Statistical Inference for Valued-Edge Networks: Generalized Exponential Random Graph Models.

B. A. Desmarais[1] and S. J. Cranmer[2]
[1]*Department of Political Science, University of Massachusetts at Amherst, Amherst, MA 01003*
[2]*Department of Political Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599*

(Dated: March 25, 2011)

Across the sciences, the statistical analysis of networks is central to the production of knowledge on relational phenomena. Because of their ability to model the structural generation of networks, exponential random graph models are a ubiquitous means of analysis. However, they are limited by an inability to model networks with valued edges. We solve this problem by introducing a class of generalized exponential random graph models capable of modeling networks whose edges are valued, thus greatly expanding the scope of networks applied researchers can subject to statistical analysis.

PACS numbers: 02.10.Ox, 89.65.Cd, 89.75.Hc

The need to analyze networks statistically transcends disciplines that have occasion to study the relationships between units. Applications in physics [1–5], computer science [6], the social sciences[7, 8], and other fields examine networks that vary in size and density, over time, and have edges with values that vary from binary ties, to counts, to bounded continuous and unbounded continuous edges. An important method for statistical inference on networks is the exponential random graph model (ERGM)[9–11], which estimates the probability of an observed network conditional on a vector of network statistics that capture the generative structures in the network. Yet the ERGM has a major limitation: it is only defined for networks with binary ties[12, 13], thus excluding a wide range of networks with valued edges (e.g., gene coexpression networks, passage time on networks of various media, monetary transactions, casualties in conflict networks).

We develop a class of generalized ERGMs (GERGMs) for inference on networks with continuous edge values, thus lifting the restriction of this methodology to a, possibly small, subset of networks. The form of our generalized model is similar to the ERGM in that it can be flexibly specified to cover a broad range of generative features. The GERGM can be estimated efficiently with a Gibbs sampler.

The strengths and limitations of the ERGM are apparent from its specification. Let Y be the n-vertex network (adjacency matrix) of interest with m edges ($m = n(n-1)$ if Y is directed and $n(n-1)/2$ if it is undirected). Y_{ij} is the edge from i to j. An ERGM of that network is specified as:

$$\mathcal{P}(Y,\boldsymbol{\theta}) = \frac{\exp\{\boldsymbol{\theta}'\,\mathbf{h}(Y)\}}{\sum_{\text{all } Y^*\in\mathcal{Y}}\exp\{\boldsymbol{\theta}'\,\mathbf{h}(Y^*)\}}, \tag{1}$$

where $\boldsymbol{\theta}$ is a parameter vector, $\mathbf{h}(Y)$ is a vector of statistics on the network, and the object of inference is the probability of the observed network among all possible permutations of the network given the network statistics. The $\mathbf{h}(Y)$ term is what gives the ERGM much of its power: this vector can contain statistics to capture the endogenous structure of connectivity in the network (statistics can be included to capture reciprocity, transitivity, cyclicality, and a wide variety of other endogenous structures) as well as the effects of exogenous covariates.

The challenges for modeling networks with valued edges are apparent from the specification in equation 1. The flexibility of the distribution comes from the lack of constraints in specifying $\mathbf{h}$; the only constraint is that $\mathbf{h}$ is finite when evaluated on any binary network. This assures that the denominator is a *convergent* sum, and therefore represents a proper normalizing constant for the distribution of networks. However, this convergence is not assured whenever $\mathbf{h}$ is finite if the support of Y is infinite. The model we derive retains the flexibility of $\mathbf{h}$ within a framework that assures a proper probability distribution for Y when Y has continuous edges.

Our generalized ERGM operates by constructing joint *continuous* distributions on networks that permit the representation of dependence features among the elements of Y through a set of statistics on the network, $\mathbf{h}(Y)$. As in the ERGM, the vector $\mathbf{h}$ can be specified to represent many forms of dependence, including transitivity (i.e., clustering), cycling, and reciprocity; an important attribute of the model because such dependence features characterize valued networks [13].

There are two specification steps in our approach to GERGMs: first, we specify a tractable joint distribution that captures the dependencies of interest on a restricted network, X, and then we transform X onto the support of Y; thus producing a probability model for Y. To illustrate these steps, begin with consideration of the restricted valued network $X \in [0,1]^m$, where m is the number of edges.

In our first specification step, $\mathbf{h}$ is formulated to represent joint features of Y in the distribution of X:

$$f_X(X,\boldsymbol{\theta}) = \frac{\exp\left[\boldsymbol{\theta}'\mathbf{h}(X)\right]}{\int_{[0,1]^m}\exp\left[\boldsymbol{\theta}'\mathbf{h}(Z)\right]dZ}, \tag{2}$$

where $\boldsymbol{\theta} \in \mathbb{R}^p$ is the parameter vector, $\mathbf{h} : [0,1]^m \to \mathbb{R}^p$, $\mathbf{h}$ is finite on $[0,1]^m$ and $h_i(\cdot)$ are the sums of subgraph products such that for every i, $\frac{\partial^2\mathbf{h}(X)}{\partial^2 X_{ij}} = 0$. This is a flexible specification because many dependence relationships can be captured by summing products over subgraphs of the network, particularly when the edges are in the unit interval[13]. For instance, networks generated by a highly reciprocal process are likely to exhibit high values of $\sum_{i<j} X_{ij}X_{ji}$, and those in which connections gravitate toward high-degree vertices exhibit high values of $\sum_i \sum_{j,k\neq i} X_{ji}X_{ki}$ (i.e., "two-stars" [14]). An important property of f_X is that when $\boldsymbol{\theta} = \mathbf{0}$, X is a network of independent uniform random variables.

In our second specification step, we apply parameterized, one-to-one, monotone increasing transformations $(G^{-1}(\cdot))$ to the m edges of the restricted network, thus transforming the restricted network X onto the support of the network of interest Y. $Y_{ij} = G_{ij}^{-1}(X_{ij}, \boldsymbol{\lambda}_{ij})$, where $\boldsymbol{\lambda}_{ij}$ parameterizes the transformation to capture marginal features of Y_{ij}. Because $dG^{-1}(X_{ij}, \boldsymbol{\lambda}_i)/dX_{ij} > 0$, the properties of multivariate transformations[15] imply that the distribution of Y is $f_Y(Y, \boldsymbol{\theta}, \boldsymbol{\Lambda}) = f_X(\boldsymbol{G}(Y, \boldsymbol{\Lambda}), \boldsymbol{\theta})|J|$, where the Jacobian matrix, J, is the matrix of first partial derivatives. Since J is a diagonal matrix, we may write the GERGM as

$$f_Y(Y, \boldsymbol{\theta}, \boldsymbol{\Lambda}) = \frac{\exp\left[\boldsymbol{\theta}'\mathbf{h}(\boldsymbol{G}(Y, \boldsymbol{\Lambda}))\right]}{\int_{[0,1]^m} \exp\left[\boldsymbol{\theta}'\mathbf{h}(Z)\right] dZ} \prod_{ij} g(Y_{ij}, \boldsymbol{\lambda}_{ij}). \quad (3)$$

A useful way to specify g is as a probability density function (i.e., G is a CDF, and G^{-1} an inverse CDF) parameterized to match the support of Y and capture features of Y such as location, scale, and dependence on covariates. This approach to specifying g has the elegant feature that the distribution contains many common models for independent and identically distributed variables as special cases when $\boldsymbol{\theta} = \mathbf{0}$. For instance, if g is a Gaussian PDF with constant variance and the mean dependent on a vector of covariates, the model reduces to that assumed in least squares regression. The GERGM also allows hypothesis tests for block restrictions (i.e., likelihood ratio or Wald tests) to test the assumption that the edges of Y are independent conditional upon $\boldsymbol{\Lambda}$.

There are two ways to interpret dependence modeling of Y via X. First, following [13], who derive an ERGM-like model for a network with discrete edges on the unit interval, X can be interpreted as a standardized relational intensity network. Second, and more directly, when g is a PDF, X is the random variable drawn from the joint distribution of the quantiles of Y. Therefore, the vectors $\mathbf{h}$ and $\boldsymbol{\theta}$ characterize the dependencies among the quantiles of Y. The latter interpretation closely resembles the process of constructing joint distributions with copula functions [16, 17]. A simple example of deriving a joint distribution through the combination of $\mathbf{h}$ and g is illustrated in figure 1, which presents the distributions of X and Y for a directed network with two vertices exhibiting a high degree of reciprocity.

Estimation of the parameters in the model is a nontrivial task. The greatest challenge in estimating $\boldsymbol{\theta}$ and $\boldsymbol{\Lambda}$ in equation 3 is that the integral in the denominator is typically intractable. Because of the polynomial structure of $\mathbf{h}$, and the fact that the variables of integration are bounded, we know that the integral is both positive and finite, meaning f_Y is a proper joint distribution. However, inference requires the approximation of the denominator.

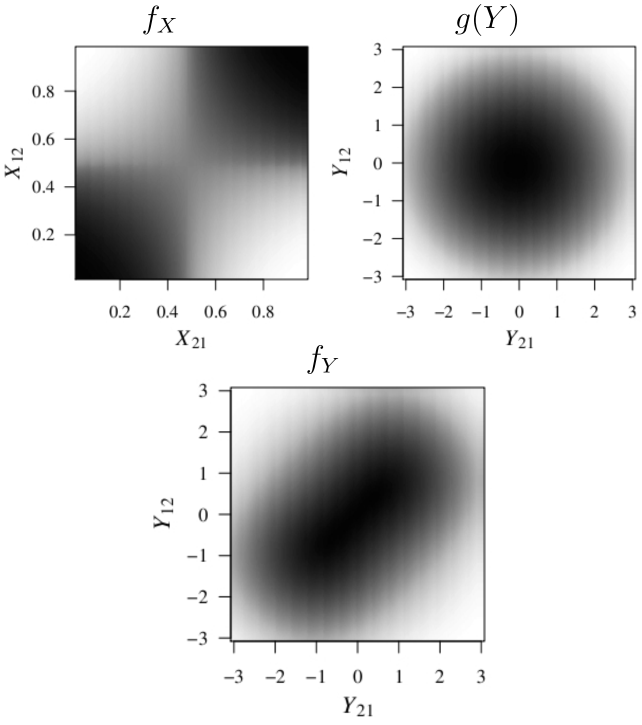


FIG. 1. Bivariate distributions for edges in a two-vertex digraph. The darker the shading, the higher the relative likelihood of a point. In this example, g is the standard normal PDF, and f_X is defined by $\mathbf{h} = \{X_{12} + X_{21}, X_{12}X_{21}\}$, and $\boldsymbol{\theta} = \{-3.5, 7\}$, representing negative density and positive reciprocity effects.

In order to approximate the denominator in equation 3, we sample from f_X using a Gibbs Sampler. To do so, we require the conditional distribution of $X_{ij}|X_{-ij}$. To simplify the notation, let $\int_{[0,1]^m} \exp\left[\boldsymbol{\theta}'\mathbf{h}(Z)\right] dZ = C(\boldsymbol{\theta})$. The conditional distribution (f_X^c) is given by

$$f_X^c(X_{ij}|\boldsymbol{\theta}) = \frac{\exp\left[X_{ij}\boldsymbol{\theta}'\frac{\partial \mathbf{h}(X)}{\partial X_{ij}}\right]}{\boldsymbol{\theta}'\left(\frac{\partial \mathbf{h}(X)}{\partial X_{ij}}\right)^{-1}\left[\exp(\boldsymbol{\theta}'\frac{\partial \mathbf{h}(X)}{\partial X_{ij}}) - 1\right]}. \quad (4)$$

We may then draw from the conditional distribution in equation 4 using the inverse CDF method. If u is a uniform (0,1) random variable, then

$$X_{ij}|X_{-ij} \sim \frac{\ln\left[1 + u\left(\exp\left[\boldsymbol{\theta}'\frac{\partial \mathbf{h}(X)}{\partial X_{ij}}\right] - 1\right)\right]}{\boldsymbol{\theta}'\frac{\partial \mathbf{h}(X)}{\partial X_{ij}}}. \quad (5)$$

When $\boldsymbol{\theta}'\frac{\partial \mathbf{h}(X)}{\partial X_{ij}} = 0$ the conditional density given in equation 4 is undefined. However, in this case, each point in the unit interval is equally likely and the conditional distribution of X_{ij} is uniform(0,1).

In order to estimate $\boldsymbol{\theta}$ and $\boldsymbol{\Lambda}$, we maximize $\ln[f_Y]$:

$$\boldsymbol{\theta}'\mathbf{h}(\boldsymbol{G}(Y,\boldsymbol{\Lambda})) + \sum_{ij} \ln\left[g(Y_{ij}|\boldsymbol{\lambda}_{ij})\right] - \ln\left[C(\boldsymbol{\theta})\right]. \qquad (6)$$

Our algorithm iteratively proceeds by maximum likelihood (ML) estimation of $\boldsymbol{\Lambda}|\boldsymbol{\theta}$ and Markov chain Monte Carlo maximum likelihood estimation (MCMC-MLE) of $\boldsymbol{\theta}|\boldsymbol{\Lambda}$ until convergence. We derive an approximation to the asymptotic variance-covariance matrix by the inverse of the negative Hessian matrix at the last iteration.

The estimation of $\boldsymbol{\Lambda}|\boldsymbol{\theta}$ is straightforward. Because $C(\boldsymbol{\theta})$ does not depend on $\boldsymbol{\Lambda}$, ML estimation of $\boldsymbol{\Lambda}|\boldsymbol{\theta}$ reduces to

$$\arg\max_{\boldsymbol{\Lambda}} \left(\boldsymbol{\theta}'\mathbf{h}(\boldsymbol{G}(Y,\boldsymbol{\Lambda})) + \sum_{ij} \ln\left[g(Y_{ij}|\boldsymbol{\lambda}_{ij})\right] \right), \qquad (7)$$

a function easy to maximize using a hill-climbing algorithm.

The estimation of $\boldsymbol{\theta}|\boldsymbol{\Lambda}$ is more involved. Let $\widehat{X} = \boldsymbol{G}(Y,\hat{\boldsymbol{\Lambda}})$ be the estimate of the intensity/quantile network given the current estimate of the transformation parameters. The second term in equation 6 does not depend on $\boldsymbol{\theta}$, so to estimate $\boldsymbol{\theta}|\boldsymbol{\Lambda}$ we find

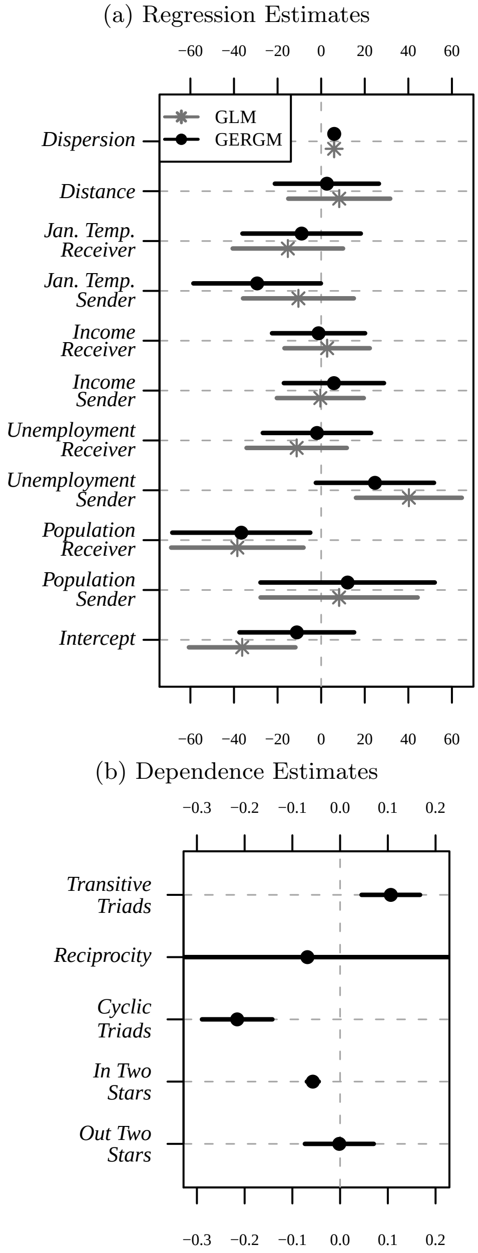
$$\arg\max_{\boldsymbol{\theta}} \left(\boldsymbol{\theta}'\mathbf{h}(\widehat{X}) - \ln\left[C(\boldsymbol{\theta})\right] \right), \qquad (8)$$

which requires an approximation of $C(\boldsymbol{\theta})$. We approximate $C(\boldsymbol{\theta})$ using MCMC-MLE; an iterative method itself. Let $\boldsymbol{\theta}^{[i-1]}$ be the previous estimate of $\boldsymbol{\theta}$, and $\tilde{\boldsymbol{X}}$ be a sample of n networks drawn from $f_X(X,\boldsymbol{\theta}^{[i-1]})$. Then, an approximation to $C(\boldsymbol{\theta})$ is given by

$$\widehat{C(\boldsymbol{\theta})} = C(\boldsymbol{\theta}^{[i-1]}) \sum_{j=1}^{n} \frac{\exp\left[\boldsymbol{\theta}'\mathbf{h}(\tilde{X}_j)\right]}{\exp\left[\boldsymbol{\theta}'^{[i-1]}\mathbf{h}(\tilde{X}_j)\right]}. \qquad (9)$$

This requires a starting value for $\boldsymbol{\theta}$. In simulation experiments, we have found the pseudolikelihood estimate ($\arg\max_{\boldsymbol{\theta}} \left(\sum_{ij} \ln\left[f_X^c(X_{ij}|\boldsymbol{\theta})\right]\right)$) to be effective in providing starting values for $\boldsymbol{\theta}$ (i.e., $\boldsymbol{\theta}^{[0]}$).

We illustrate important features of the GERGM and demonstrate its efficacy by applying it to a real-world network: domestic migration in the United States[18, 19]. We model changes in the directional migration flows between the 50 United States (as well as Washington D.C. and Puerto Rico) between 2006 and 2007. Y_{ij} is the difference between the number of people who migrated from state i to state j in 2007 and the number who migrated from i to j in 2006. These data allow us to consider the GERGM in the context of a valued network requiring transformation away from an intensity network onto a continuous unbounded support with exogenous covariates and endogenous parameters, thus making full use of the GERGM's flexibility. We use the Cauchy distribution as our g function because its thick tails capture the high empirical kurtosis (637) of the network [20]. Thus, in the case where the edges of the network are independent conditional on the covariates, this specification reduces to a generalized linear model (GLM) [21] with a Cauchy link function. Because previous work on interstate migration[22] suggests that population, unemployment, per-capita income, and mean January temperature of both the sending and receiving states are significant determinants of migration, we include the change in each of these variables from 2005 to 2006 as covariates in our GERGM. We complete our specification by


(a) Regression Estimates

(b) Dependence Estimates

FIG. 2. Estimates of the parameters: bars span 95% confidence intervals. 5,000 draws for three iterations used in the MCMC-MLE

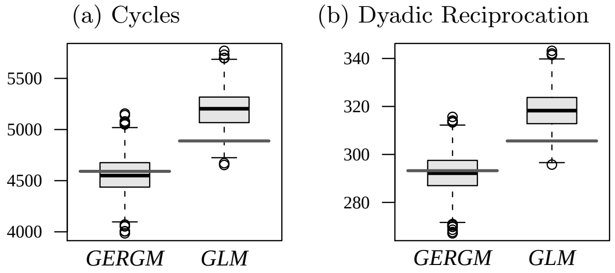


FIG. 3. Reciprocal Feature Prediction: The boxplots represent the respective dependence statistic computed on 1,000 instances of the latent intensity network drawn from each model. Let $\widehat{X}$ be the respective estimate of the intensity network obtained as the CDF evaluated at the transormation parameters ($\mathbf{\Lambda}$) for the GERGM and Cauchy GLM. Then cycles (a) is $\sum_{i<j<k} \widehat{X}_{ij}\widehat{X}_{jk}\widehat{X}_{ki} + \widehat{X}_{ik}\widehat{X}_{kj}\widehat{X}_{ji}$, and dyad reciprocation (b) is $\sum_{i<j} \widehat{X}_{ij}\widehat{X}_{ji}$. Horizontal grey bars are placed at the statistic computed on the estimated intensity network.

including endogenous dependence terms for clustering, dyadic reciprocity, generalized reciprocity (i.e., cycling – the degree to which change in flows to and from a state are correlated[23]), state level attraction, and state level repellence.

Figure 2 shows the estimates from our GERGM as well as estimates from a Cauchy GLM. A Wald test suggests the restriction of the dependence terms to zero in the regression model is inappropriate and that the GERGM provides a better fit to the data (Wald statistic = 119.19 on 5 degrees of freedom, statistically significant at the 0.001 level). The statistically significant effects for the network parameters indicate that (a) there are clustering effects in the network, (b) migration to states repels further migration, and (c) increases in migration flows from a state are not offset by increases in flows to that state. We also find a decrease in the number of people leaving warm states, a decrease in migration to states that experienced a substantial increase in population in the previous year, and evidence of an increase in migration away from states experiencing increases in unemployment.

The superior performance of the GERGM relative to the Cauchy regression is further depicted in figure 3, which gives the predicted and observed network-level reciprocity and cycling measures from the GERGM and Cauchy GLM. This figure shows that the regression does not adequately fit the lack of reciprocity in the migration network. Theoretically, it is expected that a network of change in migration would exhibit anti-reciprocity and anti-cycling. If a locale is experienceing a spike in migration to other places, that is likely indicative of some undesireable feature of said locale. This anti-reciprocal feature of the migration network cannot be integrated into the conventional regression modeling framework.

Our GERGM model greatly expands the scope of networks which can be modeled within the ERGM framework. We used this technology to analyze a real-world network and produce insights that could not be produced without the GERGM. Our general model represents a major advance in the statistical analysis of networks, and we expect it to become a common tool in disciplines spanning the sciences.

The authors thank James Fowler and Peter Mucha for useful comments. This work was supported in part by a grant from the University of Massachusetts Amherst College of Social and Behavioral Sciences.

[1] B. Karrer and M. E. J. Newman, Phys. Rev. E **83** (2010).
[2] B. Karrer and M. Newman, Phys. Rev. E **80**, 1 (2009).
[3] M. Newman, Phys. Rev. Lett. **103**, 1 (2009).
[4] D. Garlaschelli and M. I. Loffredo, Phys. Rev. Lett. **93**, 188701 (2004).
[5] G. Bianconi and A.-L. Barabási, Phys. Rev. Lett. **86**, 5632 (2001).
[6] S. Myers and J. Leskovec, in *Advances in Neural Information Processing Systems 23* (2010) pp. 1741–1749.
[7] C. T. Butts, Sociological Methodology **38**, 155 (2008).
[8] S. J. Cranmer and B. A. Desmarais, Political Analysis **19**, 66 (2011).
[9] P. W. Holland and S. Leinhardt, J. Am. Stat. Assoc. **76**, 33 (1981).
[10] J. Berg and M. Lässig, Phys. Rev. Lett. **89**, 228701 (2002).
[11] J. Park and M. E. J. Newman, Phys. Rev. E **70**, 066117 (2004).
[12] G. Robins, T. Snijders, and S. Wasserman, Psychometrica **64**, 371 (1999).
[13] D. Wyatt, T. Choudhury, and J. Bilmes, in *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence* (2010) pp. 630–636.
[14] J. Park and M. E. J. Newman, Phys. Rev. E **70**, 066146 (2004).
[15] G. Casella and R. L. Berger, *Statistical Inference* (Duxbury, Pacific Grove, CA, USA, 2001).
[16] J. P. Gleeson, Phys. Rev. E **77**, 046117 (2008).
[17] M. Sato, K. Ichiki, and T. T. Takeuchi, Phys. Rev. Lett. **105**, 251301 (2010).
[18] J. Ke, X. Chen, Z. Lin, Y. Zheng, and W. Lu, Phys. Rev. E **74**, 056102 (2006).
[19] J. Ke, Z. Lin, Y. Zheng, X. Chen, and W. Lu, Phys. Rev. Lett. **97**, 028301 (2006).
[20] I. Mizera and C. H. Mller, Statistics & Probability Letters **57**, 79 (2002).
[21] J. A. Nelder and R. W. M. Wedderburn, Journal of the Royal Statistical Society. Series A (General) **135**, 370 (1972).
[22] Y. Chun, Journal of Geographic Systems **10**, 317 (2008).
[23] L. Jian and J. K. MacKie-Mason, in *Proceedings of the 10th international conference on Electronic commerce*, ICEC '08 (ACM, New York, NY, USA, 2008) pp. 4:1–4:8.