Bayesian model determination in complex systems

Mohammadi, Abdolreza

Chapter 5

Bayesian Modelling of the Exponential Random Graph Models with Non-observed Networks

5.1 Abstract

Across the sciences, one of the main objectives is network modelling to discover complex relationships among many variables. The most promising statistical models that can be used for network modelling is the Exponential Random Graph Models (ERGMs). These models provide an insightful probabilistic model to represent a variety of structural tendencies that define complicated dependence patterns hardly modeled by other probabilistic models. However, they are restricted to the models that regarded the network as given, observed networks data. In the present paper, we develop a novel Bayesian statistical framework which combines the class of ERGMs with graphical models capable of modelling non-observed networks. Our proposed method greatly extends the scope of the ERGMs to more applied research areas. We discuss possible extensions of the method.

Key words: Bayesian inference, Exponential random graph models, Graphical models, Covariance selection, Birth-death process, Markov chain Monte Carlo, G-Wishart.

5.2 Introduction

Network modeling pervades all of science since one of the main objectives of science is to discover complex relationships among large numbers of variables. For the prevention of epidemics, social science relies on a keen understanding of interactions among individuals (20, 7). Similarly in biology, curing complex diseases requires an understanding of how "genes talk to each other"(1). One way to describe these kinds of complex relationships is by means of an abstract network. For a general overview of the statistical models and methods for network modeling, see (10).

Exponential random graph models (34, 22) are promising and flexible family of statistical models for modelling network topology. These models have been used mainly in the social science literature since they allow to statistically account for the complexity inherent in many network data (27, 26). In ERGMs, the basic assumption is that the topological structure of an observed network can be explained by the a vector of network statistics that capture the generative structures in the network (27). However, up till now, these models are restricted to the observed network data. In most applications data are not the observed networks, like in biology (e.g. microarray gene expression data and cell signaling data) in neuroscience (e.g. fMRI data). Is it possible to extend the ERGMs to those type of data? A possible solution is combining the class of ERGMs with graphical models.

Graphical models (12) provide an appealing and insightful way to obtain uncertainty estimates when inferring network structure. The close relationship between the topology of the underlying graphs and their probabilistic properties is a main aspect in graphical models, and it provides the potential tools to interpret the complex models. In this regard, Bayesian approaches provide a mainly straightforward tools, and much recent progress has been made in graphical models (9, 5, 14, 32, 16). More recently, (16) have developed a search algorithm based on birth-death MCMC approach that work with high-dimensional data. However, graphical models are powerful approaches only for estimating the underlying graph structure, they are not designed to model the network graphs.

In this paper, we develop a new Bayesian statistical framework for ERGMs, which is capable for network modeling of non-observed networks. The proposed method greatly extends the scope of the ERGMs to more applied research areas, which not limited only in social science. In our method, to apply the ERGMs to non-observed networks data, we combine the class of ERGMs with graphical models capable of modelling non-observed networks. In particular, in our Bayesian framework, we design a computationally efficient search algorithm to explore all the graph space to distinguish not only important edges

but also key features and detect the underlying graph structure. This search algorithm is based on birth-death Markov chain Monte Carlo algorithm proposed by (16) for Gaussian graphical models.

The paper is organized as follows. In section 5.3 we briefly explain the exponential random graph models and graphical models. In section 5.4 we introduce our proposed Bayesian framework for exponential graph models with non-observed networks.

5.3 Exponential families and graphical models

In this section we briefly review two probabilistic families of graphical models which we will use in our proposed methodology; for more details see e.g. (29).

5.3.1 Exponential random graph models

Exponential random graph models (ERGMs) or p^* models (34) are the families of statistical models that provide a flexible way to model the complex dependence structure of networks. They are the most popular models for social networks and also used in physics and biology (25). The aim to model data as observed networks consisting of nodes and edges, which in the social network context represent actors and relationships between these actors, respectively. There has been comparably little research on using Bayesian approach to infer the parameters of ERGMs besides recent articles by (26, 11, 2, 3). See (21) and (22) for an overview of ERGMs.

In an ERGMs, the random matrix $G = \{g_{ij}\}$ is defined over the graph space on a set of p nodes, with each variable in G representing the presence or absence of a particular edge ($g_{ij} = 1$ if there is a link from i to j, and $g_{ij} = 0$ otherwise). Edges connecting a node to itself are not allowed so $g_{ii} = 0$. For a graph (which consists of a set of edges over the set of nodes), the ERGM is then given by

$$P(G|\theta) = \frac{1}{Z(\theta)} \exp\{\theta^t S(G)\}, \tag{5.1}$$

where $\theta \in \Theta$ represents a vector of unknown parameters, and $Z(\theta)$ is a normalizing constant,

$$Z(\theta) = \sum_{G \in \mathcal{G}_p} \exp\{\theta^t S(G)\},$$

and $S(G)$ term is a network statistic of interest that gives the ERGMs much of its explana-

tory power. The vector $S(G)$ can contain statistics to capture the generative structures of connectivity in the network. It can include, for instance,

- the number of edges ($\sum_{i,j} g_{ij}$) to capture network density;
- the number of triangles ($\sum_{i,j,l} g_{ij}g_{il}g_{jl}$) to capture transitivity;
- the number of 2-stars ($\sum_{i,j,l} g_{il}g_{jl}$) where a k-star ($k > 2$) is a node with k neighbors or a node of degree k,

and the wide variety of other endogenous structures (27).

Estimating ERGMs parameters is a challenging problem, since the analytic form of the normalizing constant $Z(\theta)$ is unknown due to the combinational complexity of summing over all possible $2^{p(p-1)/2}$ graphs in $\mathcal{G}_p$ (2, 3, 8). Although, ERGMs are difficult to handle in practice, they are quite popular in the literature since they are conceived to capture the complex dependence structure of the graph and allow a reasonable interpretation of the observed data.

At the basis of this class of models, the dependence hypothesis is that edges self-organize into specific structures called configurations (e.g. triangles, n-star). Flexibility to adapt to different network contexts, there is a broad range of possible network configurations (34, 22). A positive value of $\theta_i \in \theta$ results in a tendency for the specific configuration corresponding to $S_i(G)$ to be observed in the data, otherwise it should expected by chance.

Note, in ERGMs data are observed networks, which is a strong limitation. In most of the applications, data are measured on variables, such as gene expression data and cell signalling data. The question we intend to answer is whether it is possible to extend the idea of ERGMs to those types of data? We extend the ERGMs to non-observed networks by combining it with graphical models.

5.3.2 Graphical models

Graphical models (12) use a graph concept to represent conditional dependence relationships among random variables, as non-observed networks. When observed data come from noisy measurements of the variables, then graphical models present an appealing and insightful way to describe graph-based dependencies between the random variables. A graph $G = (V, E)$ denotes a set of vertices $V = \{1, 2, ..., p\}$ – where each node corresponds with a random variable – and a set of existing edges E, and $\overline{E}$ denotes the set of non-existing edges. We are interested on undirected graphical models in where $(i, j) \in E$ is equivalent with $(j, i) \in E$, also known as Markov random fields (24). In this class of models, nodes in the graph G correspond to the random variables. The absence of an edge between two nodes determines the two corresponding variables are conditionally inde-

pendent given the remaining variables, while an edge between the two nodes indicates the conditional dependence of the variables.

Graphical models that follow the multivariate Gaussian distribution are called Gaussian graphical models (GGMs), also known as covariance selection models (4). In GGMs, zero entries in the precision matrix correspond to the absence of edges on the graph, which are conditional independence between pairs of random variables given the rest. With respect to the graph G, a zero mean Gaussian graphical model is

$$\mathcal{M}_G = \left\{ \mathcal{N}_p(0, \Sigma) \mid K = \Sigma^{-1} \in \mathbb{P}_G \right\},$$

where $\mathbb{P}_G$ denotes the space of $p \times p$ positive definite matrices with entries equal to zero for not existing edges in graph G. Let $\mathbf{X} = (X^1, ..., X^n)$ be an independent and identically distributed sample of size n from model $\mathcal{M}_G$; Then, the likelihood is

$$P(\mathbf{X}|K, G) \propto |K|^{n/2} \exp\left\{ -\frac{1}{2}\text{tr}(KU) \right\}, \tag{5.2}$$

where $U = \mathbf{X}'\mathbf{X}$.

For a p-dimensional graph, the size of graph space is in total $2^{p(p-1)/2}$, which grows supper-exponentially with the number of nodes in the graph. Thus, Bayesian inference on all graph space is severely limited by the nature of the graph space. In this regard, there are the efficient stochastic search algorithms that can explore the graph space (16, 5, 9). These types of search algorithms explore the graph space by adding or removing one edge in each step, knows as neighborhood search algorithm. These algorithms can potentially work with the graph with more than 100 nodes (9, 16, 17).

5.4 Bayesian hierarchical model for ERGMs with non-observed networks

In this section, we proposed the hierarchical Bayesian methodology to discover the underlining network structure and features which are important. We can display the hierarchical model schematically as below

$$\theta \longrightarrow G \longrightarrow K \longrightarrow \mathbf{X} = (X_1, ..., X_n).$$

Thus, we consider the joint posterior distribution of the parameters as bellow

$$P(\theta, G, K \mid \mathbf{X}) \propto P(\mathbf{X} \mid \theta, G, K)\, P(K \mid G)\, P(G \mid \theta)\, P(\theta). \tag{5.3}$$

In our methodology, we assume the observed data follows a multivariate Gaussian distribution.

5.4.1 Model for prior specification on graph

Here, by consider the idea of exponential random graph, we use a prior on the graph as follows

$$P(G \mid \theta) = \frac{1}{Z(\theta)} \exp\{\theta^t S(G)\} \tag{5.4}$$

where $S(G)$ is a vector of statistics of the graph (e.g., the number of edges, triangles, etc.) and $\theta \in \Theta$ denotes the parameter vector of the model.

5.4.2 Prior specification on precision matrix

For the prior distribution of the precision matrix, we use the G-Wishart (23) distribution. In Gaussian graphical models, the G-Wishart prior distribution is highly attractive since it is conjugate to normally distributed data and places no probability mass on zero entries of the precision matrix. The G-Wishart distribution $W_G(b, D)$, for random matrix K is defined as

$$P(K|G) = \frac{1}{I_G(b, D)} |K|^{(b-2)/2} \exp\left\{-\frac{1}{2}\text{tr}(DK)\right\}, \qquad K \in \mathbb{P}_G,$$

where $b > 2$ is a degree of freedom, D is a symmetric positive definite matrix, and $I_G(b, D)$ is a normalizing constant,

$$I_G(b, D) = \int_{\mathbb{P}_G} |K|^{(b-2)/2} \exp\left\{-\frac{1}{2}\text{tr}(DK)\right\} dK.$$

For complete graph G, the G-Wishart distribution reduces to the Wishart distribution, hence, its normalizing constant has an explicit form (18). Also, for decomposable graphs, the $I_G(b, D)$ has an explicit form (23). However, for non-decomposable graphs, the $I_G(b, D)$ has an intractable form (28).

Since the G-Wishart prior is conjugate to the likelihood (5.2), the posterior distribution

of K is

$$P(K|\mathbf{X}, G) = \frac{1}{I_G(b^*, D^*)}|K|^{(b^*-2)/2}\exp\left\{-\frac{1}{2}\text{tr}(D^*K)\right\},$$

where $b^* = b + n$ and $D^* = D + U$, that is, $W_G(b^*, D^*)$. To consider other possible prior for precision matrix see e.g. (33, 31, 30), and (35). They place no probability mass on zero entries and stable priors on the non-zero entries of the precision matrix.

5.4.3 MCMC sampling scheme

The high dimensionality of the graph G leads to the use of MCMC sampler algorithm can potentially explore all graph space. Specifically, we introduce the MCMC algorithm that simulate from the joint posterior distribution (5.3). The proposed MCMC algorithm is in three steps as follows

Step 1: Sample from θ, based on exchange algorithm (19).

Step 2: Sample from graph space, based on birth-death MCMC sampling algorithm proposed by (16).

Step 3: Sample from precision matrix, based on exact sampling algorithm form G-Wishart distribution proposed by (13).

For step 1, in section 1, We illustrate how to sample from conditional distribution of θ based on exchange algorithm (19).

For step 2, we using computationally efficient birth-death MCMC sampler proposed by (16) for Gaussian graphical models. Their algorithm explores the graph space by adding or deleting an edge in a birth or death event, in which the events are based on a continuous time birth-death Markov process. In a graph $G = (E, V)$ with precision matrix K, each edge $e = (i, j) \in E$ dies independently of others as a Poisson process with death rate $\delta_e(K)$. Since the events are independent, the overall death rate is $\delta(K) = \sum_{e\in E}\delta_e(K)$. With similar definition, each non-edge $e = (i, j) \in \overline{E}$ appears independently as a Poisson process with birth rate $\beta_e(K)$ and the overall birth rate is $\beta(K) = \sum_{e\in\overline{E}}\beta_e(K)$. The birth and death rates of edges occur in continuous time with the rates determined by the stationary distribution of the process. The algorithm is considered in such a way that the stationary distribution equals the target posterior distribution.

Since, the birth-death processes are independent Poisson processes, the time between two successive events has an exponential distribution with mean $1/(\beta(K) + \delta(K))$. This time can be considered as the process expends for any particular instance of the graph space. Therefore, the probability of birth and death events are proportional to their rates

as

$$P(\text{birth for edge } e) = \frac{\beta_e(K)}{\beta(K) + \delta(K)}, \qquad \text{for each } \ e \in \overline{E}, \tag{5.5}$$

$$P(\text{death for edge } e) = \frac{\delta_e(K)}{\beta(K) + \delta(K)}, \qquad \text{for each } \ e \in E. \tag{5.6}$$

Mohammadi and Wit (16, section 3) proof the birth-death MCMC sampling algorithm converge to the target posterior distribution by considering accordingly birth and death rates,

$$\beta_e(K) = \frac{P(G^{+e}, K^{+e}|\mathbf{X}, \theta)}{P(G, K|\mathbf{X}, \theta)}, \ \text{ for each } \ e \in \overline{E}, \tag{5.7}$$

$$\delta_e(K) = \frac{P(G^{-e}, K^{-e}|\mathbf{X}, \theta)}{P(G, K|\mathbf{X}, \theta)}, \qquad \text{for each } \ e \in E, \tag{5.8}$$

in which $G^{+e} = (V, E \cup \{e\})$, and $K^{+e} \in \mathbb{P}_{G^{+e}}$ and similarly $G^{-e} = (V, E \setminus \{e\})$, and $K^{-e} \in \mathbb{P}_{G^{-e}}$. Based on the above explanation, we summarize the proposed sampler algorithm as below.

Algorithm 5.1. Given the current state (θ, G, K):

1. Update θ, using a Metropolis step.

- **1.1.** Draw θ^* from symmetric proposal distribution $h(\theta^* \mid \theta)$
- **1.2.** Draw $G^* \sim p(G \mid \theta^*)$
- **1.3.** Propose the exchange move with probability

$$\alpha(\theta \to \theta^*) = \min\left\{1, \frac{q_{\theta^*}(G)P(\theta^*)q_\theta(G^*)}{q_\theta(G)P(\theta)q_{\theta^*}(G^*)}\right\}$$

2. Update G, conditional on θ based on birth and death process.

- **2.1.** Calculate the birth rates by equation (5.7) and $\beta(K) = \sum_{e\in\overline{E}} \beta_e(K)$,
- **2.2.** Calculate the death rates by equation (5.8) and $\delta(K) = \sum_{e\in E} \delta_e(K)$,
- **2.3.** Calculate the waiting time by $\mathcal{W}(K) = 1/(\beta(K) + \delta(K))$,
- **2.4.** Calculate the type of jump from (5.7) and (5.8)

3. Update K, conditional on the recent G, sample the new precision matrix.

In Algorithm 5.1, the first step is to sample from θ by using exchange algorithm which we explain in section 1. Then (in step two), we calculate the birth and death rates and waiting times. Based on birth and death rates we calculate the type of jump. Details of how to efficiently calculate the birth and death rates are discussed in subsection 1. Finally

in step 3, according to the new state of the jump, we sample from a new precision matrix using a direct sampling scheme from the G-Wishart distribution which proposed by (13).

Step 1: Updating parameter θ

In step 1 of algorithm 5.1, we are interested in sample from conditional distribution of θ,

$$P(\theta|G) = \frac{1}{Z(\theta)} \exp\{\theta^t S(G)\} p(\theta). \tag{5.9}$$

Sampling from this conditional distribution is difficult, since requires the evaluation of the intractable normalizing constant $Z(\theta) = \sum_{G \in \mathcal{G}_p} \exp\{\theta^t S(G)\}$.

Murray et al. (19) introduced the exchange algorithm based on exact sampling, which is designed for general MCMC algorithms in which their target posterior distributions have additional intractable normalization constant. To circumvent such an intractable normalizing constant, Caimo and Friel (2) explain how to use the concept behind the exchange algorithm.

Suppose that G is the current state of our algorithm and we would like to sample θ from (5.9), first we sample θ^* from symmetric proposal distribution $h(\theta^*|\theta)$. Then we sample G^*, which is the difficult step of the algorithm since this requires a draw from (5.4).

Note that, exchange algorithm requires an exact sampling from G. Following (2), we approximate the exact simulation by sampling G^* from $P(G|\theta^*)$ using an MCMC run that is "long enough" to get a point that can be treated as if it were simulated exactly from $P(G|\theta^*)$. They suggested that 500 iteration is a long-enough run. (6) proof that as few as 50 or 100 iterations are usually sufficient.

Computing the birth and death rates

Calculating the birth and death rates (5.7 and 5.8) is the bottleneck of our BDMCMC algorithm. Here, we explain how to calculate efficiently the death rates and for the birth rates is followed a similar manner. For more details see (15, 16).

Following (16) and some simplification, for each $e = (i, j) \in E$, we have

$$\delta_e(K) = \frac{P(G^{-e}|\theta)}{P(G|\theta)} \frac{I_G(b, D)}{I_{G^{-e}}(b, D)} \left(\frac{D^*_{jj}}{2\pi(k_{ii} - k^1_{11})}\right)^{\frac{1}{2}} H(K, D^*),$$

where

$$\frac{P(G^{-e}|\theta)}{P(G|\theta)} = e^{\theta\left[S(G^{-e}) - S(G)\right]},$$

and

$$H(K, D^*) = \exp\left\{-\frac{1}{2}\left[\text{tr}(D^*_{e,e}(K^0 - K^1)) - (D^*_{ii} - \frac{(D^*_{ij})^2}{D^*_{jj}})(k_{ii} - k^1_{11})\right]\right\},$$

in which K^0 is a 2×2 diagonal matrix that $k^0_{11} = k_{ii}$ and $k^0_{22} = K_{j,V\setminus j}(K_{V\setminus j,V\setminus j})^{-1}K_{V\setminus j,j}$ and $K^1 = K_{e,V\setminus e}(K_{V\setminus e,V\setminus e})^{-1}K_{V\setminus e,e}$. Evaluating the ratio of prior normalizing constants is the computational bottleneck for computing the death/birth rates.

Explicit form for ratio of prior normalizing constants Calculating the ratio of prior normalizing constants has been a major computational bottleneck in the Bayesian literature (28, 32, 16, 15). More recently (28) providing an explicit representation of such intractable normalizing constant and (15) implement this concept to sampling algorithm of graphical models. By using the theorem derived by (28, Theorem 3.7.), for special case which D is an identity matrix, we have

$$\frac{I_G(b, \mathbb{I}_p)}{I_{G^{-e}}(b, \mathbb{I}_p)} = 2\sqrt{\pi}\frac{\Gamma((b+d+1)/2)}{\Gamma((b+d)/2)},$$

where d denotes the number of triangles formed by the edge e and two other edges in G and $\mathbb{I}_p$ denotes a p dimensional identity matrix.

Therefore, for the case of $D = \mathbb{I}_p$, we have a simplified formula for the death rates which is given by

$$\delta_e(K) = e^{\theta\left[S(G^{-e}) - S(G)\right]}\frac{\Gamma((b+d+1)/2)}{\Gamma((b+d)/2)}(\frac{2D^*_{jj}}{(k_{ii} - k^1_{ii})})^{\frac{1}{2}}H(K, D^*),$$

5.5 Discussion

We have proposed a Bayesian methodology for exponential random graph models with non-observed networks, which opens a large toolbox to network modelling for non-observed data. By combining exponential random graph models and graph models, we have developed hierarchical Bayesian frameworks.

The Bayesian framework that we proposed here is not limited to data which follow Gaussianity assumption. One possible extension of our work could be to Gaussian copula graphical models (15).

In our Bayesian framework, we focus on undirected graphical models. We can extend our work to directed graphical models as well. This is an ongoing research subject.

References

[1] Bhadra, A. and Mallick, B. K. (2013). Joint high-dimensional bayesian variable and covariance selection with an application to eqtl analysis. *Biometrics*, 69(2):447–457.

[2] Caimo, A. and Friel, N. (2011). Bayesian inference for exponential random graph models. *Social Networks*, 33(1):41–55.

[3] Caimo, A. and Friel, N. (2013). Bayesian model selection for exponential random graph models. *Social Networks*, 35(1):11–24.

[4] Dempster, A. (1972). Covariance selection. *Biometrics*, 28(1):157–175.

[5] Dobra, A., Lenkoski, A., and Rodriguez, A. (2011). Bayesian inference for general gaussian graphical models with application to multivariate lattice data. *Journal of the American Statistical Association*, 106(496):1418–1433.

[6] Everitt, R. G. (2012). Bayesian parameter estimation for latent markov random fields and social networks. *Journal of Computational and Graphical Statistics*, 21(4):940–960.

[7] Goodreau, S. M. (2007). Advances in exponential random graph (p*) models applied to a large social network. *Social Networks*, 29(2):231–248.

[8] He, R. and Zheng, T. (2015). Glmle: graph-limit enabled fast computation for fitting exponential random graph models to large social networks. *Social Network Analysis and Mining*, 5(1):1–19.

[9] Jones, B., Carvalho, C., Dobra, A., Hans, C., Carter, C., and West, M. (2005). Experiments in stochastic computation for high-dimensional graphical models. *Statistical Science*, 20(4):388–400.

[10] Kolaczyk, E. D. (2009). *Statistical analysis of network data: methods and models*. Springer Science & Business Media.

[11] 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.

[12] Lauritzen, S. (1996). *Graphical models*, volume 17. Oxford University Press, USA.

[13] Lenkoski, A. (2013). A direct sampler for g-wishart variates. *Stat*, 2(1):119–128.

[14] Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34(3):1436–1462.

[15] Mohammadi, A., Abegaz Yazew, F., van den Heuvel, E., and Wit, E. C. (2015). Bayesian modeling of dupuytren disease using copula gaussian graphical models. *Arxiv preprint arXiv:1501.04849v2*.

[16] Mohammadi, A. and Wit, E. C. (2015a). Bayesian structure learning in sparse gaussian graphical models. *Bayesian Analysis*, 10(1):109–138.

[17] Mohammadi, A. and Wit, E. C. (2015b). Bdgraph: Bayesian structure learning of graphs in r. *arXiv preprint arXiv:1501.05108v2*.

[18] Muirhead, R. (1982). *Aspects of multivariate statistical theory*, volume 42. Wiley Online Library.

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

[20] Newman, M. E., Watts, D. J., and Strogatz, S. H. (2002). Random graph models of social networks. *Proceedings of the National Academy of Sciences*, 99(suppl 1):2566–2572.

[21] Robins, G., Pattison, P., Kalish, Y., and Lusher, D. (2007a). An introduction to exponential random graph (< i> p</i>*) models for social networks. *Social networks*, 29(2):173–191.

[22] Robins, G., Snijders, T., Wang, P., Handcock, M., and Pattison, P. (2007b). Recent developments in exponential random graph (< i> p</i>*) models for social networks. *Social networks*, 29(2):192–215.

[23] Roverato, A. (2002). Hyper inverse wishart distribution for non-decomposable graphs and its application to bayesian inference for gaussian graphical models. *Scandinavian Journal of Statistics*, 29(3):391–411.

[24] Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. CRC Press.

[25] Saul, Z. M. and Filkov, V. (2007). Exploring biological network structure using exponential random graph models. *Bioinformatics*, 23(19):2604–2611.

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

[27] 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.

[28] Uhler, C., Lenkoski, A., and Richards, D. (2014). Exact formulas for the normalizing constants of wishart distributions for graphical models. *arXiv preprint arXiv:1406.4901*.

[29] Wainwright, M. J. and Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1-2):1–305.

[30] Wang, H. (2012). Bayesian graphical lasso models and efficient posterior computation. *Bayesian Analysis*, 7(4):867–886.

[31] Wang, H. (2014). Scaling it up: Stochastic search structure learning in graphical models.

[32] Wang, H. and Li, S. (2012). Efficient gaussian graphical model determination under g-wishart prior distributions. *Electronic Journal of Statistics*, 6:168–198.

[33] Wang, H. and Pillai, N. S. (2013). On a class of shrinkage priors for covariance matrix estimation. *Journal of Computational and Graphical Statistics*, 22(3):689–707.

[34] 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.

[35] Wong, F., Carter, C. K., and Kohn, R. (2003). Efficient estimation of covariance selection models. *Biometrika*, 90(4):809–830.