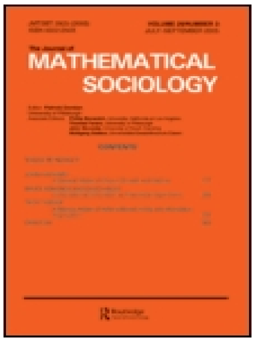

The Journal of Mathematical Sociology

ISSN: 0022-250X (Print) 1545-5874 (Online) Journal homepage: http://www.tandfonline.com/loi/gmas20

A Flexible Parameterization for Baseline Mean Degree in Multiple-Network ERGMs

Carter T. Butts & Zack W. Almquist

Journal of Mathematical Sociology, 39: 163–167, 2015

ISSN: 0022-250X print/1545-5874 online
DOI: 10.1080/0022250X.2014.967851

Routledge
Taylor & Francis Group

Letter to the Editor

A FLEXIBLE PARAMETERIZATION FOR BASELINE MEAN DEGREE IN MULTIPLE-NETWORK ERGMS

Carter T. Butts
Departments of Sociology, Statistics, and EECS, and Institute for Mathematical Behavioral Sciences, University of California, Irvine, California, USA

Zack W. Almquist
Department of Sociology, School of Statistics, and Minnesota Population Center, University of Minnesota, Minneapolis, Minnesota, USA

The conventional exponential family random graph model (ERGM) parameterization leads to a baseline density that is constant in graph order (i.e., number of nodes); this is potentially problematic when modeling multiple networks of varying order. Prior work has suggested a simple alternative that results in constant expected mean degree. Here, we extend this approach by suggesting another alternative parameterization that allows for flexible modeling of scenarios in which baseline expected degree scales as an arbitrary power of order. This parameterization is easily implemented by the inclusion of an edge count/log order statistic along with the traditional edge count statistic in the model specification.

Keywords: baseline models, exponential family random graph models (ERGMs), mean degree, model parameterization

Exponential family random graph models (ERGMs) are widely used to describe distributions on graphs, particularly social networks (Wasserman & Robins, 2005; Robins & Morris, 2007). While most work to date has focused on graphs of constant order (i.e., number of vertices), there is growing interest in the modeling of graph sets whose members vary in order, either dynamically (Almquist & Butts, 2014) or cross-sectionally (Lubbers & Snijders, 2007; Goodreau, Kitts, & Morris, 2009). While specifications in such multiple network cases vary, they have in common the use of an exponential family (ERGM) form for the conditional probability of a given graph in the set ($Y_i \in (Y_1, \ldots, Y_k)$); i.e.,

$$\Pr(Y_i = y_i | \theta, X_i, N_i) = \frac{\exp\left(\theta^T t(y_i, X_i, N_i)\right)}{\sum_{y_i' \in \mathcal{Y}_i(N_i)} \exp\left(\theta^T t\left(y_i', X_i, N_i\right)\right)},$$

Address correspondence to Carter T. Butts, University of California, Irvine, Department of Sociology, 3151 Social Science Plaza, Irvine, California, 92697-5100, USA. E-mail: buttsc@uci.edu

where Y_i is a random graph on support $\mathcal{Y}_i$ (often, the set of all graphs or digraphs of order N_i), X_i is a covariate set, N_i is the order of Y_i, $\theta \in \mathbb{R}^p$ is a parameter vector, and $t : \mathcal{Y}_i, X_i, N_i \to \mathbb{R}^p$ is a vector of sufficient statistics. Model parameterization focuses on the elements of the linear predictor, $\theta^T t(Y, X, N)$; t is generally chosen on substantive grounds (see, e.g., Robins & Pattison, 2005) and θ is inferred from data, although constraints may be placed on θ (resulting in curved families, e.g., Hunter & Handcock, 2006).

In typical ERGM applications, it is common to include a term whose statistic counts the number of edges (here denoted by M); we may thus decompose the linear predictor as

$$\theta^T t(y_i, X_i, N_i) = \phi M(y_i) + \psi^T s(y_i, X_i, N_i),$$

where ϕ is the parameter associated with the edge count, and ψ and s constitute the remainder of θ and t (respectively). Our focus is here on this first term, which sets the base tie probability within the network. Specifically, holding out all other effects (i.e., setting $\psi = 0$), the expected density of Y_i in the above parameterization is logit^{-1} ϕ, and hence constant in N_i.

Constant density is implausible in many settings, particularly as N_i becomes large (Mayhew & Levinger, 1976). An alternative parameterization was proposed by Butts (2011), which instead results in a constant expected degree:

$$\phi_\delta = \log\left(\frac{\delta}{N_i - 1 - \delta}\right), \tag{1}$$

with δ being the expected degree. (Note that this is obtained by simply rewriting the expected density in terms of the expected degree and applying the logit transform.) This parameterization is an exact version of an approximate constant mean degree parameterization proposed earlier by Krivitsky, Handcock, and Morris (2011), which arises as the sparse-graph limit of ϕ_δ. Specifically,

$$\begin{aligned}\lim_{N_i/\delta \to \infty} \phi_\delta &= \log \delta - \log N_i \\ &\equiv \phi_K,\end{aligned}$$

with ϕ_K being the Krivitsky parameterization. ϕ_K closely approximates ϕ_δ in most practical settings (see below) and is of considerable utility due to ease of implementation: simply adding a $-\log N_i$ offset to the edge term in the fitted ERGM yields the correct scaling, with the remaining estimated coefficient corresponding approximately to $\log\delta$ (i.e., the log expected degree).

The above parameterizations provide alternatives for the constant density and constant mean degree cases, respectively, but leave open the problem of specifying baselines that vary as a more general function of N_i. As we show, this can easily be done for the particular case in which expected degree scales as a power law in N_i. We begin by taking the expected degree for a given N to be equal to a degree constant multiplied by a fixed power of N; i.e., $\delta_N = \delta N_i^\gamma$. This functional form has been suggested on empirical grounds, for example, by Leskovec, Kleinberg, and

Faloutsos (2007), who refer to it as a "densification power law"; it may also be regarded as a convenient approximation to the broader class of nonlinearly varying baseline degree functions. Substituting δ_N for the expected degree in Eq. (1) yields

$$\begin{aligned}\phi_{\delta_N} &= \log\left(\frac{\delta_N}{N_i - 1 - \delta_N}\right)\\ &= \log\left(\frac{\delta N_i^{\gamma}}{N_i - 1 - \delta N_i^{\gamma}}\right)\\ &= \log\delta + \gamma\log N_i - \log\left(N_i - 1 - \delta N_i^{\gamma}\right)\end{aligned} \tag{2}$$

$$\xrightarrow{N_i^{1-\gamma}/\delta\to\infty} \log\delta + (\gamma - 1)\log N_i. \tag{3}$$

This limiting form closely resembles the Krivitsky parameterization, with the important exception that $-\log N_i$ is no longer an offset for the edge term. Instead, our ERGM form has two terms: one corresponding to ($\log\delta$) $M(Y_i)$, and another corresponding to $(\gamma - 1)$ ($\log N_i$) $M(Y_i)$ (i.e., an interaction between the log number of vertices and the number of edges). Both terms are identifiable so long as multiple graphs are observed with varying N (and the graphs are neither null nor complete).

Note that the above parameterization recovers both constant density (i.e., δN_i^1) and constant mean degree (i.e., δN_i^0) as special cases. In the former case, ϕ_{δ_N} approaches $\log\delta$ as γ approaches 1, with $\log\delta$ in turn approaching ϕ as δ/N_i approaches 0 (since $\log x \to \text{logit}\, x$ as $x \to 0$). In the latter case, it is immediate from the definitions of ϕ_{δ_N} and ϕ_K that the $\phi_{\delta_N} \to \phi_K$ as $\gamma \to 0$. More generally, $0 \leq \gamma < 1$ describes the regime in which mean degree is constant or sublinearly increasing in N, the most plausible range for most social networks. $\gamma > 1$ implies a net density increase (supralinearly increasing mean degree), which is unlikely to be sustainable over a large size range; by turns, $\gamma < 0$ implies mean degree that declines as N grows (likewise dubious in most settings). With the proviso that we remain in the sparse regime (i.e., N much larger than mean degree), the parameterization of Eq. (3) covers a wide range of cases.

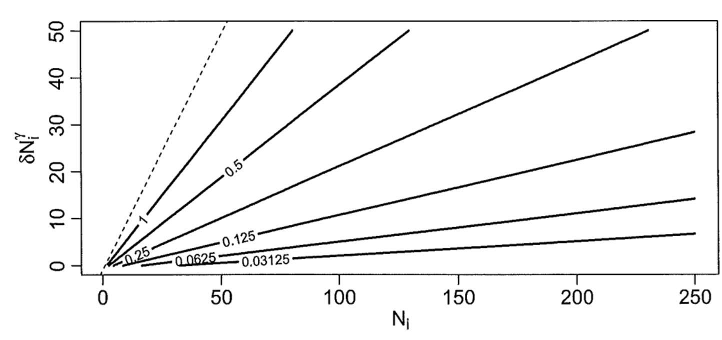

FIGURE 1 Approximation error for the density parameter in the log N parameterization, as a function of order (N_i) and expected baseline degree δN_i^{γ} (simple graph case). Dotted line indicates maximum possible degree; contour lines show absolute error versus exact parameterization.

Since the above results are given for the sparse limit, it is useful to have some sense of the approximation error involved for less sparse cases. Comparing the approximate parameterization of Eq. (3) with its exact counterpart Eq. (2), we see that the absolute error in the density parameter reduces to $\log N_i - \log(N_i - 1 - \delta N_i^{\gamma})$. Figure 1 shows the absolute error isolines for various choices of order and expected density; as the figure illustrates, the isolines have a simple form, with the absolute error being $\leq \varepsilon$ so long as the maximum expected baseline degree is $\leq N_i(1 - e^{-\varepsilon}) - 1$. The approximation error is quite small for even moderately large graphs, so long as the expected baseline degree is less than 10 or so. For networks with hundreds or thousands of nodes, the approximation error will be negligible relative to the size of the true parameter, even if the baseline mean degree is quite large in absolute terms.

To use the flexible baseline parameterization in practice, one employs the following simple procedure:

1. Add two statistics to the linear predictor, one for $M(Y_i)$ and one for ($\log N_i$) $M(Y_i)$.
2. Fit the model using standard ERGM/TERGM methods, yielding respective parameter estimates $\hat{\phi}^{(1)}_{\delta_N}$ and $\hat{\phi}^{(2)}_{\delta_N}$ for the edge count and size interaction statistics.
3. Find $\hat{\delta} = \exp\left[\hat{\phi}^{(1)}_{\delta_N}\right]$ and $\hat{\gamma} = \hat{\phi}^{(2)}_{\delta_N} + 1$.

The estimated (baseline) expected degree scaling is then $\hat{\delta} N^{\hat{\gamma}}$.

Note that step 1 is even easier than it appears: $M(Y_i)$ is just the standard edge count statistic, and adding ($\log N_i$) $M(Y_i)$ simply requires adding an edge covariate for each (j, k) edge variable in each Y_i equal to $\log N_i$. This can be performed using standard software tools (e.g., Hunter, Handcock, Butts, Goodreau, & Morris, 2008; Wang, Robins, & Pattison, 2009) without recourse to custom procedures.

The flexible parameterization usefully extends the family of baselines that can be used with multiple-network ERGMs. Obviously, the standard (fixed ϕ) parameterization is appropriate when one is willing to assume constant baseline density. When one is willing to assume constant baseline mean degree and is in the large/sparse regime, the Krivitsky et al. (2011) parameterization is a simple and appropriate choice; the curved parameterization of Eq. (1) provides an exact alternative for small and/or dense graphs. When it is unknown whether baseline mean degree scales nonlinearly with N, the flexible parameterization described here allows a straightforward method of modeling this dependence (while still capturing constant density and constant mean degree as special cases). It is hoped that the availability of this and related techniques will serve to encourage further work on the modeling of multiple networks within the ERGM framework.

FUNDING

This work is based on research supported by National Science Foundation award #IIS-1251267 and Army Research Office award #W911NF-14-1-0552.

REFERENCES

Almquist, Z. W., & Butts, C. T. (2014). Logistic network regression for scalable analysis of networks with joint edge/vertex dynamics. *Sociological Methodology*, *44*, 273–321.

Butts, C. T. (2011). Bernoulli graph bounds for general random graphs. *Sociological Methodology*, *41*, 299–345.

Goodreau, S. M., Kitts, J. A., & Morris, M. (2009). Birds of a feather, or friend of a friend?: Using exponential random graph models to investigate adolescent social networks. *Demography*, *46*, 103–125.

Hunter, D. R., & Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, *15*, 565–583.

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, *24*(3), 1–29.

Krivitsky, P. N., Handcock, M. S., & Morris, M. (2011). Adjusting for network size and composition effects in exponential-family random graph models. *Statistical Methodology*, *8*, 319–339.

Leskovec, J., Kleinberg, J., & Faloutsos, C. (2007). Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data*, *1*(1). doi:10.1145/1217299.1217301

Lubbers, M. J., & Snijders, T. A. B. (2007). A comparison of various approaches to the exponential random graph model: a reanalysis of 102 student networks in school classes. *Social Networks*, *29*, 489–507.

Mayhew, B. H., & Levinger, R. L. (1976). Size and density of interaction in human aggregates. *American Journal of Sociology*, *82*, 86–110.

Robins, G. L. & Morris, M. (2007). Advances in exponential random graph (p^*) models. *Social Networks*, *29*, 169–172.

Robins, G. L. & Pattison, P. E. (2005). Interdependencies and social processes: Dependence graphs and generalized dependence structures. In P. J. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and methods in social network analysis* (pp. 192–214). Cambridge, UK: Cambridge University Press.

Wang, P., Robins, G., & Pattison, P. (2009). PNet: Program for the simulation and estimation of exponential random graph (p^*) models [Electronic data file]. Retrieved from http://sna.unimelb.edu.au/PNet

Wasserman, S., & Robins, G. L. (2005). An introduction to random graphs, dependence graphs, and p^*. In P. J. Carrington, J. Scott & S. Wasserman (Eds.), *Models and methods in social network analysis* (pp. 148–161). Cambridge, UK: Cambridge University Press.