University of Wollongong

Research Online

Faculty of Engineering and Information
Sciences - Papers: Part A

Faculty of Engineering and Information
Sciences

1-1-2012

Marginalized exponential random graph models

Thomas F. Suesse

University of Wollongong, tsuesse@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/eispapers>

Part of the [Engineering Commons](#), and the [Science and Technology Studies Commons](#)

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Marginalized exponential random graph models

Abstract

Exponential random graph models (ERGMs) are a popular tool for modeling social networks representing relational data, such as working relationships or friendships. Data on exogenous variables relating to participants in the network, such as gender or age, are also often collected. ERGMs allow modeling of the effects of such exogenous variables on the joint distribution, specified by the ERGM, but not on the marginal probabilities of observing a relationship. In this article, we consider an approach to modeling a network that uses an ERGM for the joint distribution of the network, but then marginally constrains the fit to agree with a generalized linear model (GLM) defined in terms of this set of exogenous variables. This type of model, which we refer to as a marginalized ERGM, is a natural extension of the standard ERGM that allows a convenient population-averaged interpretation of parameters, for example, in terms of log odds ratios when the GLM includes a logistic link, as well as fast computation of marginal probabilities. Several algorithms to obtain maximum likelihood estimates are presented, with a particular focus on reducing the computational burden. These methods are illustrated using data on the working relationship between 36 partners in a New England law firm.

Keywords

graph, random, marginalized, models, exponential

Disciplines

Engineering | Science and Technology Studies

Publication Details

Suesse, T. F. (2012). Marginalized exponential random graph models. *Journal of Computational and Graphical Statistics*, 21 (4), 883-900.

This article was downloaded by: [University of Wollongong]

On: 10 February 2013, At: 16:43

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

Journal of Computational and Graphical Statistics

Publication details, including instructions for authors and subscription information:

<http://amstat.tandfonline.com/loi/ucgs20>

Marginalized Exponential Random Graph Models

Thomas Suesse^a

^a Centre for Statistical and Survey Methodology, School of Mathematics and Applied Statistics, University of Wollongong, NSW 2522, Australia

Accepted author version posted online: 23 May 2012. Version of record first published: 14 Dec 2012.

To cite this article: Thomas Suesse (2012): Marginalized Exponential Random Graph Models, Journal of Computational and Graphical Statistics, 21:4, 883-900

To link to this article: <http://dx.doi.org/10.1080/10618600.2012.694750>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://amstat.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Marginalized Exponential Random Graph Models

Thomas SUESSE

Exponential random graph models (ERGMs) are a popular tool for modeling social networks representing relational data, such as working relationships or friendships. Data on exogenous variables relating to participants in the network, such as gender or age, are also often collected. ERGMs allow modeling of the effects of such exogenous variables on the joint distribution, specified by the ERGM, but not on the marginal probabilities of observing a relationship. In this article, we consider an approach to modeling a network that uses an ERGM for the joint distribution of the network, but then marginally constrains the fit to agree with a generalized linear model (GLM) defined in terms of this set of exogenous variables. This type of model, which we refer to as a marginalized ERGM, is a natural extension of the standard ERGM that allows a convenient population-averaged interpretation of parameters, for example, in terms of log odds ratios when the GLM includes a logistic link, as well as fast computation of marginal probabilities. Several algorithms to obtain maximum likelihood estimates are presented, with a particular focus on reducing the computational burden. These methods are illustrated using data on the working relationship between 36 partners in a New England law firm. Supplementary materials for the article are available online.

Key Words: Marginalized models; Markov chain Monte Carlo; Maximum likelihood; Odds ratio; Social network.

1. INTRODUCTION

Networks, or mathematical graphs, are an important tool for representing relational data, that is, data on the existence, strength, and direction of relationships between interacting actors. Types of actors include individuals, firms, and countries. In its most basic form, a network consists of a set of n nodes and a set of edges, where nodes represent actors and edges the presence of a specific relationship between actors. The network can be represented by an $n \times n$ matrix $\mathbf{Y} = (Y_{ij})_{i,j=1}^n$, where Y_{ij} is a binary indicator, which takes the value 1 if an edge exists from i to j and is zero otherwise. By convention, $Y_{ii} = 0$. A pair of nodes is often called a *dyad*.

Thomas Suesse is Lecturer, Centre for Statistical and Survey Methodology, School of Mathematics and Applied Statistics, University of Wollongong, NSW 2522, Australia (E-mail: tsuesse@uow.edu.au).

© 2012 *American Statistical Association, Institute of Mathematical Statistics,
and Interface Foundation of North America*

Journal of Computational and Graphical Statistics, Volume 21, Number 4, Pages 883–900

DOI: [10.1080/10618600.2012.694750](https://doi.org/10.1080/10618600.2012.694750)

The most commonly used model for a network is an exponential family model (Casella and Berger 2002) of the form

$$\Pr(\mathbf{Y} = \mathbf{y}; \boldsymbol{\theta}) = \exp(\boldsymbol{\eta}(\boldsymbol{\theta})' \mathbf{Z}(\mathbf{y}) - \kappa(\boldsymbol{\theta})), \quad (1)$$

with

$$\kappa(\boldsymbol{\theta}) = \log \left\{ \sum_{\tilde{\mathbf{y}} \in S} \exp(\boldsymbol{\eta}(\boldsymbol{\theta})' \mathbf{Z}(\tilde{\mathbf{y}})) \right\},$$

where the summation is over the sample space S of the network. The vector $\boldsymbol{\theta} \in \mathbb{R}^p$ contains the model parameters, $\mathbf{Z}(\mathbf{y}) \in \mathbb{R}^q$ is a vector of network statistics, and $\kappa(\boldsymbol{\theta})$ is the normalizing constant. Here, $\boldsymbol{\eta}(\boldsymbol{\theta})$ is a mapping from $\mathbb{R}^p$ to $\mathbb{R}^q$, with $p \leq q$. There are two important subcases. For the identity map $\boldsymbol{\eta}(\boldsymbol{\theta}) = \boldsymbol{\theta}$, (1) is a canonical exponential family model, and η_i are the canonical parameters. When $\boldsymbol{\eta}(\boldsymbol{\theta})$ is nonlinear and $p < q$, (1) defines a curved exponential family model (Efron 1978). For random graphs, the first subcase is often referred to as the exponential random graph model (ERGM) and the second as the curved ERGM (CERGM). To be consistent with standard exponential family terminology, we refer to the first subcase as the canonical ERGM and the more general model specified by (1) as an ERGM. These models are currently widely used for social networks (Strauss and Ikeda 1990; Snijders 2002; Hunter and Handcock 2006; the last reference abbreviated as HH06). The first model of this type for social networks was proposed by Holland and Leinhardt (1981), and is known as the p_1 model.

There are many choices of $\mathbf{Z}(\mathbf{y})$ —for example, see Morris, Handcock, and Hunter (2008)—and most induce dependence among dyads. Maximum likelihood (ML) estimation is complicated and can usually only be achieved by a stochastic approximation of the log-likelihood using Markov chain Monte Carlo (MCMC) algorithms. However, ML estimation for canonical ERGMs based on MCMC methods often fails because of model degeneracy. CERGMs were introduced (Snijders et al. 2006; HH06) to reduce the problem of degeneracy.

Node attributes are also frequently collected, and a small number, say l , can be regarded as covariates. Let the $n \times l$ matrix $\mathbf{X}$ contain these covariates. Usually, only covariates that are exogenous are considered, that is, variables that are not influenced by the network; see HH06. Suppose the scientific interest is in modeling of the marginal probability $\Pr(Y_{ij} = 1)$, conditionally on $\mathbf{X}$, denoted by $\Pr(Y_{ij} = 1 | \mathbf{X})$. For example, how is the effect of equal gender of actors i and j , defined as $f(\mathbf{X}) = 1$ for equal gender and zero otherwise, on $\Pr(Y_{ij} = 1)$? ERGMs are not useful for marginal modeling, because generally, marginal probabilities are intractable. A naive approach ignoring the dyadic dependence is to apply a generalized linear model (GLM) (McCullagh and Nelder 1989) instead of an ERGM. It allows to calculate easily marginal probabilities and provides a convenient population-averaged interpretation of the parameters, for example, in terms of log odds ratios for the logit link. However, this approach does not account for the dyadic dependence structure of the network, which is likely to result in incorrect standard errors.

Covariate effects as $f(\mathbf{X})$ can also be accounted for by an ERGM by adding statistics, depending on $f(\mathbf{X})$, to $\mathbf{Z}(\mathbf{y})$. However, when the main interest is in exogenous effects, such

as equal gender, then interpretation of parameters, in terms of conditional log odds ratios, is difficult, as demonstrated in Section 2.

To apply a marginal model and to account for the dyadic dependence, we introduce marginalized ERGMs (MERGMs) in Section 2, combining GLMs and ERGMs. Advantages of MERGMs over ERGMs are discussed by means of an example, using the Lazega (2001) dataset. In Section 3, we derive two sets of likelihood-based estimating equations and use a Fisher scoring scheme for solving them. Details of ML estimation are described here, including two alternate methods for solving the two sets of likelihood equations in each step of the iterative process. Section 4 illustrates the proposed method on the Lazega dataset. This article finishes with a discussion.

2. MARGINALIZED ERGMs

2.1 LIMITATIONS OF ERGMs

Consider a network of collaborative working relationships between 36 partners in a New England law firm, described in detail by Lazega and Pattison (1999) and Lazega (2001). An edge between two partners exists if both partners indicate collaboration with each other. This network is undirected, that is, $\mathbf{Y} = \mathbf{Y}'$, which will always be assumed in the remainder of the article. The data also contain a number of attributes of each partner: seniority (rank number of entry into the firm), practice (litigation/corporate law), gender (male/female), and office (three offices in three different cities).

One might think of fitting an ERGM to these data. Two examples of typical network statistics $\mathbf{Z}(\mathbf{y})$ that describe an ERGM are: the number of *edges*, denoted by $E = \sum_{i < j} y_{ij}$, and the *edgewise shared partner statistic* with k common neighbors (nodes i and j are neighbors, if $Y_{ij} = 1$), denoted by EP_k , $k = 0, \dots, n - 2$, which is the number of edges that share exactly k neighbors in common. An ERGM described by $\mathbf{Z}(\mathbf{y}) = (E, EP_1, \dots, EP_{n-2})'$ has the form of a canonical ERGM. To alleviate the common problem of model degeneracy, HH06 suggested using the geometrically weighted edgewise shared partner (GWESP) distribution instead, which is characterized by

$$\text{GWESP}(\mathbf{y}, \theta_1, \theta_2) = \theta_1 \exp(\theta_2) \sum_{i=1}^{n-2} \{1 - (1 - \exp(-\theta_2)^i)\} EP_i(\mathbf{y}), \quad (2)$$

which implies that the ERGM has now the form of a CERGM.

Let $\mathbf{X}_i$ be the i th row of the matrix of covariates $\mathbf{X} = (\mathbf{X}'_1, \dots, \mathbf{X}'_n)'$. To account for the effect of the node attributes, symmetric functions $f(\mathbf{X}) = f(\mathbf{X}_i, \mathbf{X}_j)$ are often considered and the statistic $\sum_{i < j} y_{ij} f(\mathbf{X}_i, \mathbf{X}_j)$ is added to $\mathbf{Z}(\mathbf{y})$, denoted by $\mathbf{Z}(\mathbf{y}, \mathbf{X})$. The notation $\Pr(\mathbf{Y} = \mathbf{y}; \boldsymbol{\theta})$ in (1) becomes $\Pr(\mathbf{Y} = \mathbf{y} | \mathbf{X}; \boldsymbol{\theta})$. HH06 used the *main* effects: *seniority* and *practice*, defined by $f_1(\mathbf{X}_i, \mathbf{X}_j) = \text{seniority}_i + \text{seniority}_j$ and $f_2(\mathbf{X}_i, \mathbf{X}_j) = \text{practice}_i + \text{practice}_j$, respectively, and *similarity* effects of *practice* (f_3), *gender* (f_4), and *office* (f_5), where the *similarity* effect of *gender* is defined by $f_4(\mathbf{X}_i, \mathbf{X}_j) = I(\text{gender}_i = \text{gender}_j)$ ($I(\cdot)$ is the indicator function.); likewise are the similarity effects of *practice* and *office* defined.

Define the rest of the graph by $Y_{ij}^c := \mathbf{Y} \setminus Y_{ij}$. Note that (1) implies

$$\log \left(\frac{\Pr(Y_{ij} = 1 | Y_{ij}^c = y_{ij}^c, \mathbf{X})}{1 - \Pr(Y_{ij} = 1 | Y_{ij}^c = y_{ij}^c, \mathbf{X})} \right) = \boldsymbol{\eta}(\boldsymbol{\theta})' \Delta(\mathbf{Z}(\mathbf{y}, \mathbf{X}))_{ij}, \quad (3)$$

where $\Delta(\mathbf{Z}(\mathbf{y}, \mathbf{X}))_{ij}$ denotes the difference in $\mathbf{Z}(\mathbf{y}, \mathbf{X})$ between $Y_{ij} = 1$ and $Y_{ij} = 0$, while Y_{ij}^c and $\mathbf{X}$ remain fixed.

The CERGM, characterized by E (parameter η_0), $\sum_{i < j} y_{ij} f_k(\mathbf{X}_i, \mathbf{X}_j)$ (η_k), $k = 1, 2, 3, 4, 5$, and the GWESP distribution (θ_1 and θ_2), has been used by HH06 to analyze the Lazega dataset. For this model, $\boldsymbol{\eta}(\boldsymbol{\theta})' \Delta(\mathbf{Z})_{ij}$ reduces to $\eta_0 + \sum_{k=1}^5 \eta_k f_k(\mathbf{X}_i, \mathbf{X}_j) + \theta_1 \exp(\theta_2) \sum_{i=1}^{n-2} \{1 - (1 - \exp(-\theta_2)^i)\} \Delta(E P_i(\mathbf{y}))_{ij}$. Let us partition the statistics in $\mathbf{Z}$ into those for which $\Delta(\mathbf{Z})_{ij}$ is constant (denoted by $\mathbf{Z}_1$) and those for which it is not constant (denoted by $\mathbf{Z}_2$), conditionally on $\mathbf{X}$. For the described CERGM, $\mathbf{Z}_1$ contains E and $\sum_{i < j} y_{ij} f_k(\mathbf{X}_i, \mathbf{X}_j)$, and $\mathbf{Z}_2 = (\text{EP}_1, \dots, \text{EP}_{n-2})'$. Note $\Pr(Y_{ij} = 1 | Y_{ij}^c = y_{ij}^c, \mathbf{X})$ depends also on the rest of the graph through $\Delta(\mathbf{Z}_2)_{ij}$.

The parameter η_4 associated with gender can be used to make interpretation in terms of conditional odds ratios. For example, the odds of collaboration between two partners of the same gender are $\exp(\eta_4)$ times the odds of collaboration between two partners of different genders, conditional on the rest of the graph and assuming all covariates are the same. However, in general, this does not allow comparing different dyads with actually observed different f_4 across the population of 36 partners.

Let us reexpress (3) for the considered ERGM ($K = 5$) as

$$\log \left(\frac{\Pr(Y_{ij} = 1 | Y_{ij}^c = y_{ij}^c, \mathbf{X})}{1 - \Pr(Y_{ij} = 1 | Y_{ij}^c = y_{ij}^c, \mathbf{X})} \right) = \eta_0 + \sum_{k=1}^K \eta_k f_k(\mathbf{X}_i, \mathbf{X}_j) + \boldsymbol{\eta}(\boldsymbol{\theta})' \Delta(\mathbf{Z}_2(\mathbf{y}))_{ij}. \quad (4)$$

Note the analogy of (4) to generalized linear mixed models (GLMM) for binary data—for example, see Agresti (2002); here, $\boldsymbol{\eta}(\boldsymbol{\theta})' \Delta(\mathbf{Z}_2)_{ij}$ takes the role of random effects. Observations, here dyads, that share a random effect usually define a cluster. In a similar fashion, a *cluster* of dyads might be defined by those dyads that share a common value of $\boldsymbol{\eta}(\boldsymbol{\theta})' \Delta(\mathbf{Z}_2)_{ij}$. Then, η_4 can be considered as a cluster-specific effect and can be used to compare different dyads within a cluster. However, interpretation of such cluster-specific effects for comparing dyads from different clusters is difficult, because different clusters have a different baseline probability, determined by $\boldsymbol{\eta}(\boldsymbol{\theta})' \Delta(\mathbf{Z}_2)_{ij}$.

The main objective of the study is to assess the effect of equal gender across the population of 36 partners, comparing all those dyads for which we observed $f_4 = 0$ with those with $f_4 = 1$, after controlling for the other covariates. An alternative model approach that caters for this comparison is a GLM (McCullagh and Nelder 1989) of the form

$$g(\Pr(Y_{ij} = 1 | \mathbf{X}_i, \mathbf{X}_j)) = \beta_0 + \sum_{k=1}^K \beta_k f_k(\mathbf{X}_i, \mathbf{X}_j), \quad (5)$$

where $g(\cdot)$ is the link function and $\boldsymbol{\beta} = (\beta_0, \dots, \beta_K)'$ is the vector of model parameters. Here, we only consider symmetric functions depending on $\mathbf{X}_i$ and $\mathbf{X}_j$, but generally, any reasonable function $f(\mathbf{X})$ may be considered. Using the logit link and the previously defined f_k , the parameters β_k allow a population-averaged comparison using the odds ratio, averaged over the clusters defined by (4). For example, the odds of collaboration between

two partners of the same gender are $\exp(\beta_4)$ times the odds of collaboration between two partners of different genders, assuming all covariates are the same. The gender effect might be considered as a between-cluster effect, because comparisons are made between dyads of the whole population and not just between those within clusters. In analogy, a general rule of thumb is that a marginal model with population-averaged effects should be applied when the main interest is in between-cluster effects, whereas a GLMM is more useful when the focus is on cluster-specific effects (Neuhaus 1992). For the Lazega dataset, we are interested in between-cluster effects, supporting model (5). Heagerty and Zeger (2000) also noted that marginal models, such as (5), are useful in many applications, because it separates the marginal model from the joint distribution. In our case, the joint distribution is the network distribution, specified by an ERGM.

Population-averaged probabilities can be easily calculated from the GLM using the inverse link, that is, $\Pr(Y_{ij} = 1 | \mathbf{X}_i, \mathbf{X}_j) = g^{-1}(\beta_0 + \sum_{k=1}^K \beta_k f_k(\mathbf{X}_i, \mathbf{X}_j))$. For example, what is the probability of observing an edge when both partners have the same sex, both practice litigation, both work in the same office, and main seniority is $f_1 = 37$, without specifying the rest of the network? Specifying $\mathbf{X}_i$ and $\mathbf{X}_j$ determines $\Pr(Y_{ij} = 1 | \mathbf{X}_i, \mathbf{X}_j)$ and applies for all dyads with this covariate setting.

On the other hand, for ERGMs, marginal probabilities are intractable. To illustrate, note that

$$\Pr(Y_{ij} = 1 | \mathbf{X}) = \sum_{\mathbf{y} \in S: y_{ij}=1} \Pr(\mathbf{Y} = \mathbf{y} | \mathbf{X}; \theta) \quad (6)$$

is a sum over $2^{((n-1)(n-2)/2)}$ networks, often an incredible large number. Obtaining $\Pr(Y_{ij} = 1 | \mathbf{X})$ can usually only be achieved by first obtaining an MCMC sample from $\Pr(\mathbf{Y} = \mathbf{y} | \mathbf{X}; \theta)$ and then estimating for each dyad $\Pr(Y_{ij} = 1 | \mathbf{X})$ from the MCMC sample. However, using such estimates is questionable, because $\Pr(Y_{ij} = 1 | \mathbf{X})$ depends on the particular dyad ij and also on the specification of all covariates, that is, $\mathbf{X}$, and not only on $\mathbf{X}_i$ and $\mathbf{X}_j$; see Section 4. The drawback of a GLM is that it assumes independence between all dyads, usually an unrealistic assumption. Therefore, using standard software to fit GLMs might result in incorrect statistical inference. In the next subsection, we introduce MERGMs, a model approach that combines the advantages of both ERGMs and GLMs.

2.2 INTRODUCTION OF MERGMs

Exponential tilting is a method to construct an exponential family from a given univariate density $f(y)$. Tilting a random variable Y with density $f(y)$ by θ , the tilting parameter, gives the tilted distribution

$$f_\theta(y) = \exp(\theta y - \kappa(\theta)) f(y),$$

where $\kappa(\theta) = \log \mathbb{E}_Y[\exp(\theta Y)]$. Then, $f_\theta(x)$ has the form of a canonical exponential family, which consists of all *tilts* of $f(y)$, normalized by $\kappa(\theta)$ to integrate to 1. For an application of univariate tilting, see Kim and Yu (2011). Tilting also changes the moments and can be seen as a method to change a distribution's mean to a given value.

In general, the marginal probabilities of an ERGM defined by (6) do not follow a GLM, unless $\mathbf{Z}_2 = \emptyset$, because then, conditional probabilities become unconditional (marginal)

probabilities. For example, when the ERGM is only characterized by statistics that fall into $\mathbf{Z}_1$, then the ERGM is identical to a logistic regression model—compare (4) and (5). To circumvent the general inequality of ERGMs and GLMs, we apply a transformation of the ERGM such that it follows marginally the GLM given by (5). The transformed density of an ERGM has the form

$$\Pr(\mathbf{Y} = \mathbf{y} | \mathbf{X}; \Psi, \theta) = \exp(\Psi' \mathbf{y} + \eta(\theta)' \mathbf{Z}(\mathbf{y}, \mathbf{X}) - \kappa(\Psi, \theta)), \quad (7)$$

with the normalizing constant $\kappa(\Psi, \theta) = \log\{\sum_{\tilde{\mathbf{y}} \in S} \exp(\Psi' \tilde{\mathbf{y}} + \eta(\theta)' \mathbf{Z}(\tilde{\mathbf{y}}, \mathbf{X}))\}$. Here, $\mathbf{Y}$ denotes for convenience the vector variable $(Y_{12}, Y_{13}, \dots, Y_{n-1,n})'$, with associated realized vector $\mathbf{y} = (y_{12}, y_{13}, \dots, y_{n-1,n})'$. Together, (5) and (7) specify a MERGM. The vector $\Psi = (\Psi_{12}, \dots, \Psi_{n-1,n})'$ contains $N := n(n-1)/2$ nuisance parameters that are determined such that the ERGM equates marginally to the GLM. Note that the MERGM is specified by θ and β , where the vector of nuisance parameters Ψ is implicitly given by θ and β through (5) and (7). In general, this type of transformation could be considered as multivariate tilting of $\Pr(\mathbf{Y} = \mathbf{y} | \mathbf{X}; \theta)$, characterized by

$$\Pr(\mathbf{Y} = \mathbf{y} | \mathbf{X}; \Psi, \theta) = \exp(\Psi' \mathbf{y} - \kappa(\Psi)) \Pr(\mathbf{Y} = \mathbf{y} | \mathbf{X}; \theta).$$

The N nuisance parameters Ψ_{ij} could be called *tilting* parameters, one for each dyad. A similar transformation has been applied by Fitzmaurice and Laird (1993) (denoted in the remainder of the article as FL93) for a log-linear model for clustered binary data to marginally match a GLM. The equations (5) and (7) are similar to those presented by FL93, noting that their log-linear model describing the joint distribution also contains the term $\Psi' \mathbf{y}$.

To distinguish the marginal probabilities implied by (5) and those implied by (7), the latter are denoted by $\tilde{\pi}_{ij}$ and the former by π_{ij} . In the next section, when ML estimation is discussed, we solve $\tilde{\pi}(\beta) = \pi(\Psi, \theta)$ for Ψ , given the current estimates of β and θ .

3. ML ESTIMATION

3.1 LIKELIHOOD EQUATIONS AND FISHER SCORING SCHEME

Let us reexpress Equation (5) in a more compact form as

$$g(\pi) = \mathbf{X}_f \beta =: \mathbf{v}, \quad (8)$$

where $\mathbf{X}_f$ is the $N \times (K+1)$ design matrix containing the K exogenous variables $f_k(\mathbf{X}_i, \mathbf{X}_j)$ plus intercept and $\pi := (\pi_{12}, \dots, \pi_{n,n-1})'$, with $\pi_{ij} := \Pr(Y_{ij} = 1 | \mathbf{X}_i, \mathbf{X}_j)$. As an alternative to the logit link, any other standard link function $g(\cdot)$ for a binary variable, such as the probit link or complementary log-log link, are also possible.

For convenience, let us drop $\mathbf{X}$ in (7). The log-likelihood for a MERGM is

$$l(\Psi, \theta; \mathbf{y}) = \Psi' \mathbf{y} + \eta(\theta)' \mathbf{Z}(\mathbf{y}) - \kappa(\Psi, \theta).$$

Let $\nabla \eta(\theta)$ denote the $q \times p$ matrix of partial derivatives of η with respect to θ . Also, let $\mathbf{D} = \partial \pi / \partial \mathbf{v}$, for the logistic link $\mathbf{D} = \text{diag}(\text{var}(\mathbf{Y}))$.

The likelihood equations have the following forms:

$$\mathbf{X}_f' \mathbf{D} \mathbf{C}_Y^{-1}(\mathbf{y}^{\text{obs}} - \pi) = \mathbf{0} \quad (9)$$

and

$$\nabla \eta(\theta)' \{-\mathbf{C}_{\mathbf{Z}, \mathbf{Y}} \mathbf{C}_{\mathbf{Y}}^{-1} (\mathbf{y}^{\text{obs}} - \boldsymbol{\pi}) + \mathbf{z}^{\text{obs}} - \mathbb{E} \mathbf{Z}\} = \mathbf{0}, \quad (10)$$

where $\mathbf{C}_{\mathbf{A}, \mathbf{B}} := \text{cov}(\mathbf{A}, \mathbf{B})$ and $\mathbf{C}_{\mathbf{A}} := \mathbf{C}_{\mathbf{A}, \mathbf{A}}$. The notations $\mathbf{y}^{\text{obs}}$ and $\mathbf{z}^{\text{obs}}$ stand for the observed network and observed network statistics, respectively. The Appendix, provided in the online supplementary materials, shows the details of the derivation. The first set of equations (9) has the standard form of generalized estimating equations (GEE) (Liang and Zeger 1986).

The covariance of the ML estimates $(\hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\theta}}')'$ can be approximated by the inverse of the Fisher information matrix

$$\hat{\mathbf{I}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})^{-1} = \begin{pmatrix} (\mathbf{X}_f' \mathbf{D} \mathbf{C}_{\mathbf{Y}}^{-1} \mathbf{D} \mathbf{X}_f)^{-1} & \mathbf{0} \\ \mathbf{0} & (\nabla \eta(\theta)' \{\mathbf{C}_{\mathbf{Z}} - \mathbf{C}_{\mathbf{Z}, \mathbf{Y}} \mathbf{C}_{\mathbf{Y}}^{-1} \mathbf{C}_{\mathbf{Z}, \mathbf{Y}}'\} \nabla \eta(\theta))^{-1} \end{pmatrix}. \quad (11)$$

The ML estimates can be obtained by a Fisher scoring scheme, where the difference between old and new iterates $(\boldsymbol{\beta}^{\text{new}} = \boldsymbol{\beta}^{\text{old}} + \Delta(\boldsymbol{\beta})$ and $\boldsymbol{\theta}^{\text{new}} = \boldsymbol{\theta}^{\text{old}} + \Delta(\boldsymbol{\theta}))$ is given by

$$\Delta(\boldsymbol{\beta}) = (\mathbf{X}_f' \mathbf{D} \mathbf{C}_{\mathbf{Y}}^{-1} \mathbf{D} \mathbf{X}_f)^{-1} \mathbf{X}_f' \mathbf{D} \mathbf{C}_{\mathbf{Y}}^{-1} (\mathbf{y}^{\text{obs}} - \boldsymbol{\pi}) \quad (12)$$

$$\begin{aligned} \Delta(\boldsymbol{\theta}) &= (\nabla \eta(\theta)' \{\mathbf{C}_{\mathbf{Z}} - \mathbf{C}_{\mathbf{Z}, \mathbf{Y}} \mathbf{C}_{\mathbf{Y}}^{-1} \mathbf{C}_{\mathbf{Z}, \mathbf{Y}}'\} \nabla \eta(\theta))^{-1} \\ &\quad \times \nabla \eta(\theta)' \{-\mathbf{C}_{\mathbf{Z}, \mathbf{Y}} \mathbf{C}_{\mathbf{Y}}^{-1} (\mathbf{y}^{\text{obs}} - \boldsymbol{\pi}) + \mathbf{z}^{\text{obs}} - \mathbb{E} \mathbf{Z}\}. \end{aligned} \quad (13)$$

The scoring equations do not only depend on $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ but also on $\boldsymbol{\Psi}$. It requires a new iterate of $\boldsymbol{\Psi}$ to apply another iteration. There are no closed-form expressions for $\boldsymbol{\Psi}$ depending on $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. FL93 circumvented this problem by applying the iterative proportional fitting (IPF) algorithm in each step to first obtain a solution for the complete joint distribution, from which all other quantities needed in (12) and (13) can be computed. Even though they were dealing with only a few binary observations (in this case, there are only a few joint probabilities), the IPF algorithm is time-consuming. For ERGMs, there are 2^N probabilities and this method is not applicable, even if n is relatively small, say $n = 5$, then $2^N = 2^{\binom{5}{2}} = 1024$ is nearly infeasible for the IPF algorithm. Therefore, other methods need to be considered. The next subsections address how ML estimation can still be achieved and follow in a similar fashion to HH06. For illustration purposes, formulas are presented in Sections 3.2 and 3.3, which can also be found implicitly in HH06, but formulas in HH06 do not contain explicitly the additional term $\boldsymbol{\Psi}' \mathbf{y}$.

3.2 MCMC PRELIMINARIES

First define $\boldsymbol{\alpha} =: (\boldsymbol{\Psi}', \boldsymbol{\theta}')'$ and $U(\boldsymbol{\alpha}, \boldsymbol{\alpha}^0, \mathbf{y}) := \exp[(\boldsymbol{\Psi} - \boldsymbol{\Psi}^0)' \mathbf{y} + (\eta(\boldsymbol{\theta}) - \eta(\boldsymbol{\theta}^0))' \mathbf{Z}(\mathbf{y})]$. Consider two distinct $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^0$ and write

$$r(\boldsymbol{\alpha}, \boldsymbol{\alpha}^0) := l(\boldsymbol{\alpha}) - l(\boldsymbol{\alpha}^0) = \log U(\boldsymbol{\alpha}, \boldsymbol{\alpha}^0, \mathbf{y}^{\text{obs}}) - (\kappa(\boldsymbol{\alpha}) - \kappa(\boldsymbol{\alpha}^0)).$$

Then, the last term on the right-hand side can be reexpressed as

$$\exp(\kappa(\boldsymbol{\alpha}) - \kappa(\boldsymbol{\alpha}^0)) = \mathbb{E}_{\boldsymbol{\alpha}^0} U(\boldsymbol{\alpha}, \boldsymbol{\alpha}^0, \mathbf{Y}). \quad (14)$$

This expectation might be approximated by a sample $\mathbf{y}_1, \dots, \mathbf{y}_m$, which may be obtained by an MCMC algorithm, from the random graph distribution for given α^0 by

$$\hat{r}_m(\alpha, \alpha^0) := \log U(\alpha, \alpha^0, \mathbf{y}^{\text{obs}}) - \log \left[\frac{1}{m} \sum_{i=1}^m U(\alpha, \alpha^0, \mathbf{y}_i) \right].$$

The term $\hat{r}_m(\alpha, \alpha^0)$ converges almost surely to $r(\alpha, \alpha^0)$ as $m \rightarrow \infty$. For fixed sample size m , maximization of $\hat{r}_m(\alpha, \alpha^0)$ for fixed α^0 as a function of α provides an approximation of the ML estimator $\tilde{\alpha}$. This procedure, called MCMC maximum likelihood estimation (MCMCML), was pioneered by Geyer and Thompson (1992) and suggested by HH06 to fit CERGMS. Snijders (2002) considered a Robbins–Monro (1951)-type algorithm for solving the likelihood equations of an ERGM. For details of obtaining an estimate of the log-likelihood and the likelihood ratio statistics, see HH06.

3.3 OBTAINING MCMC SAMPLES

Let $\mathbf{y}^{\text{cur}}$ be the current network and assume that the aim is to sample another network. Then, we use a stochastic or deterministic process to determine a pair (i, j) and then decide whether $Y_{ij} = 1$ or $Y_{ij} = 0$. When fixing the rest of the graph ($Y_{ij}^c = y_{ij}^c$),

$$\frac{\Pr(Y_{ij} = 1 | Y_{ij}^c = y_{ij}^c)}{\Pr(Y_{ij} = 0 | Y_{ij}^c = y_{ij}^c)} = \exp(\Psi_{ij} + \eta(\theta)' \Delta(\mathbf{Z}(\mathbf{y}))_{ij}), \quad (15)$$

where $\Delta(\mathbf{Z}(\mathbf{y}))_{ij}$ is the change statistic for $\mathbf{Z}$ defined previously. This formula is slightly different from (3), because it contains additionally Ψ_{ij} .

This method is called Gibbs sampling. An alternative is the Metropolis algorithm, for which we need to propose transitions from $\mathbf{y}^{\text{cur}}$ to $\mathbf{y}^{\text{prop}}$. The algorithm accepts $\mathbf{y}^{\text{prop}}$ with probability $\min(1, \Pr(\mathbf{Y} = \mathbf{y}^{\text{prop}}) / \Pr(\mathbf{Y} = \mathbf{y}^{\text{cur}}))$. The ratio is

$$\frac{\Pr(\mathbf{Y} = \mathbf{y}^{\text{prop}})}{\Pr(\mathbf{Y} = \mathbf{y}^{\text{cur}})} = \exp[\Psi'(\mathbf{y}^{\text{prop}} - \mathbf{y}^{\text{cur}}) + \eta(\theta)'(\mathbf{Z}(\mathbf{y}^{\text{prop}}) - \mathbf{Z}(\mathbf{y}^{\text{cur}}))]. \quad (16)$$

When $\mathbf{y}^{\text{prop}}$ and $\mathbf{y}^{\text{cur}}$ only differ by a single edge, that is, $\mathbf{y}_{ij}^{\text{prop}} - \mathbf{y}_{ij}^{\text{cur}} = \pm 1$, then the right-hand side of (16) reduces to $\exp(\pm[\Psi_{ij} + \eta(\theta)' \Delta(\mathbf{Z}(\mathbf{y}))_{ij}])$; see Equation (15). When $\mathbf{y}^{\text{cur}}$ and $\mathbf{y}^{\text{prop}}$ differ substantially, then we can consider a sequence of networks, two consecutive networks only differing by one pair (i, j) , and the sequence starting with the current network and finishing with the proposed network. For each step, the ratio is a simple function of change statistics, making the ratio (16) relatively easily computable.

For our purposes, we use the R (R Development Core Team 2010) package *ergm* (Handcock et al. 2010) to simulate an MCMC sample, because all common network statistics are implemented and also the efficient calculation of change statistics makes fast simulation of the MCMC sample possible. However, *ergm* does not allow the specification of an ERGM with the term $\Psi' \mathbf{y}$ directly, so we apply a trick and specify the edge covariates as $y_{ij} \Psi_{ij}$, usually of the form $y_{ij} f(\mathbf{X}_i, \mathbf{X}_j)$, and set the corresponding canonical parameter η_{ij} to 1.

Another problem occurs when the current *ergm* package is used to obtain a sample $\mathbf{y}_1, \dots, \mathbf{y}_m$ of networks needed to fit a MERGM. The *ergm* package provides a sample of network statistics $\mathbf{Z}(\mathbf{y}_1), \dots, \mathbf{Z}(\mathbf{y}_m)$ to solve the likelihood equations for an ERGM. For the

Lazega dataset and for $m = 10,000$, this only takes a few seconds on a PC with a 2.83-GHz Intel C2Q Q9550. However, we require a sample of networks to fit MERGMs and this can take several minutes rather than seconds. Some implementation changes of the ergm package should reduce this computation time from minutes to seconds, as for ERGMs.

Handcock (2003) reported that computational failure is likely to occur if MCMC samples generated under α do not cover the observed sufficient statistics. For MERGMs, the fitted marginal GLM causes the number of edges generated by MCMC to be near the observed number of edges. Heuristically, this in turn implies that $\mathbf{Z}(\mathbf{y}_1), \dots, \mathbf{Z}(\mathbf{y}_m)$ are more likely to cover $\mathbf{Z}(\mathbf{y}^{\text{obs}})$. This is supported by our experience from generating MCMC samples, and computational failure, for example, due to degeneracy, seems less likely to occur for MERGMs compared with ERGMs.

3.4 FITTING ALGORITHM

First, there are two parameterizations of the full model to be aware of: $\alpha = (\Psi', \theta')'$ and $\zeta := (\beta', \theta')'$. Second, initial parameter values for which the MCMC sample was generated are denoted by α^0 (needs to satisfy (8) with β^0). Parameters of the k th iteration still based on this MCMC sample generated at α^0 are denoted by $\beta^{(k)}$, $\theta^{(k)}$, and $\Psi^{(k)}$ and for $k = 0$: $\beta^{(0)} := \beta^0$, $\theta^{(0)} := \theta^0$ and $\Psi^{(0)} := \Psi^0$.

The main algorithm, a Fisher scoring algorithm, is presented next, followed by a detailed explanation of the steps:

Algorithm 1 Main algorithm.

- 0 Select initial values β^0 , θ^0 , and Ψ^0 satisfying (7) and (8)
- 1 Obtain MCMC sample $\mathbf{y}_1, \dots, \mathbf{y}_m$, along with $\mathbf{Z}(\mathbf{y}_1), \dots, \mathbf{Z}(\mathbf{y}_m)$, for given α^0 , set $k := 0$, go to Step 3.
- 2 Obtain $\Psi^{(k+1)}$ that solves

$$\tilde{\pi}(\Psi^{(k+1)}, \theta^{(k)}) = \pi(\beta^{(k)})$$

for given MCMC sample generated at α^0 ; if $\widehat{\text{var}}_{\text{MC}}(\hat{r}_m)$ in Equation (17) is too large, say $\widehat{\text{var}}_{\text{MC}}(\hat{r}_m) > c$, then set $k := 0$ and obtain new MCMC sample.

- 3 Estimate $\mathbf{C}_Y$, $\mathbf{C}_{Z,Y}$, $\mathbf{C}_Z$, and $\mathbb{E}\mathbf{Z}$ from MCMC sample for given α^0 and $\alpha^{(k)}$ needed for formulas (12) and (13); see below for details of estimation of moments.
- 4 Apply iteration schemes (12) and (13); new estimates are obtained via

$$\begin{aligned}\beta^{(k+1)} &= \beta^{(k)} + \gamma_{\beta^{(k)}} \Delta(\beta^{(k)}) \\ \theta^{(k+1)} &= \theta^{(k)} + \gamma_{\theta^{(k)}} \Delta(\theta^{(k)});\end{aligned}$$

- 5 If converged, consider $\tilde{\beta} = \beta^{(k+1)}$ and $\tilde{\theta} = \theta^{(k+1)}$ as MCMCML estimates; stop.
 - 6 $k := k + 1$; return to Step 2.
-

Compared with the algorithm presented by HH06, Step 2 is new, and Step 4 is based on different likelihood equations and uses the step sizes $\gamma_{\beta^{(k)}}$ and $\gamma_{\theta^{(k)}}$, which often equal 1. Our experience has shown that step sizes should rather be smaller, for example, 0.2–0.5, making the iteration scheme more stable.

3.4.1 Computation of Expectations in Step 3 and of $\widehat{\text{var}}_{\text{MC}}(\hat{f}_m)$ in Step 2. Previously, we defined $U(\alpha, \alpha^0, \mathbf{Y})$. We write $U_1, \dots, U_m$ for $U(\alpha, \alpha^0, \mathbf{y}_1), \dots, U(\alpha, \alpha^0, \mathbf{y}_m)$ for the given MCMC sample $\mathbf{y}_1, \dots, \mathbf{y}_m$ at α^0 , and similarly $\mathbf{Z}_i$ for $\mathbf{Z}(\mathbf{y}_i)$. Define the weights

$$\omega_i^{(k)} := \frac{U_i}{\sum_{j=1}^m U_j}.$$

Clearly, these weights are equal, that is, $\omega_i^{(k)} = 1/m$, if $\alpha^{(k)} = \alpha^0$.

$$\widehat{\text{var}}_{\text{MC}}(\hat{f}_m) := \frac{1}{m^2 \bar{U}^2} \sum_{k=-K}^K (m - |k|) \phi_k, \quad (17)$$

where $\phi_k = \phi_{-k}$ is the auto-covariance of the sequence $U_1, \dots, U_m$ and $\bar{U} := \frac{1}{m} \sum_{i=1}^m U_i$. This is the same equation as in HH06. The expectations can be estimated by weighted sums given by

$$\widehat{\mathbb{E}}\mathbf{Z} = \sum_{i=1}^m \omega_i^{(k)} \mathbf{Z}_i \quad \hat{\mathbf{C}}_{\mathbf{Z}, \mathbf{Y}} = \sum_{i=1}^m \omega_i^{(k)} \mathbf{Z}_i \mathbf{Y}_i' - \mathbb{E}\mathbf{Z}(\mathbb{E}\mathbf{Y})', \quad (18)$$

and is similar for all other expectations, because generally for a function h

$$\mathbb{E}_{\alpha}\{h(\mathbf{Y})\} = \frac{\mathbb{E}_{\alpha^0}\{h(\mathbf{Y})U(\alpha, \alpha^0, \mathbf{Y})\}}{\exp(\kappa(\alpha) - \kappa(\alpha^0))},$$

which can be derived in a similar way as (14).

These estimated expectations are identical to sample means and sample covariances for a fresh MCMC sample, when $\alpha^{(k)} = \alpha^0$. Equation (18) enables estimation of moments without generating new MCMC samples. Step 2 ($\widehat{\text{var}}_{\text{MC}}(\hat{f}_m) > c$) says, when the variation in $U_1, \dots, U_m$ is too large, then we need to generate a new MCMC sample, because the old sample is too unreliable in obtaining estimates of the expectations.

3.4.2 Least-Squares Algorithm for Solving for Ψ . Step 2 of the main algorithm is complex. We need to obtain $\Psi^{(k+1)}$ that solves

$$\pi(\beta^{(k)}) = \tilde{\pi}(\Psi^{(k+1)}, \theta^{(k)}),$$

with the previously defined

$$\pi(\beta) = g^{-1}(\mathbf{X}_f \beta)$$

and

$$\tilde{\pi}(\Psi, \theta) := \sum_{\mathbf{y} \in \mathcal{S}} \mathbf{y} \exp(\Psi^T \mathbf{y} + \eta(\theta)^T \mathbf{Z}(\mathbf{y}) - \kappa(\Psi, \theta)).$$

The vector of probabilities $\pi(\beta)$ is given by the GLM (8), and $\tilde{\pi}(\Psi, \theta) [= \tilde{\pi}(\alpha)]$ is given by (7). Define $\mathbf{h}(\Psi) := \pi(\beta) - \tilde{\pi}(\Psi, \theta)$. The equation $\mathbf{h} = \mathbf{0}$ can be solved iteratively (using

index j for subiterations for the k th step of the main algorithm) by nonlinear least squares

$$\Delta(\Psi)^{(k,j)} = (\nabla \mathbf{h}(\Psi^{(k,j)})' \mathbf{W} \nabla \mathbf{h}(\Psi^{(k,j)}))^{-1} \nabla \mathbf{h}(\Psi^{(k,j)})' \mathbf{W} \mathbf{h}(\Psi^{(k,j)}), \quad (19)$$

where $\Psi^{(k,j)}$ is the current iterate of Ψ and $\nabla \mathbf{h}(\Psi^{(k,j)})$ is the derivative of $\mathbf{h}$ with respect to Ψ at $\Psi^{(k,j)}$. This derivative reduces to $\nabla \mathbf{h}(\Psi) = \partial \pi(\alpha) / \partial \Psi = \mathbf{C}_Y(\Psi)$. Matrix $\mathbf{W}$ is a weight matrix; for simplicity, let $\mathbf{W} = \mathbf{I}$. Equation (19) then reduces to

$$\Delta(\Psi^{(k,j)}) = \mathbf{C}_Y^{-1}(\Psi^{(k,j)}) \mathbf{h}(\Psi^{(k,j)}). \quad (20)$$

We suggest the following algorithm for Step 2 of the main algorithm to obtain $\Psi^{(k+1)}$ for given $\beta^{(k)}$ and $\theta^{(k)}$.

Algorithm 2 Step 2 of the main algorithm.

2.0 Given is an MCMC sample $\mathbf{y}_1, \dots, \mathbf{y}_m$ at $\alpha^0 = ((\Psi^0)', (\theta^0)')'$; set $j := 0$ and set $\Psi^{(k,j)} := \Psi^{(k)}$.

2.1 Estimate $\mathbf{C}_Y^{-1}(\Psi^{(k,j)})$ and $\mathbf{h}(\Psi^{(k,j)})$ from the MCMC sample.

2.2 Adjust step size $\gamma_{\Psi^{(j)}}$ and $\mathbf{C}_Y^{-1}$ according to $\|\mathbf{h}(\Psi^{(j)})\|$, where $\|\cdot\|$ is some norm.

2.3 Apply Equation (20) for current $\Psi^{(k,j)}$ and obtain $\Psi^{(k,j+1)}$ by

$$\Psi^{(k,j+1)} = \Psi^{(k,j)} + \gamma_{\Psi^{(k,j)}} \Delta(\Psi^{(k,j)}).$$

2.4 Update weights $\omega_i^{(j+1)}$.

2.5 If $\widehat{\text{var}}_{\text{MC}}(\hat{f}_m)$ in Equation (17) is too large, say $\widehat{\text{var}}_{\text{MC}}(\hat{f}_m) > c$, then $\Psi^0 := \Psi^{(k,j+1)}$ and obtain a new MCMC sample for Ψ^0 and $\theta^0 := \theta^{(k)}$; in this case, set $\beta^0 := \beta^{(k)}$, $j := 0$, and $k := 0$, and go to Step 2.1.

2.6 Convergence is achieved when

$$T := m(\widehat{\mathbf{h}}(\Psi^{(k,j+1)})' (\widehat{\mathbf{C}}_Y(\Psi^{(k,j+1)}))^{-1} \widehat{\mathbf{h}}(\Psi^{(k,j+1)})) \leq T_2^\epsilon [N, m-1] =: T_2^\epsilon;$$

if converged, stop and consider $\Psi^{(k+1)} := \Psi^{(k,j+1)}$ as the solution of Step 2, otherwise continue with 2.7.

2.7 Go to Step 2.1, $j := j + 1$.

Due to the estimation of $\mathbf{h}$ based on the MCMC sample, we cannot determine exactly whether indeed $\mathbf{h} = \mathbf{0}$, that is, whether $\tilde{\pi}(\Psi^{(k+1)}, \theta^{(k)}) = \pi(\beta^{(k)})$ due to the stochastic approximations. Let $\tilde{\pi}(\Psi^{(k+1)}, \theta^{(k)})$ denote the true vector of probabilities and $\hat{\pi}(\Psi^{(k+1)}, \theta^{(k)}) := \sum_{i=1}^m \mathbf{y}_i / m$, the estimate based on the MCMC sample $\mathbf{y}_1, \dots, \mathbf{y}_m$. We know $\sqrt{m} \hat{\pi}(\Psi, \theta) \sim \text{Bin}(\tilde{\pi}(\Psi, \theta), m)$; hence, $\sqrt{m} \hat{\pi}(\Psi, \theta) \sim_d N(\tilde{\pi}(\Psi, \theta), \mathbf{C}_Y)$. In Step 2.6, we apply Hotelling's T^2 test, since $\mathbf{C}_Y$ also needs to be estimated. $T_2^\epsilon := T_2^\epsilon [N, m-1]$ is the ϵ -quantile of the T^2 -distribution with parameters N and $m-1$. Note $\frac{a-b+1}{ab} T_{a,b}^2 = F_{a,b-a+1}$, where $F_{a,b}$ is the F -distribution with parameters a and b . We chose $\epsilon = 0.1$. Be aware that $\mathbf{y}_1, \dots, \mathbf{y}_m$ are not independent due to the MCMC technique; therefore, the T^2 -distribution applies only approximately. This test serves as a stopping criterion.

Step 2.2 includes a step size γ_{Ψ} . As for γ_{β} and γ_{θ} in the main algorithm, this step size ideally equals 1, but we found that tuning these and other parameters made the algorithm more stable. The norm $\|\mathbf{h}(\Psi)\|$ can be any distance measure. For our implementation, we use T , because this Mahalanobis distance is already used as a stopping criterion in Step 2.6. When T is large, say $T > 2 \cdot T_2^{\epsilon}$, then we use $\gamma_{\Psi} = 0.5$; otherwise, the step size is reduced in each step by, say, 5% until it reaches a minimal step size, say 0.005.

This might be exactly the opposite of what one might expect, but a large step size (e.g., 0.9) near the solution resulted frequently in a big jump away from the solution. We also modified $\mathbf{C}_Y^{-1}$ in Equation (20) according to $\|\mathbf{h}(\Psi)\|$, that is, $\mathbf{C}_Y$ is replaced by

$$\lambda \mathbf{C}_Y + (1 - \lambda) \text{var}(\mathbf{Y}),$$

where $\text{var}(\mathbf{Y})$ is the diagonal matrix with variance of $\mathbf{Y}$ on its diagonal. Parameter $\lambda \in [0, 1]$ should be 1 when T is close to T_2^{ϵ} , say $T < 1.2 \cdot T_2^{\epsilon}$. If far apart, say $T > 2 \cdot T_2^{\epsilon}$, then we set $\lambda = 0$. If $1.2 \cdot T_2^{\epsilon} \leq T \leq 2 \cdot T_2^{\epsilon}$, then $\lambda = 1 - \max\{(T - T_2^{\epsilon})/T_2^{\epsilon}, 0\}$. This tuning worked well for the Lazega (2001) dataset.

3.4.3 Starting Values. Step 2 aims to find a Ψ that solves $\mathbf{h}(\Psi) = \mathbf{0}$ for given β and θ , that is, finding a Ψ effectively such that $\mathbf{h}(\Psi) \approx \mathbf{0}$ would speed up the algorithm in Step 2 dramatically. In the first iteration of the main algorithm, one can start with $\eta(\theta^0) = \mathbf{0}$, implying unrealistically that dyads are independent, because then, Ψ^0 is easily determined, for example, for the logit link $\Psi^0 \equiv \mathbf{X}_f \beta^0$. However, the next step of the main algorithm will usually require a large $\Delta\theta$, which implies that Ψ^0 is far away from solving $\mathbf{h}(\Psi^0) = \mathbf{0}$. Alternatively, one may first obtain an initial estimate θ^0 by fitting an ERGM and then use standard GLM routines to obtain an initial estimate β^0 . This brings us back to the question of how to find efficiently Ψ with $\mathbf{h}(\Psi) \approx \mathbf{0}$ before applying the proposed algorithm in Step 2.

3.4.4 Sequential Solving for Ψ . An alternate method to the one proposed in Step 2 could be to solve each component of $\mathbf{h} = \mathbf{0}$, that is, $\tilde{\pi}_{ij}(\Psi^{(k+1)}, \theta^{(k)}) = \pi_{ij}(\beta^{(k)})$, directly as a function of $\Psi_{ij}^{(k+1)}$. The marginal probability $\tilde{\pi}_{ij} = \sum_{\mathbf{y} \in S: Y_{ij}=1} \Pr(\mathbf{Y} = \mathbf{y}; \theta)$ is difficult to compute, as it needs an estimate of the normalizing constant κ .

Let us express the marginal probability implied by an ERGM as

$$\tilde{\pi}_{ij} = \sum_{\mathbf{y} \in S: Y_{ij}=1} \Pr(Y_{ij} = 1 | Y_{ij}^c = y_{ij}^c) \Pr(Y_{ij}^c = y_{ij}^c),$$

which can be estimated from the MCMC sample by [see Equation (15)]

$$\tilde{\pi}_{ij} \approx \sum_{k=1}^m \Pr((\mathbf{Y}_k)_{ij} = 1 | (\mathbf{Y}_k)_{ij}^c = (\mathbf{y}_k)_{ij}^c) = \sum_{k=1}^m \text{expit}(\Psi_{ij} + \eta(\theta)' \Delta(\mathbf{Z}(\mathbf{y}_k))_{ij}), \quad (21)$$

where $\mathbf{y}_k$ is the k th network of the MCMC sample and $(\mathbf{y}_k)_{ij}$ is the dyad ij of the k th network. The function $\text{expit}(x) = \exp(x)/(1 + \exp(x))$ is the inverse of the logit link. Assume that we would know all $\Delta(\mathbf{Z}(\mathbf{y}_k))_{ij}$ for the MCMC sample, then we simply need to solve

$$\pi_{ij}(\beta) - \sum_{k=1}^m \omega_i \times \text{expit}(\Psi_{ij} + \eta(\theta)' \Delta(\mathbf{Z}(\mathbf{y}_k))_{ij}) = 0, \quad (22)$$

which can be solved very quickly for each Ψ_{ij} , given β and θ , by standard optimization routines. After we have solved (22) for the pair ij , we can proceed to solve for the pair uv , but weights ω_i need to be updated due to the change from $\Psi_{ij}^{(k)}$ to $\Psi_{ij}^{(k+1)}$, resulting from solving Equation (22).

When $\widehat{\text{var}}_{\text{MCMC}}(\hat{r}_m)$ in Equation (17) is too large, say $\widehat{\text{var}}_{\text{MCMC}}(\hat{r}_m) > c$, then we need to obtain a new MCMC sample for $\Psi^0 := \Psi^{(k+1)}$ and $\theta^0 := \theta^{(k)}$.

There are two main issues with this algorithm. The first is that we need to solve for N parameters in each step k , and each large change in Ψ might require generating a new MCMC sample. The second issue is the computation of all $\Delta(\mathbf{Z}(\mathbf{y}_k))_{ij}$ of the complete MCMC sample. We computed this for each network of the MCMC sample directly, that is, N change statistics for each of the m , say $m = 10,000$, networks. Even though we implemented it efficiently, it is still not efficient enough, taking roughly 40 min for the Lazega (2001) dataset.

However, there is a relatively easy solution. When the MCMC sample is created, each sampled network $\mathbf{y}_k$ differs by the predecessor $\mathbf{y}_{k-1}$ by a few dyads. Assume that $\Delta(\mathbf{Z}(\mathbf{y}_{k-1}))_{ij}$ are known. Then, the change statistics $\Delta(\mathbf{Z}(\mathbf{y}_k))_{ij}$ can be computed from $\Delta(\mathbf{Z}(\mathbf{y}_{k-1}))_{ij}$ and the knowledge of the dyads Y_{ij} that have changed. The only challenge remaining is its implementation in the existing *ergm* package, along with an output of the networks $\mathbf{y}_1, \dots, \mathbf{y}_m$, as previously mentioned.

3.5 MONTE CARLO ERROR

Approximating the ML estimates $\hat{\zeta}$ (previously defined by $\zeta = (\beta', \theta')'$) by $\tilde{\zeta}$ incurs another error, the Monte Carlo error. Equivalently, let $\hat{\alpha}$ and $\tilde{\alpha}$ denote the estimates for the alternate parameterization $\alpha = (\Psi', \theta')'$.

Following the same arguments as in HH06, the approximate covariance matrix for $\tilde{\zeta}$ is

$$\frac{1}{m} \hat{\mathbf{I}}(\tilde{\zeta})^{-1} \tilde{\mathbf{V}} \hat{\mathbf{I}}(\tilde{\zeta})^{-1},$$

where $\hat{\mathbf{I}}(\tilde{\zeta})$ is the estimated Fisher information matrix; see Equation (11). Covariance matrix $\mathbf{V}$ is defined by

$$\tilde{\mathbf{V}} := \frac{1}{m^2} \left[\sum_{i=1}^m U(\tilde{\alpha}, \alpha^0, \mathbf{y}_i) \right]^2 \sum_{k=-K}^K \hat{\xi}_k,$$

where $\hat{\xi}_k$ is the sample lag- k auto-covariance matrix of the sequence $\mathbf{W}_1(\tilde{\alpha})$, $\mathbf{W}_2(\tilde{\alpha})$, $\dots$, $\mathbf{W}_m(\tilde{\alpha})$, with

$$\mathbf{W}_i(\alpha) = \begin{pmatrix} \mathbf{X}_f' \mathbf{D} \mathbf{C}_Y^{-1} (\mathbf{y}^{\text{obs}} - \mathbf{y}_i) \times U_i(\alpha, \alpha_0) \\ \nabla \eta(\theta)' \{ -\mathbf{C}_{Z,Y} \mathbf{C}_Y^{-1} (\mathbf{y}^{\text{obs}} - \mathbf{y}_i) + \mathbf{z}^{\text{obs}} - \mathbb{E} \mathbf{Z}_i \} \times U_i(\alpha, \alpha_0) \end{pmatrix}.$$

4. EXAMPLE—LAZEGA DATASET

The Lazega dataset, along with a detailed description, is available at http://www.stats.ox.ac.uk/~snijders/siena/Lazega_lawyers_data.htm. An edge between

Table 1. Estimates (Est.) for the ERGM with and without GWESP distribution (logistic regression) for the Lazega dataset, along with standard errors (S.E.) and p -values

Parameters	ERGM without GWESP (logistic regression)			Odds ratio $\exp(\beta_k)$	ERGM with GWESP		
	Est.	(S.E.)	p -value		Est.	(S.E.)	p -value
Edges (β_0)	−8.306	(0.953)	$< 2e - 16$	0.002	−7.551	(0.833)	$< 2e - 16$
Main seniority (β_1)	0.044	(0.009)	$8.9e - 07$	1.045	0.033	(0.008)	$1.6e - 05$
Main practice (β_2)	0.902	(0.163)	$3.1e - 08$	2.464	0.615	(0.134)	$5.7e - 06$
Sim practice (β_3)	0.879	(0.231)	0.00014	2.408	0.811	(0.205)	$8.3e - 05$
Sim gender (β_4)	1.128	(0.348)	0.00121	3.089	0.866	(0.305)	0.0083
Sim office (β_5)	1.653	(0.254)	$7.6e - 11$	5.222	1.361	(0.220)	$1.1e - 09$
GWESP (θ_1)	—	—	—	—	0.456	(0.041)	$< 2e - 16$
GWESP (θ_2)	—	—	—	—	0.656	(0.151)	$1.8e - 05$

two partners of a Northeastern U.S. corporate law firm exists, that is, $Y_{ij} = 1$, if both partners indicate collaboration with each other. The data also contain a number of attributes for each partner: seniority (1, 2, . . . , 36: rank number of entry into the firm), practice (litigation = 1, corporate law = 2), gender (1 = man; 2 = woman) and office (1 = Boston; 2 = Hartford; 3 = Providence).

Table 1 shows the fitting results for the ERGM with the statistics $y_{ij}f_k(\mathbf{X}_i, \mathbf{X}_j); k = 1, \dots, 5$ and the edges statistics E , and for the same ERGM with additional GWESP distribution, characterized by parameters θ_1 and θ_2 ; see Section 2 for how the functions $f_k(\mathbf{X}_i, \mathbf{X}_j)$ and the network statistics were defined. The ERGM without the GWESP distribution is equivalent to a standard logistic regression model of the form

$$\log \left(\frac{\pi_{ij}}{1 - \pi_{ij}} \right) = \beta_0 + \sum_{k=1}^5 \beta_k f_k(\mathbf{X}_i, \mathbf{X}_j).$$

The fitting results for the MERGM based on the proposed algorithm are presented in Table 2. We started with sampling $m = 10,000$ networks and finished with $m = 30,000$ to obtain higher accuracy. We also increased the step size (also known as thinning factor)

Table 2. Estimates (Est.) for the marginalized ERGM with GWESP distribution for the Lazega dataset, along with standard errors (S.E.) and p -values

Variables	Est.	Accounting for MCMC error		Not accounting for MCMC error		Odds ratio $\exp(\beta_k)$
		(S.E.)	p -value	(S.E.)	p -value	
Edges (β_0)	−7.383	(1.609)	$4.7e - 06$	(1.350)	$3.3e - 08$	0.0006
Main seniority (β_1)	0.039	(0.014)	0.00294	(0.012)	0.00066	1.0395
Main practice (β_2)	0.764	(0.347)	0.01400	(0.275)	0.00280	2.1475
Sim practice (β_3)	0.880	(0.262)	0.00041	(0.271)	0.00061	2.4109
Sim gender (β_4)	0.783	(0.483)	0.05272	(0.429)	0.03421	2.1882
Sim office (β_5)	1.757	(0.393)	$4.6e - 06$	(0.369)	$1.2e - 06$	5.7962
GWESP (θ_1)	0.878	(0.288)	0.00120	(0.269)	0.00058	—
GWESP (θ_2)	0.863	(0.214)	$3.1e - 05$	(0.212)	$2.7e - 05$	—

from 1000 at the beginning of the algorithm to 3000 at the end of the algorithm for the MCMC chain to obtain an MCMC sample with less dependence. Our proposed algorithm to fit a MERGM needed roughly 20 main iterations, and each iteration needed roughly 10–50 subiterations to solve for Ψ with the proposed least-squares algorithm, in total roughly 10 hr. We often created a new MCMC sample, even though this was not needed because $\widehat{\text{var}}_{\text{MC}}(\hat{f}_m)$ was small. Table 2 also shows the standard errors and p -values when not accounting for the MCMC error. Ignoring this error might lead to too-small standard errors.

The β_k coefficients (for convenience, now called β_k instead of η_k ; see Section 2) of the ERGM with GWESP distribution, see Table 1, can be used to make interpretation in terms of “conditional” log odds ratios, conditional on the rest of the graph. For example, the odds of collaboration between two partners of the same gender are $\exp(1.128) = 3.089$ times the odds of collaboration between two partners of different genders, conditional on the rest of the graph and assuming that all other covariates are the same. The logistic regression model and the MERGM allow an “unconditional” or population-averaged interpretation using the odds ratios, unconditionally on the rest of the graph. The odds of collaboration between two partners of the same gender are $\exp(0.7831) = 2.188$ times those for partners of different genders, assuming that all other covariates are the same.

Let us now focus on the computation of marginal probabilities. Suppose $f_1 = 3$ (ranks 1 and 2), $f_2 = 3$ and $f_3 = 0$ (implies practices 1 and 2), $f_4 = 1$ (same gender), and $f_5 = 1$ (same office). For the MERGM, we obtain a unique estimate $\hat{\Pr}(Y_{ij} = 1 | \mathbf{X}_i, \mathbf{X}_j) = \text{expit}(-7.383 + 3 \cdot 0.039 + 3 \cdot 0.764 + 0 \cdot 0.880 + 1 \cdot 0.783 + 1 \cdot 1.757) = \text{expit}(-2.434) = 0.0806$.

For the ERGM with GWESP distribution, an MCMC sample with $m = 1000$ networks is obtained. One dyad has this covariate setting and we obtain $\hat{\Pr}(Y_{ij} = 1 | \mathbf{X}) = 0.041$ as an estimate from the MCMC sample. Let us consider another setting with $f_1 = 10$, while f_2, f_3, f_4 , and f_5 remain unchanged. Three dyads have this setting, the estimated marginal probabilities are 0.056, 0.050, and 0.062. We might use an average of these three numbers as a population estimate. For the MERGM, we obtain $\hat{\Pr}(Y_{ij} = 1) = \text{expit}(-2.161) = 0.103$. We might also be interested in a hypothetical covariate setting that is not present in the current sample. For the MERGM, we only need to apply the inverse link for a given set of covariates. For the ERGM, we need to first choose dyads for which we change the observed covariate setting to the desired setting. Then, a new MCMC sample can be obtained to approximate the marginal probabilities.

This illustrates that for ERGMs, the marginal probabilities are only uniquely defined for a particular dyad and a particular $\mathbf{X}$, whereas for a MERGM, they are uniquely defined for a particular covariate setting of partners i and j , that is, $\mathbf{X}_i$ and $\mathbf{X}_j$.

Another important point needs to be made. Tables 1 and 2 might suggest that both approaches, ERGM and MERGM, give roughly similar estimates β_k . Simpson’s paradox (Agresti 2002, p. 51) is a special case where unconditional and conditional effects can be in opposite directions, showing that parameters can be quite different, and that one does not imply the other.

Table 3 shows the deviances for the models considered in this article. In this example, the ERGM with the GWESP distribution provides a better fit than the MERGM. While the fit of the ERGM is better than the fit of the MERGM, it does not imply the ERGM is preferred over the MERGM when our primary focus is on the marginal model. Importantly,

Table 3. Deviances for null model, logistic model, MERGM, and ERGM

Model	Residual deviance	Deviance	Residual df	<i>p</i> -value
(a) Intercept model	598.78	—	—	—
(b) Logistic regression	501.80	96.98	5	0.000
(c) MERGM	482.31	19.49	2	5.858e − 05 (compared with (b))
(d) ERGM	469.09	13.22	0	7.890e − 08 (compared with (b))

in contrast to logistic regression, the MERGM allows to make valid statistical inference, based on the marginal model, because the MERGM accounts for the typical network-induced dependence structure.

5. DISCUSSION

In this article, we consider a GLM for the marginal probabilities $\Pr(Y_{ij} = 1 | \mathbf{X}_i, \mathbf{X}_j)$ of a network, with covariates defined by the node attributes $\mathbf{X}_i$ and $\mathbf{X}_j$ and where the (joint) distribution of the network is specified by an ERGM to account for the network-typical dependence structure. We motivated the approach by our interest in $\Pr(Y_{ij} = 1 | \mathbf{X}_i, \mathbf{X}_j)$. Such marginalized modeling approaches are quite common—for example, FL93 proposed a marginal GLM with a log-linear model describing the joint distribution. Heagerty and Zeger (2000) proposed a GLMM that is subjected to a marginal GLM and argued that the separation of the marginal model from the joint distribution is advantageous in many applications. So far, ML estimation for such a marginalized approach for ERGMs has not been addressed, because it was assumed that such a problem was infeasible.

The main advantages of MERGMs over ERGMs are that interpretation of parameters associated with exogenous effects of the marginal model is generally easier, and that marginal probabilities are easily obtainable via the inverse link.

For example, when fitting a MERGM to the Lazega (2001) dataset, the estimated similarity effect of gender was 0.7831; hence, the odds of collaboration of partners with the same gender are $\exp(0.7831) = 2.1882$ higher than the odds for partners with different genders, assuming all covariates are the same. The current methodology for ERGMs only allows a conditional interpretation for the sample network, conditional on the rest of the network. Such a conditional interpretation only allows comparisons of dyads within a cluster, where a cluster of dyads is defined as the set of dyads ij for which $\Delta(\mathbf{Z}_2)_{ij}$ is equal. This interpretation is clearly difficult. In contrast, the marginal model allows comparing dyads across the population of 36 partners, avoiding the difficult interpretation in terms of clusters.

When comparing results for logistic regression and MERGMs, increased standard errors relative to logistic regression can be observed for the proposed approach. A standard GLM does not give proper standard errors, but a MERGM does. This is similar to wrongly applying a linear model to clustered data while ignoring the dependence within clusters. Estimated standard errors using such a naive approach assuming independence will be smaller than standard errors using a GLMM.

The drawback of the proposed methodology is that the estimation method is relatively complex. We outlined some algorithms that make estimation possible, but it is very

time-consuming, currently a matter of hours. However, we also outlined simple solutions to reduce the computation time. The existing MCMC sampler of the *ergm* package needs only minor modifications to make this happen. We believe such a modification is the best option to make use of the seemingly hundreds of implemented network statistics and to avoid unnecessary reimplementations.

Another disadvantage of the proposed methodology is that it can only be applied for relatively small networks, those with a small n , because an $\binom{n}{2} \times \binom{n}{2}$ matrix has to be inverted. Modern computers allow inversion of matrices with $\binom{n}{2} \approx 1000\text{--}10,000$. Therefore, an MERGM cannot be fitted for very large networks. The same problem applies to ERGMs, but to a lesser extent. ERGMs have been fitted for $n \approx 2000$ nodes, and clearly, this number of nodes is too large to fit a MERGM.

We do not claim that the proposed algorithm is numerically optimal. Instead, we hope that future research for more efficient algorithms will be stimulated by our article. Another thought posed by the article and question for further consideration is whether multiple independent and multiple dependent (repeated) networks can be fitted. However, we continue to evaluate these questions, and ideally, would like to report outcomes at a later time.

We believe the proposed MERGM approach is preferable over the existing ERGM approach if the main focus is on the relationship between the probability of observing an edge and exogenous effects. On the contrary, if the main focus is on the relationship between the probability of observing a particular edge in the sample network and the rest of the graph, then the standard ERGM approach is appropriate and exogenous effects might be added to improve the model.

SUPPLEMENTARY MATERIALS

R Code: R programs that can be used to replicate the ML fitting for the Lazega dataset.

Please read file README.txt contained in the zip file for more details. (Suesse.zip, zip archive)

Appendix: Gives the derivation of the likelihood equations (9) and (10). (suesse.appendix.pdf)

ACKNOWLEDGMENTS

The author thanks Prof David Steel, Prof Ray Chambers, Steve Hurley, Tung Pham (Ph.D.), and Ph.D. student Sarah Neville for their helpful comments, as well as the referees. The research was supported by the Australian Research Council (ARC) Discovery grant LX0883143.

[Received February 2011. Revised March 2012.]

REFERENCES

- Agresti, A. (2002), *Categorical Data Analysis* (2nd ed.), Wiley Series in Probability and Statistics, New York: Wiley. [886,897]
- Casella, G., and Berger, R. (2002), *Statistical Inference* (2nd ed.), Pacific Grove, CA: Thomson Learning. [884]
- Efron, B. (1978), "Geometry of Exponential Families," *The Annals of Statistics*, 6, 362–376. [884]

- Fitzmaurice, G. M., and Laird, N. M. (1993), "A Likelihood-Based Method for Analyzing Longitudinal Binary Responses," *Biometrika*, 80, 141–151. [888]
- Geyer, C. J., and Thompson, E. A. (1992), "Constrained Monte-Carlo Maximum-Likelihood for Dependent Data," *Journal of the Royal Statistical Society, Series B*, 54, 657–699. [890]
- Handcock, M. (2003), "Assessing Degeneracy in Statistical Models of Social Networks," Technical report, University of Washington. Available at <http://www.csss.washington.edu/Papers>. [891]
- Handcock, M., Hunter, D. R., Butts, C., Goodreau, S. M., and Morris, M. (2010), "ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. Version 2.2-4." Available at <http://statnetproject.org> [890]
- Heagerty, P. J., and Zeger, S. L. (2000), "Marginalized Multilevel Models and Likelihood Inference," *Statistical Science*, 15, 1–19. [887,898]
- Holland, P. W., and Leinhardt, S. (1981), "An Exponential Family of Probability Distributions for Directed Graphs," *Journal of the American Statistical Association*, 76, 33–50. [884]
- Hunter, D. R., and Handcock, M. S. (2006), "Inference in Curved Exponential Family Models for Networks," *Journal of Computational and Graphical Statistics*, 15, 565–583. [884]
- Kim, J. K., and Yu, C. L. (2011), "A Semiparametric Estimation of Mean Functionals With Nonignorable Missing Data," *Journal of the American Statistical Association*, 106, 157–165. [887]
- Lazega, E. (2001), *The Collegial Phenomenon: The Social Mechanism of Cooperation Among Peers in a Corporate Law Partnership*, Oxford: Oxford University Press. [885,894,895,898]
- Lazega, E., and Pattison, P. (1999), "Multiplexity, Generalized Exchange and Cooperation in Organizations: A Case Study," *Social Networks*, 21, 67–90. [885]
- Liang, K. Y., and Zeger, S. L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13–22. [889]
- McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models* (2nd ed.), New York: Chapman & Hall. [884,886]
- Morris, M., Handcock, M., and Hunter, D. (2008), "Specification of Exponential-Family Random Graph Models: Terms and Computational Aspects," *Journal of Statistical Software*, 24, 1–24. [884]
- Neuhaus, J. M. (1992), "Statistical Methods for Longitudinal and Clustered Designs With Binary Responses," *Statistical Methods in Medical Research*, 1, 249–273. [887]
- R Development Core Team (2010), "R: A Language and Environment for Statistical Computing," available online at <http://www.R-project.org>. [890]
- Robbins, H., and Monro, S. (1951), "A Stochastic Approximation Method," *Annals of Mathematical Statistics*, 22, 400–407. [890]
- Snijders, T. (2002), "Markov Chain Monte Carlo Estimation of Exponential Random Graph Models," *Journal of Social Structure*, 1–40. [884,890]
- Snijders, T., Pattison, P., Robins, G., and Handcock, M. (2006), "New Specifications for Exponential Random Graph Models," *Sociological Methodology*, 36, 99–153. [884]
- Strauss, D., and Ikeda, M. (1990), "Pseudolikelihood Estimation for Social Networks," *Journal of the American Statistical Association*, 85, 204–212. [884]