.

\* \* \*

Overall, we still obtain reasonable estimates for all specifications, even if the panel length is smaller than  $N$ .

### 5.3 Zachary's Karate Club: Undirected Network

In this last simulation, we consider Zachary's karate club network ( $N = 34$ ), as shown in [Figure 7](#). This network is considerably larger ( $34 \cdot (34 - 1)/2 = 561$  free elements in the adjacency matrix) than the previous cases. Furthermore, a symmetric adjacency matrix poses extra problems when it comes to sampling procedure, since specialised functions that calculates the likelihood ratio efficiently are needed.

One simulation group is presented here:

Group K1:  $\mathbf{X}_1$  is an  $N \times 2$  matrix of exogenous variables. The first column is generated from a normal distribution, the second column is generated from a Bernoulli distribution.  $\mathbf{X}_2$  is an  $N \times 1$  matrix identical to the first column of  $\mathbf{X}_1$ . The endogenous effect of this group is  $\lambda = 0.05$ .

Group K2: The exogenous variables  $\mathbf{X}_1$  and  $\mathbf{X}_2$  are generated with the same specifications as in [Group K1](#). The only difference is that the endogenous effect of this group is  $\lambda = 0.1$ .

Four [ERGM](#) terms (`edges`, `kstar(2)`, `kstar(3)`, and `triangle`) are incorporated in the estimation. We focus on how the size of the endogenous effect changes our posterior inference.

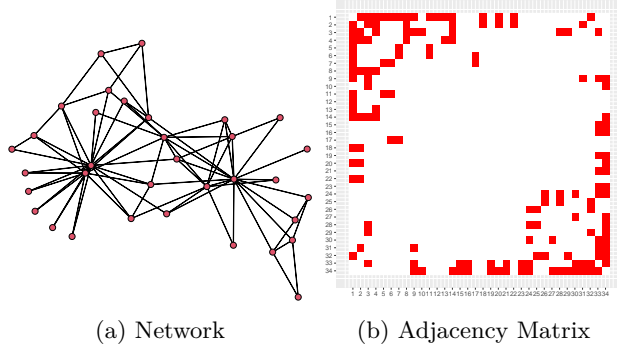


Figure 7: Zachary's Karate Club

The results are presented in Table 3 and Figure 8. The retrieval of  $\mathbf{W}$  is surprisingly good in both groups: even with only panel length of 10, one can clearly see that much of the network ties are correctly retrieved. Comparing Group K1 with Group K2, one can see that the posterior distributions of the parameters and adjacency matrices in Group K2 are less concentrated, leading to slightly larger estimation in  $\sigma^2$ , although the estimations of  $\sigma^2$  are poor in both groups. This results is expected, since a larger  $\lambda$  implies a more drastic change in the term  $(\mathbf{I} - \lambda\bar{\mathbf{W}})^{-1}$ , which contributes to the large  $\sigma^2$ . This is a well-known result in SAR studies.

On another note, considering only *sparsity* is arguably better than considering more ERGM terms in this case. That is, only include  $\theta_{\text{edges}}$  in the ERGM prior. This is probably due to the fact that the posterior  $\theta$  in the full specification are highly correlated in certain directions. It is clear that the posterior `kstar(2)` and `kstar(3)` estimates would be highly correlated. In fact, in Group K1, the Pearson correlation between `kstar(2)` and `kstar(3)` is about 0.8. Similar to the problem of multi-collinearity, this introduces a larger variation in the posterior distribution of other parameters, especially the posterior of adjacency matrix  $\mathbf{W}$ . In general, a more sophisticated prior specification does not necessary lead to better performance, especially with  $\mathbf{W}$ . Nevertheless, which prior network formation process to specify should depend on the research interest, not necessary on posterior performance.

## 6 Discussion

In this paper, we present a novel method to estimate SAR models that introduces a network formation process as a prior distortion on network  $\mathbf{W}$ . This approach combines two goals: first, this approach is a natural extension of well-known high-dimensional methods, making estimation of the high-dimensional parameter  $\mathbf{W}$  possible in situations frequentist methods fail; second, this approach infuses microfoundation in the prior, which can be interpreted as the econometricians prior belief in network formation process, and this makes the recovery of network character of interest logical and systematic. In the simulation studies performed above, we have shown that this approach works remarkably well in retrieving links in the unknown network, given the relatively short panel.

However, significant results for the direct effect  $\lambda$  cannot be reliably obtained. In traditional SAR estimation, we are essentially estimating the  $\pi(\lambda | \mathbf{W}, \{\mathbf{y}_t\})$ ; however, in our case of unknown network, we are estimating  $\pi(\lambda, \mathbf{W} | \{\mathbf{y}_t\})$ . That is, the probability distribution of  $\lambda$  is influenced largely by  $\mathbf{W}$ , meaning that small disturbances

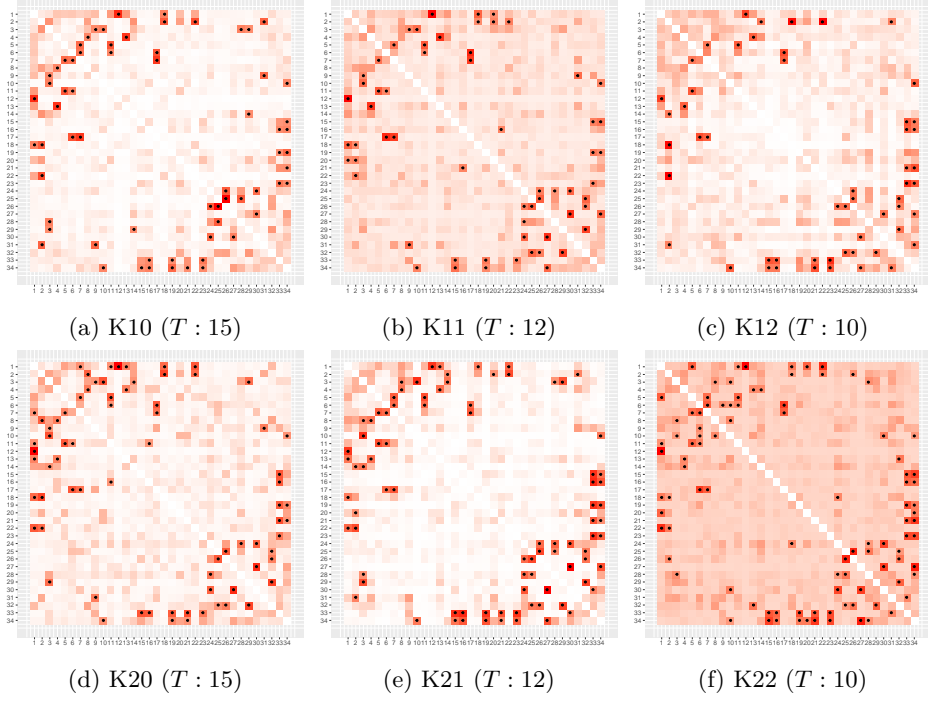


Figure 8: Posterior Adjacency Matrices.

of  $\mathbf{W}$ . This poses a problem not only for unknown network **SAR** in our case, but also questions the robustness of known network **SAR**'s. Even when the true  $\lambda$  is relatively large, we face the same problem as in known network **SAR** models: a trade-off between the estimation of  $\lambda$  and  $\sigma^2$ . With large  $\lambda$ , we often have poor estimation of  $\sigma^2$ , as the matrix  $(\mathbf{I} - \lambda\bar{\mathbf{W}})^{-1}$  becomes more unstable when  $\lambda$  is large.

Lastly, we mention two possible extensions of this Bayesian approach for future work: First, the convergence of the posterior distribution to the true parameter of our proposed estimator regarding **SAR** with unknown networks could be investigated. Second, although **ERGM** can be infused with microfoundation, it is not straightforward to interpret the coefficients  $\theta$  in search of microfoundation. One can extend the result of this paper to other probabilistic structure on social networks, e.g., any *strategic network formation* model that yields a probability on the space of adjacency matrices can be used. And since many strategic network formation models only depends on local information (dyadic information), the computation of posterior  $\mathbf{W}$  could be quite efficient. ■

Spec.	Param.	True	Prior Dist.	Post. Mean	95% Credible Interval	
					$Q_{2.5\%}$	$Q_{97.5\%}$
P10 $T = 12$	$\lambda$	0.1	Unif(-1, 1)	0.0028	-0.0088	0.0135
	$\sigma^2$	2	$\mathcal{IG}(1, 1)$	2.1017	1.6062	2.7360
	$\beta_1$	1	$\mathcal{N}(0, 30)$	1.0012	0.9848	1.0165
	$\beta_2$	0.5	$\mathcal{N}(0, 30)$	0.5715	0.5580	0.5847
	$\theta_{\text{edges}}$	-2	$\mathcal{N}(-2, 0.25)$	-2.1791	-2.7672	-1.6394
	$\theta_{\text{mutual}}$	2	$\mathcal{N}(2, 0.25)$	1.9970	1.1036	2.8564
P11 $T = 12$	$\lambda$	0.1	Unif(-1, 1)	0.0340	-0.0349	0.1004
	$\sigma^2$	2	$\mathcal{IG}(1, 1)$	4.5427	3.2691	6.2023
	$\beta_1$	1	$\mathcal{N}(0, 30)$	0.9946	0.9564	1.0339
	$\beta_2$	0.5	$\mathcal{N}(0, 30)$	0.5909	0.5348	0.6494
	$\theta_{\text{edges}}$	-2	$\mathcal{N}(-2, 1)$	-1.7810	-2.4095	-1.2490
	$\theta_{\text{mutual}}$	2	$\mathcal{N}(2, 1)$	1.0451	-0.1856	2.2240
P12 $T = 12$	$\lambda$	0.1	Unif(-1, 1)	0.0323	-0.0355	0.0964
	$\sigma^2$	2	$\mathcal{IG}(1, 1)$	4.5741	3.3396	6.2076
	$\beta_1$	1	$\mathcal{N}(0, 30)$	0.9944	0.9584	1.0334
	$\beta_2$	0.5	$\mathcal{N}(0, 30)$	0.5930	0.5386	0.6543
	$\theta_{\text{edges}}$	-2	$\mathcal{N}(0, 1)$	-1.5410	-2.0779	-1.0458
	$\theta_{\text{mutual}}$	2	$\mathcal{N}(0, 1)$	0.4882	-0.6913	1.6313
P13 $T = 12$	$\lambda$	0.1	Unif(-1, 1)	0.0333	-0.0311	0.0991
	$\sigma^2$	2	$\mathcal{IG}(1, 1)$	4.6069	3.3528	6.1911
	$\beta_1$	1	$\mathcal{N}(0, 30)$	0.9937	0.9572	1.0308
	$\beta_2$	0.5	$\mathcal{N}(0, 30)$	0.5929	0.5382	0.6487
	$\theta_{\text{edges}}$	-2	$\mathcal{N}(0, 1)$	-1.4252	-1.8560	-1.0364
P14 $T = 12$	$\lambda$	0.1	Unif(-1, 1)	0.0346	-0.0318	0.1028
	$\sigma^2$	2	$\mathcal{IG}(1, 1)$	4.5940	3.3532	6.1414
	$\beta_1$	1	$\mathcal{N}(0, 30)$	0.9928	0.9531	1.0317
	$\beta_2$	0.5	$\mathcal{N}(0, 30)$	0.5922	0.5343	0.6494
	$\theta_{\text{edges}}$	-2	$\mathcal{N}(0, 10)$	-1.4813	-1.9315	-1.0646
P20 $T = 12$	$\lambda$	0.1	Unif(-1, 1)	0.0996	0.0495	0.1504
	$\sigma^2$	2	$\mathcal{IG}(1, 1)$	2.1766	1.6172	2.9838
	$\beta_1$	1	$\mathcal{N}(0, 30)$	1.0013	0.9792	1.0227
	$\beta_2$	0.5	$\mathcal{N}(0, 30)$	0.1887	0.1727	0.2040
	$\theta_{\text{edges}}$	-2	$\mathcal{N}(-2, 0.25)$	-1.9680	-2.8537	-1.1959
	$\theta_{\text{mutual}}$	2	$\mathcal{N}(2, 0.25)$	1.5528	0.4588	2.5846
P21 $T = 12$	$\lambda$	0.1	Unif(-1, 1)	0.0996	0.0501	0.1505
	$\sigma^2$	2	$\mathcal{IG}(1, 1)$	2.2045	1.6301	3.0302
	$\beta_1$	1	$\mathcal{N}(0, 30)$	1.0017	0.9804	1.0234
	$\beta_2$	0.5	$\mathcal{N}(0, 30)$	0.1887	0.1727	0.2043
	$\theta_{\text{edges}}$	-2	$\mathcal{N}(-2, 1)$	-1.7415	-3.0813	-0.7993
	$\theta_{\text{mutual}}$	2	$\mathcal{N}(2, 1)$	0.9022	-0.7581	2.3364
P22 $T = 12$	$\lambda$	0.1	Unif(-1, 1)	0.0978	0.0470	0.1466
	$\sigma^2$	2	$\mathcal{IG}(1, 1)$	2.2190	1.6317	3.1347
	$\beta_1$	1	$\mathcal{N}(0, 30)$	1.0026	0.9795	1.0253
	$\beta_2$	0.5	$\mathcal{N}(0, 30)$	0.1891	0.1736	0.2049
	$\theta_{\text{edges}}$	-2	$\mathcal{N}(0, 1)$	-1.7877	-3.4018	-0.7241
	$\theta_{\text{mutual}}$	2	$\mathcal{N}(0, 1)$	0.2623	-1.3760	1.7652
P23 $T = 12$	$\lambda$	0.1	Unif(-1, 1)	0.0930	0.0454	0.1403
	$\sigma^2$	2	$\mathcal{IG}(1, 1)$	2.2601	1.6619	3.2626
	$\beta_1$	1	$\mathcal{N}(0, 30)$	1.0042	0.9813	1.0287
	$\beta_2$	0.5	$\mathcal{N}(0, 30)$	0.1894	0.1735	0.2056
	$\theta_{\text{edges}}$	-2	$\mathcal{N}(0, 1)$	-1.9940	-4.0807	-0.8732
P24 $T = 12$	$\lambda$	0.1	Unif(-1, 1)	0.1001	0.0534	0.1472
	$\sigma^2$	2	$\mathcal{IG}(1, 1)$	2.2279	1.6419	3.0942
	$\beta_1$	1	$\mathcal{N}(0, 30)$	1.0026	0.9812	1.0250
	$\beta_2$	0.5	$\mathcal{N}(0, 30)$	0.1884	0.1734	0.2043
	$\theta_{\text{edges}}$	-2	$\mathcal{N}(0, 10)$	-1.4521	-3.5933	-0.2893

Table 1: Posterior Statistics



Spec.	Param.	True	Prior Dist.	Post. Mean	95% Credible Interval	
					$Q_{2.5\%}$	$Q_{97.5\%}$
S10 $T = 18$	$\lambda$	0.1	$\text{Unif}(-1, 1)$	0.0320	-0.0857	0.1964
	$\sigma^2$	2	$\mathcal{IG}(1, 1)$	2.8549	1.9332	5.6295
	$\beta_{1,1}$	2	$\mathcal{N}(0, 30)$	2.0123	1.9648	2.0617
	$\beta_{1,2}$	2	$\mathcal{N}(0, 30)$	2.0984	1.1329	2.7002
	$\beta_{2,1}$	0.5	$\mathcal{N}(0, 30)$	0.7005	0.4930	0.8812
	$\beta_{2,2}$	0.5	$\mathcal{N}(0, 30)$	0.1332	-5.8634	3.3570
	$\theta_{\text{edges}}$	-1.8	$\mathcal{N}(-3, 1)$	-0.9080	-1.6285	-0.0488
	$\theta_{\text{mutual}}$	2.3	$\mathcal{N}(2, 1)$	1.2242	0.3165	2.1608
	$\theta_{\text{ctriad}}$	-0.1	$\mathcal{N}(-0.5, 1)$	-0.2491	-0.7051	0.1311
S11 $T = 12$	$\lambda$	0.1	$\text{Unif}(-1, 1)$	0.0306	-0.0906	0.1989
	$\sigma^2$	2	$\mathcal{IG}(1, 1)$	4.6437	1.7827	11.3241
	$\beta_{1,1}$	2	$\mathcal{N}(0, 30)$	2.0015	1.9525	2.0572
	$\beta_{1,2}$	2	$\mathcal{N}(0, 30)$	4.2848	1.2563	10.8403
	$\beta_{2,1}$	0.5	$\mathcal{N}(0, 30)$	0.6090	0.0223	0.9359
	$\beta_{2,2}$	0.5	$\mathcal{N}(0, 30)$	0.9627	-12.6832	10.0607
	$\theta_{\text{edges}}$	-1.8	$\mathcal{N}(-3, 1)$	-1.2569	-4.7459	0.3174
	$\theta_{\text{mutual}}$	2.3	$\mathcal{N}(2, 1)$	0.6115	-1.0884	1.8689
	$\theta_{\text{ctriad}}$	-0.1	$\mathcal{N}(-0.5, 1)$	-0.3708	-1.9857	0.2509
S12 $T = 8$	$\lambda$	0.1	$\text{Unif}(-1, 1)$	0.0327	-0.0749	0.1773
	$\sigma^2$	2	$\mathcal{IG}(1, 1)$	4.1386	1.9170	9.6382
	$\beta_{1,1}$	2	$\mathcal{N}(0, 30)$	1.9911	1.9232	2.0518
	$\beta_{1,2}$	2	$\mathcal{N}(0, 30)$	2.5672	0.8796	6.8146
	$\beta_{2,1}$	0.5	$\mathcal{N}(0, 30)$	0.6376	0.2932	0.8542
	$\beta_{2,2}$	0.5	$\mathcal{N}(0, 30)$	0.5532	-10.3836	7.4464
	$\theta_{\text{edges}}$	-1.8	$\mathcal{N}(-3, 1)$	-0.8364	-2.1052	0.2629
	$\theta_{\text{mutual}}$	2.3	$\mathcal{N}(2, 1)$	0.8155	-0.3281	2.0061
	$\theta_{\text{ctriad}}$	-0.1	$\mathcal{N}(-0.5, 1)$	-0.3789	-1.3651	0.1746
S20 $T = 18$	$\lambda$	0.1	$\text{Unif}(-1, 1)$	0.0933	-0.0515	0.2552
	$\sigma^2$	2	$\mathcal{IG}(1, 1)$	6.0217	3.3168	10.7843
	$\beta_{1,1}$	2	$\mathcal{N}(0, 30)$	1.9934	1.9176	2.0710
	$\beta_{1,2}$	0.5	$\mathcal{N}(0, 30)$	0.8951	-0.1525	1.8155
	$\beta_{2,1}$	2	$\mathcal{N}(0, 30)$	2.0583	1.5653	2.4986
	$\beta_{2,2}$	0.5	$\mathcal{N}(0, 30)$	-0.7854	-5.3961	3.1025
	$\theta_{\text{edges}}$	-1.8	$\mathcal{N}(-3, 1)$	-1.2451	-1.8671	-0.5679
	$\theta_{\text{mutual}}$	2.3	$\mathcal{N}(2, 1)$	1.0436	0.1724	1.9152
	$\theta_{\text{ctriad}}$	-0.1	$\mathcal{N}(-0.5, 1)$	0.0069	-0.3277	0.2604
S21 $T = 12$	$\lambda$	0.1	$\text{Unif}(-1, 1)$	0.0339	-0.1151	0.2276
	$\sigma^2$	2	$\mathcal{IG}(1, 1)$	6.1242	3.2807	13.9441
	$\beta_{1,1}$	2	$\mathcal{N}(0, 30)$	2.0333	1.8949	2.1770
	$\beta_{1,2}$	0.5	$\mathcal{N}(0, 30)$	0.9477	-0.1826	2.3512
	$\beta_{2,1}$	2	$\mathcal{N}(0, 30)$	2.2310	1.7853	2.5866
	$\beta_{2,2}$	0.5	$\mathcal{N}(0, 30)$	1.4372	-3.9601	5.0759
	$\theta_{\text{edges}}$	-1.8	$\mathcal{N}(-3, 1)$	-1.1667	-1.7794	-0.5208
	$\theta_{\text{mutual}}$	2.3	$\mathcal{N}(2, 1)$	0.8031	-0.0575	1.6794
	$\theta_{\text{ctriad}}$	-0.1	$\mathcal{N}(-0.5, 1)$	0.0580	-0.2533	0.2849
S22 $T = 8$	$\lambda$	0.1	$\text{Unif}(-1, 1)$	0.0775	-0.1160	0.3386
	$\sigma^2$	2	$\mathcal{IG}(1, 1)$	12.0367	6.1852	22.6555
	$\beta_{1,1}$	2	$\mathcal{N}(0, 30)$	2.0389	1.8837	2.2204
	$\beta_{1,2}$	0.5	$\mathcal{N}(0, 30)$	0.7332	-2.9808	5.7202
	$\beta_{2,1}$	2	$\mathcal{N}(0, 30)$	2.3419	1.6713	2.7921
	$\beta_{2,2}$	0.5	$\mathcal{N}(0, 30)$	-6.8160	-27.5867	11.7373
	$\theta_{\text{edges}}$	-1.8	$\mathcal{N}(-3, 1)$	-0.6298	-1.3749	0.0000
	$\theta_{\text{mutual}}$	2.3	$\mathcal{N}(2, 1)$	0.4093	-0.2683	1.2248
	$\theta_{\text{ctriad}}$	-0.1	$\mathcal{N}(-0.5, 1)$	0.0528	-0.1713	0.2361

Table 2: Posterior Statistics

Spec.	Param.	True	Prior Dist.	Post. Mean	95% Credible Interval	
					$Q_{2.5\%}$	$Q_{97.5\%}$
K10 $T = 15$	$\lambda$	0.05	$\text{Unif}(-1, 1)$	0.2761	-0.3237	0.9442
	$\sigma^2$	5	$\mathcal{IG}(1, 1)$	40.1132	8.4908	173.1849
	$\beta_{1,1}$	2	$\mathcal{N}(0, 30)$	2.0150	1.7000	2.2317
	$\beta_{1,2}$	1	$\mathcal{N}(0, 30)$	0.2775	-10.1314	8.5981
	$\beta_2$	0.5	$\mathcal{N}(0, 30)$	-0.1082	-1.7671	1.3477
	$\theta_{\text{edges}}$	-3	$\mathcal{N}(-3, 1)$	-2.3436	-3.3285	-1.1674
	$\theta_{2\text{-star}}$	0.2	$\mathcal{N}(0.2, 1)$	0.7999	0.1206	1.5804
	$\theta_{3\text{-star}}$	-0.1	$\mathcal{N}(-0.1, 1)$	-0.4000	-1.1722	-0.0340
	$\theta_{\text{triangle}}$	0.3	$\mathcal{N}(0.3, 1)$	0.2646	-0.8123	1.2890
K11 $T = 12$	$\lambda$	0.05	$\text{Unif}(-1, 1)$	0.0292	-0.3401	0.4165
	$\sigma^2$	5	$\mathcal{IG}(1, 1)$	37.3011	8.3389	149.4285
	$\beta_{1,1}$	2	$\mathcal{N}(0, 30)$	2.0682	1.8333	2.4211
	$\beta_{1,2}$	1	$\mathcal{N}(0, 30)$	1.0926	-4.0489	7.2829
	$\beta_2$	0.5	$\mathcal{N}(0, 30)$	0.4879	-0.4604	1.3605
	$\theta_{\text{edges}}$	-3	$\mathcal{N}(-3, 1)$	-2.4652	-3.6032	-1.2811
	$\theta_{2\text{-star}}$	0.2	$\mathcal{N}(0.2, 1)$	0.7811	-0.0684	1.5686
	$\theta_{3\text{-star}}$	-0.1	$\mathcal{N}(-0.1, 1)$	-0.2669	-1.0970	1.1224
	$\theta_{\text{triangle}}$	0.3	$\mathcal{N}(0.3, 1)$	0.2900	-0.8121	1.4029
K12 $T = 10$	$\lambda$	0.05	$\text{Unif}(-1, 1)$	-0.2268	-0.8450	0.3245
	$\sigma^2$	5	$\mathcal{IG}(1, 1)$	46.6107	8.3781	158.1703
	$\beta_{1,1}$	2	$\mathcal{N}(0, 30)$	2.0760	1.6324	2.6014
	$\beta_{1,2}$	1	$\mathcal{N}(0, 30)$	2.3291	-2.9775	10.1066
	$\beta_2$	0.5	$\mathcal{N}(0, 30)$	1.1367	-0.2382	2.9732
	$\theta_{\text{edges}}$	-3	$\mathcal{N}(-3, 1)$	-2.4565	-3.3930	-1.3636
	$\theta_{2\text{-star}}$	0.2	$\mathcal{N}(0.2, 1)$	0.7695	0.1184	1.4427
	$\theta_{3\text{-star}}$	-0.1	$\mathcal{N}(-0.1, 1)$	-0.3250	-1.0001	-0.0413
	$\theta_{\text{triangle}}$	0.3	$\mathcal{N}(0.3, 1)$	0.2568	-0.7657	1.2202
K20 $T = 15$	$\lambda$	0.1	$\text{Unif}(-1, 1)$	0.0527	-0.2202	0.3424
	$\sigma^2$	5	$\mathcal{IG}(1, 1)$	45.0312	8.3686	140.6322
	$\beta_{1,1}$	2	$\mathcal{N}(0, 30)$	1.9974	1.5802	2.5150
	$\beta_{1,2}$	1	$\mathcal{N}(0, 30)$	0.8264	-5.0927	8.6789
	$\beta_2$	0.5	$\mathcal{N}(0, 30)$	0.6450	0.1342	0.9109
	$\theta_{\text{edges}}$	-3	$\mathcal{N}(-3, 1)$	-2.4708	-3.4504	-1.2638
	$\theta_{2\text{-star}}$	0.2	$\mathcal{N}(0.2, 1)$	0.7538	0.1035	1.4728
	$\theta_{3\text{-star}}$	-0.1	$\mathcal{N}(-0.1, 1)$	-0.3072	-0.9901	-0.0301
	$\theta_{\text{triangle}}$	0.3	$\mathcal{N}(0.3, 1)$	0.2463	-0.6884	1.1994
K21 $T = 12$	$\lambda$	0.1	$\text{Unif}(-1, 1)$	0.0216	-0.3382	0.3432
	$\sigma^2$	5	$\mathcal{IG}(1, 1)$	43.4672	7.9732	178.8634
	$\beta_{1,1}$	2	$\mathcal{N}(0, 30)$	2.0573	1.6866	2.4233
	$\beta_{1,2}$	1	$\mathcal{N}(0, 30)$	1.0191	-3.1973	4.3400
	$\beta_2$	0.5	$\mathcal{N}(0, 30)$	0.6581	0.0032	1.3475
	$\theta_{\text{edges}}$	-3	$\mathcal{N}(-3, 1)$	-2.4819	-3.3739	-1.3568
	$\theta_{2\text{-star}}$	0.2	$\mathcal{N}(0.2, 1)$	0.7527	0.1402	1.5161
	$\theta_{3\text{-star}}$	-0.1	$\mathcal{N}(-0.1, 1)$	-0.2971	-0.9783	-0.0365
	$\theta_{\text{triangle}}$	0.3	$\mathcal{N}(0.3, 1)$	0.2581	-0.7139	1.2428
K22 $T = 10$	$\lambda$	0.1	$\text{Unif}(-1, 1)$	0.0178	-0.2256	0.3442
	$\sigma^2$	5	$\mathcal{IG}(1, 1)$	40.5321	9.3359	142.4679
	$\beta_{1,1}$	2	$\mathcal{N}(0, 30)$	2.0410	1.8122	2.2568
	$\beta_{1,2}$	1	$\mathcal{N}(0, 30)$	1.6281	-3.8057	8.4873
	$\beta_2$	0.5	$\mathcal{N}(0, 30)$	0.6790	-0.0492	1.2230
	$\theta_{\text{edges}}$	-3	$\mathcal{N}(-3, 1)$	-2.5016	-3.4187	-1.4754
	$\theta_{2\text{-star}}$	0.2	$\mathcal{N}(0.2, 1)$	0.7723	0.1618	1.4633
	$\theta_{3\text{-star}}$	-0.1	$\mathcal{N}(-0.1, 1)$	-0.2847	-0.8029	-0.0374
	$\theta_{\text{triangle}}$	0.3	$\mathcal{N}(0.3, 1)$	0.2489	-0.7229	1.1626

Table 3: Posterior Statistics

## Acronyms

<b>ADS</b>	Adaptive Directional Sampler. 7
<b>DAG</b>	Directed Acyclic Graph. 1, 5, 6
<b>DGP</b>	Data Generation Process. 8
<b>ER</b>	Erdős-Rényi. 2, 4
<b>ERGM</b>	Exponential Random Graph Model. 1, 3–5, 7–9, 11–14, 22, 23
<b>EX</b>	Exchange Algorithm. 4, 6, 7, 22
<b>G-G</b>	Griddy-Gibbs Algorithm. 6
<b>HMC</b>	Hamiltonian Monte Carlo. 6, 7, 22
<b>M-H</b>	Metropolis-Hastings Algorithm. 6
<b>MAP</b>	Maximum A Posteriori. 8
<b>MC-MLE</b>	Markov Chain Maximum Likelihood. 4
<b>MCMC</b>	Markov Chain Monte Carlo. 4, 6, 11, 22
<b>MLE</b>	Maximum Likelihood Estimator. 7
<b>MPLE</b>	Maximum Pseudolikelihood. 4
<b>SAR</b>	Spatial Autoregressive Model. 1–3, 5, 6, 8, 13, 14

## References

- Anselin, Luc (2009). “Thirty years of spatial econometrics”. In: *Papers in Regional Science* 89(1), pp. 3–25. URL: <https://doi.org/10.1111/j.1435-5957.2010.00279.x>.
- Blume, Lawrence E. et al. (2015). “Linear Social Interactions Models”. In: *Journal of Political Economy* 123(2), pp. 444–496. DOI: 10.1086/679496. eprint: <https://doi.org/10.1086/679496>. URL: <https://doi.org/10.1086/679496>.
- Bonaldi, Pietro, Ali Hortaçsu, and Jakub Kastl (Aug. 2015). *An Empirical Analysis of Funding Costs Spillovers in the EURO-zone with Application to Systemic Risk*. NBER Working Papers 21462. National Bureau of Economic Research, Inc. URL: <https://ideas.repec.org/p/nbr/nberwo/21462.html>.
- Caimo, Alberto and Nial Friel (2011). “Bayesian inference for exponential random graph models”. In: *Social Networks* 33(1), pp. 41–55. ISSN: 0378-8733. DOI: <https://doi.org/10.1016/j.socnet.2010.09.004>. URL: <https://www.sciencedirect.com/science/article/pii/S0378873310000493>.
- Caimo, Alberto and Nial Friel (2014). *Bergm: Bayesian Exponential Random Graphs in R*. arXiv: 1201.2770 [stat.CO].
- Caner, Mehmet and Hao Helen Zhang (2014). “Adaptive Elastic Net for Generalized Methods of Moments”. In: *Journal of Business & Economic Statistics* 32(1). PMID: 24570579, pp. 30–47. DOI: 10.1080/07350015.2013.836104. eprint: <https://doi.org/10.1080/07350015.2013.836104>. URL: <https://doi.org/10.1080/07350015.2013.836104>.
- de Paula, Áureo, Imran Rasul, and Pedro Souza (2023). *Identifying Network Ties from Panel Data: Theory and an Application to Tax Competition*. arXiv: 1910.07452 [econ.EM]. URL: <https://arxiv.org/abs/1910.07452>.
- Geyer, Charles J. and Elizabeth A. Thompson (1992). “Constrained Monte Carlo Maximum Likelihood for Dependent Data”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 54(3), pp. 657–699. ISSN: 00359246. URL: <http://www.jstor.org/stable/2345852> (visited on 01/06/2024).

- Goldsmith-Pinkham, Paul and Guido W. Imbens (2013). “Social Networks and the Identification of Peer Effects”. In: *Journal of Business & Economic Statistics* 31(3), pp. 253–264. DOI: [10.1080/07350015.2013.801251](https://doi.org/10.1080/07350015.2013.801251). eprint: <https://doi.org/10.1080/07350015.2013.801251>. URL: <https://doi.org/10.1080/07350015.2013.801251>.
- Handcock, Mark S. (2003). “Assessing Degeneracy in Statistical Models of Social Networks”. CSSS Working Paper no. 39.
- Handcock, Mark S. et al. (2023). *ergm: Fit, Simulate and Diagnose Exponential-Family Models for Networks*. R package version 4.5.0. The Statnet Project (<https://statnet.org>). URL: <https://CRAN.R-project.org/package=ergm>.
- Hsieh, Chih-Sheng and Hans van Kippersluis (2018). “Smoking initiation: Peers and personality”. In: *Quantitative Economics* 9(2), pp. 825–863. DOI: <https://doi.org/10.3982/QE615>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.3982/QE615>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.3982/QE615>.
- Hunter, David R., Steven M. Goodreau, and Mark S. Handcock (2013). “ergm.userterms: A Template Package for Extending statnet”. In: *Journal of Statistical Software* 52(2), pp. 1–25. DOI: [10.18637/jss.v052.i02](https://doi.org/10.18637/jss.v052.i02). URL: <https://www.jstatsoft.org/index.php/jss/article/view/v052i02>.
- Hunter, David R., Mark S. Handcock, et al. (2008). “ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks”. In: *Journal of Statistical Software* 24(3), pp. 1–29. DOI: [10.18637/jss.v024.i03](https://doi.org/10.18637/jss.v024.i03). URL: <https://www.jstatsoft.org/index.php/jss/article/view/v024i03>.
- Krisztin, Tamás and Philipp Piribauer (2022). “A Bayesian approach for estimation of weight matrices in spatial autoregressive models”. In: *Spatial Economic Analysis* 18 (1), pp. 44–63. DOI: <https://doi.org/10.1080/17421772.2022.2095426>.
- Lam, Clifford and Pedro C.L. Souza (2020). “Estimation and Selection of Spatial Weight Matrix in a Spatial Lag Model”. In: *Journal of Business & Economic Statistics* 38(3), pp. 693–710. DOI: [10.1080/07350015.2019.1569526](https://doi.org/10.1080/07350015.2019.1569526). eprint: <https://doi.org/10.1080/07350015.2019.1569526>. URL: <https://doi.org/10.1080/07350015.2019.1569526>.
- Lee, Lung-fei (2003). “Best Spatial Two-Stage Least Square Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances”. In: *Econometric Reviews* 22(4), pp. 307–335.
- Lee, Lung-fei (2004). “Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models”. In: *Econometrica* 72(6), pp. 1899–1925. URL: <http://www.jstor.org/stable/3598772>.
- LeSage, James and R. Kelley Pace (2009). *Introduction to Spatial Econometrics*. Taylor & Francis Group. ISBN: 978-1-4200-6424-7.
- Madigan, David and Adrian E. Raftery (1994). “Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam’s Window”. In: *Journal of the American Statistical Association* 89(428), pp. 1535–1546. ISSN: 01621459. URL: <http://www.jstor.org/stable/2291017> (visited on 06/08/2023).
- Manresa, Elena (2016). *Estimating the Structure of Social Interactions Using Panel Data*. Manuscript.
- Manski, Charles F. (July 1993). “Identification of Endogenous Social Effects: The Reflection Problem”. In: *The Review of Economic Studies* 60(3), pp. 531–542. ISSN: 0034-6527. DOI: [10.2307/2298123](https://doi.org/10.2307/2298123). eprint: <https://academic.oup.com/restud/article-pdf/60/3/531/4468725/60-3-531.pdf>. URL: <https://doi.org/10.2307/2298123>.

- Meinshausen, Nicolai and Peter Bühlmann (2006). “High-dimensional graphs and variable selection with the Lasso”. In: *The Annals of Statistics* 34(3), pp. 1436–1462. DOI: [10.1214/009053606000000281](https://doi.org/10.1214/009053606000000281). URL: <https://doi.org/10.1214/009053606000000281>.
- Mele, Angelo (2017). “A Structural Model of Dense Network Formation”. In: *Econometrica* 85(3), pp. 825–850. DOI: <https://doi.org/10.3982/ECTA10400>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA10400>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA10400>.
- Neal, Radford M. (2011). “MCMC using Hamiltonian dynamics”. In: *Handbook of Markov Chain Monte Carlo*. Ed. by Steve Brooks et al. Chapman and Hall/CRC. ISBN: 9780429138508. DOI: <https://doi.org/10.48550/arXiv.1206.1901>.
- Ord, Keith (1975). “Estimation Methods for Models of Spatial Interaction”. In: *Journal of the American Statistical Association* 70(349), pp. 120–126. URL: <https://doi.org/10.2307/2285387>.
- Pakman, Ari and Liam Paninski (2015). *Auxiliary-variable Exact Hamiltonian Monte Carlo Samplers for Binary Distributions*. arXiv: [1311.2166](https://arxiv.org/abs/1311.2166) [stat.CO].
- Rinaldo, Alessandro, Stephen E. Fienberg, and Yi Zhou (2009). “On the geometry of discrete exponential families with application to exponential random graph models”. In: *Electronic Journal of Statistics* 3(none), pp. 446–484. DOI: [10.1214/08-EJS350](https://doi.org/10.1214/08-EJS350). URL: <https://doi.org/10.1214/08-EJS350>.
- Ritter, Christian and Martin A. Tanner (1992). “Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler”. In: *Journal of the American Statistical Association* 87(419), pp. 861–868. ISSN: 01621459. URL: <http://www.jstor.org/stable/2290225> (visited on 01/07/2024).
- Robins, Garry et al. (2007). “An introduction to exponential random graph ( $p^*$ ) models for social networks”. In: *Social Networks* 29(2). Special Section: Advances in Exponential Random Graph ( $p^*$ ) Models, pp. 173–191. ISSN: 0378-8733. DOI: <https://doi.org/10.1016/j.socnet.2006.08.002>. URL: <https://www.sciencedirect.com/science/article/pii/S0378873306000372>.
- Rose, Christiern D. (Oct. 2017). “Identification of peer effects through social networks using variance restrictions”. In: *The Econometrics Journal* 20(3), S47–S60. ISSN: 1368-4221. DOI: [10.1111/ectj.12101](https://doi.org/10.1111/ectj.12101). eprint: <https://academic.oup.com/ectj/article-pdf/20/3/S47/28311585/ectj0s47.pdf>. URL: <https://doi.org/10.1111/ectj.12101>.
- Smirnov, Oleg and Luc Anselin (2001). “Fast maximum likelihood estimation of very large spatial autoregressive models: A characteristic polynomial approach”. English (US). In: *Computational Statistics and Data Analysis* 35(3), pp. 301–319. ISSN: 0167-9473. DOI: [10.1016/S0167-9473\(00\)00018-9](https://doi.org/10.1016/S0167-9473(00)00018-9).
- Strauss, David and Michael Ikeda (1990). “Pseudolikelihood Estimation for Social Networks”. In: *Journal of the American Statistical Association* 85(409), pp. 204–212. DOI: [10.1080/01621459.1990.10475327](https://doi.org/10.1080/01621459.1990.10475327). eprint: <https://www.tandfonline.com/doi/pdf/10.1080/01621459.1990.10475327>. URL: <https://www.tandfonline.com/doi/abs/10.1080/01621459.1990.10475327>.
- Watts, Duncan J. and Steven H. Strogatz (1998). “Collective dynamics of ‘small-world’ networks”. In: *Nature* (393), pp. 440–442. DOI: <https://doi.org/10.1038/30918>.
- Yuan, Ming and Yi Lin (Mar. 2007). “Model selection and estimation in the Gaussian graphical model”. In: *Biometrika* 94(1), pp. 19–35. ISSN: 0006-3444. DOI: [10.1093/biomet/asm018](https://doi.org/10.1093/biomet/asm018). eprint: <https://academic.oup.com/biomet/article-pdf/94/1/19/617853/asm018.pdf>. URL: <https://doi.org/10.1093/biomet/asm018>.

## A Derivation of Posteriors

### A.1 Posterior Density of $\lambda$

The posterior density of  $\lambda$  assumes no well-known form. With prior  $\lambda \sim \text{Unif}(-1, 1)$ , the posterior density of  $\lambda$  is proportional to the likelihood function:

$$\pi(\lambda | \{\mathbf{y}_t\}, \mathbf{W}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}) \propto f(\{\mathbf{y}_t\} | \lambda, \mathbf{W}, \boldsymbol{\beta}, \sigma^2) = \prod_t f(\mathbf{y}_t | \lambda, \mathbf{W}, \boldsymbol{\beta}, \sigma^2).$$

### A.2 Posterior Density of $\boldsymbol{\beta}$

With prior  $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}_\beta, \mathbf{V}_\beta)$ , the posterior density of  $\boldsymbol{\beta}$  follows normal distribution with mean

$$\left( \mathbf{V}_\beta^{-1} + \frac{1}{\sigma^2} \sum_t \mathbf{H}_t^\top \mathbf{H}_t \right)^{-1} \left( \mathbf{V}_\beta^{-1} \boldsymbol{\mu}_\beta + \frac{1}{\sigma^2} \sum_t \mathbf{H}_t^\top \mathbf{M} \mathbf{y}_t \right)$$

and variance

$$\left( \mathbf{V}_\beta^{-1} + \frac{1}{\sigma^2} \sum_t \mathbf{H}_t^\top \mathbf{H}_t \right)^{-1}.$$

### A.3 Posterior Density of $\sigma^2$

With prior  $\sigma^2 \sim \mathcal{IG}(a_{\sigma^2}, b_{\sigma^2})$ , the posterior density of  $\sigma^2$  follows Inverse Gamma distribution with parameters

$$a_{\sigma^2} + \frac{1}{2}TN \quad \text{and} \quad b_{\sigma^2} + \frac{1}{2} \sum_t \|\mathbf{M} \mathbf{y}_t - \mathbf{H}_t \boldsymbol{\beta}\|_2^2$$

respectively.

### A.4 Posterior Probability of $\mathbf{W}$

The posterior probability of  $\mathbf{W}$  assumes no well-known form, it is proportional to

$$\begin{aligned} \pi(\mathbf{W} | \{\mathbf{y}_t\}, \lambda, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}) &\propto f(\{\mathbf{y}_t\} | \lambda, \mathbf{W}, \boldsymbol{\beta}, \sigma^2) \pi(\mathbf{W} | \boldsymbol{\theta}) \\ &\propto \det(\mathbf{M}) \exp \left( -\frac{1}{2\sigma^2} \sum_t \|\mathbf{M} \mathbf{y}_t - \mathbf{H}_t \boldsymbol{\beta}\|_2^2 \right) \exp(\mathbf{s}(\mathbf{W})^\top \boldsymbol{\theta}). \end{aligned}$$

### A.5 Posterior Density of $\boldsymbol{\theta}$

The posterior density of  $\boldsymbol{\theta}$  assumes no well-known form, it is proportional to

$$\begin{aligned} \pi(\boldsymbol{\theta} | \{\mathbf{y}_t\}, \lambda, \mathbf{W}, \boldsymbol{\beta}, \sigma^2) &\propto \pi(\mathbf{W} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \\ &\propto \frac{1}{z(\boldsymbol{\theta})} \exp(\mathbf{s}(\mathbf{W})^\top \boldsymbol{\theta}) \exp \left( -\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu}_\theta)^\top \mathbf{V}_\theta^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}_\theta) \right). \end{aligned}$$

## B Efficient Sampling of Posterior $\mathbf{W}$

The main challenge of the sampling procedure is to sample the posterior of  $\mathbf{W}$ . Although the posterior of  $\mathbf{W}$  follows **ERGM** (as shown in [subsection A.5](#)), the functional form of the sufficient statistics are quite complex, making sampling the posterior of  $\mathbf{W}$  impractical with standard **ERGM** packages such as **ergm** (Handcock et al., 2023) in R.<sup>12</sup>

We propose to use **HMC**, a method that is well-known in physics, to sample the posterior of  $\mathbf{W}$  (See, e.g., Neal (2011) for an introduction to **HMC**). Recent developments extended the capabilities of **HMC** to sample from discrete distributions. In particular, Pakman and Paninski (2015) proposed an **HMC** algorithm that samples high-dimension binary distributions, which, similar to other **MCMC** methods, requires only the likelihood ratio to be known.

In our case, the posterior likelihood ratio is

$$\begin{aligned} \log \left( \frac{\pi(\tilde{\mathbf{W}} | \{\mathbf{y}_t\}, \lambda, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta})}{\pi(\mathbf{W} | \{\mathbf{y}_t\}, \lambda, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta})} \right) &= \log \left( \frac{f(\{\mathbf{y}_t\} | \lambda, \tilde{\mathbf{W}}, \boldsymbol{\beta}, \sigma^2) \pi(\tilde{\mathbf{W}} | \boldsymbol{\theta})}{f(\{\mathbf{y}_t\} | \lambda, \mathbf{W}, \boldsymbol{\beta}, \sigma^2) \pi(\mathbf{W} | \boldsymbol{\theta})} \right) \\ &= T \log \left( \frac{\det(\mathbf{I} - \lambda \tilde{\mathbf{W}})}{\det(\mathbf{I} - \lambda \mathbf{W})} \right) \end{aligned} \quad (4)$$

$$+ \frac{1}{2\sigma^2} \sum_t \left[ \|\mathbf{M}\mathbf{y}_t - \mathbf{H}_t\boldsymbol{\beta}\|_2^2 - \|\tilde{\mathbf{M}}\mathbf{y}_t - \mathbf{H}_t\boldsymbol{\beta}\|_2^2 \right] \quad (5)$$

$$+ [\mathbf{s}(\tilde{\mathbf{W}}) - \mathbf{s}(\mathbf{W})]^\top \boldsymbol{\theta}. \quad (6)$$

Since the algorithm attempts to update the binary matrix one entry at a time, calculating of terms (4), (5), and (6) can be easily optimized. Krisztin and Piribauer (2022) optimized the calculation of determinant ratio term (4) with the matrix inversion lemma.

In this paper, optimization for (6) is implemented. This amounts to writing a custom function for each sufficient statistic specified in the **ERGM** prior. The difference functions for **ERGM** terms **edges**, **mutual**, **ctriad** (cyclic triple), **ttriad** (transitive triple), **kstar**, and **triangle** are implemented.

## C Implementation Details

The sampling procedure is implemented in R and C++. R is used as the user interface, whereas all the computationally intensive procedures are implemented in C++. There are only two R dependencies: **Rcpp** (use C++ in R) and **RcppArmadillo** (linear algebra library).

The discrete **HMC** algorithm is a direct adaptation from Pakman and Paninski (2015); the **EX** algorithm is rewritten in C++ following the description in Caimo and Friel (2011). I implemented custom **ERGM** statistics and sampler.<sup>13</sup> Although the **ergm** package is not used in this paper, the naming of **ERGM** terms follows their convention. A list of implemented **ERGM** terms is presented in [Table C1](#), or one can refer to the file `ERGM_stat.h` to see the detailed implementa-

<sup>12</sup> There is an R package **ergm.userterms** that allows users to define custom **ERGM** terms (See Hunter, Goodreau, and Handcock (2013)). However, as noted in [introduction](#) from Handcock et al. (2023), coding new **ERGM** terms requires not only knowledge in the C language, but also in building R from source.

<sup>13</sup>It is possible to write custom **ERGM** terms for the **ergm** package. See [footnote 12](#).



Term	Description
<code>edges</code>	number of edges
<code>kstar(2)</code>	number of 2-stars
<code>kstar(3)</code>	number of 3-stars
<code>triangle</code>	number of triangles
<code>ctriad</code>	number of cyclic triples
<code>ttriad</code>	number of transitive triples

Table C1: Implemented **ERGM** terms.

tion. All the source code and the  $\text{\TeX}$  file for this paper can be found at <https://github.com/jessekelighine/bayesian-unknown-network-sar>.

All calculations in this paper are performed with computation resources from [Behavioral and Data Science Research Center](#) at National Taiwan University. Since we utilize Gibbs sampling to sample the posterior distribution, all programs are run single-threaded. The time it took to run each estimation depends on the panel length  $T$  and number of nodes  $N$ . Roughly, each estimation in [subsection 5.1](#) took about one hour; each estimation in [subsection 5.2](#) took about three hours; each estimation in [subsection 5.3](#) took about two days.