# BSTERGM: Bayesian Separable-Temporal Exponential family Graph Models

Choi, Seokjun

14 Dec. 2020

**Abstract**

Separable Temporal Exponential Random Graph Models (STERGMs) 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 STERGMs (BSTERGMs): STERGMs 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

Nowadays, new kinds of data having special forms emerge in many field. These new data are not confined by a series of values of real numbers as traditional data are in the past. The data may have non-number type of value, special structure, or even integrated different type of values with model-specific relation. These data are called complex data, and an attention of statistical analysis on the data increases over time.

As one form of the complex data, networks get highlighted attention among complex data for statistical analysis. Many phenomena can be represented by some relational structure, and the structure can be expressed effectively by a form of a network. For example, social networks including both offline and online can be thought as a sort of networks. In this view, some social aspects of relation between actors can be modeled with an object of social science including an analysis of power structure and a summary of intimacy structure on given group. Not only that, application to a transmitted disease leads a modeling using real data in epidemiology.

Since statistical models on network provide not only lots of possibility of application but also interesting and useful features on observed network,

statistical models have important role. The statistical model gives information about local relations like formation or dissolution between two actors as well as global features of the whole network such as probability that a specific structure realizes. Not only that, the statistical models investigate a complex behavior of network including how the global features of an observed network may be related to local network structure. During the investigation, statistical model can capture existing dependency on a given network and provide a tool to understand what certain structural features are more common on rare.

Exponential-family random graph models(ERGMs) may be the most useful tools to analyze cross-sectional networks and their dependencies. [Frank and Strauss, 1986], [Wasserman and Pattison, 1996], [Snijders, 2002] and [Robins et al., 2007] proposed the models and apply them to analyze dependent structures, and the Bayesian approaches of the models were started by [Caimo and Friel, 2011] and [Caimo and Friel, 2013]. In the ERGMs, The dependencies are captured using a series of local features called network statistics from classical graph-theory. Since these models play important role in this paper, I will introduce more precisely.

Next natural question is a way to model a network sequence, or the evolution of network over time. For instance, one may want to model social network structures for 1 year, or simulation of a disease infection based on the sequential observations in the past. The first exploration applying ERGMs was [Robins and Pattison, 2001]. After them, a notable contributions were of [Hanneke and Xing, 2006] and [Hanneke et al., 2010], the Temporal ERGMs(TERGMs). In the TERGM, ERGMs are used by modelling the transition probability between times. Moreover, [Krivitsky and Handcock, 2014] extend the TERGMs to the Separable Temporal ERGMs(STERGMs), considering an separation between duration - how long ties tend to last - and incidence - the rate at which new ties are formed - from prevalence - for example, the total number of ties - of the network properties. Imposing a conditional independence condition between the duration and incidence, the models can be extended effectively using similar fitting methods with TERGMs without adding much complexity.

Although the STERGMs give us strength when it comes to interpret the fitted parameter, a Bayesian approach for the STERGMs has not been conducted. It is partly because of a hardship coming from the huge computation in a parameter estimation method for the model. Nonetheless, the Bayesian approach of STERGMs provides the statistical inference method following the standard Bayesian way, so the approach has huge advantage of convenience for statistical works for fitted parameter and predictions to next time network structure. Thus, in this paper, I specify the bayesian separable temporal exponential random graph models(BSTERGMs) and propose a fitting

algorithm for the model parameter. I also discuss ways of inference, the diagnostics procedure, and checking the goodness of fit.

In chapter 2, I review the background meterials including ERGMs, BERGMs, TERGMs, and STERGMs, and in chapter 3, I show the detail setting of BSTERGMs. In chapter 4, I propose the fitting method for BSTERGM. In chapter 5, I explain ways to conduct diagnostics, statistical inference, and goodness of fit check. In chapter 6, I illustrate an application of BSTERGMs to network data of scholars and faculty from the UCLA medical working project team. Finally, I discuss some extension and ways to supplement the BSTERGM further for future research.

## 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 $Y : \Omega \to \mathcal{Y}$ 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 $i, j = 1, 2, ..., n$ 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 $i, j = 1, 2, ..., n$.

## 2.2 ERGMs: Exponential random graph models

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 (ERGMs) ( [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, ERGMs 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 ERGMs.

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 more detail of fitting algorithm, see [Hunter and Handcock, 2006] the detail on fitting algorithm using MCMCMLE. You can also see [Hunter et al., 2008b] addressing the ergm package in R with the MCMCMLE or the classic method using psudolikelihood.

## 2.3 BERGMs: Bayesian ERGMs

[Caimo and Friel, 2011] and [Caimo and Friel, 2013] extend ERGMs to bayesian setting. Putting prior $p(\theta)$ to the parameter $\theta$, the model becomes

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

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. This situation is said to doubly-intractable constant problem( [Koskinen, 2004]), 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 [Caimo and Friel, 2011] for details of fitting methods of BERGMs and [Murray et al., 2012] for the exchange algorithm.

## 2.4  TERGMs: Temporal Exponential family Random Graphs Models

Let's consider a sequence of random graph $\{Y_1, Y_2, ..., Y_T\}(T \in \mathbb{N})$ and assume that we are interested in the dynamics or evolution of the network sequence. The temporal Temporal Exponential family Random Graphs Models(TERGMs) proposed by [Hanneke and Xing, 2006] and [Hanneke et al., 2010] is a model for our interest if we permit a Markov assumption

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

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 $Y_2, ..., Y_T(T \in \mathbb{N})$ be random graphs in $\mathcal{Y}$. Set a distribution on $\mathcal{Y} \times ... \times \mathcal{Y}$ $(T - 1$ 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 $t = 2, ..., T$, $\theta \in \mathbb{R}^p$, 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 TERGMs.

The estimation of $\theta$ can be performed by MCMCMLE, but the estimation suffers severe degeneracy problem. Furthermore, When we interpret the estimated TERGMs 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  STERGMs: Separable-Temporal Exponential family Random Graphs Models

One solution to remedy the problems of TERGMs is the Separable-Temporal Exponential family Random Graphs Models, or simply STERGMs. 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 $\mathcal{Y} \times ... \times \mathcal{Y}$ $(T - 1$ 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 $Y_2, ..., Y_T$ 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 $Y_t^+ : \Omega \to \mathcal{Y}^+|_t$, $Y_t^- : \Omega \to \mathcal{Y}^-|_t$, $y_t^+ \in \mathcal{Y}^+|_t$ 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

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

and the conditional independence between $Y_t^+$ and $Y_t^-$ for all $t = 2, ..., T$, 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 STERGMs.

## 2.6 Network Statistics for ERGMs and derived models

Now, I briefly discuss the terms of network statistics. As a usual, $s(y)$ of ERGMs or $s(y_i, y_{t-1})$ of TERGMs and STERGMs 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 $|y|$. Following [Hunter and Handcock, 2006], common candidates are distributions defined by some sense of structure of given graph. For a given $i, 1 \le i \le n - 1$, $D_i(y)$ 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 $i, 1 \le i \le n - 2$,

$P_i(y)$ 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,

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

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

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

and

$$|y| = P_0(y) + P_1(y) + ... + P_{n-2}(y) = \frac{1}{2} \sum_{i=1}^{n-1} i D_i(y)$$

A function of the sufficient statistics can be candidates for the model statistics. [Snijders et al., 2006] 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 by [Hunter, 2007], [Hunter and Handcock, 2006], [Snijders et al., 2006] and [Hunter et al., 2008a]. 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 a arguments like $s(y_t, y_{t-1})$. One of common choices for these statistics

7

is the bundle of sufficient statistics of first argument $y_t$. Another common choice is using a difference, $s(y_t, y_{t-1}) = s'(y_t) - s'(y_{t-1})$ where $s'$ is a network statistics of given network. Although there are lots of other statistics related to ERGMs, I omit them and end this subsection here. If you need, see [Morris et al., 2008a] for the ERGMs, and [Hanneke and Xing, 2006]'s chapter 2.1 for temporal the ERGMs.

# 3  Bayesian STERGMs

A goal is to convert the STERGMs to the Bayesian setting. With priors $p(\theta^+), p(\theta^-)$ over $\theta^+, \theta^-$, the Bayesian model for the STERGMs can be made. I call this model Bayesian Separable Temporal Exponential family Random Graph Models, or the BSTERGMs. More explicitly, for given the number of nodes $n$, let $Y_1 = y_1 \in \mathcal{Y}$ be given and $Y_2, ..., Y_T$ be random graphs (of $\mathcal{Y}$). As the STERGMs, 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

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

and the conditional independence(separability) between $Y_t^+$ and $Y_t^-$ for all $t = 2, ..., T$. 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, the BSTERGMs also have similar kinds of problems of the BERGMs. 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    Estimation

Suppose that a sequence of graph samples $y_1, ..., y_T$ is observed. (then, $y_2^+, ..., y_T^+$ and $y_2^-, ..., y_T^-$ are uniquely determined.) Then how can we find the posterior distribution of $\theta^+, \theta^-$ with the observation using the BSTERGMs? To solve the doubly-intractable constant problem, I propose an novel algorithm using the exchange algorithm, one of the special MCMC techniques by [Murray et al., 2012], extending the algorithm of the BERGMs'( [Caimo and Friel, 2011]). Although MCMC types of approach cannot give an analytic expression of posterior distribution, it gives posterior samples as much as we want, and using the samples we can conduct statistical inference on given data.

The key idea is running two MCMC chain to make independent sample to cancel the constant parts when calculating the MCMC accept-reject ratio. As a motivation, consider the ordinary MCMC ratio for $m$-th MCMC iteration in the BSTERGMs context to show doubly intractable constant problem explicitly.

With symmetric proposal distribution,

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

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

$$\pi_{ordinary} = \frac{\frac{exp(\theta_*^+ s(y_t^+, y_{t-1}))}{c(\theta_*^+, y_{t-1})} \frac{exp(\theta_*^- s(y_t^-, y_{t-1}))}{c(\theta_*^-, y_{t-1})} p(\theta_*^+)p(\theta_*^-)}{\frac{exp(\theta_{m-1}^+ s(y_t^+, y_{t-1}))}{c(\theta_{m-1}^+, y_{t-1})} \frac{exp(\theta_{m-1}^- s(y_t^-, y_{t-1}))}{c(\theta_{m-1}^-, y_{t-1})} p(\theta_{m-1}^+)p(\theta_{m-1}^-)}$$

Since each c depends on unique subscript of $\theta$ respectively, the constants are not canceled out.

If we can generate simulated samples $y_{ex}|y_{t-1}, \theta_{m-1}$ and $y_{ex}|y_{t-1}, \theta_*$ and calculate their probability following the BSTERGMs, this situation can be

resolved. At this point, applying the key idea mentioned above, multiply the probability terms in order to exclude the constant terms. Then we get the exchange algorithm's ratio,

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

$$= \frac{\frac{exp(\theta_*^+ s(y_t^+,y_{t-1}))}{c(\theta_*^+,y_{t-1})}\frac{exp(\theta_*^- s(y_t^-,y_{t-1}))}{c(\theta_*^-,y_{t-1})}p(\theta_*^+)p(\theta_*^-)}{\frac{exp(\theta_{m-1}^+ s(y_t^+,y_{t-1}))}{c(\theta_{m-1}^+,y_{t-1})}\frac{exp(\theta_{m-1}^- s(y_t^-,y_{t-1}))}{c(\theta_{m-1}^-,y_{t-1})}p(\theta_{m-1}^+)p(\theta_{m-1}^-)} \quad \frac{\frac{exp(\theta_{m-1}^+ s(y_{ex}^+,y_{t-1}))}{c(\theta_{m-1}^+,y_{t-1})}\frac{exp(\theta_{m-1}^- s(y_{ex}^-,y_{t-1}))}{c(\theta_{m-1}^-,y_{t-1})}}{\frac{exp(\theta_*^+ s(y_{ex}^+,y_{t-1}))}{c(\theta_*^+,y_{t-1})}\frac{exp(\theta_*^- s(y_{ex}^-,y_{t-1}))}{c(\theta_*^-,y_{t-1})}}$$

$$= \frac{exp(\theta_*^+ s(y_t^+,y_{t-1}))exp(\theta_*^- s(y_t^-,y_{t-1}))}{exp(\theta_{m-1}^+ s(y_t^+,y_{t-1}))exp(\theta_{m-1}^- s(y_t^-,y_{t-1}))} \frac{exp(\theta_{m-1}^+ s(y_t^+,y_{t-1}))exp(\theta_{m-1}^- s(y_{ex}^-,y_{t-1}))}{exp(\theta_*^+ s(y_{ex}^+,y_{t-1}))exp(\theta_*^- s(y_{ex}^-,y_{t-1}))} \frac{p(\theta_*^+)p(\theta_*^-}{p(\theta_{m-1}^+)p(\theta_n^-}$$

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

The last expression shows that all constants are canceled out. This finding leads us to the fitting algorithm for the BSTERGMs using the exchange algorithm.

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

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

2. Select a lag $(t-1,t)$ randomly on $2 \le t \le T$.

3. Generate an exchange graph $y_{ex} \in \mathcal{Y}|_t$ (with $y_{ex}^+, y_{ex}^-$) at the $\theta_*^+,\theta_*^-$.

4. Calculate network statistics $s(y_t^+,y_{t-1}), s(y_{ex}^+,y_{t-1}), s(y_t^-,y_{t-1}), s(y_{ex}^-,y_{t-1})$.

5. Calculate the exchange MCMC ratio $\pi$ at the lag,

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

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

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

10

If $s(.,.)$ and $\theta$ can be multi-dimension (say, of $\mathbb{R}^p$,) replace the multiplication of the exponent parts to dot product between two vectors. In addition, when implementing the algorithm, I strongly recommend to use $log\pi$ instead of $\pi$ to avoid an underflow problem.

$$log\pi = (\theta_*^+ - \theta_{m-1}^+)(s'(y_t^+) - s'(y_{ex}^+)) + (\theta_*^- - \theta_{m-1}^-)(s'(y_t^-) - s'(y_{ex}^-)) + log\frac{P(\theta_*^+, \theta_*^-)}{P(\theta_{m-1}^+, \theta_{m-1}^-)}$$

To generate the exchange graph $y_{ex}$ for time $t$ at the second part of main chain, a generative algorithm at a given parameter points is needed. However, since the normalizing constant is still unknown, we cannot generate the exchange algorithm directly using the model's expression of the distribution. Thus, I use one more MCMC chain to get the exchange samples without calculating the constants.

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

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.) Oterwise, 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

The ERGMs and all derived models of the ERGMs are suffered degeneracy problem: all nodes of exchange sample become fully connected or totally isolated as the MCMC chain runs. If the situation is worse, the problem cause an explosion to $\infty$ or $-\infty$ of the fitted parameters. The BSTERGMs are not exception. In other words, the diagnostic process is very important part for the BSTERGMs.

Since two MCMC chains are run in the fitting algorithm of the BSTERGMs, two diagnostic tasks should be proceeded.

The main chain produces the posterior samples of parameter $\{\theta_i^+\}_{i=1}^M$, $\{\theta_i^-\}_{i=1}^M$ in $\mathbb{R}^p, p \in \mathbb{N}$, which are not different with running ordinary MCMC algorithm. Depict traceplots of each parameter chain to check the convergence and the mixing. If needed, cut samples of burn-in period and do thinning. Plotting sample autocorrelation for each lag is also important. Calculate ESS if it is needed.

However, diagnostics of auxiliary chains is challenging. One reason of the difficulty is that there are $M$ auxiliary chains in our whole procedure. Practically, it is not possible to all the chains as the number of iteration $M$ of main chain is large. Thus here is a guideline as a compromise: it suffices to check only one auxiliary chain of the last iteration of the main chain. Second reason comes from the outputs' form of the chains. They are not vectors in $\mathbb{R}^p$ but network structures. Since we cannot draw traceplots of these network sequences, we need other ways for the diagnostics on them. One way is calculateing network statistics $s(.,.)$ in the model and drawing traceplots of them. On the traceplots, check whether each network statistics converges or not. In summary, for auxiliary chain,

- Calculate network statistics of all graphs produced by the last auxiliary chain.

- Depict traceplots of each statistics to check the convergence and the mixing.

If the network statistics do not converges, raise the number of iteration $K$ of each auxiliary chain and run the fitting algorithm again.

Here are some tips for remedy when facing a problem. If parameters explode to infinity the number of iteration of main chain becomes large, check whether the number of iteration of exchange chain is too small. Moreover, take account of the model itself. If you put linearly dependent network statistics simultaneously in the model, the parameter may not be identifiable. If you set the model with network statistics of hugh different scale but similar MCMC proposal distribution, the MCMC chain tends to fail to converge. For example, if you put k-star of specific order and edge count to the model at the same time but use same proposal like $N(0, \sigma^2)$ for some $\sigma^2 > 0$, MCMC chain does not converge easily because the parameter of edge count should be changed by relatively wide range even if the parameter of k-star moves slightly. Finally, the convergence of the BSTERGMs is very sensitive to initial point of main chain. If the main chain starts at the region of degeneracy in the parameter space, it is a hopeless one to converge the chain properly.

## 5.2 Model inference

Following standard inference methods of the Bayesian statistics, we can conduct the statistical inference on the parameters of the BSTERGMs. From the posterior samples generated by the exchange algorithm for the BSTERGMs, we can get

- A outlining shape of the posterior of $\theta^+, \theta^- | y_1, y_2, ..., y_T$ by histogram.

- An approximated summary statistics: mean, mode, variance and so on.

- An approximated $p$-th quantile and probability interval

Moreover, we already have a generative algorithm for a network at the specific parameter point. In fact, it is same to the algorithm used in the auxiliary chain. Thus, we can predict the form of network at $T + 1, T + 2, ...$ using posterior sample following the standard Bayesian prediction method. For example, for predicting networks structure at $T + 1$,

- 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 of the generated networks and get some summary statistics on them.

## 5.3 Goodness of Fit

To evaluate the goodness of fit for the BSTERGMs and to make sure the posterior $\theta^+, \theta^-$ is fitted properly, we can use the auxiliary chain algorithm once again.

**Algorithm 5.1** (GOF procedure). For $t = 2, ..., T$
and For $j = 1, ..., J$

1. sample $\theta_j^+, \theta_j^-$ from the estimate of posterior.

2. simulate $y_j$ using the auxiliary chain under $y_{t-1}$.

3. calculate $s(y_j, y_{t-1})$, some higher degree statistics (eg. Node-degree dist & Edgewise Shared Partner dist)

Take $p$-th and $(1-p)$-th quantiles of each element of $s(y_j, y_{t-1})$s respectively and check whether the network statistics of observed data is covered by the range of te quantiles or not. In addition, draw the box-plot of calculated $s(y_j, y_{t-1})$ and add points or a line with the network statistics of observed data.

# 6 Example: UCLA medical working project - team science data

As an application, I consider the Bayesian separable temporal exponential family model with UCLA medical working project's team science data. These data consist of three sequence, and each sequence has three time points and for each time, there is a network with 52 nodes corresponding 29 scholar members and 23 faculty members.

I consider 2 dimentional model as follows.

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

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

and

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

where $s(z_1, z_2)^T = (|z_1| - |z_2|)$ for $z_1, z_2 \in \mathcal{Y}$ with 52 nodes. In other words, the network statistic in the model is difference of the number of edges for both the formation model and the disolution model.

I use the ordinary normal prior $p(\theta^+) \, \mathcal{N}(0, 100)$ and $p(\theta^-) \, \mathcal{N}(0, 100)$, and set the proposal distribution of main chain $\epsilon \, \mathcal{N}(0, 0.02)$. The auxiliary chains consist of 1000 iterations and the main chain consists of 50000 iteration. After generating the chain, I cut first 10000 samples for a burn-in period and thin the sample by 10.

The posterior estimates for each network and histograms are follows.

For the diagnostics, I attach traceplots and autocorrelation plots.

In addition, here are traceplots of the number of edges in last auxiliary chain. They shows that all chains are not degenerated at last parameter point.

Finally, I append the results of goodness of fit.

# 7 Summary and Discussion

In this paper, I have proposed the Bayesian STERGMs for a sequence of networks and a framework for statistical estimation and inference using the standard Bayesian approach for the model. I also have shown a application of the BSTERGMs.

Since the models are based on ERGMs, it can be extended to more general networks including a variety of network statistics. Moreover, the model can

be applied to weighted graph and use the weights as covariates of the model. Plus, the BSTERGMs are using the exchange algorithm, so it is flexible enough to be integrated with other methods to help the MCMC chains to converge and run faster. For example, the 'tie no tie'(TNT) algorithm used by ergm package ( [Hunter et al., 2008b]) can be applied to enhance the performance of the auxiliary chain of the BSTERGMs.

A modeling by the BSTERGMs for large networks is possible, but a problem of computational inefficiency rises. If given observed network has lots of nodes, the estimation algorithm for the BSTERGM needs huge number of auxiliary chain itertation to sample a proper exchange network at a given parameter point. In addition, the calculation of network statistics in the model cause another problem for large networks, especially when the model has network statistics that high complexity of computation. For example, statistics based on edgewise shared partner distributions needs $O(n^3)$, so computation cost increases rapidly as the number of nodes $n$ grows and the auxiliary chains become bottlenecks of the whole computation procedure.

Another important issue related in the BSTERGMs is degeneracy as I have mentioned at the section of model diagnostics. The BSTERGMs are very sensitive to the initial point of main MCMC chain, so searching a region for non-degenerate initial parameter point in the parameter space is very important. Often, however, the work finding the region is too difficult for the BSTERGMs, partly because of the high number of parameters in the models. Since the BSTERGMs has two parts for $\theta^-, \theta^+$ in the setting of the models by the their definition, the dimension of parameter space increases easily although each formation and dissolution model does not have too many terms respectively. What makes the situation worse is not only the high dimension of parameter space but also the shape of convergent region of the parameter. In many case, the region is geometrically curved due to an implicit relation between several network statistics in the models and interaction between the formation model and the dissolution model. Thus, the MCMC algorithm may fail to exploring the whole parameter space, especially when the space has high dimension, and suffer non-convergence problem. Imposing some model level constraint or using special prior may be a remedy, but finding the proper constraint or the prior distribution will often be a more difficult problem than original one. Thus, further research is needed to resolve these issues.

**Supplements**

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

# References

[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.

[Hanneke et al., 2010] Hanneke, S., Fu, W., Xing, E. P., et al. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics*, 4:585–605.

[Hanneke and Xing, 2006] Hanneke, S. and Xing, E. P. (2006). Discrete temporal models of social networks. In *ICML Workshop on Statistical Network Analysis*, pages 115–125. Springer.

[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.

[Murray et al., 2012] Murray, I., Ghahramani, Z., and MacKay, D. (2012). Mcmc for doubly-intractable distributions. *arXiv preprint arXiv:1206.6848.*

[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, 2002] Snijders, T. A. (2002). Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2):1–40.

[Snijders, 2005] Snijders, T. A. (2005). Models for longitudinal network data. *Models and methods in social network analysis*, 1:215–247.

[Snijders et al., 2006] Snijders, T. A., Pattison, P. E., Robins, G. L., and Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological methodology*, 36(1):99–153.

[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.