On the Bayesian Estimation for ERGM

Bo-An Chen, Hsian-Ming Huang

December 12, 2023

Final Presentation for Bayesian Statistical Methods

Table of contents

Introduction

Main Research Questions

- Is double-Metropolis Hasting (DMH) algorithm *implementable* for exponential random graph models (ERGM)?
- What are the consequences of *insufficient* model selections?

Motivation

- Network formation is Markov-dependent, and we want to understand the dynamics.
- Networks in ERGM formulation is a doubly intractable statistical model.
- Exchange algorithm (a Bayesian approach) is asymptotically consistent for ERGM, but inefficient.

Model Setup

ERGM is a model describing a graph $\mathbf{y}$ with a probability in a general exponential form forms:

$$\pi(\mathbf{y}|\theta) = \frac{1}{z(\theta)} \exp\left(\sum_{d=1}^{D} \theta_d s_d(\mathbf{y})\right) = \frac{\exp\left(\theta^{\mathrm{T}} \mathbf{s}(\mathbf{y})\right)}{z(\theta)} \tag{1}$$

where the sufficient statistics in our study are:

$$\theta^{\mathrm{T}} \mathbf{s}(\mathbf{y}) = \theta_\nu \nu(\mathbf{y}) + \theta_\rho \rho(\mathbf{y}) + \theta_\tau \tau(\mathbf{y})$$

where

- $\nu(\mathbf{y}) = \sum_{i,j} y_{ij}$ # of edges
- $\rho(\mathbf{y}) = \sum_i \sum_{k>j} \sum_{j \neq i,k} y_{ij} y_{jk}$ # of two-stars
- $\tau(\mathbf{y}) = 3 \sum_i \sum_{k>j} \sum_{j \neq i,k} y_{ij} y_{jk} y_{ki}$ # of triangles

Previous & Potential Approaches

- MCMC estimation (Snijders, 2002)
- DMH & Exchange algorithm approach (Caimo and Friel, 2013)
- Approximate Bayesian Computation (Yin and Butts, 2020)

Our Model and Algorithms

DMH Algorithm

1. **for** $t = 1$ **to** T **do**
2. Generate $\theta' \sim h(\cdot|\theta)$
3. Sample an auxiliary variable $\mathbf{y}' \sim \pi(\mathbf{y}'|\theta')$ using an exact sampler
4. Compute
$$r(\theta, \theta', \mathbf{y}'|\mathbf{y}) = \frac{\pi(\mathbf{y}|\theta')\pi(\theta')}{\pi(\mathbf{y}|\theta)\pi(\theta)} \frac{\pi(\mathbf{y}'|\theta)}{\pi(\mathbf{y}'|\theta')} \frac{h(\theta|\theta', \mathbf{y})}{h(\theta'|\theta, \mathbf{y})}$$
5. Draw $u \sim \textbf{Uniform}(0, 1)$
6. **if** $u < r$ **then** set $\theta = \theta'$
7. **end for**

The normalizing constant $z(\theta)$ is canceled by introducing $\pi(\mathbf{y}'|\theta)$. This is the exchange algorithm proposed by Liang (2010).

Exact Sampler: Markov dependencies

On a tie level, we can describe the probability to form a chosen edge between note i and j as:

$$logit(y_{ij} = 1 | y_{ij}^c) = \theta' \delta(y_{ij}) \quad (2)$$

where

- y_{ij} is a random variable state of the actor pair i, j.
- y_{ij}^c is the complement of y_{ij}
- $\delta(y_{ij})$ is the vector of the "change statistics", i.e. $\delta(y_{ij}) = g(y_{ij}^+) - g(y_{ij}^-)$, where g denotes the state of tie y_{ij}.

Exact Sampler: MH Algorithm

1. set the true parameter $\hat{\theta}$
2. for t = 1 to T do
3.   randomly choose i, j
4.   define $\delta(y_{ij})$ as

$$(1, deg(i) + deg(j) - 2 \times \sum_{i} \sum_{j<i} \sum_{k \neq i,j} y_{ik} y_{jk}, \sum_{i} \sum_{j<i} \sum_{k \neq i,j} y_{ik} y_{jk})$$

5.   compute $\hat{\theta}' \delta(y_{ij})$
6.   calculate

$$log(p) = (-1)^{y_{ij}} \cdot log(\frac{exp(\hat{\theta}' \delta(y_{ij}))}{1 - exp(\hat{\theta}' \delta(y_{ij}))}) \tag{3}$$

7.   if $\min(log(p), 0) > log(\text{Uni}(0, 1))$:
8. $y_{ij} = 1 - y_{ij}$

DMH Improvements i

1. A beta-binomial is first used to approach θ_0^{init}
2. Mixed adaptive MCMC is proposed by a linear combination of random walk and covariance.
3. Hamiltonian MCMC may not work as we cannot derive the gradient for the likelihood (even if we could, it cannot be applied across different phase spaces).
4. We reject highly degenerated networks, i.e. networks generated from any θ with an acceptance rate lower than 0.02.

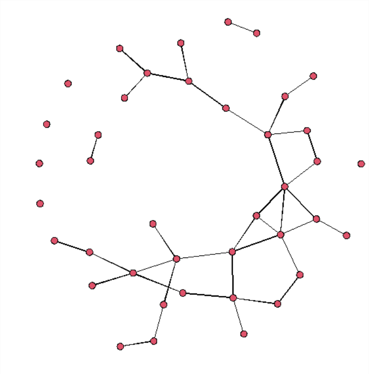
Simulation Results

Network generation and estimation

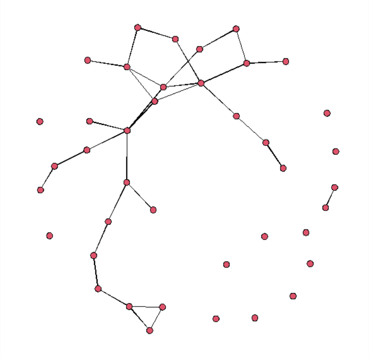
Given the parameter value, we use both `ergm` and our code to generate two networks.
We consider 4 cases of estimation:

1. `ergm` simulated network + `ergm` and `bergm` parameter estimation
2. `ergm` simulated network + our DMH code
3. our simulated network + our DMH code
4. our simulated network + `ergm` and `bergm` parameter estimation

Network Generation



(a) ergm. edge: 38; 2-star: 72; triangle: 3



(b) our code. edge: 32; 2-star: 57; triangle: 1

Figure 1: Parameter value: -3.5, 0.1, 0.5

Case1: `ergm` network + `ergm` and `bergm` estimation

ergm: estimate with Monte Carlo MLE

- R Code

```
model <- sim_network ~ edges + kstar(2) + triangle
est_ergm <- ergm(model)
summary(est_ergm)
```

- Result

Parameter	Mean	Std.
θ_0 (edge)	-2.8456	0.4797
θ_1 (2-star)	-0.0762	0.1339
θ_2 (triangle)	1.1726	0.6722

Case1: `ergm` network + `ergm` and `bergm` estimation

bergm: using exchange algorithm and parallel adaptive direction sampler to improve the mixing of Markov Chains (related to the gamma and nchains)

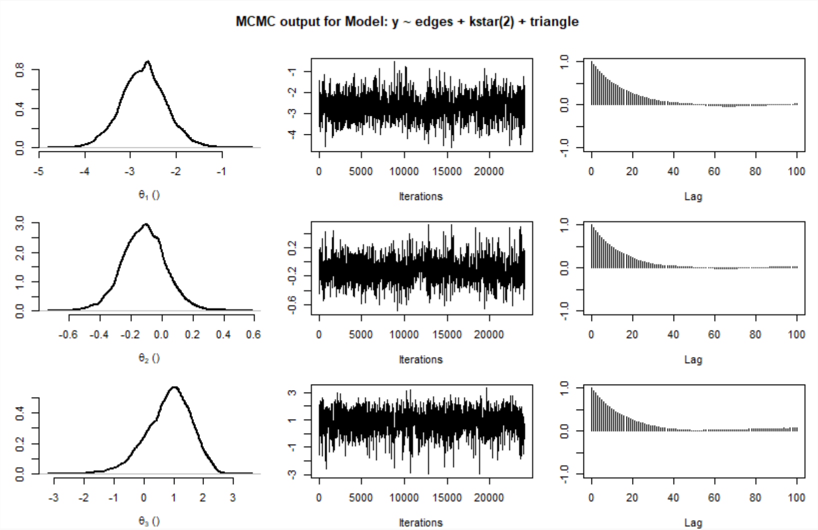
- R Code

```
model <- sim_network ~ edges + kstar(2) + triangle
est_bergm <- bergm(model, burn.in = 500,
                   main.iters = 3000, aux.iters = 2500,
                   nchains = 8, gamma = 0.6)
summary(est_bergm)
```

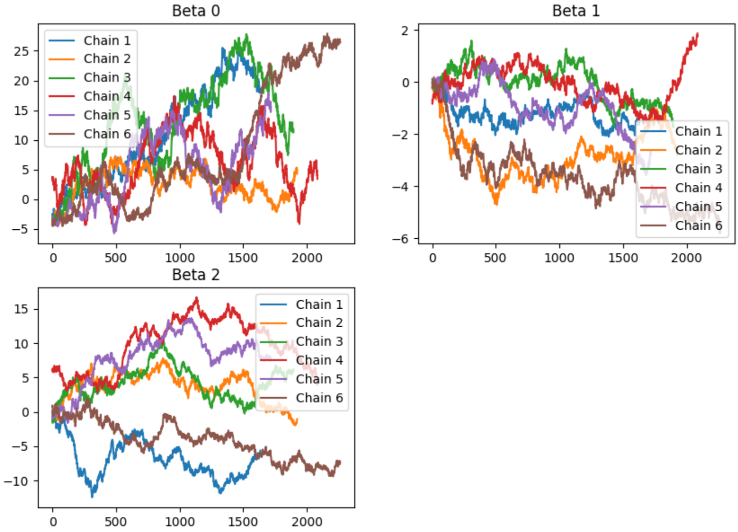
- Result

Parameter	Mean	Std.
θ_0 (edge)	-3.14444889	0.4874545
θ_1 (2-star)	-0.04660349	0.1659005
θ_2 (triangle)	-0.12884212	1.3766226

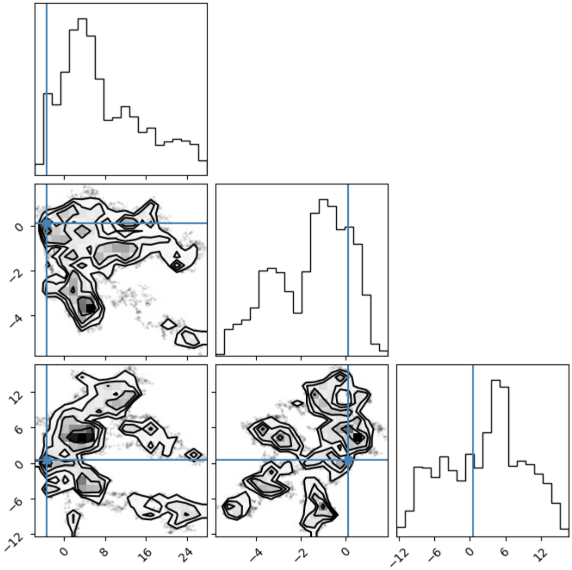
MCMC Diagnostic bergm



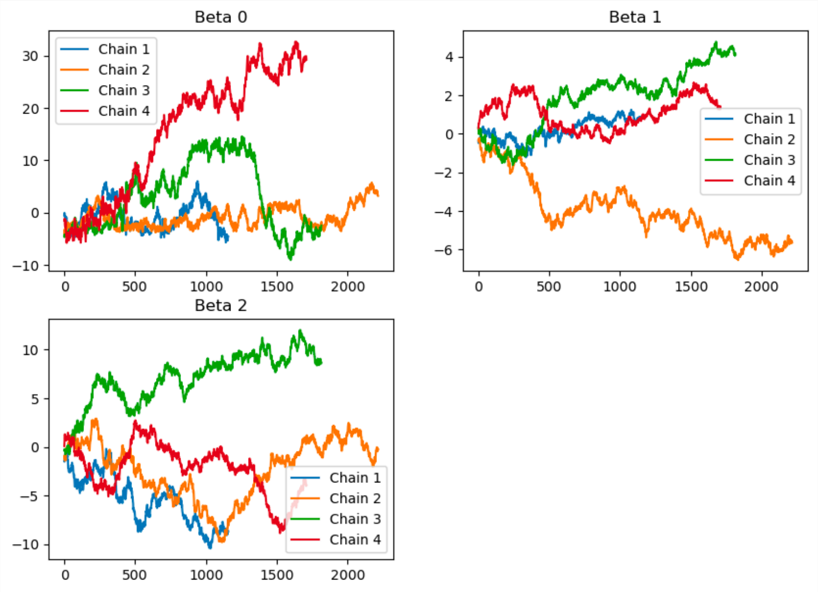
Case2: `ergm` simulated network + our DMH code



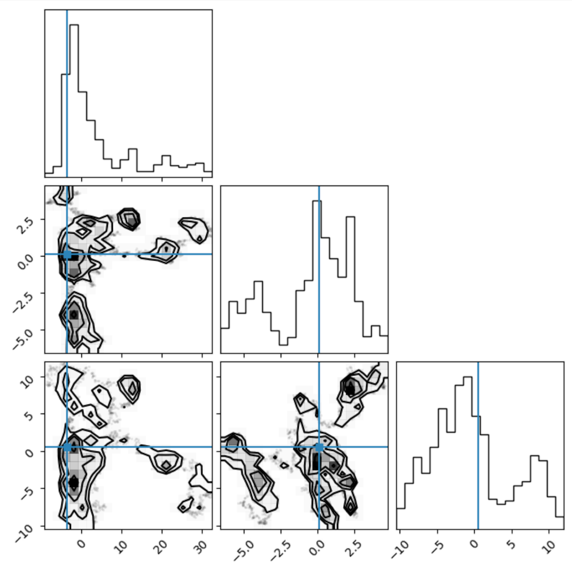
Case2: `ergm` simulated network + our DMH code



Case3: our simulated network + our DMH code



Case3: our simulated network + our DMH code



Case4: our simulated network + `ergm` and `bergm` estimation

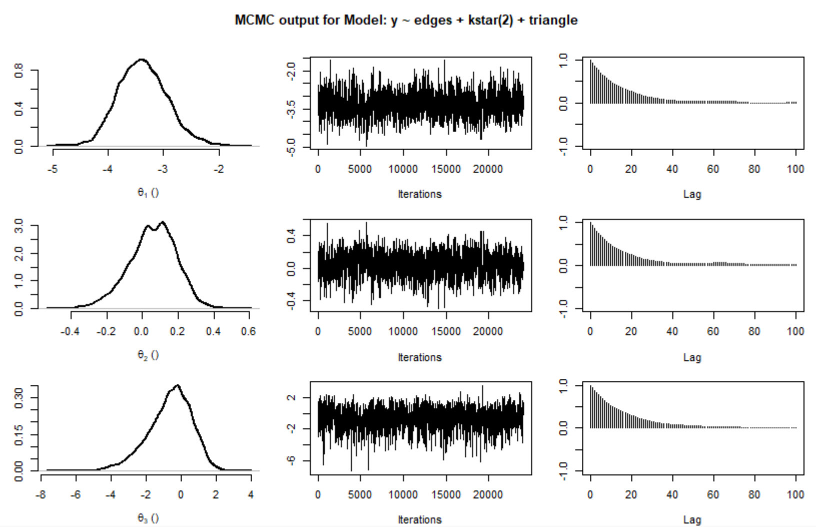
ergm

Parameter	Mean	Std.
θ_0 (edge)	-3.53935	0.41757
θ_1 (2-star)	0.11874	0.11842
θ_2 (triangle)	-0.04484	1.09843

bergm

Parameter	Mean	Std.
θ_0 (edge)	-3.36917773	0.4344070
θ_1 (2-star)	0.05895917	0.1324414
θ_2 (triangle)	-0.64722396	1.2797613

MCMC Diagnostic `bergm`



Discussions

Exact Sampler: Problems

- How many iterations are considered *asymptotically* large enough? 30000? 120000?
- Instability and degeneracies, i.e. different θ could lead to the same configurations (Schweinberger, 2012).
- Hunter et al. (2008) argues that the k-star and triangles are in nature degenerate, which implies the parameters in the model may be unidentifiable.

Exact Sampler: Realization

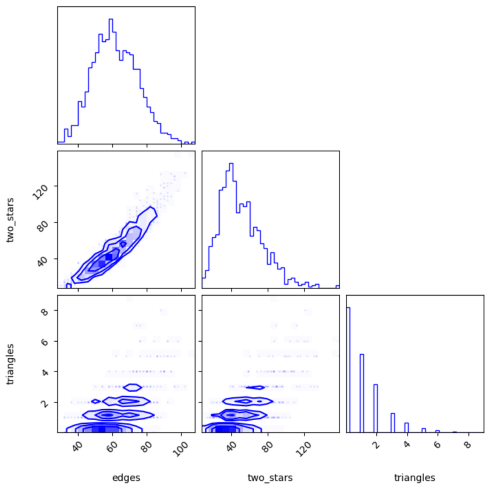


Figure 3: Distribution of sampled edges, two-stars, triangles from parameter: (-3.5, 0.1, 0.5). 1000 samples, 30000 iterations

The consequence of degeneracy i

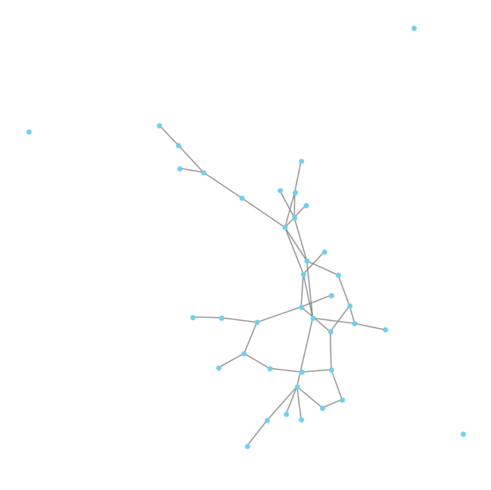


Figure 4: edges: 42, two-stars: 95, triangles: 1; Parameter: (-5, -0.2, 5)

The consequence of degeneracy ii

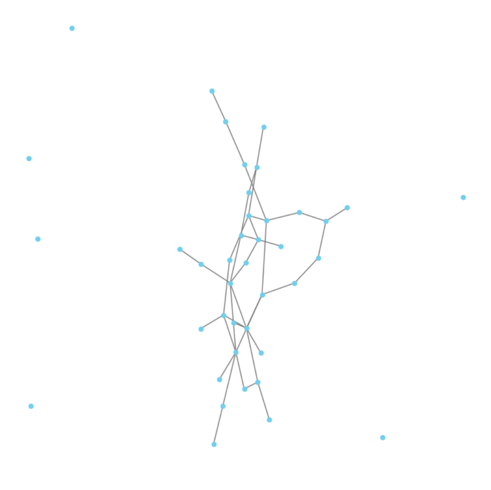


Figure 5: edges: 42, two-stars: 80, triangles: 1; Parameter: (-3.5, 0.1, 0.5)

DMH: Differences

- *bergm* has implemented Gibbs sampler for proposing new θ, while we used MH sampler
    - How to derive the marginal density?
    - Any accepted θ may be good enough?
- Our auxiliary network generating iterations is 30000, while we chose 2500 iterations for *bergm*. Which generates good-quality networks?

DMH Improvements ii

1. Hyperparameters for acceptance rate and step size are updated [under construction].
2. **Simulated annealing**: We define temperature as a metric related to the observed $s(y)$ and set up a tolerance for restarting and mixing, naively related to **approximate Bayes computation (ABC)** (Albert, Künsch, and Scheidegger, 2015) [under construction].

Future Works

- Is it possible to target the parameters in the formation mechanism directly via ABC? Could it avoid degeneracy (partially yes).
- Without nodal (heterogeneous) effects, the missing-not-at-random data may be applicable within this framework.
- Model selection either on the ERGM family or the dynamical system by Bayesian statistics should be further established.
- We will optimize the hyperparameter updating rules, as well as construct priors with empirical Bayes.
- Examine a version of the Gibbs sampler

Albert, C., H.R. Künsch, and A. Scheidegger (2015). "A simulated annealing approach to approximate Bayes computations". In: *Statistical Computing* 25, pp. 1217–1232.

Caimo, A. and N. Friel (2013). "Bayesian model selection for exponential random graph models". In: *Social Networks* 35.1, pp. 11–24. ISSN: 0378-8733.

Hunter, David R. et al. (2008). "ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks". In: *Journal of Statistical Software* 24.3.

Liang, Faming (Sept. 2010). "A double Metropolis–Hastings sampler for spatial models with intractable normalizing constants". In: *Journal of Statistical Computation and Simulation* 80, pp. 1007–1022.

Schweinberger, Michael (2012). **“Instability, Sensitivity, and Degeneracy of Discrete Exponential Families”.** In: *Journal of the American Statistical Association* 107.499, pp. 1361–1370.

Snijders, Tom A. B. (2002). **“Markov Chain Monte Carlo Estimation of Exponential Random Graph Models”.** In: *J. Soc. Struct.* 3.

Yin, Fan and Carter T. Butts (2020). *Kernel-based Approximate Bayesian Inference for Exponential Family Random Graph Models.* arXiv: `2004.08064 [stat.CO]`.