

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 Ω 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. This situation is said to doubly-intractable constant problem, and it causes serious challenge to estimate the parameter vector θ in real data analysis. For more detail on fitting algorithm using MCMCMLE, see Hunter and Handcock(2006).

2.3 BERGM: Bayesian ERGM

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

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

Since not only the doubly-intractable constant problem still exists but also the $p(y)$ cannot be evaluated directly, the estimation of θ needs more advanced algorithm than the standard MCMC procedure. The solution of Caimo and Friel is using the exchange algorithm.

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 θ 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 θ 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

Definition 2.1 (STERGM: Separable-Temporal Exponential family Random Graphs Models). 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 (on \mathcal{Y}). Set a distribution on $\mathcal{Y} \times \dots \times \mathcal{Y}$ ($T - 1$ folds) by following way:

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 $c(\theta^+, y_{t-1}) = \sum_{y^+ \in \mathcal{Y}^+} \exp((\theta^+)^T s(y^+, y_{t-1}))$, $c(\theta^-, y_{t-1}) = \sum_{y^- \in \mathcal{Y}^-} \exp((\theta^-)^T s(y^-, y_{t-1}))$.

Then, defining operations $+$, $-$ on \mathcal{Y} following the boolean algebra edgewise-ly, set y_t to

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

Additionally, assume that

- The first-order Markov assumption: $P(Y_2, \dots, Y_T | Y_1) = P(Y_2 | Y_1) \dots P(Y_T | Y_{T-1})$
- The separability: the conditional independence between Y_t^+ and Y_t^- for all $t = 2, \dots, T$; thus, $P(Y_t = y_t | Y_{t-1} = y_{t-1}; \theta^+, \theta^-) = P(Y_t^+ = y_t^+ | Y_{t-1} = y_{t-1}; \theta^+) P(Y_t^- = y_t^- | Y_{t-1} = y_{t-1}; \theta^-)$

Models which have a such form called the STERGM.

2.6 Network Statistics for BSTERGM

As a usual, $s(y_t, y_{t-1})$ would be chosen by a difference of network statistics, $s'(y_t) - s'(y_{t-1})$. A common candidates of s' are:

- the number of edges
- node degree distribution (each order)
- edgewise shared partner distribution (each order)
- dyadwise shared partner distribution (each order)

- k-star distribution
- triangle distribution

In fact, a bundle of all order of node degree distribution and all order of edgewise shared partner distribution form sufficient statistics of a graph. Also, we can set s' by a function of these statistics.

3 Bayesian STERGM

To convert the STERGM to the Bayesian setting, put priors $p(\theta^+), p(\theta^-)$ over θ^+, θ^- and take the Bayes theorem as a inference rule. Then, the posterior of θ^+, θ^- becomes

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

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

Remarks:

- We cannot compute the normalizing constants $c(\theta^+, y_{t-1}), c(\theta^-, y_{t-1})$ practically because we need to sum up too many terms.
- The constants are doubly intractable: they depend on θ^+, θ^- , the parameters of a model. Thus, we cannot use an ordinary MCMC algorithm to get the posterior sample.

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,) and we want to fit the observation using BSTERGM.

How do we find the posterior distribution of θ^+, θ^- ?

Note that an ordinary MCMC algorithm does not work in our situation, because the constant part remains when calculating the MCMC ratio. To solve this doubly-intractable constant problem, we should use more special MCMC technique, the exchange MCMC algorithm.

Here is the algorithm of our main chain for the exchange algorithm.

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

1. Propose candidates θ_*^+, θ_*^- 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 θ_*^+, θ_*^- .

4. Calculate the exchange MCMC ratio π 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 π 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 θ as θ^+ or θ^- according to the case. Calculate the MCMC ratio ϕ ,

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

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 θ^+, θ^- , 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 θ_s^+, θ_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 [16]

References

- [1] 2013.
- [2] Carter T Butts. 4. a relational event framework for social action. *Sociological Methodology*, 38(1):155–200, 2008.
- [3] Carter T Butts. network: a package for managing relational data in r. *Journal of statistical software*, 24(2):1–36, 2008.
- [4] Alberto Caimo and Nial Friel. Bayesian inference for exponential random graph models. *Social Networks*, 33(1):41–55, 2011.
- [5] Alberto Caimo and Nial Friel. Bayesian model selection for exponential random graph models. *Social Networks*, 35(1):11–24, 2013.
- [6] Ove Frank. Statistical analysis of change in networks. *Statistica Neerlandica*, 45(3):283–293, 1991.
- [7] Ove Frank and David Strauss. Markov graphs. *Journal of the american Statistical association*, 81(395):832–842, 1986.
- [8] Charles J Geyer and Elizabeth A Thompson. Constrained monte carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 54(3):657–683, 1992.

- [9] Mark S Handcock. *Statistical models for social networks: Inference and degeneracy*. na, 2003.
- [10] Mark S Handcock and Krista J Gile. Modeling social networks from sampled data. *The Annals of Applied Statistics*, 4(1):5, 2010.
- [11] Mark S Handcock, David R Hunter, Carter T Butts, Steven M Goodreau, and Martina Morris. statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of statistical software*, 24(1):1548, 2008.
- [12] Paul W Holland and Samuel Leinhardt. A dynamic model for social networks. *Journal of Mathematical Sociology*, 5(1):5–20, 1977.
- [13] David R Hunter. Curved exponential family models for social networks. *Social networks*, 29(2):216–230, 2007.
- [14] David R Hunter, Steven M Goodreau, and Mark S Handcock. Goodness of fit of social network models. *Journal of the American Statistical Association*, 103(481):248–258, 2008.
- [15] David R Hunter and Mark S Handcock. Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15(3):565–583, 2006.
- [16] David R Hunter, Mark S Handcock, Carter T Butts, Steven M Goodreau, and Martina Morris. ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of statistical software*, 24(3):nihpa54860, 2008.
- [17] Matthew O Jackson and Alison Watts. The evolution of social and economic networks. *Journal of economic theory*, 106(2):265–295, 2002.
- [18] Laura M Koehly, Steven M Goodreau, and Martina Morris. Exponential family models for sampled and census network data. *Sociological Methodology*, 34(1):241–270, 2004.
- [19] JH Koskinen. Bayesian analysis of exponential random graphs—estimation of parameters and model selection. Report, Research Report 2004: 2, Department of Statistics, Stockholm University, 2004.
- [20] Johan H Koskinen, Garry L Robins, and Philippa E Pattison. Analysing exponential random graph (p-star) models with missing data using bayesian data augmentation. *Statistical Methodology*, 7(3):366–384, 2010.
- [21] Pavel N Krivitsky and Mark S Handcock. A separable model for dynamic networks. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 76(1):29, 2014.

- [22] Pavel N Krivitsky, Mark S Handcock, and Martina Morris. Adjusting for network size and composition effects in exponential-family random graph models. *Statistical methodology*, 8(4):319–339, 2011.
- [23] Martina Morris, Mark S Handcock, and David R Hunter. Specification of exponential-family random graph models: terms and computational aspects. *Journal of statistical software*, 24(4):1548, 2008.
- [24] Martina Morris, Mark S Handcock, and David R Hunter. Specification of exponential-family random graph models: terms and computational aspects. *Journal of statistical software*, 24(4):1548, 2008.
- [25] Garry Robins and Philippa Pattison. Random graph models for temporal processes in social networks. *Journal of Mathematical Sociology*, 25(1):5–41, 2001.
- [26] Garry Robins, Pip Pattison, Yuval Kalish, and Dean Lusher. An introduction to exponential random graph (p^*) models for social networks. *Social networks*, 29(2):173–191, 2007.
- [27] Tom AB Snijders. The statistical evaluation of social network dynamics. *Sociological methodology*, 31(1):361–395, 2001.
- [28] Tom AB Snijders. Models for longitudinal network data. *Models and methods in social network analysis*, 1:215–247, 2005.
- [29] Tom AB Snijders, Christian EG Steglich, and Michael Schweinberger. Modeling the co-evolution of networks and behavior. *Longitudinal models in the behavioral and related sciences*, 31(4):41–71, 2007.
- [30] Tom AB Snijders, Gerhard G Van de Bunt, and Christian EG Steglich. Introduction to stochastic actor-based models for network dynamics. *Social networks*, 32(1):44–60, 2010.
- [31] Stanley Wasserman and Philippa Pattison. Logit models and logistic regressions for social networks: I. an introduction to markov graphs andp. *Psychometrika*, 61(3):401–425, 1996.