

# BSTERGM: Bayesian Separable-Temporal Exponential family Graph Models

Choi, Seokjun

14 Dec. 2020

## Abstract

Separable Temporal Exponential Random Graph Models (STERGM) are useful models to describe changes in dynamic networks' structures evolving over time. The Bayesian-inference based approach for the model, however, had not been developed yet. In this paper, I propose the Bayesian STERGM (BSTERGM): STERGM with the Bayesian setting, and I illustrate the computational algorithm that generates posterior samples from the model dealing with the doubly-intractable constant problem. Furthermore, I show the interpretability of the model in the Bayesian context.

## 1 Introduction

## 2 Background

### 2.1 Random graphs

For statistical analysis on relations between subjects of data, I will introduce terminologies and notations for denoting components related to networks, which are a representation of these type of data.

We say each subject on the data to a node. They are same as a vertex in graph theory. In addition, we say a tie joining two nodes to an edge. A tie can be directed or not. These two components have fundamental role describing the data. Define a graph or a network to  $y = (V, E)$ , where  $V$  is a set of node and  $E$  is a set of edges of a network.

For simplicity, I will deal with a situation without weight on each node or edge. Then, a graph  $y$  can be represented by simpler form. Assume that the number of node  $n \in \mathbb{N}$  on data is given. Then, for  $\mathcal{B} = \{0, 1\}$ , the set  $\mathcal{Y} \subset \mathcal{B}^{n^2}$  is a set of graphs of  $n$  nodes. Now, a graph  $y \in \mathcal{Y}$  can be represented by a  $n \times n$  binary matrix with elements  $y_{ij} = 1$  if there is a tie from  $i$ -th node to  $j$ -th node, otherwise  $y_{ij} = 0$ .

If edges of  $y$  have directions, then  $y$  is called a directed graph. Otherwise,  $y$  is called a undirected graph. These are obvious that, for given  $n \in \mathbb{N}$ ,  $|\mathcal{Y}| = 2^{n(n-1)}$  if  $\mathcal{Y}$  is the set of all directed graphs not permitting self-connecting edges, and  $|\mathcal{Y}| = 2^{n(n-1)/2}$  if  $\mathcal{Y}$  is the set of all graphs of undirected one. Thus, the size of  $\mathcal{Y}$  grows exponentially when  $n$  increases.

I extend the concept of graphs to be able to describe a situation that connections among nodes are random. Let  $\Omega$  be an event set of all possible connected situations among given nodes. We say  $Y : \Omega \rightarrow \mathcal{Y}$  is a random variable for a graph, or a random graph. For a random graph  $Y \in \mathcal{Y}$ , denote the edge between  $i$ -th node and  $j$ -th node by  $Y_{ij}$  for  $i, j = 1, 2, \dots, n$  satisfying  $Y_{ij} = 1$  if the edge is connected. Otherwise,  $Y_{ij} = 0$ . Note that the  $Y_{ij}$ s are also random variables.

Let me notate a realization of random graph by  $y$  and its edges by  $y_{ij}$  for  $i, j = 1, 2, \dots, n$ .

## 2.2 Exponential random graph model: ERGM

From the definition of random graph, one of the natural questions is how we can model a distribution on  $\mathcal{Y}$ . As a answer, the Exponential random graph models (ERGM) (Wasserman and Pattison(1996) and Robins et al.(2007)) gives a classical but flexible scheme to model for time-invariant networks represented as a single directed or undirected graph. Moreover, ERGM provide a statistical approach to the systematic exploration of several local features simultaneously and how these local structures interact to form a global network.

Specifically, following the notation of previous section, let  $\mathcal{Y}$  be a set of random graphs with  $n$  nodes and  $Y \in \mathcal{Y}$  be a random graph. Set a distribution on  $\mathcal{Y}$  to

$$P(Y = y; \theta) = \frac{\exp(\theta^T s(y))}{c(\theta)}$$

for some  $\theta \in \mathbb{R}^p$ , where  $s(y) \in \mathbb{R}^p$  is a vector which is part of  $y$ 's sufficient network statistics, and  $c(\theta) \in \mathbb{R}$  is a normalizing constant satisfying  $c(\theta) = \sum_{y \in \mathcal{Y}} \exp(\theta^T s(y))$ . Models which have such a form called the ERGM.

Note that the form of the right-hand side is the canonical form of exponential family distributions. That is why the model is called exponential. Moreover, observe that the normalizing constant cannot be directly calculated because the size of  $\mathcal{Y}$  grows exponentially as  $n$  grows. Thus in real data analysis, the calculation of  $c(\theta)$  should be avoided. For example, see the detail on fitting algorithm using MCMCMLE.

### 2.3 BERGM: Bayesian ERGM

Caimo and Friel(2012) extend ERGM to bayesian setting. Putting prior  $p(\theta)$  to the parameter  $\theta$ , the model becomes

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{\exp(\theta^T s(y)) p(\theta)}{c(\theta) p(y)}$$

Observe that not only  $c(\theta)$  but also the  $p(y)$  cannot be evaluated directly because of the difficulty of summation when  $n$  grows and of integration of  $p(y)$ . Moreover, the constant depends on the parameter  $\theta$ , so ordinary MCMC technique cannot be used and This situation is said to doubly-intractable constant problem, and causes serious challenge to estimate the parameter vector  $\theta$ . Thus, The estimation of  $\theta$  needs more advanced algorithm than the standard MCMC procedure. The solution of Caimo and Friel is using the exchange algorithm. See

### 2.4 TERGM: Temporal Exponential family Random Graphs Models

Let's consider a sequence of random graph  $\{Y_1, Y_2, \dots, Y_T\} (T \in \mathbb{N})$  and assume that we are interested in the dynamics or evolution of the network sequence. The temporal Temporal Exponential family Random Graphs Models(TERGM) proposed by Hanneke, Fu and Xing(2009) is a model for our interest if we permit a Markov assumption

$$P(Y_2, Y_3, \dots, Y_T | Y_1) = P(Y_2 | Y_1) P(Y_3 | Y_2) \dots P(Y_T | Y_{T-1})$$

and focus on the transition probability along index of the sequence. For convenience, I refer the index as a time.

Specifically, let  $\mathcal{Y}$  be a set of graphs with  $n$  nodes. Let  $Y_1 = y_1 \in \mathcal{Y}$  be given and  $Y_2, \dots, Y_T (T \in \mathbb{N})$  be random graphs in  $\mathcal{Y}$ . Set a distribution on  $\mathcal{Y} \times \dots \times \mathcal{Y}$  ( $T - 1$  folds) to

$$P(Y_t = y_t | Y_{t-1} = y_{t-1}; \theta) = \frac{\exp(\theta^T s(y_t, y_{t-1}))}{c(\theta, y_{t-1})}$$

where  $t = 2, \dots, T$ ,  $\theta \in \mathbb{R}^p$ , sufficient network statistics  $s(y_t, y_{t-1}) \in \mathbb{R}^n$ , and a normalizing constant  $c(\theta, y_{t-1}) = \sum_{y \in \mathcal{Y}} \exp(\theta^T s(y, y_{t-1}))$ . Models which have a such form called the TERGM.

The estimation of  $\theta$  can be performed by MCMCMLE, but the estimation suffers severe degeneracy problem. Furthermore, When we interpret the estimated TERGM parameters, we cannot separate the generation and duration because the higher value of  $\theta$  increases both. These raise demand for modified models to remedy the degeneracy and enable separable interpretation.

## 2.5 STERGM: Separable-Temporal Exponential family Random Graphs Models

One solution to remedy the problems of TERGM is the Separable-Temporal Exponential family Random Graphs Models, or simply STERGM. The key idea is to separate the formation models and dissolution model for decoupling an effect on edge's duration of model parameter from edge's incidence when setting a distribution on  $\mathcal{Y} \times \dots \times \mathcal{Y}$  ( $T - 1$  folds,  $T \in \mathbb{N}$ ), where  $\mathcal{Y}$  is a set of graphs with given  $n$  nodes.

Precisely, Let  $Y_1 = y_1 \in \mathcal{Y}$  be given and  $Y_2, \dots, Y_T$  be random graphs (on  $\mathcal{Y}$ ). Let  $\mathcal{Y}^+|_t$  be a subset of  $\mathcal{Y}$  consisting all graphs which have equal or additional edges comparing to  $y_{t-1}$ . Likewise, let  $\mathcal{Y}^-|_t$  be a subset of  $\mathcal{Y}$  consisting all graphs which have equal or sparse edges comparing to  $y_{t-1}$ . Next, for  $Y_t^+ : \Omega \rightarrow \mathcal{Y}^+|_t$ ,  $Y_t^- : \Omega \rightarrow \mathcal{Y}^-|_t$ ,  $y_t^+ \in \mathcal{Y}^+|_t$  and  $y_t^- \in \mathcal{Y}^-|_t$ , set

$$P(Y_t^+ = y_t^+ | Y_{t-1} = y_{t-1}; \theta^+) = \frac{\exp((\theta^+)^T s(y_t^+, y_{t-1}))}{c(\theta^+, y_{t-1})}$$

$$P(Y_t^- = y_t^- | Y_{t-1} = y_{t-1}; \theta^-) = \frac{\exp((\theta^-)^T s(y_t^-, y_{t-1}))}{c(\theta^-, y_{t-1})}$$

for some  $\theta^+, \theta^- \in \mathbb{R}^p$ ,  $s(y_t^+, y_{t-1}), s(y_t^-, y_{t-1}) \in \mathbb{R}^n$ , which are parts of sufficient network statistics, and normalizers  $c(\theta^+, y_{t-1}) = \sum_{y^+ \in \mathcal{Y}^+} \exp((\theta^+)^T s(y^+, y_{t-1}))$ , and  $c(\theta^-, y_{t-1}) = \sum_{y^- \in \mathcal{Y}^-} \exp((\theta^-)^T s(y^-, y_{t-1}))$ .

Then, defining operations  $+$ ,  $-$  on  $\mathcal{Y}$  following the edgewise boolean algebra, (For detail, Define binary operations on the edge values  $\{0, 1\}$  by  $+$  and  $-$  by  $1 + 1 = 1, 1 + 0 = 0 + 1 = 1, 0 + 0 = 0$  and  $1 - 1 = 0, 1 - 0 = 1, 0 - 1 = 0, 0 - 0 = 0$ ) set the combined network  $y_t$  to

$$y_t = y_t^+ - (y_{t-1} - y_t^-) = y_t^- + (y_t^+ - y_{t-1})$$

Additionally, assume the first-order Markov assumption

$$P(Y_2, \dots, Y_T | Y_1) = P(Y_2 | Y_1) \dots P(Y_T | Y_{T-1})$$

and the conditional independence between  $Y_t^+$  and  $Y_t^-$  for all  $t = 2, \dots, T$ , called the separability. Then we get

$$P(Y_t = y_t | Y_{t-1} = y_{t-1}; \theta^+, \theta^-) = P(Y_t^+ = y_t^+ | Y_{t-1} = y_{t-1}; \theta^+) P(Y_t^- = y_t^- | Y_{t-1} = y_{t-1}; \theta^-)$$

Models which have a such form called the STERGM.

## 2.6 Network Statistics for ERGM and its derivative models

Now, I briefly discuss the terms of network statistics. As a usual,  $s(y)$  of ERGM or  $s(y_i, y_{t-1})$  of TERGM and STERGM would reflect the structure of given network without identifying each node index. Common candidates of the statistics are here.

Let  $y \in \mathcal{Y}$  and  $n \in \mathbb{N}$  be the number of nodes of  $y$ . The simplest candidate is the number of edges, denoted by  $|y|$ . Another common candidates are distributions defined by some sense of structure of given graph. For a given  $i, 1 \leq i \leq n-1$ ,  $D_i(y)$  is defined to be a number of nodes in  $y$  whose the number of edges incident to the node equals  $i$ , and called  $i$ -th order node degree distribution. For a given  $i, 1 \leq i \leq n-2$ ,  $P_i(y)$  is defined to be a number of dyads  $(j, k), 1 \leq j, k \leq n$  in  $y$  whose  $j$  and  $k$  have nodes connected by an edge each other and they share exactly  $i$  connected nodes in common, and called  $i$ -th order edgewise shared partner distribution. Note that, the bundle of all order of node degree distribution and all order of edgewise shared partner distribution,

$$(D_1(y), D_2(y), \dots, D_{n-1}(y), P_1(y), P_2(y), \dots, P_{n-2}(y))$$

forms sufficient statistics of a graph. And there are relation that

$$n = D_0(y) + D_1(y) + \dots + D_{n-1}(y)$$

and

$$|y| = D_0(y) + D_1(y) + \dots + D_{n-1}(y) = \frac{1}{2} \sum_{i=1}^{n-1} i D_i(y)$$

A function of the sufficient statistics can be candidates for the model statistics. They proposed graph statistics including  $k$ -star  $S_k(y)$  for  $1 \leq k \leq n-1$  and  $k$ -triangles  $T_k$  for  $1 \leq k \leq n-2$ . The  $k$ -star consists of a node together with a set of  $k$  of its connected nodes, and the  $k$ -triangle consists of  $k$  triangles that share one common edge. They can be expressed by the definition of node degree distributions and edgewise shared partner distributions,

$$S_1 = |y|, S_k = \sum_{i=1}^{n-1} \binom{i}{k} D_i(y), k \geq 2$$

and

$$T_1 = \frac{1}{3} \sum_{i=1}^{n-2} i P_i(y), T_k = \sum_{i=1}^{n-2} \binom{i}{k} P_i(y), k \geq 2$$

Another interesting statistics are geometrically weighted value of distributions defined earlier. For example, geometrically weighted (node) degree

with parameter  $\tau$  is defined by

$$GWD(y; \tau) = e^\tau \sum_{k=1}^{n-2} (1 - (1 - e^{-\tau})^k) D_k(y)$$

and geometrically weighted shared partner distribution with parameter  $\tau$  is defined by

$$GWESP(y; \tau) = e^\tau \sum_{k=1}^{n-2} (1 - (1 - e^{-\tau})^k) P_k(y)$$

In the context of temporal models, parameters in the model take two network as arguments like  $s(y_t, y_{t-1})$ . One of common choices for these statistics is the bundle of sufficient statistics of first argument  $y_t$ , or Although there are lots of other statistics related to ERGM, I omit them and end this subsection here.

### 3 Bayesian STERGM

A goal is to convert the STERGM to the Bayesian setting. With priors  $p(\theta^+), p(\theta^-)$  over  $\theta^+, \theta^-$ , the bayesian model for STERGM can be made. I call this model Bayesian Separable Temporal Exponential family Random Graph Model, or BSTERGM. More explicitly, for given the number of nodes  $n$ , let  $Y_1 = y_1 \in \mathcal{Y}$  be given and  $Y_2, \dots, Y_T$  be random graphs (of  $\mathcal{Y}$ ). As STERGM, define  $\mathcal{Y}^+|_t, \mathcal{Y}^-|_t$  be subsets of  $\mathcal{Y}$  consisting all graphs which have equal or additional edges comparing to  $y_{t-1}$ , and which have equal or sparse edges comparing to  $y_{t-1}$  respectively. Then, for  $Y_t^+$  of  $\mathcal{Y}^+|_t$  and  $Y_t^-$  of  $\mathcal{Y}^-|_t$ ,  $y_t^+ \in \mathcal{Y}^+|_t$  and  $y_t^- \in \mathcal{Y}^-|_t$ , set

$$P(Y_t^+ = y_t^+ | Y_{t-1} = y_{t-1}; \theta^+) = \frac{\exp((\theta^+)^T s(y_t^+, y_{t-1}))}{c(\theta^+, y_{t-1})}$$

$$P(Y_t^- = y_t^- | Y_{t-1} = y_{t-1}; \theta^-) = \frac{\exp((\theta^-)^T s(y_t^-, y_{t-1}))}{c(\theta^-, y_{t-1})}$$

for some  $\theta^+, \theta^- \in \mathbb{R}^p$ ,  $s(y_t^+, y_{t-1}), s(y_t^-, y_{t-1}) \in \mathbb{R}^n$ , which are parts of sufficient network statistics, and normalizers  $c(\theta^+, y_{t-1}) = \sum_{y^+ \in \mathcal{Y}^+} \exp((\theta^+)^T s(y^+, y_{t-1}))$ , and  $c(\theta^-, y_{t-1}) = \sum_{y^- \in \mathcal{Y}^-} \exp((\theta^-)^T s(y^-, y_{t-1}))$ .

Then, with the edgewise boolean algebra, set the combined network  $y_t$  to

$$y_t = y_t^+ - (y_{t-1} - y_t^-) = y_t^- + (y_t^+ - y_{t-1})$$

and assume that the first-order Markov property

$$P(Y_2, \dots, Y_T | Y_1) = P(Y_2 | Y_1) \dots P(Y_T | Y_{T-1})$$

and the conditional independence(separability) between  $Y_t^+$  and  $Y_t^-$  for all  $t = 2, \dots, T$ . Then we get

$$\theta^+ \sim p(\theta^+), \theta^- \sim p(\theta^-)$$

$$P(Y_t = y_t | Y_{t-1} = y_{t-1}; \theta^+, \theta^-) = P(Y_t^+ = y_t^+ | Y_{t-1} = y_{t-1}; \theta^+) P(Y_t^- = y_t^- | Y_{t-1} = y_{t-1}; \theta^-)$$

The purpose of Bayesian inference is to get information of the posterior distribution of  $\theta^+, \theta^-$ . By the Bayes rule, the posterior can be expressed by

$$P(\theta^+, \theta^- | y_t, y_{t-1}) = \frac{P(Y_t^+ = y_t^+ | y_{t-1}, \theta^+) P(Y_t^- = y_t^- | y_{t-1}, \theta^-) P(\theta^+) P(\theta^-)}{c(\theta^+, y_{t-1}) c(\theta^-, y_{t-1}) p(y_t | y_{t-1})}$$

where  $y_t = y_t^+ - (y_{t-1} - y_t^-) = y_t^- + (y_t^+ - y_{t-1})$  as set above.

Note that, BSTERGM also have similar kinds of problems of BERGM. The normalizing constants  $c(\theta^+, y_{t-1})$ ,  $c(\theta^-, y_{t-1})$  practically cannot be computed because we need to sum up too many terms. The constants are doubly intractable: they depend on  $\theta^+, \theta^-$ , the parameters of a model. Thus, we cannot use an ordinary MCMC algorithm to get the posterior sample in this model, because the constant part remains when calculating the MCMC ratio.

## 4 Bayesian Estimation

Suppose that a sequence of graph samples  $y_1, \dots, y_T$  is observed. (then,  $y_2^+, \dots, y_T^+$  and  $y_2^-, \dots, y_T^-$  are uniquely determined.) Then how can we find the posterior distribution of  $\theta^+, \theta^-$  with the observation using BSTERGM? To solve this doubly-intractable constant problem, I propose an algorithm using the exchange MCMC algorithm, one of the special MCMC techniques.

The idea is running two MCMC chain to make independent sample to cancel the constant parts when calculating the MCMC accept-reject ratio. To begin with, consider the MCMC ratio in the BSTERGM context. With symmetric proposal distribution,

$$\pi = \frac{P(y_t^+ | y_{t-1}, \theta_*^+) P(y_t^- | y_{t-1}, \theta_*^-) p(\theta_*^+) p(\theta_*^-)}{P(y_t^+ | y_{t-1}, \theta_{m-1}^+) P(y_t^- | y_{t-1}, \theta_{m-1}^-) p(\theta_{m-1}^+) p(\theta_{m-1}^-)}$$

where  $\theta^*$  is proposed parameter in an iteration of MCMC. Plugging specific model form to above equation,

$$\pi =$$

**Algorithm 4.1** (the main chain). Let  $y_1, \dots, y_T$  be given. For  $m = 1, \dots, M$ ,

1. Propose candidates  $\theta_*^+, \theta_*^-$  from  $\epsilon(\cdot | \theta_{m-1}^+, \theta_{m-1}^-)$ .

2. Select a lag  $(t-1, t)$  randomly on  $2 \leq t \leq T$ .
3. Generate an exchange graph  $y_{ex} \in \mathcal{Y}|_t$  (with  $y_{ex}^+, y_{ex}^-$ ) at the  $\theta_*^+, \theta_*^-$ .
4. Calculate the exchange MCMC ratio  $\pi$  at the lag,

$$\pi = \frac{P(y_t^+|y_{t-1}, \theta_*^+)P(y_t^-|y_{t-1}, \theta_*^-)p(\theta_*^+, \theta_*^-)}{P(y_t^+|y_{t-1}, \theta_{m-1}^+)P(y_t^-|y_{t-1}, \theta_{m-1}^-)p(\theta_{m-1}^+, \theta_{m-1}^-)} \frac{P(y_{ex}^+|y_{t-1}, \theta_{m-1}^+)P(y_{ex}^-|y_{t-1}, \theta_{m-1}^-)}{P(y_{ex}^+|y_{t-1}, \theta_*^+)P(y_{ex}^-|y_{t-1}, \theta_*^-)}$$

5. With probability  $\min(\pi, 1)$ , accept the proposal and put  $(\theta_m^+, \theta_m^-) = (\theta_*^+, \theta_*^-)$ .  
Otherwise, reject the proposal and put  $(\theta_m^+, \theta_m^-) = (\theta_{m-1}^+, \theta_{m-1}^-)$ .

Observe that, the  $\pi$  has no normalizing constant terms because they are canceled out by added exchange terms.  $\log \pi = (\theta_*^+ - \theta_{m-1}^+)(s(y_t^+, y_{t-1}) - s(y_{ex}^+, y_{t-1})) + (\theta_*^- - \theta_{m-1}^-)(s(y_t^-, y_{t-1}) - s(y_{ex}^-, y_{t-1})) + \log \frac{P(\theta_*^+, \theta_*^-)}{P(\theta_{m-1}^+, \theta_{m-1}^-)} = (\theta_*^+ - \theta_{m-1}^+)(s'(y_t^+) - s'(y_{ex}^+)) + (\theta_*^- - \theta_{m-1}^-)(s'(y_t^-) - s'(y_{ex}^-)) + \log \frac{P(\theta_*^+, \theta_*^-)}{P(\theta_{m-1}^+, \theta_{m-1}^-)}$

We need one more thing. To generate the exchange graph  $y_{ex}$  for time  $t$  (of the second part of main chain), we should have a generative algorithm at a given parameter points. However, we still do not know the normalizing constant. Thus, I use one more MCMC chain.

**Algorithm 4.2** (the auxiliary chain). Let  $\theta^+, \theta^-, y_{t-1}$  be given. For  $k = 1, \dots, K$ ,

1. Select one edge randomly from the  $y_{k-1}$ , say,  $y_{k-1;ij}$ .
2. Propose a new graph  $y_*$  with switching the  $y_{ij}$  value from  $y_{k-1}$ :  
If  $y_{k-1;ij} = 1$ , then set  $y_{*;ij} = 0$ (dissolution case.) Otherwise, If  $y_{k-1;ij} = 0$ , then set  $y_{*;ij} = 1$ (formation case.)  
(If sample graphs are undirected, switch  $y_{ji}$  simultaneously.)
3. Take  $\theta$  as  $\theta^+$  or  $\theta^-$  according to the case. Calculate the MCMC ratio  $\phi$ ,

$$\phi = \frac{P(Y_t = y_*|y_{t-1}, \theta)}{P(Y_t = y_{k-1}|y_{t-1}, \theta)} = \frac{\exp(\theta^T s(y_*, y_{t-1}))}{\exp(\theta^T s(y_{k-1}, y_{t-1}))} = \exp(\theta^T (s'(y_*) - s'(y_{k-1})))$$

4. With probability  $\min(\phi, 1)$ , accept the proposal and put  $y_k = y_*$ .  
Otherwise, reject the proposal and put  $y_k = y_{k-1}$ .

After  $K$  iteration, use the last network as the exchange sample in the second part of the main algorithm.



## 5 Diagnostics, inference, and Bayesian GOF

### 5.1 Model diagnostics

Since we run two kinds of MCMC chains, we should proceed two diagnostic task.

The main chain produces the posterior samples of parameter, so the procedure is same as ordinary MCMC case.

- Cut burn-in period. Do thinning if it is needed.
- Depict traceplots of each parameter chain to check the convergence and the mixing.
- Depict autocorrelation plot. Calculate ESS if it is needed.

Next, checking all auxiliary chains is practically irritating one. In general, it is suffice to check the auxiliary chain of the last iteration (of the main chain) with statistics included in the model.

- Calculate network statistics of all graphs produced by the last auxiliary chain.
- Depict traceplots of each statistics to check the convergence and the mixing.

### 5.2 Model inference

Since we have posterior samples by running the main chain as many as we want, we can do basic inference procedure.

- A outlining shape of the posterior of  $\theta^+, \theta^- | y_1, y_2, \dots, y_T$  by histogram.
- An approximated summary statistics: mean, mode, variance, ...
- An approximated quantile and probability interval

Moreover, we already have a generative algorithm at the specific parameter point, we can predict the form of network at  $T + 1, T + 2, \dots$  using posterior sample following standard Bayesian method. For example, for predicting  $T + 1$ ,

- Run K iteration using auxiliary chain at  $y_T$  with each posterior sample points.
- Take the each network as a predicted result.

If you need, calculate some network statistics for the results and get a summary statistics of them.

### 5.3 Goodness of Fit

To evaluate the goodness of fit for the posterior  $\theta^+, \theta^-$ , we can use the auxiliary chain algorithm once again.

**Algorithm 5.1** (GOF procedure). For  $t=2, \dots, T$   
For  $s=1, \dots, S$

1. sample  $\theta_s^+, \theta_s^-$  from the estimate of posterior.
2. simulate  $y_s$  using the auxiliary chain under  $y_{t-1}$ .
3. calculate  $g(y_s)$ , some higher degree statistics (eg. Node-degree dist & Edgewise Shared Partner dist)

Draw the box-plot of  $g(y_s)$  and compare with  $g(y_t)$ .

## 6 Example: Scholars and faculty members' experiences in online social networks

## 7 Summary and Discussion

### Supplements

You can find the C++ implementation (using Armadillo: see <http://arma.sourceforge.net/>) of BSTERGM fitting, diagnostic, and GOF algorithms at my Github page: <https://github.com/letsjdosth/BayesianSTERGM>.

testcite [Hunter et al., 2008b]

### References

- [RN1, 2013] (2013).
- [Butts, 2008a] Butts, C. T. (2008a). 4. a relational event framework for social action. *Sociological Methodology*, 38(1):155–200.
- [Butts, 2008b] Butts, C. T. (2008b). network: a package for managing relational data in r. *Journal of statistical software*, 24(2):1–36.
- [Caimo and Friel, 2011] Caimo, A. and Friel, N. (2011). Bayesian inference for exponential random graph models. *Social Networks*, 33(1):41–55.
- [Caimo and Friel, 2013] Caimo, A. and Friel, N. (2013). Bayesian model selection for exponential random graph models. *Social Networks*, 35(1):11–24.

- [Frank, 1991] Frank, O. (1991). Statistical analysis of change in networks. *Statistica Neerlandica*, 45(3):283–293.
- [Frank and Strauss, 1986] Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81(395):832–842.
- [Geyer and Thompson, 1992] Geyer, C. J. and Thompson, E. A. (1992). Constrained monte carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 54(3):657–683.
- [Handcock, 2003] Handcock, M. S. (2003). *Statistical models for social networks: Inference and degeneracy*. na.
- [Handcock and Gile, 2010] Handcock, M. S. and Gile, K. J. (2010). Modeling social networks from sampled data. *The Annals of Applied Statistics*, 4(1):5.
- [Handcock et al., 2008] Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., and Morris, M. (2008). statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of statistical software*, 24(1):1548.
- [Holland and Leinhardt, 1977] Holland, P. W. and Leinhardt, S. (1977). A dynamic model for social networks. *Journal of Mathematical Sociology*, 5(1):5–20.
- [Hunter, 2007] Hunter, D. R. (2007). Curved exponential family models for social networks. *Social networks*, 29(2):216–230.
- [Hunter et al., 2008a] Hunter, D. R., Goodreau, S. M., and Handcock, M. S. (2008a). Goodness of fit of social network models. *Journal of the American Statistical Association*, 103(481):248–258.
- [Hunter and Handcock, 2006] Hunter, D. R. and Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15(3):565–583.
- [Hunter et al., 2008b] Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., and Morris, M. (2008b). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of statistical software*, 24(3):nihpa54860.
- [Jackson and Watts, 2002] Jackson, M. O. and Watts, A. (2002). The evolution of social and economic networks. *Journal of economic theory*, 106(2):265–295.
- [Koehly et al., 2004] Koehly, L. M., Goodreau, S. M., and Morris, M. (2004). Exponential family models for sampled and census network data. *Sociological Methodology*, 34(1):241–270.

- [Koskinen, 2004] Koskinen, J. (2004). Bayesian analysis of exponential random graphs-estimation of parameters and model selection. Report, Research Report 2004: 2, Department of Statistics, Stockholm University.
- [Koskinen et al., 2010] Koskinen, J. H., Robins, G. L., and Pattison, P. E. (2010). Analysing exponential random graph (p-star) models with missing data using bayesian data augmentation. *Statistical Methodology*, 7(3):366–384.
- [Krivitsky and Handcock, 2014] Krivitsky, P. N. and Handcock, M. S. (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 76(1):29.
- [Krivitsky et al., 2011] Krivitsky, P. N., Handcock, M. S., and Morris, M. (2011). Adjusting for network size and composition effects in exponential-family random graph models. *Statistical methodology*, 8(4):319–339.
- [Morris et al., 2008a] Morris, M., Handcock, M. S., and Hunter, D. R. (2008a). Specification of exponential-family random graph models: terms and computational aspects. *Journal of statistical software*, 24(4):1548.
- [Morris et al., 2008b] Morris, M., Handcock, M. S., and Hunter, D. R. (2008b). Specification of exponential-family random graph models: terms and computational aspects. *Journal of statistical software*, 24(4):1548.
- [Robins and Pattison, 2001] Robins, G. and Pattison, P. (2001). Random graph models for temporal processes in social networks. *Journal of Mathematical Sociology*, 25(1):5–41.
- [Robins et al., 2007] Robins, G., Pattison, P., Kalish, Y., and Lusher, D. (2007). An introduction to exponential random graph (p\*) models for social networks. *Social networks*, 29(2):173–191.
- [Snijders, 2001] Snijders, T. A. (2001). The statistical evaluation of social network dynamics. *Sociological methodology*, 31(1):361–395.
- [Snijders, 2005] Snijders, T. A. (2005). Models for longitudinal network data. *Models and methods in social network analysis*, 1:215–247.
- [Snijders et al., 2007] Snijders, T. A., Steglich, C. E., and Schweinberger, M. (2007). Modeling the co-evolution of networks and behavior. *Longitudinal models in the behavioral and related sciences*, 31(4):41–71.
- [Snijders et al., 2010] Snijders, T. A., Van de Bunt, G. G., and Steglich, C. E. (2010). Introduction to stochastic actor-based models for network dynamics. *Social networks*, 32(1):44–60.
- [Wasserman and Pattison, 1996] Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks: I. an introduction to markov graphs andp. *Psychometrika*, 61(3):401–425.