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

陳捷
Jesse Chieh Chen

2024-05-24

Contents

Motivation

- Econometricians are interested in networks, but network data are hard to come by.
- In many cases, although the network structure is unknown, we have a good idea about its characteristics (e.g., small world, mutuality).

Question

How can we effectively incorporate these characteristics in estimation?

Literature

In other fields:

- Causal Inference:
  1. Madigan and Raftery (1994): DAG causal graph recovery
  2. Yuan and Lin (2007): recover covariance relations

In Econometrics:

- Frequentist attempts: High-dimensional Methods
  1. Meinshausen and Bühlmann (2006): Lasso
  2. Lam and P. C. Souza (2020): Lasso
  3. de Paula, Rasul, and P. Souza (2023): Identification without sparsity, elastic net
- Bayesian attempt: Krisztin and Piribauer (2022)
  1. Simple network priors (Erdős-Rényi type)
  2. No motivation from network formation

Spatial Autoregressive Model

SAR

$$\underset{(N\times 1)}{\mathbf{y}} = \lambda \underset{(N\times N)}{\bar{\mathbf{W}}} \mathbf{y} + \underset{(N\times K_1)}{\mathbf{X}_1} \underset{(K_1\times 1)}{\boldsymbol{\beta}_1} + \bar{\mathbf{W}}\mathbf{X}_2 \underset{(K_2\times 1)}{\boldsymbol{\beta}_2} + \mathbf{u}$$

- $\mathbf{y}$: vector of outcome variables
- $\bar{\mathbf{W}}$: row-normalized adjacency matrix
- $\mathbf{X}_1, \mathbf{X}_2$: exogenous variables
- $\mathbf{u}$: Gaussian error

1. λ: peer effect or endogenous effect
2. $\boldsymbol{\beta}_1$: exogenous effect
3. $\boldsymbol{\beta}_2$: contextual effect

Why Bayesian?

- There are two problems with SAR:
  1. The network is endogenously formed
  2. The network is, in many cases, unknown
- It is natural to specify a **network formation model**.
- Hence, the likelihood combined with the network formation probability gives the **complete** picture.
- This calls for a Bayesian approach.

$$\underbrace{f(\mathbf{y} \mid \lambda, \mathbf{W}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)}_{\text{SAR likelihood}} \cdot \underbrace{\pi(\mathbf{W})}_{\substack{\text{network}\\\text{formation}}}$$

- In this research, we focus on ERGM priors.

Exponential Random Graph Model

ERGM

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

- $\mathrm{s}(\mathbf{W})$: vector of sufficient statistics,
- $\boldsymbol{\theta}$: vector of natural parameters
- $z(\boldsymbol{\theta})$: normalizing constant.

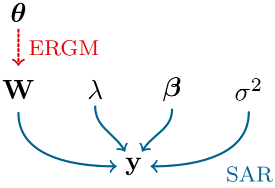
1. Flexibility: $\mathrm{s}(\boldsymbol{\theta})$ can be anything
   - Erdős-Rényi (Poisson), homophily models, etc
2. Tractability: known algorithms to sample $\mathbf{W}$ and recover $\boldsymbol{\theta}$
3. Interpretability: some microfoundation

Likelihood Function

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

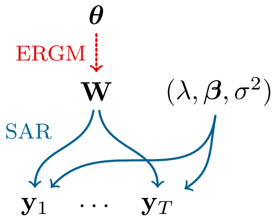
- $\boldsymbol{\beta}$ is the vector of parameters $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)^\top$
- $\mathbf{M}$ is the reduced form matrix $\mathbf{I} - \lambda\mathbf{W}$
- $\mathbf{H}$ is the stacked exogenous variables $[\mathbf{X}_1, \mathbf{WX}_2]$
- $\|\mathbf{x}\|_2^2$ is the Euclidean norm $\mathbf{x}^\top\mathbf{x}$

Priors



$$
\begin{cases}
\boldsymbol{\beta} & \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \mathbf{V}_{\boldsymbol{\beta}}) \\
\sigma^2 & \sim \mathcal{IG}(a_{\sigma^2}, b_{\sigma^2}) \\
\lambda & \sim \mathrm{Unif}(-1, 1) \\
\pi(\mathbf{W} \mid \boldsymbol{\theta}) & \propto \exp(\mathbf{s}(\mathbf{W})^\top \boldsymbol{\theta}) \qquad \text{(ERGM)} \\
\boldsymbol{\theta} & \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \mathbf{V}_{\boldsymbol{\theta}})
\end{cases}
$$

Priors: Panel Data



- One observation of $\mathbf{y}$ contains too little information.
- We assume we have iid observations $\{\mathbf{y}_t, \mathbf{X}_t\}$ with likelihood function:

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

Sampling: $\boldsymbol{\beta}$ and σ^2

- With prior $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \mathbf{V}_{\boldsymbol{\beta}})$, the posterior of $\boldsymbol{\beta}$ also follows normal distribution. (Normal-Normal conjugacy)
- With prior $\sigma^2 \sim \mathcal{IG}(a_{\sigma^2}, b_{\sigma^2})$, the posterior density of σ^2 also follows Inverse Gamma distribution. (Inverse Gamma-Normal conjugacy)

Sampling: λ

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

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

- A simple Metropolis-Hastings Algorithm is used to sample posterior λ.
- One can also use Griddy-Gibbs Algorithm.

Sampling: $\mathbf{W}$

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

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

- Since this is a high-dimensional sampling problem, we use a discrete variant of Hamiltonian Monte Carlo to sample posterior $\mathbf{W}$.

Sampling: $\boldsymbol{\theta}$

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

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

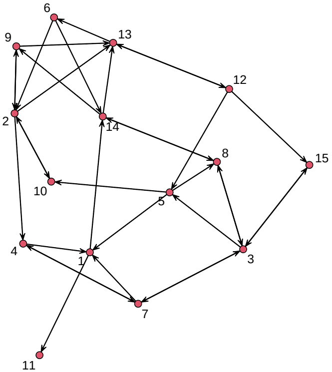
- Since $z(\boldsymbol{\theta})$ is unknown, we use Exchange Algorithm to obtain posterior samples of $\boldsymbol{\theta}$.

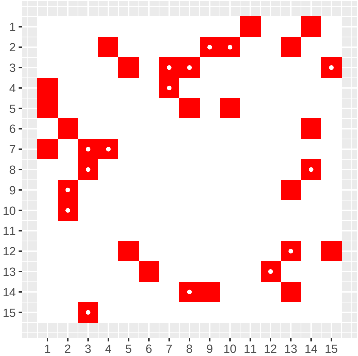
Gibbs Sampling

Initialize $(\lambda, \mathbf{W}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta})$
while true do
  $\boldsymbol{\beta}' \leftarrow$ sample from $\pi(\boldsymbol{\beta} \mid \{\mathbf{y}_t\}, \lambda, \mathbf{W}, \sigma^2)$ exactly
  $\sigma^{2\prime} \leftarrow$ sample from $\pi(\sigma^2 \mid \{\mathbf{y}_t\}, \lambda, \mathbf{W}, \boldsymbol{\beta}')$ exactly
  $\lambda' \leftarrow$ sample from $\pi(\lambda \mid \{\mathbf{y}_t\}, \mathbf{W}, \boldsymbol{\beta}', \sigma^{2\prime})$ with Metropolis-Hastings Algorithm (M-H)
  $\mathbf{W}' \leftarrow$ sample from $\pi(\mathbf{W} \mid \{\mathbf{y}_t\}, \lambda', \boldsymbol{\beta}', \sigma^{2\prime}, \boldsymbol{\theta})$ with Hamiltonian Monte Carlo (HMC)
  $\boldsymbol{\theta}' \leftarrow$ sample from $\pi(\boldsymbol{\theta} \mid \mathbf{W}')$ with Exchange Algorithm (EX)
  $(\lambda, \mathbf{W}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}) \leftarrow (\lambda', \mathbf{W}', \boldsymbol{\beta}', \sigma^{2\prime}, \boldsymbol{\theta}')$
end while

- All sampling algorithms are implemented in `R` and `C++`

Basic ERGM





- This network is generated from ERGM with $\texttt{edges} = -2$ and $\texttt{mutual} = 2$.
- Simulated data is generated for $T = 12$.

Group P1 and P2

- **Group P1**: $\mathbf{X}_1$ is an $N \times 1$ matrix of exogenous variables generated from a normal distribution and $\mathbf{X}_2$ equals $\mathbf{X}_1$.
- **Group P2**: $\mathbf{X}_1$ is an $N \times 2$ matrix of exogenous variables generated from a normal distribution.

Basic ERGM: Results

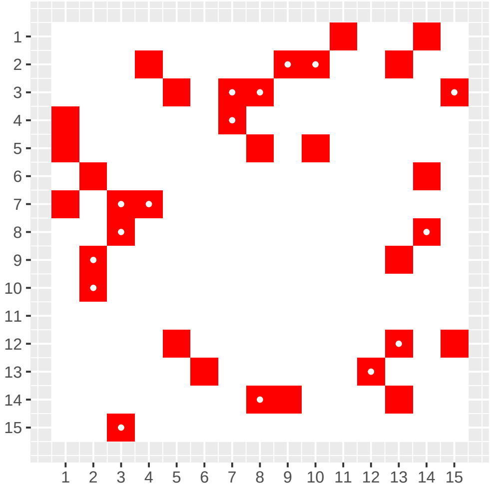


Figure: True $\mathbf{W}$

Basic ERGM: Results P10 $\mathbf{W}$

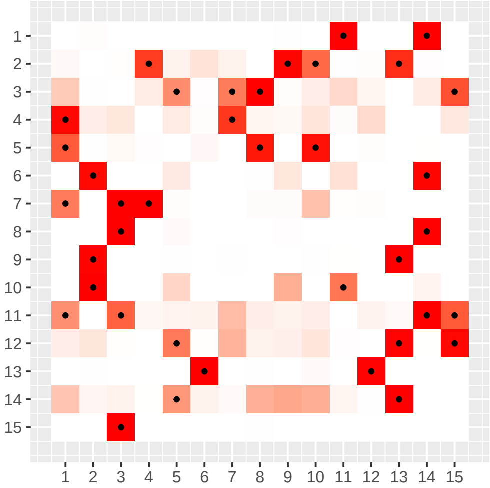


Figure: Posterior Mean $\mathbf{W}$

Basic ERGM: Results P11 $\mathbf{W}$

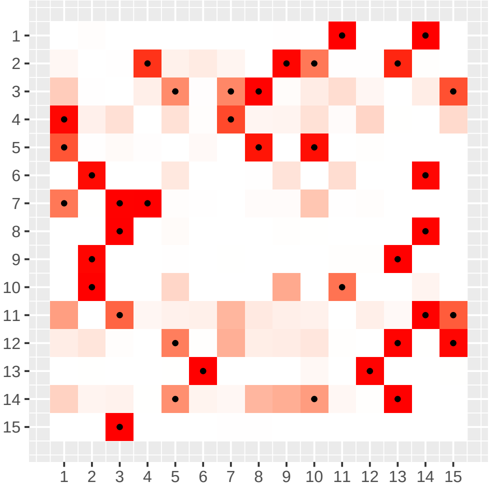


Figure: Posterior Mean $\mathbf{W}$

Basic ERGM: Results P12 $\mathbf{W}$

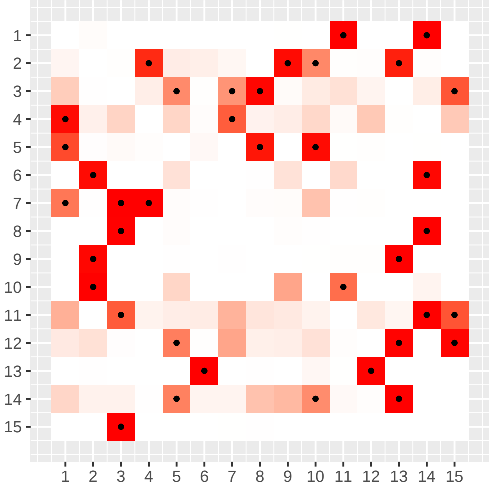


Figure: Posterior Mean $\mathbf{W}$

Basic ERGM: Results P13 $\mathbf{W}$

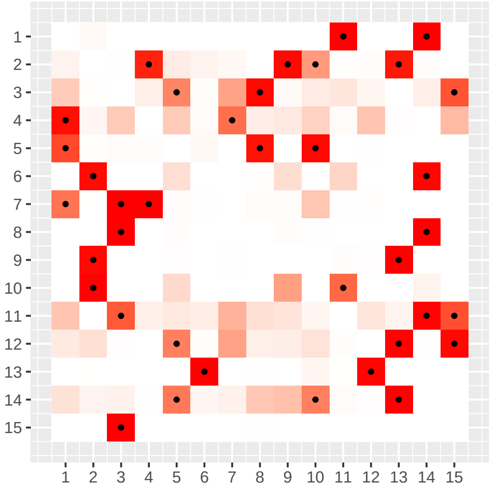


Figure: Posterior Mean $\mathbf{W}$

Basic ERGM: Results P14 $\mathbf{W}$

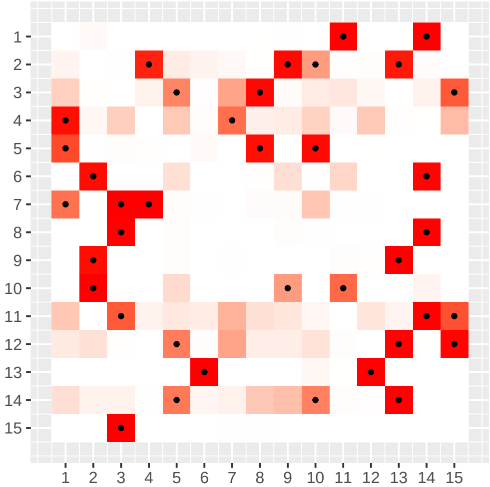


Figure: Posterior Mean $\mathbf{W}$

Basic ERGM: Results P1

Spec.	Param.	True	Prior Dist.	Post. Mean	95% Credible Interval $Q_{2.5\%}$	$Q_{97.5\%}$
P12 $T = 12$	λ	0.1	Unif$(-1, 1)$	0.0323	−0.0355	0.0964
	σ^2	2	$\mathcal{IG}(1, 1)$	4.5741	3.3396	6.2076
	$\boldsymbol{\beta}_1$	1	$\mathcal{N}(0, 30)$	0.9944	0.9584	1.0334
	$\boldsymbol{\beta}_2$	0.5	$\mathcal{N}(0, 30)$	0.5930	0.5386	0.6543
	$\boldsymbol{\theta}_{\texttt{edges}}$	−2	$\mathcal{N}(0, 1)$	−1.5410	−2.0779	−1.0458
	$\boldsymbol{\theta}_{\texttt{mutual}}$	2	$\mathcal{N}(0, 1)$	0.4882	−0.6913	1.6313
P13 $T = 12$	λ	0.1	Unif$(-1, 1)$	0.0333	−0.0311	0.0991
	σ^2	2	$\mathcal{IG}(1, 1)$	4.6069	3.3528	6.1911
	$\boldsymbol{\beta}_1$	1	$\mathcal{N}(0, 30)$	0.9937	0.9572	1.0308
	$\boldsymbol{\beta}_2$	0.5	$\mathcal{N}(0, 30)$	0.5929	0.5382	0.6487
	$\boldsymbol{\theta}_{\texttt{edges}}$	−2	$\mathcal{N}(0, 1)$	−1.4252	−1.8560	−1.0364

Basic ERGM: Results P20 $\mathbf{W}$

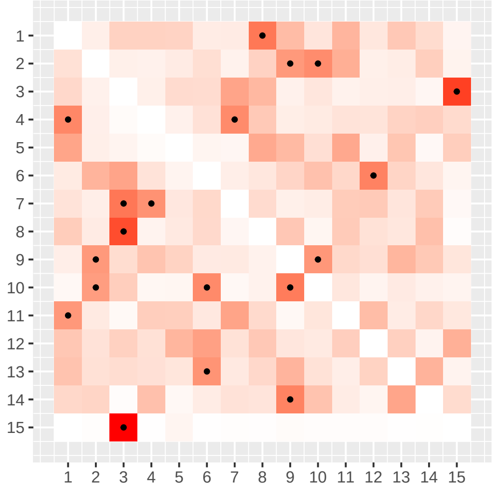


Figure: Posterior Mean $\mathbf{W}$

Basic ERGM: Results P21 $\mathbf{W}$

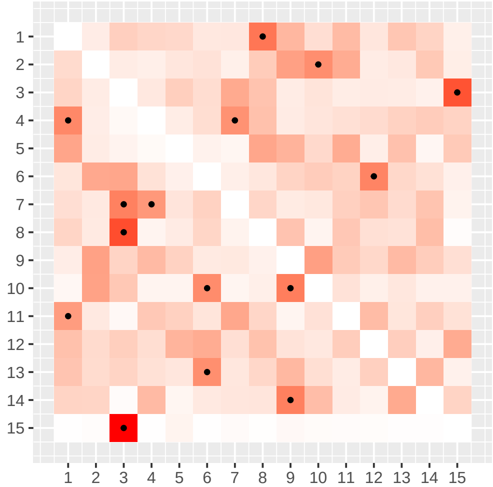


Figure: Posterior Mean $\mathbf{W}$

Basic ERGM: Results P22 $\mathbf{W}$

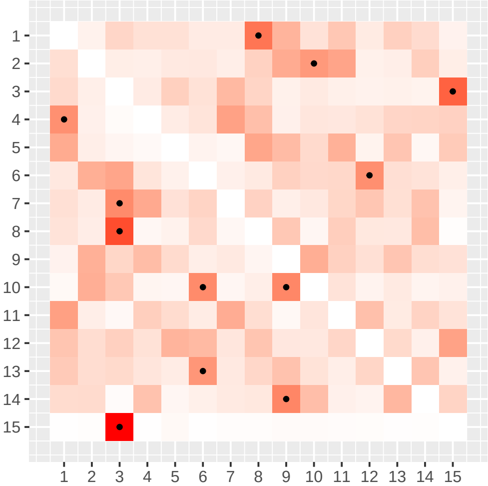


Figure: Posterior Mean $\mathbf{W}$

Basic ERGM: Results P23 $\mathbf{W}$

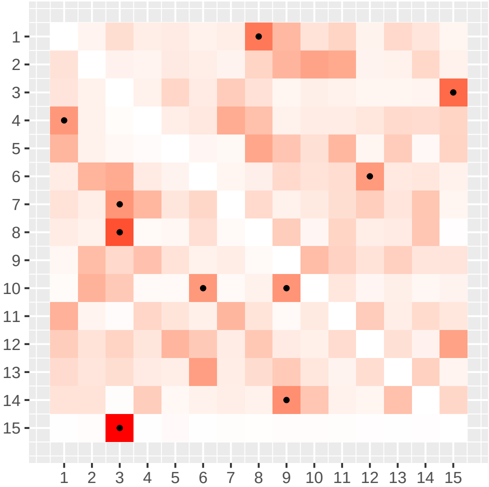


Figure: Posterior Mean $\mathbf{W}$

Basic ERGM: Results P24 $\mathbf{W}$

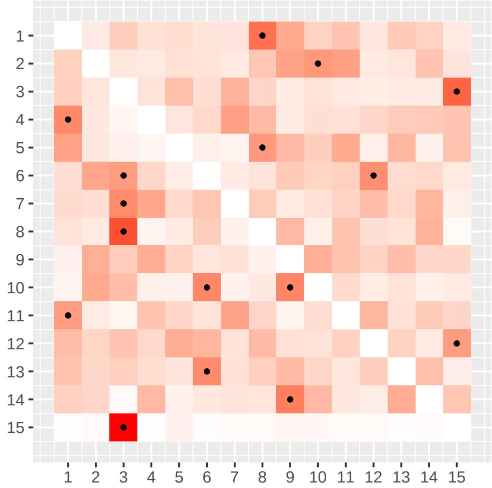


Figure: Posterior Mean $\mathbf{W}$

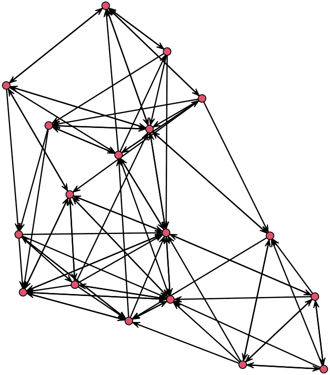
Basic ERGM: Results P2

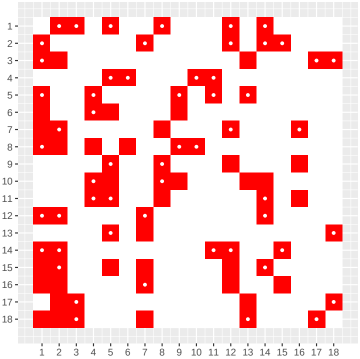
Spec.	Param.	True	Prior Dist.	Post. Mean	95% Credible Interval $Q_{2.5\%}$	$Q_{97.5\%}$
P22 $T=12$	λ	0.1	$\text{Unif}(-1,1)$	0.0978	0.0470	0.1466
	σ^2	2	$\mathcal{IG}(1,1)$	2.2190	1.6317	3.1347
	$\boldsymbol{\beta}_1$	1	$\mathcal{N}(0,30)$	1.0026	0.9795	1.0253
	$\boldsymbol{\beta}_2$	0.5	$\mathcal{N}(0,30)$	0.1891	0.1736	0.2049
	$\boldsymbol{\theta}_{\texttt{edges}}$	-2	$\mathcal{N}(0,1)$	-1.7877	-3.4018	-0.7241
	$\boldsymbol{\theta}_{\texttt{mutual}}$	2	$\mathcal{N}(0,1)$	0.2623	-1.3760	1.7652
P23 $T=12$	λ	0.1	$\text{Unif}(-1,1)$	0.0930	0.0454	0.1403
	σ^2	2	$\mathcal{IG}(1,1)$	2.2601	1.6619	3.2626
	$\boldsymbol{\beta}_1$	1	$\mathcal{N}(0,30)$	1.0042	0.9813	1.0287
	$\boldsymbol{\beta}_2$	0.5	$\mathcal{N}(0,30)$	0.1894	0.1735	0.2056
	$\boldsymbol{\theta}_{\texttt{edges}}$	-2	$\mathcal{N}(0,1)$	-1.9940	-4.0807	-0.8732

Basic ERGM: Results Summary

- Significant results are obtained for λ, however, this depends heavily on the model specification.
  - In general, $\pi(\lambda \mid \mathbf{W}, \{\mathbf{y}_t\})$ is very different from $\pi(\lambda, \mathbf{W} \mid \{\mathbf{y}_t\})$
- The spatial matrix $\mathbf{W}$ is well-recovered by the approach.
- Inference of $\boldsymbol{\theta}$ is as expected, but only edges obtain consistently significant results.
- Our model out performs specifications where only $\boldsymbol{\theta}_{\texttt{edges}}$ is specified in the prior, but only by a little.

Sampson's Monk





- We use ERGM with `edges` $= -3$, `mutual` $= 2$, and `ctriad` $= -0.5$ as prior.

Group S1 and S2

1. **Group S1**: $\mathbf{X}_1 = \mathbf{X}_2$ are $N \times 2$ matrix of exogenous variables. In this group, the exogenous effects are larger than endogenous effects.
2. **Group S2**: $\mathbf{X}_1 = \mathbf{X}_2$ are $N \times 2$ matrix of exogenous variables. In this group, the some exogenous effects are smaller than some endogenous effects.

Sampson's Monk: Results

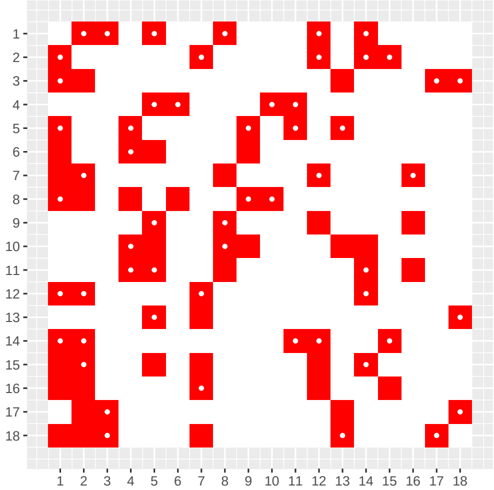


Figure: True $\mathbf{W}$

Sampson's Monk: Results S10 $\mathbf{W}$

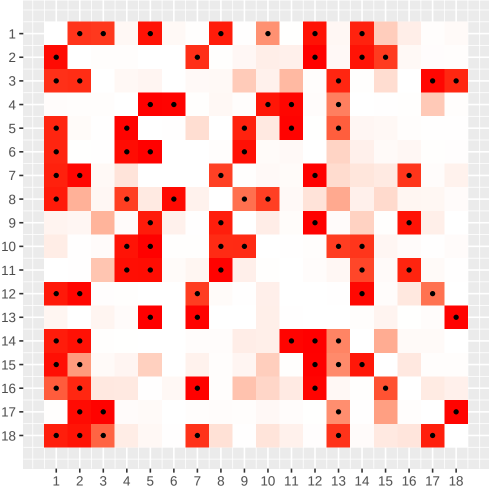


Figure: Posterior Mean $\mathbf{W}$

Sampson's Monk: Results S11 $\mathbf{W}$

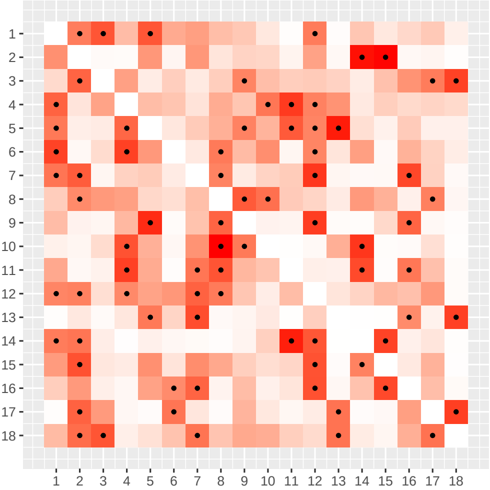


Figure: Posterior Mean $\mathbf{W}$

Sampson's Monk: Results S12 $\mathbf{W}$

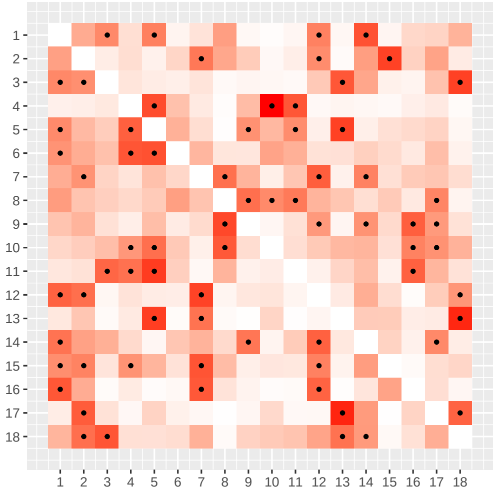


Figure: Posterior Mean $\mathbf{W}$

Sampson's Monk: Results S1

Spec.	Param.	True	Prior Dist.	Post. Mean	95% Credible Interval $Q_{2.5\%}$	$Q_{97.5\%}$
S10 $T = 18$	λ	0.1	$\text{Unif}(-1, 1)$	0.0320	−0.0857	0.1964
	σ^2	2	$\mathcal{IG}(1, 1)$	2.8549	1.9332	5.6295
	$\boldsymbol{\beta}_{1,1}$	2	$\mathcal{N}(0, 30)$	2.0123	1.9648	2.0617
	$\boldsymbol{\beta}_{1,2}$	2	$\mathcal{N}(0, 30)$	2.0984	1.1329	2.7002
	$\boldsymbol{\beta}_{2,1}$	0.5	$\mathcal{N}(0, 30)$	0.7005	0.4930	0.8812
	$\boldsymbol{\beta}_{2,2}$	0.5	$\mathcal{N}(0, 30)$	0.1332	−5.8634	3.3570
	$\boldsymbol{\theta}_{\texttt{edges}}$	−1.8	$\mathcal{N}(-3, 1)$	−0.9080	−1.6285	−0.0488
	$\boldsymbol{\theta}_{\texttt{mutual}}$	2.3	$\mathcal{N}(2, 1)$	1.2242	0.3165	2.1608
	$\boldsymbol{\theta}_{\texttt{ctriad}}$	−0.1	$\mathcal{N}(-0.5, 1)$	−0.2491	−0.7051	0.1311
S11 $T = 12$	λ	0.1	$\text{Unif}(-1, 1)$	0.0306	−0.0906	0.1989
	σ^2	2	$\mathcal{IG}(1, 1)$	4.6437	1.7827	11.3241
	$\boldsymbol{\beta}_{1,1}$	2	$\mathcal{N}(0, 30)$	2.0015	1.9525	2.0572
	$\boldsymbol{\beta}_{1,2}$	2	$\mathcal{N}(0, 30)$	4.2848	1.2563	10.8403
	$\boldsymbol{\beta}_{2,1}$	0.5	$\mathcal{N}(0, 30)$	0.6090	0.0223	0.9359
	$\boldsymbol{\beta}_{2,2}$	0.5	$\mathcal{N}(0, 30)$	0.9627	−12.6832	10.0607
	$\boldsymbol{\theta}_{\texttt{edges}}$	−1.8	$\mathcal{N}(-3, 1)$	−1.2569	−4.7459	0.3174
	$\boldsymbol{\theta}_{\texttt{mutual}}$	2.3	$\mathcal{N}(2, 1)$	0.6115	−1.0884	1.8689
	$\boldsymbol{\theta}_{\texttt{ctriad}}$	−0.1	$\mathcal{N}(-0.5, 1)$	−0.3708	−1.9857	0.2509

Sampson's Monk: Results S20 $\mathbf{W}$

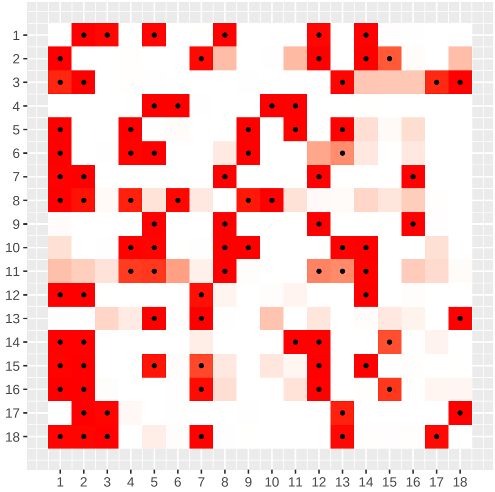


Figure: Posterior Mean $\mathbf{W}$

Sampson's Monk: Results S21 $\mathbf{W}$

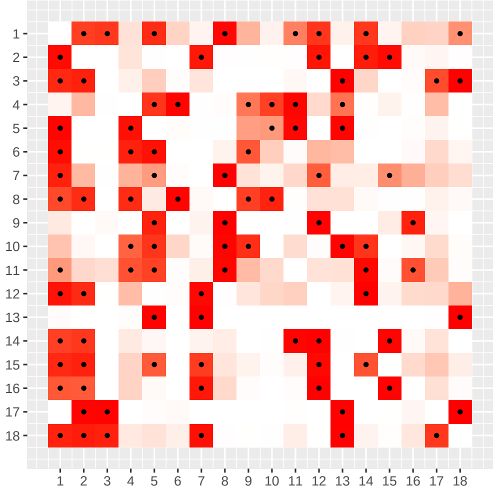


Figure: Posterior Mean $\mathbf{W}$

Sampson's Monk: Results S22 $\mathbf{W}$

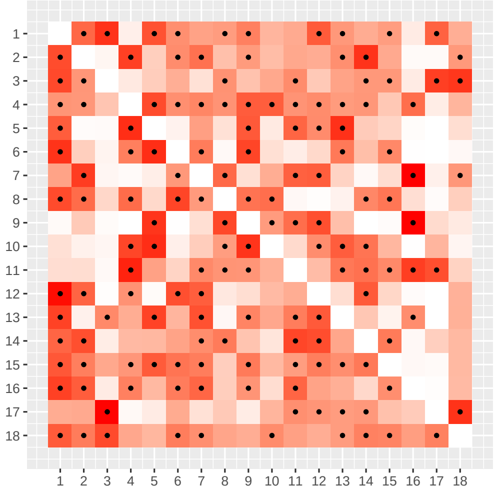


Figure: Posterior Mean $\mathbf{W}$

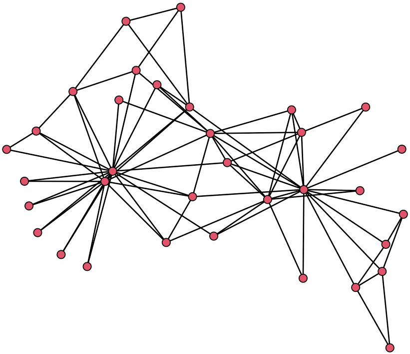
Sampson's Monk: Results S2

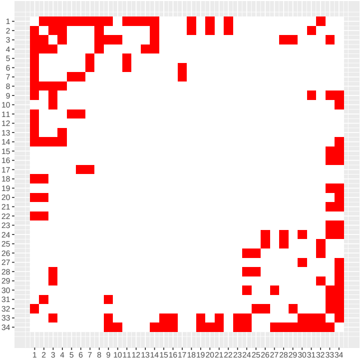
Spec.	Param.	True	Prior Dist.	Post. Mean	95% Credible Interval $Q_{2.5\%}$	$Q_{97.5\%}$
S20 $T = 18$	λ	0.1	$\text{Unif}(-1, 1)$	0.0933	−0.0515	0.2552
	σ^2	2	$\mathcal{IG}(1, 1)$	6.0217	3.3168	10.7843
	$\boldsymbol{\beta}_{1,1}$	2	$\mathcal{N}(0, 30)$	1.9934	1.9176	2.0710
	$\boldsymbol{\beta}_{1,2}$	0.5	$\mathcal{N}(0, 30)$	0.8951	−0.1525	1.8155
	$\boldsymbol{\beta}_{2,1}$	2	$\mathcal{N}(0, 30)$	2.0583	1.5653	2.4986
	$\boldsymbol{\beta}_{2,2}$	0.5	$\mathcal{N}(0, 30)$	−0.7854	−5.3961	3.1025
	$\boldsymbol{\theta}_{\texttt{edges}}$	−1.8	$\mathcal{N}(-3, 1)$	−1.2451	−1.8671	−0.5679
	$\boldsymbol{\theta}_{\texttt{mutual}}$	2.3	$\mathcal{N}(2, 1)$	1.0436	0.1724	1.9152
	$\boldsymbol{\theta}_{\texttt{ctriad}}$	−0.1	$\mathcal{N}(-0.5, 1)$	0.0069	−0.3277	0.2604
S21 $T = 12$	λ	0.1	$\text{Unif}(-1, 1)$	0.0339	−0.1151	0.2276
	σ^2	2	$\mathcal{IG}(1, 1)$	6.1242	3.2807	13.9441
	$\boldsymbol{\beta}_{1,1}$	2	$\mathcal{N}(0, 30)$	2.0333	1.8949	2.1770
	$\boldsymbol{\beta}_{1,2}$	0.5	$\mathcal{N}(0, 30)$	0.9477	−0.1826	2.3512
	$\boldsymbol{\beta}_{2,1}$	2	$\mathcal{N}(0, 30)$	2.2310	1.7853	2.5866
	$\boldsymbol{\beta}_{2,2}$	0.5	$\mathcal{N}(0, 30)$	1.4372	−3.9601	5.0759
	$\boldsymbol{\theta}_{\texttt{edges}}$	−1.8	$\mathcal{N}(-3, 1)$	−1.1667	−1.7794	−0.5208
	$\boldsymbol{\theta}_{\texttt{mutual}}$	2.3	$\mathcal{N}(2, 1)$	0.8031	−0.0575	1.6794
	$\boldsymbol{\theta}_{\texttt{ctriad}}$	−0.1	$\mathcal{N}(-0.5, 1)$	0.0580	−0.2533	0.2849

Sampson's Monk: Results Summary

- The posterior $\mathbf{W}$ is more noisy when the network is more busy.
- Similar to the first simulation study, no significant result for λ.
- Posterior inferences for $\boldsymbol{\theta}$ is as expected: more significant results when the panel is longer.
- Matrix degeneracy is a problem in this simulation.

Zachary's Karate Cub





- For this undirected network, we use ERGM with `edges` $= -2$, `2-star` $= 0.2$, `3-star` $= -0.1$, and `triangle` $= 0.3$.

Group K1 and K2

- **Group K1**: $\mathbf{X}_1$ is an $N \times 2$ matrix of exogenous variables. The endogenous effect of this group is $\lambda = 0.05$.
- **Group K2**: $\mathbf{X}_1$ is an $N \times 2$ matrix of exogenous variables. The endogenous effect of this group is $\lambda = 0.1$.

Zachary's Karate Club: Results

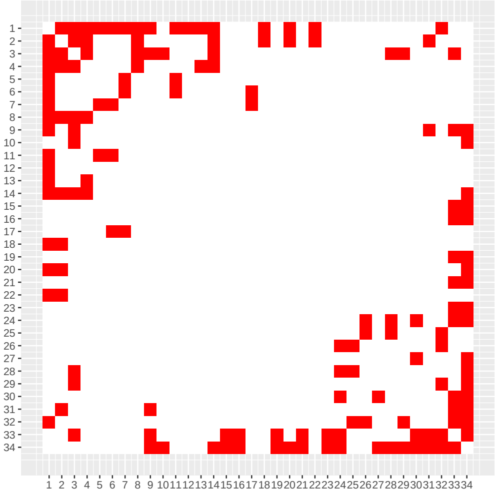


Figure: True $\mathbf{W}$

Zachary's Karate Club: Results K10 $\mathbf{W}$

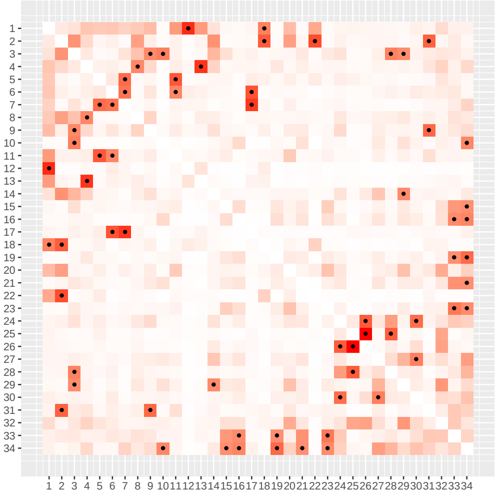


Figure: Posterior Mean $\mathbf{W}$

Zachary's Karate Club: Results K11 $\mathbf{W}$

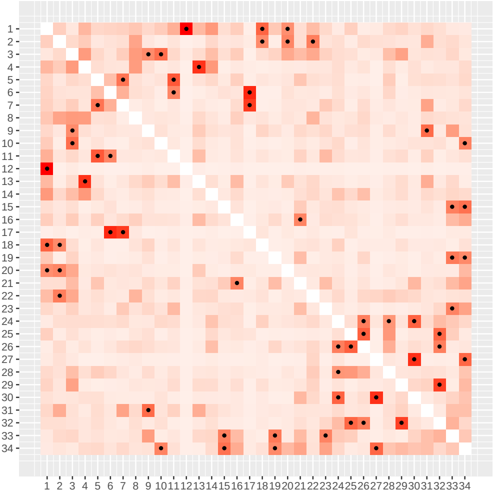


Figure: Posterior Mean $\mathbf{W}$

Zachary's Karate Club: Results K12 $\mathbf{W}$

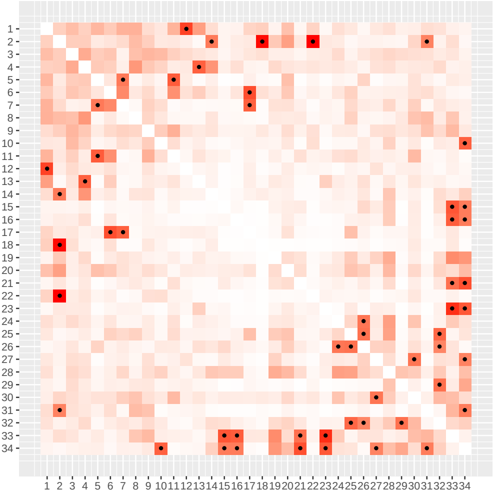


Figure: Posterior Mean $\mathbf{W}$

Zachary's Karate Club: Results K1

Spec.	Param.	True	Prior Dist.	Post. Mean	95% Credible Interval $Q_{2.5\%}$	$Q_{97.5\%}$
K10 $T = 15$	λ	0.05	Unif$(-1, 1)$	0.2761	-0.3237	0.9442
	σ^2	5	$\mathcal{IG}(1, 1)$	40.1132	8.4908	173.1849
	$\boldsymbol{\beta}_{1,1}$	2	$\mathcal{N}(0, 30)$	2.0150	1.7000	2.2317
	$\boldsymbol{\beta}_{1,2}$	1	$\mathcal{N}(0, 30)$	0.2775	-10.1314	8.5981
	$\boldsymbol{\beta}_2$	0.5	$\mathcal{N}(0, 30)$	-0.1082	-1.7671	1.3477
	$\boldsymbol{\theta}_{\texttt{edges}}$	-3	$\mathcal{N}(-3, 1)$	-2.3436	-3.3285	-1.1674
	$\boldsymbol{\theta}_{\texttt{2-star}}$	0.2	$\mathcal{N}(0.2, 1)$	0.7999	0.1206	1.5804
	$\boldsymbol{\theta}_{\texttt{3-star}}$	-0.1	$\mathcal{N}(-0.1, 1)$	-0.4000	-1.1722	-0.0340
	$\boldsymbol{\theta}_{\texttt{triangle}}$	0.3	$\mathcal{N}(0.3, 1)$	0.2646	-0.8123	1.2890
K11 $T = 12$	λ	0.05	Unif$(-1, 1)$	0.0292	-0.3401	0.4165
	σ^2	5	$\mathcal{IG}(1, 1)$	37.3011	8.3389	149.4285
	$\boldsymbol{\beta}_{1,1}$	2	$\mathcal{N}(0, 30)$	2.0682	1.8333	2.4211
	$\boldsymbol{\beta}_{1,2}$	1	$\mathcal{N}(0, 30)$	1.0926	-4.0489	7.2829
	$\boldsymbol{\beta}_2$	0.5	$\mathcal{N}(0, 30)$	0.4879	-0.4604	1.3605
	$\boldsymbol{\theta}_{\texttt{edges}}$	-3	$\mathcal{N}(-3, 1)$	-2.4652	-3.6032	-1.2811
	$\boldsymbol{\theta}_{\texttt{2-star}}$	0.2	$\mathcal{N}(0.2, 1)$	0.7811	-0.0684	1.5686
	$\boldsymbol{\theta}_{\texttt{3-star}}$	-0.1	$\mathcal{N}(-0.1, 1)$	-0.2669	-1.0970	1.1224
	$\boldsymbol{\theta}_{\texttt{triangle}}$	0.3	$\mathcal{N}(0.3, 1)$	0.2900	-0.8121	1.4029

Zachary's Karate Club: Results K20 $\mathbf{W}$

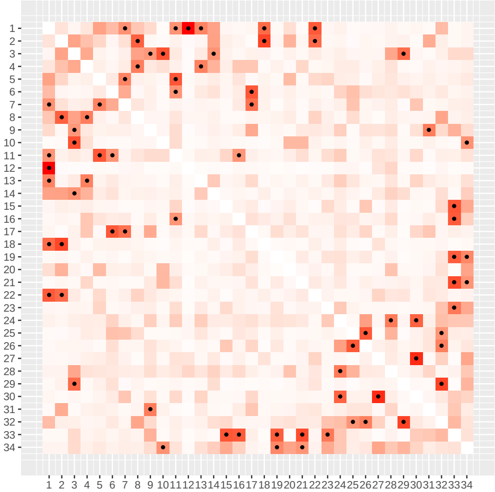


Figure: Posterior Mean $\mathbf{W}$

Zachary's Karate Club: Results K21 $\mathbf{W}$

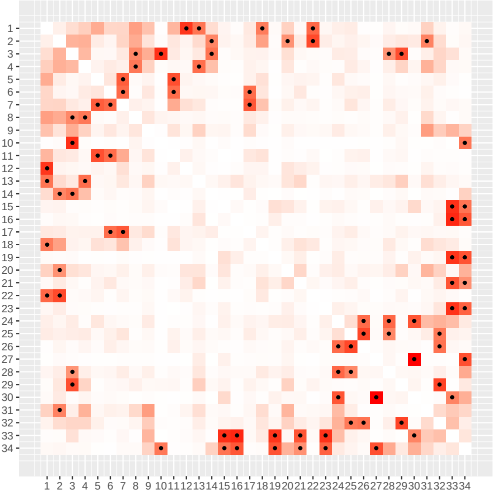


Figure: Posterior Mean $\mathbf{W}$

Zachary's Karate Club: Results K22 $\mathbf{W}$

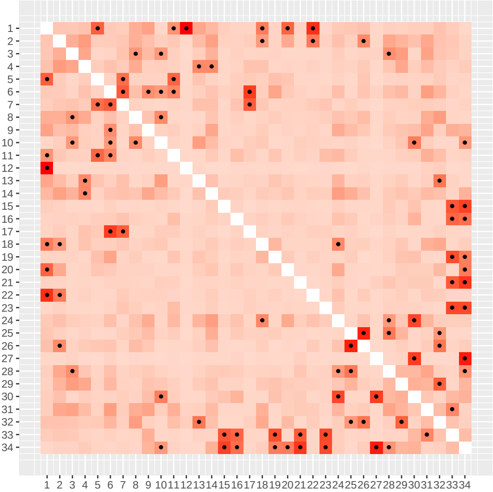


Figure: Posterior Mean $\mathbf{W}$

Zachary's Karate Club: Results K2

Spec.	Param.	True	Prior Dist.	Post. Mean	95% Credible Interval $Q_{2.5\%}$	$Q_{97.5\%}$
K20 $T = 15$	λ	0.1	Unif$(-1, 1)$	0.0527	-0.2202	0.3424
	σ^2	5	$\mathcal{IG}(1, 1)$	45.0312	8.3686	140.6322
	$\boldsymbol{\beta}_{1,1}$	2	$\mathcal{N}(0, 30)$	1.9974	1.5802	2.5150
	$\boldsymbol{\beta}_{1,2}$	1	$\mathcal{N}(0, 30)$	0.8264	-5.0927	8.6789
	$\boldsymbol{\beta}_2$	0.5	$\mathcal{N}(0, 30)$	0.6450	0.1342	0.9109
	$\boldsymbol{\theta}_{\texttt{edges}}$	-3	$\mathcal{N}(-3, 1)$	-2.4708	-3.4504	-1.2638
	$\boldsymbol{\theta}_{\texttt{2-star}}$	0.2	$\mathcal{N}(0.2, 1)$	0.7538	0.1035	1.4728
	$\boldsymbol{\theta}_{\texttt{3-star}}$	-0.1	$\mathcal{N}(-0.1, 1)$	-0.3072	-0.9901	-0.0301
	$\boldsymbol{\theta}_{\texttt{triangle}}$	0.3	$\mathcal{N}(0.3, 1)$	0.2463	-0.6884	1.1994
K21 $T = 12$	λ	0.1	Unif$(-1, 1)$	0.0216	-0.3382	0.3432
	σ^2	5	$\mathcal{IG}(1, 1)$	43.4672	7.9732	178.8634
	$\boldsymbol{\beta}_{1,1}$	2	$\mathcal{N}(0, 30)$	2.0573	1.6866	2.4233
	$\boldsymbol{\beta}_{1,2}$	1	$\mathcal{N}(0, 30)$	1.0191	-3.1973	4.3400
	$\boldsymbol{\beta}_2$	0.5	$\mathcal{N}(0, 30)$	0.6581	0.0032	1.3475
	$\boldsymbol{\theta}_{\texttt{edges}}$	-3	$\mathcal{N}(-3, 1)$	-2.4819	-3.3739	-1.3568
	$\boldsymbol{\theta}_{\texttt{2-star}}$	0.2	$\mathcal{N}(0.2, 1)$	0.7527	0.1402	1.5161
	$\boldsymbol{\theta}_{\texttt{3-star}}$	-0.1	$\mathcal{N}(-0.1, 1)$	-0.2971	-0.9783	-0.0365
	$\boldsymbol{\theta}_{\texttt{triangle}}$	0.3	$\mathcal{N}(0.3, 1)$	0.2581	-0.7139	1.2428

Zachary's Karate Club: Results Summary

- A larger λ leads to a more volatile estimation of the parameters.
- A more complex ERGM does not necessarily perform better, this is perhaps due to high correlation in some dimensions in the posterior of $\boldsymbol{\theta}$. In fact, the correlation between posterior $\boldsymbol{\theta}_{\texttt{kstar(2)}}$ and $\boldsymbol{\theta}_{\texttt{kstar(3)}}$ is about 0.8, which is to be expected.

Conclusion

- In this paper, we present a novel method to estimate SAR models that introduces a network formation process/probability as a prior distortion on network $\mathbf{W}$.
- This approach brings two main benefits:
  1. It is a natural extension of high-dimensional methods to more complex restrictions.
  2. It has more economic interpretability compared to common high-dimensional restrictions and can be easily extended for different scenarios.
- The main difficulty lies in efficient sampling algorithms and a data set that is sufficiency informative.
- The effect of the most interest λ cannot be reliably recovered.

ERGM Terms

Term	Description
`edges`	number of edges
`kstar(2)`	number of 2-stars
`kstar(3)`	number of 3-stars
`triangle`	number of triangles
`ctriad`	number of cyclic triples
`ttriad`	number of transitive triples

Table: Implemented ERGM terms.

Loss Functions

1. A common loss function is the mean-square loss.
   This loss corresponds to the posterior mean decision.
2. A more reasonable loss for in our case for $\mathbf{W}$ is miss-categorization loss:

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

   This loss corresponds to the decision:

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