Social Networks 46 (2016) 11–28

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Bayesian exponential random graph models with nodal random effects

S. Thiemichen [a,*], N. Friel [b], A. Caimo [c], G. Kauermann [a]

[a] Institut für Statistik, Ludwigs-Maximilians-Universität München, Germany
[b] School of Mathematics and Statistics and Insight: The National Centre for Data Analytics, University College Dublin, Ireland
[c] School of Mathematical Sciences, Dublin Institute of Technology, Ireland

ARTICLE INFO

Article history:
Available online 1 March 2016

Keywords:
Exponential random graph models
Bayesian inference
Random effects
Network analysis

ABSTRACT

We extend the well-known and widely used exponential random graph model (ERGM) by including nodal random effects to compensate for heterogeneity in the nodes of a network. The Bayesian framework for ERGMs proposed by Caimo and Friel (2011) yields the basis of our modelling algorithm. A central question in network models is the question of model selection and following the Bayesian paradigm we focus on estimating Bayes factors. To do so we develop an approximate but feasible calculation of the Bayes factor which allows one to pursue model selection. Three data examples and a small simulation study illustrate our mixed model approach and the corresponding model selection.

1. Introduction

The analysis of network data is an emerging field in statistics which is challenging both model-wise and computationally. Recently Goldenberg et al. (2010), Hunter et al. (2012), Fienberg (2012), and Salter-Townshend et al. (2012), respectively, published comprehensive survey articles discussing statistical approaches, challenges and developments in network data analysis. We also refer to the monograph of Kolaczyk (2009) for a comprehensive introduction to the field.

In this paper, we consider networks represented as a $n \times n$ dimensional adjacency matrix $\boldsymbol{Y}$, where the element $Y_{ij} = 1$, if an edge exists between vertex i and vertex j, and $Y_{ij} = 0$ otherwise, with $i, j \in \{1, \ldots, n\}$ and $i \neq j$, that is there is no connection from a vertex to itself. With n we denote the number of vertices in the network and for simplicity we assume undirected edges, that is $Y_{ij} = Y_{ji}$. Therefore, the matrix $\boldsymbol{Y}$ is symmetric and for simplicity it is sufficient to consider the upper triangle of $\boldsymbol{Y}$ only, that is $Y_{ij}, j > i$. Our approach equally applies to non-symmetric adjacency matrices corresponding to directed graphs. A concrete realization of $\boldsymbol{Y}$ is denoted with $\boldsymbol{y}$.

With respect to the available statistical models for modelling cross-sectional network data one may roughly distinguish between two strands, (a) models which explain the existence of an edge purely with external nodal covariates or random effects and (b) models where the existence of an edge also depends on the local network structure. The first strand of models is phrased as p_1 and p_2 models tracing back to Holland and Leinhardt (1981). Specifically, in the p_1 model we set

$$\text{logit}[\mathbb{P}(Y_{ij} = 1)] = \log\left\{\frac{\mathbb{P}(Y_{ij} = 1)}{1 - \mathbb{P}(Y_{ij} = 1)}\right\} = \alpha_i + \alpha_j + \boldsymbol{z}_{ij}^t\boldsymbol{\beta}, \tag{1}$$

where $\boldsymbol{z}_{ij}$ denotes a set of covariates relating to the vertices i and j and α_i and α_j are nodal effects, here assuming undirected edges. Since the number of parameters increases with increasing network size n, van Duijn et al. (2004) proposed to replace the α parameters in (1) by random effects, see also Zijlstra et al. (2006). This yields the p_2 model

$$\text{logit}[\mathbb{P}(Y_{ij} = 1|\boldsymbol{\phi})] = \phi_i + \phi_j + \boldsymbol{z}_{ij}^t\beta, \tag{2}$$

$$\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)^t \sim N(0, \sigma_\phi^2 I_n) \tag{2}$$

with I_n as n dimensional unit matrix. A general principle with this approach is that vertices (or actors in the network, respectively) are not considered as homogeneous but heterogeneous, though their heterogeneity is not observable but latent and expressed in the node specific random effects ϕ_i, $i = 1, \ldots, n$.

Both, the p_1 and the p_2 model lie within the classical generalized linear (mixed) model framework which allows estimation using standard statistical software. The p_2 models also allow for Bayesian estimation approaches, see for example Gill and Swartz (2004).

* Corresponding author at: Institut für Statistik, Ludwigs-Maximilians-Universität München, Ludwigstr. 33, 80539 München, Germany. Tel.: +49 089 2180 2232; fax: +49 089 2180 5040.

http://dx.doi.org/10.1016/j.socnet.2016.01.002
0378-8733/

The second strand in statistical network modelling is based on the so called exponential random graph model (ERGM) proposed by Frank and Strauss (1986). Here we model directly the network using the likelihood function

$$\mathbb{P}(\boldsymbol{Y}=\boldsymbol{y}|\boldsymbol{\theta})=f(\boldsymbol{y}|\boldsymbol{\theta})=\frac{q_{\boldsymbol{\theta}}(\boldsymbol{y})}{\kappa(\boldsymbol{\theta})}=\frac{\exp\{\boldsymbol{\theta}^t s(\boldsymbol{y})\}}{\kappa(\boldsymbol{\theta})}, \tag{3}$$

where $\boldsymbol{\theta}=(\theta_0, \ldots, \theta_p)^t$ is the vector of model parameters and $s(\boldsymbol{y})=(s_0(\boldsymbol{y}), \ldots, s_p(\boldsymbol{y}))^t$ is a vector of sufficient network statistics like the number of edges or two-stars in a network, see for example Snijders et al. (2006). In Eq. (3) the term $\kappa(\boldsymbol{\theta})$ denotes the normalizing constant, that is

$$\kappa(\boldsymbol{\theta})=\sum_{\boldsymbol{y}\in\mathcal{Y}}\exp\{\boldsymbol{\theta}^t s(\boldsymbol{y})\}$$

and is accordingly the sum over $2^{\binom{n}{2}}$ potential undirected graphs and therefore numerically intractable, except for very small graphs. Early fitting approaches are based on the pseudolikelihood idea proposed by Strauss and Ikeda (1990). More advanced are MCMC based routines proposed by Hunter and Handcock (2006) based on the work of Geyer and Thompson (1992). A fully Bayesian approach to estimate ERGMs has been developed by Caimo and Friel (2011).

Model (3) allows for a conditional interpretation by focusing on the occurrence of a single edge between two nodes. To be specific we obtain

$$\text{logit}[\mathbb{P}(Y_{ij}=1|Y_{kl},(k,l)\neq(i,j);\boldsymbol{\theta})]=\boldsymbol{\theta}^t s_{ij}(\boldsymbol{y}), \tag{4}$$

where $s_{ij}(\boldsymbol{y})$ denotes the vector of so called change statistics

$$s_{ij}(\boldsymbol{y})=s(y_{ij}=1,y_{kl},(k,l)\neq(i,j))-s(y_{ij}=0,y_{kl},(k,l)\neq(i,j)).$$

We refer to Robins et al. (2007a,b), and the rather recent work of Lusher et al. (2013) for a deeper discussion of exponential random graph models.

Contrasting Eq. (4) with the p_1 and p_2 model given in Eqs. (1) and (2) it becomes obvious that the ERGM in contrast to the p_1 and p_2 models take the network structure into account while considering the nodes to be homogeneous. When modelling network data this means that all possible heterogeneity in the network nodes (that is the actors in the network) is included as covariates in the model and influence the (global) structure of the network. Since homogeneity of the nodes have led from p_1 to p_2 models, we want to pursue the same modelling exercise by allowing for latent node specific heterogeneity in exponential random graph models. To do so, we combine the p_2 model (2) with the ERGM (4) towards

$$\text{logit}[\mathbb{P}(Y_{ij}=1|Y_{kl},(k,l)\neq(i,j);\boldsymbol{\theta},\phi_i,\phi_j)]=\boldsymbol{\theta}^t s_{ij}(\boldsymbol{y})+\phi_i+\phi_j \tag{5}$$

with $\boldsymbol{\phi}=(\phi_1, \ldots, \phi_n)^t$ and $\phi_i \overset{\text{i.i.d.}}{\sim} N(\mu_\phi,\sigma_\phi^2)$, $i=1,\ldots,n$. The parameter μ_ϕ captures the average propensity in the network for forming a tie. Therefore θ_0, which is usually the parameter associated with the edges statistic, is excluded from $\boldsymbol{\theta}$, i.e. $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_p)^t$ here. In terms of the likelihood function for the whole network we obtain from (5)

$$\mathbb{P}(\boldsymbol{Y}=\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\phi})=f(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\phi})=\frac{q_{\boldsymbol{\theta},\boldsymbol{\phi}}(\boldsymbol{y})}{\kappa(\boldsymbol{\theta},\boldsymbol{\phi})}=\frac{\exp\{\boldsymbol{\theta}^t s(\boldsymbol{y})+\boldsymbol{\phi}^t t(\boldsymbol{y})\}}{\kappa(\boldsymbol{\theta},\boldsymbol{\phi})}, \tag{6}$$

where $t(\boldsymbol{y})$ contains the degree statistics of the n vertices, i.e. $t_i(\boldsymbol{y})=\sum_{j=1}^n y_{ij}$, *for* $i=1,\ldots,n$. That is we fit an exponential random graph model with random, node specific effects accounting for heterogeneity. The model in Eqs. (5) and (6) falls in the general class of Exponential-family Random Network Models proposed by Fellows and Handcock (2012) but unlike their model we treat the node specific effect as latent and we pursue a fully Bayesian estimation. We also refer to Krivitsky et al. (2009) who propose a model with actor specific random effects based on a latent cluster model. The authors also propose node specific random effects. We follow this line and give further interpretability of the effects. A central issue in model extensions is the question of model selection. We emphasize this point in the paper by comparing models with and without nodal effects using the Bayes factor as model selection criterion. However, calculation of the Bayes factor suffers from the above mentioned problem in exponential random graph models in that the normalization constant $\kappa(\cdot)$ is numerically infeasible. We therefore propose an approximate calculation of the Bayes factor and show in a simulation study its usability for model selection.

For estimation and model selection of model (6) we extend the fully Bayesian approach from Caimo and Friel (2011). The developed estimation routine is based on the numerical work of Caimo and Friel (2014) with their R (R Core Team, 2015) package `Bergm` (see http://cran.r-project.org/web/packages/Bergm). Our algorithms for model fitting and selection will be included in the `Bergm` package.

The paper is organized as follows. In Section 2 we derive a fully Bayesian formulation of the model. This is followed by a detailed description of the MCMC based estimation routine. Section 3 deals with the issue of model selection using Bayes factors. Three data examples and some simulation results are presented in Section 4. Finally Section 5 concludes with a discussion.

2. Bayesian model formulation and estimation

Before proposing a fully Bayesian formulation for model (6) bear in mind that the normalizing constant $\kappa(\boldsymbol{\theta},\boldsymbol{\phi})$ is numerically infeasible to calculate except for small networks so that numerically demanding simulation based fitting routines need to be employed. We follow a fully Bayesian approach by imposing a prior distribution on $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_p)^t$. The posterior of interest for the Bayesian exponential random graph model with nodal random effects in (6) then becomes

$$p(\boldsymbol{\theta},\boldsymbol{\phi},\mu_\phi,\sigma_\phi^2|\boldsymbol{y})=\frac{f(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\phi})p(\boldsymbol{\theta})p(\boldsymbol{\phi}|\mu_\phi,\sigma_\phi^2)p(\mu_\phi)p(\sigma_\phi^2)}{p(\boldsymbol{y})}, \tag{7}$$

where $p(\boldsymbol{\theta})$ is the prior distribution of $\boldsymbol{\theta}$ and $p(\boldsymbol{\phi}|\mu_\phi,\sigma_\phi^2)$ the prior for the random nodal effects $\boldsymbol{\phi}$. We assume the nodal effects to be independent and identically normally distributed, that is

$$\phi_i \sim N(\mu_\phi,\sigma_\phi^2), \quad \text{for} \quad i=1,\ldots,n$$

and accordingly we use $\boldsymbol{\theta}\sim N(0,\rho^2 I_p)$, with I_p denoting the p-dimensional unity matrix and ρ^2 chosen such that the prior distribution is flat. For the hyper prior distribution $p(\mu_\phi)$ of the mean μ_ϕ we assume a normal distribution centred at 0, that is

$$\mu_\phi \sim N(0,\tau^2).$$

The hyper prior $p(\sigma_\phi^2)$ of the variance σ_ϕ^2 is assumed to be an inverse gamma distribution, that is

$$\sigma_\phi^2 \sim IG(a,b).$$

Finally, the parameters τ^2, a and b are all constants and chosen in a way that results in flat hyper prior distributions. Fig. 1 illustrates this Bayesian model formulation.

It is important to note, that the posterior distribution in (7) is so-called doubly-intractable. This is because, firstly, it is not possible to evaluate the posterior density (7) due to $p(\boldsymbol{y})$, the marginal likelihood or evidence, being intractable. Secondly, it is also numerically infeasible to calculate the normalizing constant $\kappa(\boldsymbol{\theta},\boldsymbol{\phi})$ in the likelihood $f(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\phi})$ except for very small network graphs. Similar to the algorithm proposed by Caimo and Friel (2011) we use the so-called exchange algorithm from Murray et al. (2006) to draw samples from