Krivitsky, P. N. (2012). Exponential-family random graph models for valued networks. Electronic Journal of Statistics, 6, 1100-1128.

Exponential-Family Random Graph Models for Valued Networks

Pavel N. Krivitsky

Abstract

Exponential-family random graph models (ERGMs) provide a principled and flexible way to model and simulate features common in social networks, such as propensities for homophily, mutuality, and friend-of-a-friend triad closure, through choice of model terms (sufficient statistics). However, those ERGMs modeling the more complex features have, to date, been limited to binary data: presence or absence of ties. Thus, analysis of valued networks, such as those where counts, measurements, or ranks are observed, has necessitated dichotomizing them, losing information and introducing biases.

In this work, we generalize ERGMs to valued networks. Focusing on modeling counts, we formulate an ERGM for networks whose ties are counts and discuss issues that arise when moving beyond the binary case. We introduce model terms that generalize and model common social network features for such data and apply these methods to a network dataset whose values are counts of interactions.

Keywords: p-star model; transitivity; weighted network; count data; maximum likelihood estimation; Conway–Maxwell–Poisson distribution

1 Introduction

Networks are used to represent and analyze phenomena ranging from sexual partnerships (Morris and Kretzschmar, 1997), to advice giving in an office (Lazega and Pattison, 1999), to friendship relations (Goodreau, Kitts, and Morris, 2008b; Newcomb, 1961), to international relations (Ward and Hoff, 2007), to scientific collaboration, and many other domains (Goldenberg, Zheng, Fienberg, and Airoldi, 2009). More often than not, the relations of interest are not strictly dichotomous in the sense that all present relations are effectively equal to each other. For example, in sexual partnership networks, some ties are short-term while others are long-term or marital; friendships and acquaintance have degrees of strength, as do international relations; and while a particular individual seeking advice might seek it from some coworkers but not others, he or she will likely do it in some specific order and weight advice of some more than others.

Network data with valued relations come in many forms. Observing messages (Freeman and Freeman, 1980; Diesner and Carley, 2005), instances of personal interaction (Bernard, Killworth, and Sailer, 1979–1980), or counting co-occurrences or common features of social actors (Zachary, 1977; Batagelj and Mrvar, 2006) produce relations in the form of counts. Measurements, such as duration of interaction (Wyatt, Choudhury, and Bilmes, 2009) or volume of trade (Westveld and Hoff, 2011) produce relations in the form of (effectively) continuous values. Observations of states of alliance and war (Read, 1954) produce signed relationships. Sociometric surveys often produce ranks in addition to binary measures of affection (Sampson, 1968; Newcomb, 1961; Bernard et al., 1979–1980; Harris, Florey, Tabor, Bearman, Jones, and Udry, 2003).

Exponential-family random graph models (ERGMs) are generative models for networks which postulate an exponential family over the space of networks of interest (Holland and Leinhardt, 1981; Frank and Strauss, 1986), specified by their sufficient statistics (Morris, Handcock, and Hunter, 2008), or, as with Frank and Strauss (1986), by their conditional independence structure leading to sufficient statistics (Besag, 1974). These sufficient statistics typically embody the features of the network of interest that are believed to be significant to the social process which had produced it, such as degree distribution (e.g., propensity towards monogamy in sexual partnership networks), homophily (i.e., "birds of a feather flock together"), and triad-closure bias (i.e., "a friend of a friend is a friend") . (Morris et al., 2008)

A major limitation of ERGMs to date has been that they have been applied almost exclusively to binary relations: a relationship between a given actor i and a given actor j is either present or absent. This is a serious limitation: valued network data have to be dichotomized for ERGM analysis, an approach which loses information and may introduce biases. (Thomas and Blitzstein, 2011)

Some extensions of ERGMs to specific forms of valued ties have been formulated: to networks with polytomous tie values, represented as a constrained three-way binary array by Robins, Pattison, and Wasserman (1999) and more directly by Wyatt et al. (2009; 2010); to multiple binary networks by Pattison and Wasserman (1999); and the authors are also aware of some preliminary work by Handcock (2006) on ERGMs for signed network data. Rinaldo, Fienberg, and Zhou (2009) discussed binary ERGMs as a special case and a motivating application of their developments in geometry of discrete exponential families.

A broad exception to this limitation has been a subfamily of ERGMs that have the property that the ties and their values are stochastically independent given the model parameters. Unlike the dependent case, the likelihoods for these models have can often be expressed as generalized linear or nonlinear models, and they tend to have tractable normalizing constants, which allows them to more easily be embedded in a hierarchical framework. Thus, to represent common properties of social networks, such as actor heterogeneity, triad-closure bias, and clustering, latent class and position models have been used and extended to valued networks. (Hoff, 2005; Krivitsky, Handcock, Raftery, and Hoff, 2009; Mariadassou, Robin, and Vacher, 2010)

In this work, we generalize the ERGM framework to directly model valued networks, particularly networks with count dyad values, while retaining much of the flexibility and interpretability of binary ERGMs, including the above-described property in the case when tie values are independent under the model. In Section 2 we review conventional ERGMs and describe their traits that valued ERGMs should inherit. In Section 3, we describe the framework that extends the model class to networks with counts as dyad values and discuss additional considerations that emerge when each dyad's sample space is no longer binary. In Section 4 we give some details and caveats of our implementation of these models and briefly address the issue of ERGM degeneracy as it pertains to count data. Applying ERGMs requiers one to specify and interpret sufficient statistics that embody network features of interest, all the while avoiding undesirable phenomena such as ERGM degeneracy. Thus, in Section 5, we introduce and discuss statistics to represent a variety of features commonly found in social networks, as well as features specific to networks of counts. In Section 6 we use these statistics to model social forces that affect the structure of a network of counts of social contexts of interactions among members of a divided karate club and a network of counts of conversations among members of a fraternity. Finally, in Section 7, we discuss generalizing ERGMs to other types of valued data.

2 ERGMs for binary data

In this section, we define notation, review the (potentially curved) exponential-family random graph model and identify those of its properties that we wish to retain when generalizing.

2.1 Notation and binary ERGM definition

Let N be the set of actors in the network of interest, assumed known and fixed for the purposes of this paper, and let $n \equiv |N|$ be its cardinality, or the number of actors in the network. For the purposes of this paper, let a *dyad* be defined as a (usually distinct) pair of actors, ordered if the network of interest is directed, unordered if not, between whom a relation of interest may exist, and let $\mathbb{Y}$ be the set of all dyads. More concretely, if the network of interest is directed, $\mathbb{Y} \subseteq N \times N$, and if it is not, $\mathbb{Y} \subseteq \{\{i, j\} : (i, j) \in N \times N\}$. In many problems, a relation of interest cannot exist between an actor and itself (e.g., a friendship network), or actors are partitioned into classes with relations only existing between classes (e.g., bipartite networks of actors attending events), in which case $\mathbb{Y}$ is a proper subset of $N \times N$, excluding those pairs (i, j) between which there can be no relation of interest.

Further, let the set of possible networks of interest (the sample space of the model) $\mathcal{Y} \subseteq 2^{\mathbb{Y}}$, the power set of the dyads in the network. Then a network $\boldsymbol{y} \in \mathcal{Y}$, can be considered a set of ties (i, j). Again, in some problems, there may be additional constraints on $\mathcal{Y}$. A common example of such constraints are degree constraints induced by the survey format (Harris et al., 2003; Goodreau et al., 2008b).

Using notation similar to that of Hunter and Handcock (2006) and Krivitsky, Handcock, and Morris (2011), define an exponential family random graph model to have the form

$$\mathrm{Pr}_{\boldsymbol{\theta};\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{Y} = \boldsymbol{y}|\boldsymbol{x}) = \frac{\exp\left(\boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \boldsymbol{g}(\boldsymbol{y}; \boldsymbol{x})\right)}{\kappa_{\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{\theta}; \boldsymbol{x})},\ \boldsymbol{y} \in \mathcal{Y}, \tag{1}$$

for random network variable $\boldsymbol{Y}$ and its realization $\boldsymbol{y}$; model parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ (for parameter space $\boldsymbol{\Theta} \subseteq \mathbb{R}^q$) and its mapping to canonical parameters $\boldsymbol{\eta} : \boldsymbol{\Theta} \to \mathbb{R}^p$; a vector of sufficient statistics $\boldsymbol{g} : \mathcal{Y} \to \mathbb{R}^p$, which may also depend on data $\boldsymbol{x}$, assumed fixed and known; and a normalizing constant (in

$\boldsymbol{y}$) $\kappa_{\boldsymbol{\eta},\boldsymbol{g}} : \mathbb{R}^q \to \mathbb{R}$ which ensures that (1) sum to 1 and thus has the value

$$\kappa_{\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{\theta};\boldsymbol{x}) = \sum_{\boldsymbol{y}'\in\mathcal{Y}} \exp\left(\boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \boldsymbol{g}(\boldsymbol{y}';\boldsymbol{x})\right).$$

Here, we have given the most general case defined by Hunter and Handcock (2006): a frequently used special case is $q = p$ and $\boldsymbol{\eta}(\boldsymbol{\theta}) = \boldsymbol{\theta}$, so the exponential family is linear. For notational simplicity, we will omit $\boldsymbol{x}$ for the remainder of this paper, as $\boldsymbol{g}$ incorporates it implicitly.

2.2 Properties of binary ERGM

2.2.1 Conditional distributions and change statistics

Snijders, Pattison, Robins, and Handcock (2006), Hunter, Handcock, Butts, Goodreau, and Morris (2008b), Krivitsky et al. (2011), and others define *change statistics*, which emerge when considering the probability of a single dyad having a tie given the rest of the network and provide a convenient local interpretation of ERGMs. To summarize, define the p-vector of change statistics

$$\boldsymbol{\Delta}_{i,j}\boldsymbol{g}(\boldsymbol{y}) \equiv \boldsymbol{g}(\boldsymbol{y} + (i,j)) - \boldsymbol{g}(\boldsymbol{y} - (i,j)),$$

where $\boldsymbol{y} + (i,j)$ is the network $\boldsymbol{y}$ with edge or arc (i,j) added if absent (and unchanged if present) and $\boldsymbol{y} - (i,j)$ is the network $\boldsymbol{y}$ with edge or arc (i,j) removed if present (and unchanged if absent). Then, through cancellations,

$$\Pr_{\boldsymbol{\theta};\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{Y}_{i,j} = 1 | \boldsymbol{Y} - (i,j) = \boldsymbol{y} - (i,j)) = \text{logit}^{-1}\left(\boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \boldsymbol{\Delta}_{i,j}\boldsymbol{g}(\boldsymbol{y})\right).$$

It is often the case that the form of $\boldsymbol{\Delta}_{i,j}\boldsymbol{g}(\boldsymbol{y})$ is simpler than that of $\boldsymbol{g}(\boldsymbol{y})$ both algebraically and computationally. For example, the change statistic for edge count $|\boldsymbol{y}|$ is simply 1, indicating that a unit increase in $\boldsymbol{\eta}_{|\boldsymbol{y}|}(\boldsymbol{\theta})$ will increase the conditional log-odds of a tie by 1, while the change statistic for the number of triangles in a network is $\left|\boldsymbol{y}_i \cap \boldsymbol{y}_j\right|$, the number of neighbors i and j have in common, suggesting that, a positive coefficient on this statistic will increase the odds of a tie between i and j exponentially in the number of common neighbors. Hunter et al. (2008b) and Krivitsky et al. (2011) offer a further discussion of change statistics and their uses, and Snijders et al. (2006) and Schweinberger (2011) use them to diagnose degeneracy in ERGMs. It would be desirable for a generalization of ERGM to valued networks to facilitate similar local interpretation.

Furthermore, the conditional distribution serves as the basis for maximum pseudo-likelihood estimation (MPLE) for these models. (Strauss and Ikeda, 1990)

2.2.2 Relationship to logistic regression

If the model has the property of *dyadic independence* discussed in the Introduction, or, equivalently, the change statistic $\boldsymbol{\Delta}_{i,j}\boldsymbol{g}(\boldsymbol{y})$ is constant in $\boldsymbol{y}$ (but may vary for different (i,j)), the model trivially reduces to logistic regression. In that case, the MLE and the MPLE are equivalent. (Strauss and Ikeda, 1990) Similarly, it may be a desirable trait for valued generalizations of ERGMs to also reduce to GLM for dyad-independent choices of sufficient statistics.

3 ERGM for counts

We now define ERGMs for count data and discuss the issues that arise in the transition.

3.1 Model definition

Define N, n, and $\mathbb{Y}$ as above. Let $\mathbb{N}_0$ be the set of natural numbers and 0. Here, we focus on counts with no *a priori* upper bound — or counts best modeled thus. Instead of defining the sample space $\mathcal{Y}$ as a subset of a power set, define it as $\mathcal{Y} \subseteq \mathbb{N}_0^{\mathbb{Y}}$, a set of mappings that assign to each dyad $(i,j) \in \mathbb{Y}$ a count. Let $\boldsymbol{y}_{i,j} = \boldsymbol{y}(i,j) \in \mathbb{N}_0$ be the value associated with dyad (i,j).

A (potentially curved) ERGM for a random network of counts $\boldsymbol{Y} \in \mathcal{Y}$ then has the pmf

$$\Pr_{\boldsymbol{\theta};h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{Y} = \boldsymbol{y}) = \frac{h(\boldsymbol{y}) \exp\left(\boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \boldsymbol{g}(\boldsymbol{y})\right)}{\kappa_{h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{\theta})}, \tag{2}$$

where the normalizing constant

$$\kappa_{h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{y} \in \mathcal{Y}} h(\boldsymbol{y}) \exp\left(\boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \boldsymbol{g}(\boldsymbol{y})\right),$$

with $\boldsymbol{\eta}$, $\boldsymbol{g}$, and $\boldsymbol{\theta}$ defined as above, and

$$\boldsymbol{\Theta} \subseteq \boldsymbol{\Theta}_{\mathrm{N}} = \{\boldsymbol{\theta}' \in \mathbb{R}^q : \kappa_{h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{\theta}') < \infty\} \tag{3}$$

(Barndorff-Nielsen, 1978, pp. 115–116; Brown, 1986, pp. 1–2), with $\boldsymbol{\Theta}_{\mathrm{N}}$ being the natural parameter space. Notably, constraint (3) is trivial for binary networks, since their sample space is finite, whereas for counts (3) may constitute a relatively complex constraint.

For the remainder of this paper, we will focus on linear ERGMs, so unless otherwise noted, $p = q$ and $\boldsymbol{\eta}(\boldsymbol{\theta}) \equiv \boldsymbol{\theta}$.

3.2 Reference measure

In addition to the specification of the sufficient statistics $\boldsymbol{g}$ and, for curved families, mapping $\boldsymbol{\eta}$ of model parameters to canonical parameters, an ERGM for counts depends on the specification of the function $h : \mathcal{Y} \to [0, \infty)$. Formally, along with the sample space, it specifies the reference measure: the distribution relative to which the exponential form is specified. For binary ERGMs, h is usually not specified explicitly, though in some ERGM applications, such as models with offsets (Krivitsky et al., 2011, for example) and profile likelihood calculations of Hunter et al. (2008b), the terms with fixed parameters are implicitly absorbed into h.

For valued network data in general, and for count data in particular, specification of h gains a great deal of importance, setting the baseline shape of the dyad distribution and constraining the parameter space. Consider a very simple $p = 1$ model with $\boldsymbol{g}(\boldsymbol{y}) = (\sum_{(i,j)\in\mathbb{Y}} \boldsymbol{y}_{i,j})$, the sum of all dyad values. If $h(\boldsymbol{y}) = 1$ (i.e., discrete uniform), the resulting family has the pmf

$$\Pr_{\boldsymbol{\theta};h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{Y} = \boldsymbol{y}) = \frac{\exp\left(\boldsymbol{\theta} \sum_{(i,j)\in\mathbb{Y}} \boldsymbol{y}_{i,j}\right)}{\kappa_{h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{\theta})} = \prod_{(i,j)\in\mathbb{Y}} \frac{\exp\left(\boldsymbol{\theta}\boldsymbol{y}_{i,j}\right)}{1 - \exp\left(\boldsymbol{\theta}\right)},$$

giving the dyadwise distribution $\boldsymbol{Y}_{i,j} \overset{\text{i.i.d.}}{\sim} \text{Geometric}(p = 1 - \exp(\boldsymbol{\theta}))$, with $\boldsymbol{\theta} < 0$ by (3). On the other hand, suppose that, instead, $h(\boldsymbol{y}) = \prod_{(i,j)\in\mathbb{Y}} (\boldsymbol{y}_{i,j}!)^{-1}$. Then,

$$\Pr_{\boldsymbol{\theta};h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{Y} = \boldsymbol{y}) = \frac{\exp\left(\boldsymbol{\theta} \sum_{(i,j)\in\mathbb{Y}} \boldsymbol{y}_{i,j}\right)}{\kappa_{h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{\theta}) \prod_{(i,j)\in\mathbb{Y}} \boldsymbol{y}_{i,j}!} = \prod_{(i,j)\in\mathbb{Y}} \frac{\exp\left(\boldsymbol{\theta}\boldsymbol{y}_{i,j}\right)}{\boldsymbol{y}_{i,j}! \exp\left(\boldsymbol{\theta}\right)},$$

giving $\boldsymbol{Y}_{i,j} \overset{\text{i.i.d.}}{\sim} \text{Poisson}(\mu = \exp(\boldsymbol{\theta}))$, with $\boldsymbol{\Theta}_{\mathrm{N}} = \mathbb{R}$. The shape of the resulting distributions for a fixed mean is given in Figure 1.

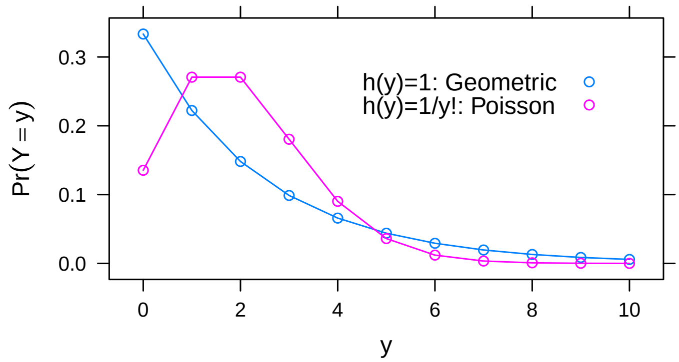


Figure 1: Effect of h on the shape of the distribution. (The mean is fixed at 2.)

The reference measure h thus determines the support and the basic shape of the ERGM distribution. For this reason, we define a *geometric-reference ERGM* to have the form (2) with $h(\boldsymbol{y}) = 1$ and a *Poisson-reference ERGM* to have $h(\boldsymbol{y}) = \prod_{(i,j)\in\mathbb{Y}}(\boldsymbol{y}_{i,j}!)^{-1}$.

Note that this does not mean that any Poisson-reference ERGM will, even under dyadic independence, be dyadwise Poisson. We discuss the sufficient conditions for this in Section 5.2.1.

4 Inference and implementation

As exponential families, valued ERGMs, and ERGMs for counts in particular, inherit the inferential properties of discrete exponential families in general and binary ERGMs in particular, including calculation of standard errors and analysis of deviance. They also inherit the caveats. For example, the Wald test results based on standard errors depend on asymptotics which are questionable for ERGMs with complex dependence structure (Hunter and Handcock, 2006), so we confirm the most important of the results using a simple Monte Carlo test: we fit a nested model without the statistic of interest and simulate its distribution under such a model. The quantile of the observed value of the statistic of interest can then be used as a more robust P-value.

At the same time, generalizing ERGMs to counts raises additional inferential issues. In particular, the infinite sample space of counts means that the constraint (3) is not always trivially satisfied, which results in some valued ERGM specifications not fulfilling regularity conditions. We give an example of this in Section 5.2.3 and Appendix B. These issues also lead to additional computational issues.

4.1 Computational issues

The greatest practical difficulty associated with likelihood inference on these models is usually that the normalizing constant $\kappa_{h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{\theta})$ is intractable, its exact evaluation requiring integration over the sample space $\mathcal{Y}$. However, the exponential-family nature of model also means that, provided a method exists to simulate realizations of networks from the model of interest given a particular $\boldsymbol{\theta}$, the methods of Geyer and Thompson (1992) for fitting exponential families with intractable normalizing constants and, more specifically, their application to ERGMs by Hunter and Handcock (2006), may be used. These methods rely on network sufficient statistics rather than networks themselves and can thus be used with little modification. More concretely, the ratio of two normalizing constants evaluated at $\boldsymbol{\theta}'$ and $\boldsymbol{\theta}$ can be expressed as

$$\begin{aligned}
\frac{\kappa_{h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{\theta}')}{\kappa_{h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{\theta})} &= \frac{\sum_{\boldsymbol{y}\in\mathcal{Y}} h(\boldsymbol{y})\exp\left(\boldsymbol{\eta}(\boldsymbol{\theta}')\cdot\boldsymbol{g}(\boldsymbol{y})\right)}{\kappa_{h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{\theta})} \\
&= \frac{\sum_{\boldsymbol{y}\in\mathcal{Y}} h(\boldsymbol{y})\exp\left((\boldsymbol{\eta}(\boldsymbol{\theta}')-\boldsymbol{\eta}(\boldsymbol{\theta}))\cdot\boldsymbol{g}(\boldsymbol{y})\right)\exp\left(\boldsymbol{\eta}(\boldsymbol{\theta})\cdot\boldsymbol{g}(\boldsymbol{y})\right)}{\kappa_{h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{\theta})} \\
&= \sum_{\boldsymbol{y}\in\mathcal{Y}} \exp\left((\boldsymbol{\eta}(\boldsymbol{\theta}')-\boldsymbol{\eta}(\boldsymbol{\theta}))\cdot\boldsymbol{g}(\boldsymbol{y})\right)\frac{h(\boldsymbol{y})\exp\left(\boldsymbol{\eta}(\boldsymbol{\theta})\cdot\boldsymbol{g}(\boldsymbol{y})\right)}{\kappa_{h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{\theta})} \\
&= \mathrm{E}_{\boldsymbol{\theta};h,\boldsymbol{\eta},\boldsymbol{g}}\left(\exp\left((\boldsymbol{\eta}(\boldsymbol{\theta}')-\boldsymbol{\eta}(\boldsymbol{\theta}))\cdot\boldsymbol{g}(\boldsymbol{Y})\right)\right),
\end{aligned}$$

so given a sample $\boldsymbol{Y}^{(1)},\ldots,\boldsymbol{Y}^{(S)}$ from an initial guess $\boldsymbol{\theta}$, it can be estimated

$$\frac{\kappa_{h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{\theta}')}{\kappa_{h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{\theta})} \approx \sum_{s=1}^{S}\exp\Big((\boldsymbol{\eta}(\boldsymbol{\theta}')-\boldsymbol{\eta}(\boldsymbol{\theta}))\cdot\boldsymbol{g}(\boldsymbol{Y}^{(s)})\Big).$$

Another method for fitting ERGMs, taking advantage of the equivalence of the method of moments to the maximum likelihood estimator for linear exponential families, was implemented by Snijders (2002), using the algorithm by Robbins and Monro (1951) for simulated statistics to fit the model. This approach also trivially extends to valued ERGMs.

Furthermore, because the normalizing constant (if it is finite) is thus accommodated by the fitting algorithm, we may focus on the unnormalized density for the purposes of model specification and interpretation. Therefore, for the remainder of this paper, we specify our models up to proportionality, as Geyer (1999) suggests.

That (3) is not trivially satisfied for all $\boldsymbol{\theta} \in \mathbb{R}^q$ presents an additional computational challenge: even for relatively simple network models, $\boldsymbol{\Theta}$ may have a nontrivial shape. For example, even a simple geometric-reference ERGM

$$\Pr_{\boldsymbol{\theta};h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{Y} = \boldsymbol{y}) \propto \prod_{(i,j)\in\mathbb{Y}} \exp\left(\boldsymbol{\theta} \cdot \boldsymbol{x}_{i,j}\boldsymbol{y}_{i,j}\right),$$

a geometric GLM with a covariate p-vector $\boldsymbol{x}_{i,j}$, has parameter space

$$\boldsymbol{\Theta} = \{\boldsymbol{\theta}' \in \mathbb{R}^p : \forall_{(i,j)\in\mathbb{Y}} \boldsymbol{\theta} \cdot \boldsymbol{x}_{i,j} < 0\},$$

an intersection of up to $|\mathbb{Y}|$ half-spaces (linear constraints). Models with complex dependence structure may have less predictable parameter spaces, and, due to the nature of the algorithm of Hunter and Handcock (2006), the only general way to detect whether a guess for $\boldsymbol{\theta}$ had strayed outside of $\boldsymbol{\Theta}$ may be by diagnostics on the simulation. Bayesian inference with improper priors faces a similar problem, and addressing it in the context of ERGMs is a subject for future work. For this paper, we focus on models in which parameter spaces are provably unconstrained or have very simple constraints.

We base our implementation on the `R` package `ergm` for fitting binary ERGMs. (Handcock, Hunter, Butts, Goodreau, Krivitsky, and Morris, 2010) The design of that package separates the specification of model sufficient statistics from the specification of the sample space of networks (Hunter et al., 2008b), and so we implement our models by substituting in a Metropolis-Hastings sampler that implements our $\mathcal{Y}$ and h of interest. (A simple sampling algorithm for realizations from a Poisson-reference ERGM, optimized for zero-inflated data, is described in Appendix A.) This implementation will be incorporated into a future public release of `ergm`.

4.2 Model degeneracy

Application of ERGMs has long been associated with a complex of problems collectively referred to as "degeneracy". (Handcock, 2003; Rinaldo et al., 2009; Schweinberger, 2011) Rinaldo et al., in particular, list three specific, interrelated, phenomena: 1) when a parameter configuration — even the MLE — induces a distribution for which only a small number of possible networks have non-negligible probabilities, and these networks are often very different from each other (e.g., a sparser-than-observed graph and a complete graph) for an effectively bimodal distribution; 2) when the MLE is hard to find by the available MCMC methods; and 3) when the probability of the observed network under the MLE is relatively low — the observed network is, effectively, between the modes. This bimodality and concentration is often a consequence of the model inducing overly strong positive dependence among dyad values. For example, Snijders et al. (2006) use change statistics to show that under models with positive coefficients on triangle and k-star ($k \geq 2$) counts — the classic "degenerate" ERGM terms — every tie added to the network increases the conditional odds of several other ties and does not decrease the odds of any, creating what Snijders et al. call an "avalanche" toward the complete graph, which emerges as by far the highest-probability realization. (More concretely, under a model with a triangle count with coefficient $\boldsymbol{\theta}_{\triangle}$, adding a tie (i, j) increases the conditional odds of as many ties as i and j have neighbors by $\exp(\boldsymbol{\theta}_{\triangle})$.) Adjusting other parameters, such as density, down to obtain the expected level of sparsity close to that of the observed graph merely induces the bimodal distribution of Phenomenon 1.

An infinite sample space makes Phenomenon 1, as such, unlikely, because the "avalanche" does not have a maximal graph in which to concentrate. However, it does not preclude excessive dependence inducing a bimodal distribution at the MLE, even if neither mode is remotely degenerate in the probabilistic sense. The observed network being between these modes, this may lead to Phenomenon 3, and, due to the nature of the estimation algorithms, such a situation may, indeed, lead to failing estimation — Phenomenon 2.

In this work, we seek to avoid this problem by constructing statistics that prevent the "avalanche" by limiting dependence or employing counterweights to reduce it. (An example of the former approach is the modeling of transitivity in Section 5.2.6, and an example of the latter is the centering in the within-actor covariance statistic developed in Section 5.2.5.) Formal diagnostics developed to date, such as those of Schweinberger (2011) do not appear to be directly applicable to models with infinite sample spaces, so we rely on MCMC diagnostics (Goodreau, Handcock, Hunter, Butts, and Morris, 2008a) instead.

5 Statistics and interpretation for count data

In this section, we develop sufficient statistics for count data to represent network features that may be of interest and discuss their interpretation. In particular, unless otherwise noted, we focus on the Poisson-reference ERGM without complex constraints: $\mathcal{Y} = \mathbb{N}_0^{\mathbb{Y}}$ and $h(\boldsymbol{y}) = \prod_{(i,j)\in\mathbb{Y}}(\boldsymbol{y}_{i,j}!)^{-1}$.

5.1 Interpretation of model parameters

The sufficient statistics of the binary ERGMs and valued ERGMs alike embody the structural properties of the network that are of interest. The tools available for interpreting them are similar as well.

5.1.1 Expectations of sufficient statistics

In a linear ERGM, if $\boldsymbol{\Theta}_{\mathrm{N}}$ is an open set, then, for every $k \in 1..p$, and holding $\boldsymbol{\theta}_{k'}$, $k' \neq k$, fixed, it is a general exponential family property that the expectation $\mathrm{E}_{\boldsymbol{\theta};h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{g}_k(\boldsymbol{Y}))$ is strictly increasing in $\boldsymbol{\theta}_k$. (Barndorff-Nielsen, 1978, pp. 120–121) Thus, if the statistic $\boldsymbol{g}_k$ is a measurement of some feature of interest of the network (e.g., magnitude of counts, interactions between or within a group, isolates, triadic structures), a greater value of $\boldsymbol{\theta}_k$ results in a distribution of networks with more of the feature measured by $\boldsymbol{g}_k$ present.

5.1.2 Discrete change statistic and conditional distribution

Binary ERGM statistics have a "local" interpretation in the form of *change statistics* summarized in Section 2.2.1, we describe similar tools for "local" interpretation of ERGMs for counts here.

Define the set of networks

$$\mathcal{Y}_{i,j}(\boldsymbol{y}) \equiv \{\boldsymbol{y}' \in \mathcal{Y} : \forall_{i',j'\in\mathbb{Y}\backslash\{(i,j)\}}\boldsymbol{y}'_{i',j'} = \boldsymbol{y}_{i',j'}\}.$$

That is, $\mathcal{Y}_{i,j}(\boldsymbol{y})$ is the set of networks such that all dyads but the focus dyad (i,j) are fixed to their values in $\boldsymbol{y}$ while (i,j) itself may vary over its possible values; and define $\boldsymbol{y}_{(i,j)=k} \in \{\boldsymbol{y}' \in \mathcal{Y}_{i,j}(\boldsymbol{y}) : \boldsymbol{y}'_{i,j} = k\}$ (a singleton set) to be the network with non-focus dyads fixed and focus dyad set to k. Then, let the *discrete change statistic*

$$\boldsymbol{\Delta}_{i,j}^{k_1 \to k_2} \boldsymbol{g}(\boldsymbol{y}) \equiv \boldsymbol{g}(\boldsymbol{y}_{(i,j)=k_2}) - \boldsymbol{g}(\boldsymbol{y}_{(i,j)=k_1}).$$

This statistic emerges when taking the ratio of probabilities of two networks that are identical except for a single dyad value:

$$\frac{\Pr_{\boldsymbol{\theta};h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{Y}_{i,j} = \boldsymbol{y}_{(i,j)=k_2} | \boldsymbol{Y} \in \mathcal{Y}_{i,j}(\boldsymbol{y}))}{\Pr_{\boldsymbol{\theta};h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{Y}_{i,j} = \boldsymbol{y}_{(i,j)=k_1} | \boldsymbol{Y} \in \mathcal{Y}_{i,j}(\boldsymbol{y}))} = \frac{h_{i,j}(k_2)}{h_{i,j}(k_1)} \exp\left(\boldsymbol{\theta} \cdot \boldsymbol{\Delta}_{i,j}^{k_1 \to k_2}(\boldsymbol{y})\right),$$

where $h_{i,j} : \mathbb{N}_0 \to \mathbb{R}$ is the component of h associated with dyad (i,j), such that $h(\boldsymbol{y}) \equiv \prod_{(i,j)\in\mathbb{Y}} h_{i,j}(\boldsymbol{y}_{i,j})$, if it can be thus factored. For a Poisson-reference ERGM, $h_{i,j}(k) = (k!)^{-1}$. This may be used to assess the effect of a particular ERGM term on the *decay rate* of the ratios of probabilities of successive values of dyads (Shmueli, Minka, Kadane, Borle, and Boatwright, 2005) and on the shape of the dyadwise conditional distribution: the conditional distribution of a dyad $(i,j) \in \mathbb{Y}$, given all other dyads $(i',j') \in \mathbb{Y}\backslash\{(i,j)\}$,

$$\begin{aligned}\Pr_{\boldsymbol{\theta};h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{Y}_{i,j} = \boldsymbol{y}_{i,j} | \boldsymbol{Y} \in \mathcal{Y}_{i,j}(\boldsymbol{y})) &= \frac{h_{i,j}(\boldsymbol{y}_{i,j}) \exp\left(\boldsymbol{\theta} \cdot \boldsymbol{g}(\boldsymbol{y})\right)}{\sum_{\boldsymbol{y}' \in \mathcal{Y}_{i,j}(\boldsymbol{y})} h(\boldsymbol{y}'_{i,j})_{i,j} \exp\left(\boldsymbol{\theta} \cdot \boldsymbol{g}(\boldsymbol{y}')\right)} \\ &= \frac{h_{i,j}(\boldsymbol{y}_{i,j}) \exp\left(\boldsymbol{\theta} \cdot \boldsymbol{\Delta}_{i,j}^{k_0 \to \boldsymbol{y}_{i,j}} \boldsymbol{g}(\boldsymbol{y})\right)}{\sum_{k \in \mathbb{N}_0} h_{i,j}(k) \exp\left(\boldsymbol{\theta} \cdot \boldsymbol{\Delta}_{i,j}^{k_0 \to k} \boldsymbol{g}(\boldsymbol{y})\right)},\end{aligned}$$

for an arbitrary baseline k_0.

5.2 Model specification statistics

We now propose some specific model statistics to represent common network structural properties and distributions of counts.

5.2.1 Poisson modeling

We begin with statistics that produce Poisson-distributed dyads and model network phenomena that can be represented in a dyad-independent manner.

As a binary ERGM reduces to a logistic regression model under dyadic independence, a Poisson-reference ERGM may reduce to a Poisson regression model.

In a Poisson-reference ERGM, the normalizing constant has a simple closed form if $\boldsymbol{g}(\boldsymbol{y}')$ is linear in $\boldsymbol{y}'_{i,j}$ and does not depend on any other dyads $\boldsymbol{y}'_{i',j'}$, $(i',j') \neq (i,j)$:

$$\forall_{\boldsymbol{y}\in\mathcal{Y}}\forall_{\boldsymbol{y}'_{i,j}\in\mathbb{N}_0}\boldsymbol{\Delta}_{i,j}^{0\to\boldsymbol{y}'_{i,j}}\boldsymbol{g}(\boldsymbol{y}) = \boldsymbol{y}'_{i,j}\boldsymbol{x}_{i,j}. \tag{4}$$

for $\boldsymbol{x}_{i,j} \equiv \boldsymbol{\Delta}_{i,j}^{k\to k+1}\boldsymbol{g}(\boldsymbol{y})$ for any $k \in \mathbb{N}_0$. Then,

$$\boldsymbol{Y}_{i,j} \overset{\text{ind.}}{\sim} \text{Poisson}\left(\mu = \exp\left(\boldsymbol{\theta}\cdot\boldsymbol{\Delta}_{i,j}^{0\to1}\boldsymbol{g}(\boldsymbol{y})\right)\right),$$

giving a Poisson log-linear model, and $\boldsymbol{\Delta}_{i,j}^{0\to1}\boldsymbol{g}$ effectively becomes the covariate vector for $\boldsymbol{Y}_{i,j}$. (If $\boldsymbol{g}(\boldsymbol{y}')$ is linear in $\boldsymbol{y}'_{i,j}$ but does depend on other dyads — $\boldsymbol{x}_{i,j}$ in (4) depends on $\boldsymbol{y}'_{i',j'}$ but not on $\boldsymbol{y}'_{i,j}$ itself — the dyad distribution is conditionally Poisson but not marginally so. An example of this arises in Section 5.2.4.)

Morris et al. (2008) describe many dyad-independent sufficient statistics for binary ERGMs. All of them have the general form

$$\boldsymbol{g}_k(\boldsymbol{y}) \equiv \sum_{(i,j)\in\mathbb{Y}} \boldsymbol{y}_{i,j}\boldsymbol{x}_{i,j,k}, \tag{5}$$

where $\boldsymbol{x}_{i,j,k} \equiv \boldsymbol{\Delta}_{i,j}\boldsymbol{g}_k$ and $\boldsymbol{x}_{i,j,k}$ may be viewed as exogenous (to the model) covariates in a logistic regression for each tie. They could then be used to model a variety of patterns for degree heterogeneity and mixing among actors over (assumed) exogenous attributes. For example, for a uniform homophily model, $\boldsymbol{x}_{i,j,k}$ may be an indicator of whether i and j belong to the same group. If $\boldsymbol{y}_{i,j}$ are counts, these statistics induce a Poisson regression type model (for a Poisson-reference ERGM), where the effect of a unit increase in some $\boldsymbol{\theta}_k$ on dyad (i,j) is to increase its expectation by a factor of $\exp(\boldsymbol{x}_{i,j,k})$. Krivitsky et al. (2009) use this type of model Slovenian periodical "co-readerships" (Batagelj and Mrvar, 2006) — numbers of readers who report reading each pair of periodicals of interest — using as exogenous covariates the class of periodical (daily, weekly, regional, etc.) and the overall readership levels of each periodical.

Curved (i.e., $\boldsymbol{\eta}(\boldsymbol{\theta}) \neq \boldsymbol{\theta}$, $p > q$, and $\boldsymbol{\eta}$ not a linear mapping) ERGMs, in which the $\boldsymbol{g}$ satisfy (4) and dyadic independence, may induce nonlinear

Poisson regression. An example of this is the likelihood component of some latent space network models, with latent space positions being treated as free parameters: for example, the likelihoods of the Poisson models of Hoff (2005) and Krivitsky et al. (2009) are special cases of such an ERGM, with $\boldsymbol{\eta}(\boldsymbol{\theta}) = \left(\boldsymbol{\eta}_{i,j}(\boldsymbol{\theta})\right)_{(i,j)\in\mathbb{Y}}$ and $\boldsymbol{g}(\boldsymbol{y}) = \left(\boldsymbol{y}_{i,j}\right)_{(i,j)\in\mathbb{Y}}$ (i.e., the sufficient statistic is the network), and $\boldsymbol{\eta}_{i,j}(\boldsymbol{\theta})$ mapping latent space positions and other parameters contained in $\boldsymbol{\theta}$ to the logarithms of dyad means (i.e., the dyadwise canonical parameters).

5.2.2 Zero modification

We now turn to model terms that may reshape the distribution of the counts away from Poisson. Social networks tend to be sparse, and larger networks of similar nature tend to be more sparse (Krivitsky et al., 2011). If the interactions among the actors are counted, it is often the case that if two actors interact at all, they interact multiple times. This leads to dyadwise distributions that are zero-inflated relative to Poisson.

These features of sparsity can be modeled using statistics developed for binary ERGMs, applied to a network produced by thresholding the counts (at 1, for zero-modification). For example, a Poisson-reference ERGM with $p = 2$ and

$$\boldsymbol{g}(\boldsymbol{y}) = \left(\sum_{(i,j)\in\mathbb{Y}} \boldsymbol{y}_{i,j}, \sum_{(i,j)\in\mathbb{Y}} 1_{\boldsymbol{y}_{i,j}>0}\right)^{\top}$$

has dyadwise distribution

$$\Pr_{\boldsymbol{\theta};h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{Y} = \boldsymbol{y}) \propto \prod_{(i,j)\in\mathbb{Y}} \exp\left(\boldsymbol{\theta}_1 \boldsymbol{y}_{i,j} + \boldsymbol{\theta}_2 1_{\boldsymbol{y}_{i,j}>0}\right) / \boldsymbol{y}_{i,j}!.$$

This is a parametrization of a zero-modified Poisson distribution (Lambert, 1992), though not a commonly used one, with the probability of 0 being $(1+\exp(\boldsymbol{\theta}_2)(\exp(\exp(\boldsymbol{\theta}_1))-1))^{-1}$ and nonzero values being distributed (conditionally on not being 0) Poisson($\mu = \exp(\boldsymbol{\theta}_1)$), both reducing to Poisson's when $\boldsymbol{\theta}_2 = 0$. Notably, the probability of 0 decreases as $\boldsymbol{\theta}_1$ increases, rather than being solely controlled by $\boldsymbol{\theta}_2$.

5.2.3 Dispersion modeling

Consider the social network of face-to-face conversations among people living in a region. A typical individual will likely not interact at all with vast majority of others, have one-time or infrequent interaction with a large number of others (e.g., with clerks or tellers), and a lot of interaction with a relatively small number of others (e.g., family, coworkers). Some of this may be accounted for by information about social roles and preexisting relationships, but if such information is not available, this leads to a highly overdispersed distribution relative to Poisson, or even zero-inflated Poisson. Overdispersed counts are often modeled using the negative binomial distribution. (McCullagh and Nelder, 1989, p. 199) However, the negative binomial distribution with an unknown dispersion parameter is not an exponential family, making it difficult to fit using our inference techniques. We thus discuss two purely exponential-family approaches for dealing with non-Poisson-dispersed interaction counts in general and overdispersed counts in particular.

Conway–Maxwell–Poisson Distribution Conway–Maxwell–Poisson (CMP) distribution (Shmueli et al., 2005) is an exponential family for counts, able to represent both under- and overdispersion: adding a sufficient statistic of the form

$$\boldsymbol{g}_{\text{CMP}}(\boldsymbol{y}) = \sum_{(i,j)\in\mathbb{Y}} \log(\boldsymbol{y}_{i,j}!), \tag{6}$$

to a Poisson-reference ERGM otherwise fulfilling conditions for Poisson regression described in Section 5.2.1 turns a Poisson regression model for dyads into a CMP regression model.

Its coefficient, $\boldsymbol{\theta}_{\text{CMP}}$, constrained by (3) to $\boldsymbol{\theta}_{\text{CMP}} \leq 1$, controls the degree of dispersion: $\boldsymbol{\theta}_{\text{CMP}} = 0$ retains the Poisson distribution; $\boldsymbol{\theta}_{\text{CMP}} < 0$ induces underdispersion relative to Poisson, approaching the Bernoulli distribution as $\boldsymbol{\theta}_{\text{CMP}} \to -\infty$; and $\boldsymbol{\theta}_{\text{CMP}} > 0$ induces overdispersion, attaining the geometric distribution at $\boldsymbol{\theta}_{\text{CMP}} = 1$, its most overdispersed point.

Normally, the greatest hurdle associated with using CMP is that its normalizing constant does not, in general, have a known closed form. In our case, because intractable normalizing constants are already accommodated by the methods of Section 4, so using CMP in this setting requires no additional effort.

At the same time, CMP is neither regular nor steep (per Appendix B), so the properties of its estimators are not guaranteed, particularly for highly overdispersed data. We have found experimentally that counts as dispersed as geometric distribution or more so often cause the fitting methods of Section 4 to fail.

Variance-like parameters Some control over the variance can be attained by adding a statistic of the form $\boldsymbol{g}(\boldsymbol{y}) = \sum_{(i,j)\in\mathbb{Y}} \boldsymbol{y}_{i,j}^a$, $a \neq 1$. Statistics with $a > 1$, such as $\boldsymbol{g}(\boldsymbol{y}) = \sum_{(i,j)\in\mathbb{Y}} \boldsymbol{y}_{i,j}^2$, suffer the same problem as a Strauss point process (Kelly and Ripley, 1976): for any $\theta, \epsilon > 0$, $\lim_{y\to\infty} \exp(\theta y^{1+\epsilon})/y! = \infty$, leading to (3) constraining $\theta \leq 0$, able to represent only *underdispersion*.

Thus, we propose to model dispersion by adding a statistic of the form

$$\boldsymbol{g}(\boldsymbol{y}) = \sum_{(i,j)\in\mathbb{Y}} \boldsymbol{y}_{i,j}^{1/2} = \sum_{(i,j)\in\mathbb{Y}} \sqrt{\boldsymbol{y}_{i,j}}. \tag{7}$$

To the extent that the counts are Poisson-like, the square root is a variance-stabilizing transformation (McCullagh and Nelder, 1989, p. 196). Then, a model with $p = 2$ and dyadwise sufficient statistic

$$\boldsymbol{g}(\boldsymbol{y}) = \left(\sum_{(i,j)\in\mathbb{Y}} \sqrt{\boldsymbol{y}_{i,j}}, \sum_{(i,j)\in\mathbb{Y}} \boldsymbol{y}_{i,j} \right)^{\top} \tag{8}$$

may be viewed as a modeling the first and second moments of $\sqrt{\boldsymbol{y}_{i,j}}$. That the highest-order term is still on the order of $\boldsymbol{y}_{i,j}$ guarantees that $\boldsymbol{\Theta} = \mathbb{R}^p$ — a practical advantage over CMP.

As with CMP, the normalizing constant is intractable. To explore the shape of this distribution, we fixed $\boldsymbol{\theta}_1$ at each of a range of values and found $\boldsymbol{\theta}_2$s such that the induced distribution had the expected value of 1. We then simulated from the fit. The estimated pmf for each configuration and the comparison with the geometric distribution with the same expectation is given in Figure 2. Smaller coefficients on (7) ($\boldsymbol{\theta}_1$) correspond to greater dispersion, with coefficients on dyad sum ($\boldsymbol{\theta}_2$) increasing to compensate, and vice versa, with $\boldsymbol{\theta}_1 = 0$ corresponding to a Poisson distribution. As the dispersion increases, the mean is preserved in part by increasing $\Pr(\boldsymbol{Y}_{i,j} = 0)$ and, for sufficiently high values of $\boldsymbol{y}_{i,j}$, the geometric distribution still dominates. Thus, there is a trade-off between the convenience of a model without complex constraints on the parameter space and the ability to model greater dispersion. In practice, if the substantive reasons for overdispersion are due

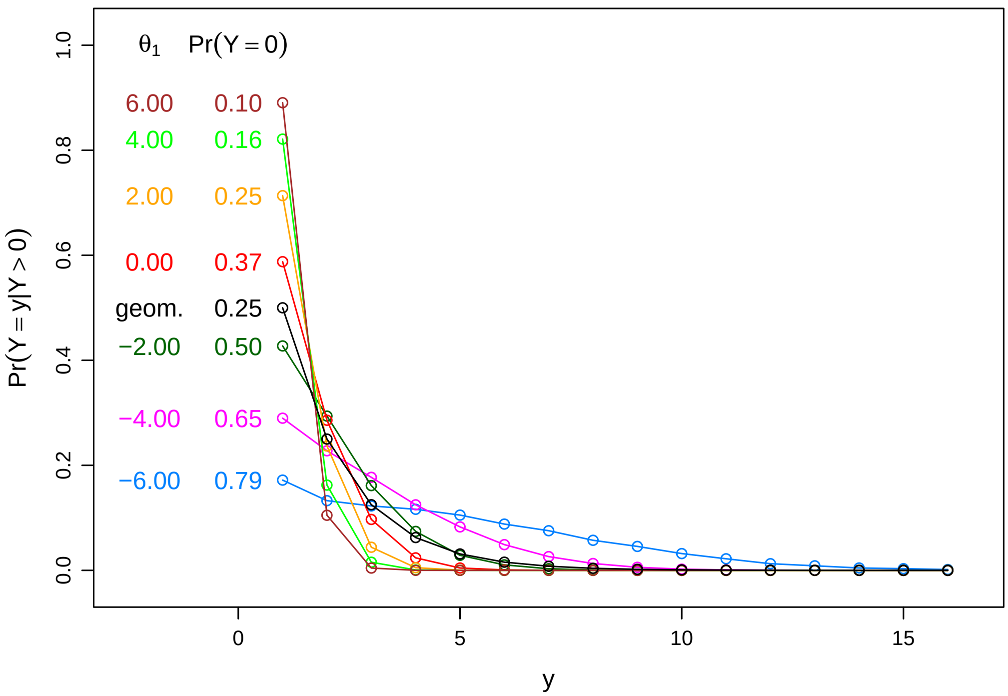


Figure 2: Dyadwise distributions attainable by the model (8). Because $\Pr(Y = 0)$ varies greatly for different $\boldsymbol{\theta}_1$ yet can be adjusted separately by an appropriate model term, we plot the probabilities conditional on $Y > 0$.

to unaccounted-for heterogeneity, the latter might not be a serious disadvantage, and excess zeros can be compensated for by a term from Section 5.2.2.

5.2.4 Mutuality

Many directed networks, such as friendship nominations, exhibit *mutuality* — that, other things being equal, if a tie (i, j) exists, a tie (j, i) is more likely to exist as well — and binary ERGMs can model this phenomenon using a sufficient statistic $\boldsymbol{g}(\boldsymbol{y}) = \sum_{(i,j)\in\mathbb{Y},i<j} \boldsymbol{y}_{i,j}\boldsymbol{y}_{j,i} = \sum_{(i,j)\in\mathbb{Y},i<j} \min(\boldsymbol{y}_{i,j}, \boldsymbol{y}_{j,i})$, counting the number of reciprocated ties. (Holland and Leinhardt, 1981) Other sufficient statistics that can model it include $\boldsymbol{g}(\boldsymbol{y}) = \sum_{(i,j)\in\mathbb{Y},i<j} 1_{\boldsymbol{y}_{i,j}\neq\boldsymbol{y}_{j,i}}$ and $\boldsymbol{g}(\boldsymbol{y}) = \sum_{(i,j)\in\mathbb{Y},i<j} 1_{\boldsymbol{y}_{i,j}=\boldsymbol{y}_{j,i}}$, the counts of asymmetric and symmetric dyads,

respectively. (Morris et al., 2008)

In the presence of an edge count term, these three are simply different parametrizations of the same distribution family:

$$\boldsymbol{y}_{i,j}\boldsymbol{y}_{j,i} = \frac{(\boldsymbol{y}_{i,j} + \boldsymbol{y}_{j,i}) - 1_{\boldsymbol{y}_{i,j} \neq \boldsymbol{y}_{j,i}}}{2} = \frac{(\boldsymbol{y}_{i,j} + \boldsymbol{y}_{j,i}) - 1 + 1_{\boldsymbol{y}_{i,j} = \boldsymbol{y}_{j,i}}}{2}.$$

Nevertheless, these three different statistics suggest two major ways to generalize the terms to count data: by evaluating a product or a minimum of the values, or by evaluating their similarity or difference. We discuss them in turn.

Product It is tempting to model mutuality for count data in the same manner as for binary data, with $\boldsymbol{y}_{i,j}$ and $\boldsymbol{y}_{j,i}$ being values rather than indicators. For example, a simple model with overall dyad mean and reciprocity terms, with $p = 2$ and

$$\boldsymbol{g}(\boldsymbol{y}) = \left(\sum_{(i,j) \in \mathbb{Y}} \boldsymbol{y}_{i,j}, \sum_{(i,j) \in \mathbb{Y}, i<j} \boldsymbol{y}_{i,j}\boldsymbol{y}_{j,i} \right)^{\mathsf{T}}$$

would have a conditional Poisson distribution:

$$\boldsymbol{Y}_{i,j} = \boldsymbol{y}_{i,j} | \boldsymbol{Y} \in \mathcal{Y}_{i,j}(\boldsymbol{y}) \sim \text{Poisson}\left(\mu = \exp\left(\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2 \boldsymbol{y}_{j,i}\right)\right),$$

a desirable property. However, because for any $c > 0$, $\lim_{y \to \infty} \exp(cy^2)/(y!)^2 = \infty$, for $\boldsymbol{\theta}_2 > 0$, representing positive mutuality, (3) is not fulfilled. (Note that the expected value of $\boldsymbol{Y}_{i,j}$ is *exponential* in the value of $\boldsymbol{Y}_{j,i}$ and vice versa. Again, a Strauss point process exhibits a similar problem. (Kelly and Ripley, 1976))

Geometric mean As with dispersion, the problem can be alleviated by using the geometric mean of $\boldsymbol{y}_{i,j}$ and $\boldsymbol{y}_{j,i}$ instead of their product. As in Section 5.2.3, this choice may be justified as an analog of covariance on variance-stabilized counts. This changes the shape of the distribution in ways that are difficult to interpret: if

$$\boldsymbol{g}(\boldsymbol{y}) = \left(\sum_{(i,j) \in \mathbb{Y}} \boldsymbol{y}_{i,j}, \sum_{(i,j) \in \mathbb{Y}, i<j} \sqrt{\boldsymbol{y}_{i,j}\boldsymbol{y}_{j,i}} \right)^{\mathsf{T}},$$

then

$$\Pr_{\boldsymbol{\theta};h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{Y}_{i,j} = \boldsymbol{y}_{i,j} | \boldsymbol{Y} \in \mathcal{Y}_{i,j}(\boldsymbol{y})) \propto \exp\left(\boldsymbol{\theta}_1 \boldsymbol{y}_{i,j} + (\boldsymbol{\theta}_2 \sqrt{\boldsymbol{y}_{j,i}}) \sqrt{\boldsymbol{y}_{i,j}}\right) / \boldsymbol{y}_{i,j}!,$$

and, with nonzero $\boldsymbol{y}_{j,i}$, the probabilities of greater values of $\boldsymbol{Y}_{i,j}$ are inflated by more. The analogy to covariance further suggests centering the statistic:

$$\boldsymbol{g}(\boldsymbol{y}) = \sum_{(i,j)\in\mathbb{Y}, i<j} (\sqrt{\boldsymbol{y}_{i,j}} - \overline{\sqrt{\boldsymbol{y}}})(\sqrt{\boldsymbol{y}_{j,i}} - \overline{\sqrt{\boldsymbol{y}}}),$$

for

$$\overline{\sqrt{\boldsymbol{y}}} = \frac{1}{|\mathbb{Y}|} \sum_{(i',j')\in\mathbb{Y}} \sqrt{\boldsymbol{y}_{i',j'}}. \tag{9}$$

Minimum An alternative generalization is to take the minimum of the two values. For example, if

$$\boldsymbol{g}(\boldsymbol{y}) = \left(\sum_{(i,j)\in\mathbb{Y}} \boldsymbol{y}_{i,j}, \sum_{(i,j)\in\mathbb{Y}, i<j} \min(\boldsymbol{y}_{i,j}, \boldsymbol{y}_{j,i}) \right)^{\mathsf{T}},$$

then

$$\Pr_{\boldsymbol{\theta};h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{Y}_{i,j} = \boldsymbol{y}_{i,j} | \boldsymbol{Y} \in \mathcal{Y}_{i,j}(\boldsymbol{y})) \propto \exp\left(\boldsymbol{\theta}_1 \boldsymbol{y}_{i,j} + \boldsymbol{\theta}_2 \min(\boldsymbol{y}_{i,j} - \boldsymbol{y}_{j,i}, 0)\right) / \boldsymbol{y}_{i,j}!. \tag{10}$$

Thus, a possible interpretation for this term is that the conditional probability for a particular value of $\boldsymbol{Y}_{i,j}$, $\boldsymbol{y}_{i,j}$ is deflated by $\exp(\boldsymbol{\theta}_2)$ for every unit by which $\boldsymbol{y}_{i,j}$ is less than $\boldsymbol{y}_{j,i}$. In a sense, $\boldsymbol{y}_{j,i}$ "pulls up" $\boldsymbol{y}_{i,j}$ to its level and vice versa.

Negative difference Generalizing the concept of similarity between $\boldsymbol{y}_{i,j}$ and $\boldsymbol{y}_{j,i}$ leads to a statistic of difference between their values. We negate it so that a positive coefficient value leads to greater mutuality. Then,

$$\boldsymbol{g}(\boldsymbol{y}) = \left(\sum_{(i,j)\in\mathbb{Y}} \boldsymbol{y}_{i,j}, \sum_{(i,j)\in\mathbb{Y}, i<j} - \left|\boldsymbol{y}_{i,j} - \boldsymbol{y}_{j,i}\right| \right)^{\mathsf{T}}, \tag{11}$$

and

$$\Pr_{\boldsymbol{\theta};h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{Y}_{i,j} = \boldsymbol{y}_{i,j} | \boldsymbol{Y} \in \mathcal{Y}_{i,j}(\boldsymbol{y})) \propto \exp\left(\boldsymbol{\theta}_1 \boldsymbol{y}_{i,j} - \boldsymbol{\theta}_2 \left|\boldsymbol{y}_{i,j} - \boldsymbol{y}_{j,i}\right|\right) / \boldsymbol{y}_{i,j}!,$$

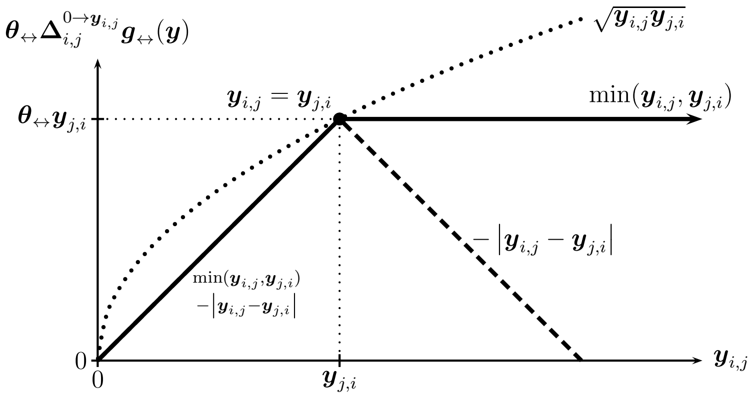


Figure 3: Effect of proposed mutuality statistics ($\boldsymbol{g}_{\leftrightarrow}$) with parameter $\boldsymbol{\theta}_{\leftrightarrow} > 0$ on the distribution of $\boldsymbol{Y}_{i,j}$, given that $\boldsymbol{Y}_{j,i} = \boldsymbol{y}_{j,i}$. Whereas the $\min(\boldsymbol{y}_{i,j}, \boldsymbol{y}_{j,i})$ statistic deflates the probabilities of those values of $\boldsymbol{y}_{i,j}$ that are less than $\boldsymbol{y}_{j,i}$, thus inflating all of those of $\boldsymbol{y}_{i,j}$ above or equal to it, thus "pulling $\boldsymbol{Y}_{i,j}$ up", the $-\left|\boldsymbol{y}_{i,j} - \boldsymbol{y}_{j,i}\right|$ statistic deflates the probabilities in both directions away from $\boldsymbol{y}_{j,i}$, thus inflating those that are the closest, "pulling $\boldsymbol{Y}_{i,j}$ in". $\sqrt{\boldsymbol{y}_{i,j}\boldsymbol{y}_{j,i}}$ inflates greater values of $\boldsymbol{y}_{i,j}$ in general, inflating by more for greater $\sqrt{\boldsymbol{y}_{j,i}}$.

so the conditional probability of a particular $\boldsymbol{y}_{i,j}$ is deflated by $\exp(\boldsymbol{\theta}_2)$ for every unit difference from $\boldsymbol{y}_{j,i}$, in either direction. Thus, $\boldsymbol{y}_{j,i}$ "pulls in" $\boldsymbol{y}_{i,j}$ and vice versa. Of course, other differences (e.g., squared difference) are also possible.

We use the discrete change statistic to visualize the differences between these variants in Figure 3, plotting the $\boldsymbol{\theta}_{\leftrightarrow}\boldsymbol{\Delta}_{i,j}^{0\to\boldsymbol{y}_{i,j}}\boldsymbol{g}_{\leftrightarrow}(\boldsymbol{y})$ summand of

$$\log \frac{\Pr_{\boldsymbol{\theta};h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{Y}_{i,j} = \boldsymbol{y}_{i,j} | \boldsymbol{Y} \in \mathcal{Y}_{i,j}(\boldsymbol{y}))}{\Pr_{\boldsymbol{\theta};h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{Y}_{i,j} = 0 | \boldsymbol{Y} \in \mathcal{Y}_{i,j}(\boldsymbol{y}))} = \boldsymbol{\theta} \cdot \boldsymbol{\Delta}_{i,j}^{0\to\boldsymbol{y}_{i,j}}\boldsymbol{g}(\boldsymbol{y})$$

for each variant. Lastly, while the conditional distributions, and hence the parameter interpretations for the minimum and the negative difference statistic, are different, models induced by (10) and (11) are also reparametrizations of each other: $\min(\boldsymbol{y}_{i,j}, \boldsymbol{y}_{j,i}) = \frac{1}{2}\left((\boldsymbol{y}_{i,j} + \boldsymbol{y}_{j,i}) - \left|\boldsymbol{y}_{i,j} - \boldsymbol{y}_{j,i}\right|\right)$.

5.2.5 Actor heterogeneity

Another property found in social networks is that different individuals have different overall propensities to have ties: activity, popularity, and (undirected) sociality. Some of this heterogeneity may be accounted for by exogenous covariates. For the unaccounted-for heterogeneity, two major approaches have been used: *conditional*, in which actor-specific parameters are added to the model to absorb it, with dyadic independence typically assumed given these parameters, and *marginal*, in which statistics are added that represent the effects of heterogeneity on the overall network features. Examples of the conditional approach include the very first exponential-family model for networks, the p_1, which included a fixed effect for every actor (Holland and Leinhardt, 1981); and the p_2 model and latent space models, which used using random effects (van Duijn, Snijders, and Zijlstra, 2004; Hoff, 2005; Krivitsky et al., 2009; Mariadassou et al., 2010). The marginal approach includes the k-star statistics for $k \geq 2$ (Frank and Strauss, 1986), which, for a fixed network density, become more prevalent as heterogeneity increases at the cost of often inducing degeneracy; alternating k-stars and geometrically weighted degree statistics (Snijders et al., 2006; Hunter and Handcock, 2006), which attempt to remedy the degeneracy of k-stars; and statistics such as the square root degree activity/popularity, which sum of each actors' degree taken to 3/2 power, which also increases with greater heterogeneity, but not as rapidly as k-stars do (Snijders, van de Bunt, and Steglich, 2010), avoiding degeneracy. In the conditional approach, using fixed effects lacks parsimony and using random effects creates a problem with a doubly-intractable normalizing constant, beyond the scope of this paper, so we develop a marginal approach here.

Actor heterogeneity may be viewed marginally as positive within-actor correlation among the dyad values. Following the discussion in the previous sections, we propose a form of pooled within-actor covariance of variance-stabilized dyad values, scaled to the same magnitude as the dyad sum:

$$\boldsymbol{g}(\boldsymbol{y}) = \sum_{i \in N} \frac{1}{n-2} \sum_{j,k \in \mathbb{Y}_{i\rightarrow} \wedge j<k} (\sqrt{\boldsymbol{y}_{i,j}} - \overline{\sqrt{\boldsymbol{y}}})(\sqrt{\boldsymbol{y}_{i,k}} - \overline{\sqrt{\boldsymbol{y}}}), \tag{12}$$

for $\mathbb{Y}_{i\rightarrow}$ being the set of actors to who whom i may have ties ($\equiv \{j' : (i,j') \in \mathcal{Y}\}$) and $\overline{\sqrt{\boldsymbol{y}}}$ defined as in (9). This statistic would increase with greater out-tie heterogeneity, an analogous statistic could be specified for in-tie heterogeneity, and dropping the directionality would produce an undirected version of this statistic.

We have considered other variants, including the uncentered version, in which each summand in (12) is simply $\sqrt{\boldsymbol{y}_{i,j}\boldsymbol{y}_{i,k}}$. We found that in undirected networks in particular, such a model term can induce a degeneracy-like bimodal distribution of networks. (This is likely because in undirected networks, the positive dependence is not contained within each actor, so subtracting $\sqrt{\boldsymbol{y}}$ serves as a counterweight to avert the "avalanche".)

5.2.6 Triad-closure bias

We now turn to the question of how to represent triad-closure bias — friend-of-a-friend effects — in count data. As with mutuality, merely multiplying values of the dyads in a triad leads to a model which cannot have positive triad closure bias. In addition, ERGM sufficient statistics that take counts over triads often exhibit degeneracy. (Schweinberger, 2011) For these reasons, we describe a family of statistics that sum over dyads instead. Wyatt et al. (2010) use a generalization of the curved geometrically-weighted edgewise shared partners (GWESP) statistic (Hunter and Handcock, 2006), though it is not clear whether it is suitable for data with an infinite sample space. We thus describe a more conservative family of statistics.

One term used to model triad closure in binary dynamic networks by Snijders et al. (2010) is the *transitive ties* effect, the most conservative special case of the GWESP (Hunter and Handcock, 2006) statistic. This statistic counts the number of ties (i, j) such that there exists at least one path of length 2 (*two-path*) between them — a third actor k such that $\boldsymbol{y}_{i,k} = \boldsymbol{y}_{k,j} = 1$. (Unlike the triangle count, each dyad may contribute at most $+1$ to the statistic, no matter how many such ks exist.)

One generalization of this statistic to counts is

$$\boldsymbol{g}(\boldsymbol{y}) = \sum_{(i,j)\in\mathbb{Y}} \min\left(\boldsymbol{y}_{i,j}, \max_{k\in N}\left(\min(\boldsymbol{y}_{i,k}, \boldsymbol{y}_{k,j})\right)\right). \tag{13}$$

Intuitively, define the strength of a two-path from i to j to be the minimum of the values along the path. The statistic is then the sum over the dyads (i, j) of the minimum of the value of (i, j) and the value of the strongest two-path between. The interpretation is thus somewhat analogous to that of the minimum mutuality statistic, with $\boldsymbol{y}_{j,i}$ replaced by $\max_{k\in N}(\min(\boldsymbol{y}_{i,k}, \boldsymbol{y}_{k,j}))$. The motivation for using minimum, as opposed to negative absolute difference, to combine the two-path value with the focus dyad value is that the intuitive notion of friend-of-a-friend effect that this statistic embodies suggests that while the presence of a mutual friend may increase the probability or expected value of a particular friendship (i.e., "pull it up"), it should not limit it (i.e., "pull it in") as an absolute difference would. These interpretations are somewhat oversimplified: it is just as true that a positive coefficient on this statistic results in $\boldsymbol{y}_{i,j}$ "pulling up" the potential two-paths between i and j.

In a directed network, (13) would model transitive (hierarchical) triads, while

$$\boldsymbol{g}(\boldsymbol{y}) = \sum_{(i,j)\in\mathbb{Y}} \min\left(\boldsymbol{y}_{i,j}, \max_{k\in N}\left(\min(\boldsymbol{y}_{j,k}, \boldsymbol{y}_{k,i})\right)\right)$$

would model cyclical (antihierarchical) triads.

The statistic (13) is a fairly conservative one, less likely to induce excessive dependence and bimodality, at the cost of sensitivity. More generally, one may specify a triadic statistic using three functions: first, $v_{\text{2-path}} : \mathbb{N}_0^2 \to \mathbb{R}$, how the "value" of a two-path $i \to j \to k$ is computed from its constituent segments; second, $v_{\text{combine}} : \mathbb{R}^{n-2} \to \mathbb{R}$, how the values of the possible two-paths from i to j are combined with each other to compute the strength of the pressure on i and j to close the triad or increase their interaction; and third, $v_{\text{affect}} : \mathbb{N}_0 \times \mathbb{R} \to \mathbb{R}$ how this pressure affects $\boldsymbol{Y}_{i,j}$. Given these,

$$\boldsymbol{g}(\boldsymbol{y}) = \sum_{(i,j)\in\mathbb{Y}} v_{\text{affect}}\left(\boldsymbol{y}_{i,j}, v_{\text{combine}}\left(v_{\text{2-path}}(\boldsymbol{y}_{i,k}, \boldsymbol{y}_{k,j})_{k\in N\setminus\{i,j\}}\right)\right). \tag{14}$$

Thus, for example, one could set v_{combine} to sum its arguments rather than take their maximum, or one can replace taking the minimum with taking a geometric mean. We illustrate the difference it makes in Section 6.2.

6 Examples

6.1 Example 1: Social relations in a karate club

In this application, we use a Poisson-reference ERGM to compare impacts of social forces — transitivity and homophily — on the structure of a valued network of interactions between members of a university karate club. Zachary (1977) reported observations of social relations in a university karate club with membership that varied between 50 and 100. The actors — 32 ordinary club members and officers, the club president ("John A."), and the part-time instructor ("Mr. Hi") — were the ones who consistently interacted outside of the club. Over the course of the study, the club divided into two factions, and, ultimately, split into two clubs, one led by Hi and the other by John and the original club's officers. The split was driven by a disagreement over whether Hi could unilaterally change the level of compensation for his services. We pose a similar question to Goodreau et al. (2008b): is the structure at the time of observation driven by faction allegiance or by transitivity ("friend-of-a-friend" effects)?

Zachary identifies the faction with which each of the 34 actors was aligned and how strongly and reports, for each pair of actors, the count of social contexts in which they interacted. The 8 contexts considered were academic classes at the university; Hi's private karate studio in his night classes; Hi's private karate studio where he taught on weekends; student-teaching at Hi's studio; the university rathskeller (bar) located near the karate club; a bar located near the university campus; open karate tournaments in the area; and intercollegiate karate tournaments. The highest number of contexts of interaction for a pair of individuals that was observed was 7. The network is visualized in Figure 4.

We begin with a Poisson-reference ERGM. Empirically, this network is more sparse than a Poisson density for dyad values would suggest: the mean number of dyadwise contexts of interaction ($\sum_{(i,j)\in\mathbb{Y}} \boldsymbol{y}_{i,j}/\left|\mathbb{Y}\right|$) is 0.41, for which a Poisson distribution predicts an expected density ($\mathrm{E}(\sum_{(i,j)\in\mathbb{Y}} 1_{\boldsymbol{Y}_{i,j}>0}/\left|\mathbb{Y}\right|)$) of 0.34, whereas the observed density is 0.14. Given that two individuals interact, the interaction for a given pair of individuals is likely to be dependent across the social contexts counted so the counts are likely to be over- or under-dispersed. Thus, as a baseline, we model the values as a zero-modified Conway–Maxwell–Poisson (Shmueli et al., 2005) distribution using the following sufficient statistics:

baseline propensity to have ties: number of dyads with nonzero value;
baseline intensity of interactions: sum of dyad values; and
CMP dispersion: a statistic of the form (6).

In modeling the structure of the interactions, we represent differential propensity of the faction leaders to interact, the effect of differences in faction membership, and triad-closure bias using the following sufficient statistics:

intensity of Hi's interaction: sum of dyad values incident on Hi;
intensity of John's interaction: sum of dyad values incident on John;
similarity (negative difference) in faction membership: a statistic of the form (5) with $\boldsymbol{x}_{i,j,k} = -\left|m_i - m_j\right|$, where m_i is the faction membership code of actor i; and

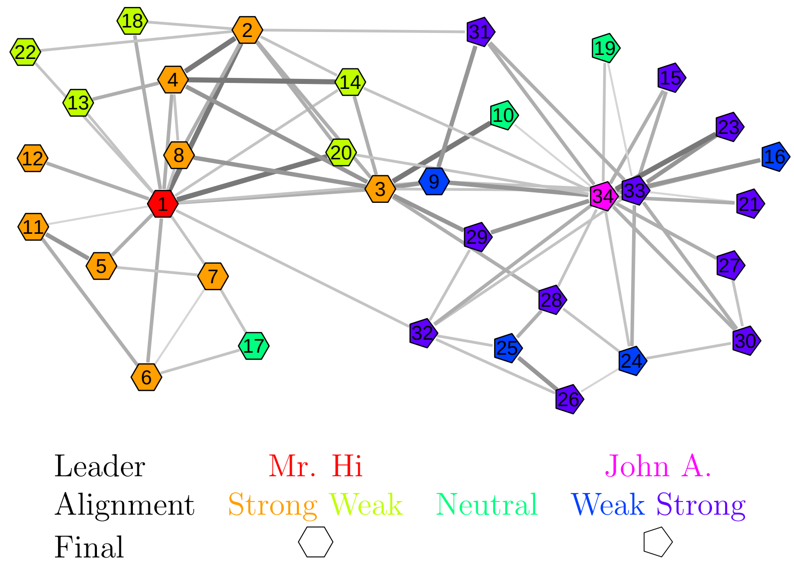


Figure 4: The sociogram of the Zachary (1977) karate club network. The color of each plotting symbol shows each actor's faction alignment and its shape shows which of the two clubs the actor ultimately joined. Darker and thicker lines correspond to more social context of interaction.

Table 1: Results from fitting the models to Karate Club network

	Estimates (Std. Errors)		
Term	Faction	Transitivity	Full
Dispersion	$\mathbf{-2.55}$ (0.57)	$\mathbf{-1.87}$ (0.61)	$\mathbf{-2.33}$ (0.60)
Ties	$\mathbf{-7.76}$ (0.99)	$\mathbf{-7.29}$ (1.04)	$\mathbf{-7.54}$ (1.01)
Baseline intensities	**3.97** (0.68)	**2.88** (0.75)	**3.64** (0.74)
Hi's intensities	**0.80** (0.15)	**0.50** (0.12)	**0.71** (0.15)
John's intensities	**0.80** (0.14)	**0.54** (0.12)	**0.72** (0.16)
Faction similarity	**0.27** (0.04)		**0.25** (0.04)
Transitivity		**0.21** (0.09)	0.11 (0.09)

Coefficients statistically significant at $\alpha = 0.05$ are **bolded**.
Standard errors incorporate the uncertainty introduced by approximating of the likelihood using MCMC (Hunter and Handcock, 2006).

transitivity of intensities: the statistic (13).
Faction memberships m_i are coded as follows: strongly Hi's as -2, weakly Hi's as -1, neutral as 0, weakly John's as $+1$, and strongly John's as $+2$. We fit three models: the full model, with all of the above-described terms, the model excluding transitivity ("Faction"), and the model excluding faction membership ("Transitivity").

Table 1 gives the results for the three fits. MCMC diagnostics, described by Goodreau et al. (2008a), show adequate mixing and networks simulated from these fits have, on average, statistics equal to the observed sufficient statistics. The CMP dispersion estimates for all three models are negative and highly significant, very far from the non-open boundary of the parameter space at $\boldsymbol{\theta}_k \leq 1$, so the lack of steepness is unlikely to be problematic in this case. The estimated value of the dispersion parameter for the full model (-2.33) suggests strong underdispersion relative to zero-modified Poisson and the rest of the model: it implies that the estimated "denominator" is $(\boldsymbol{y}_{i,j}!)^{1-(-2.33)} = (\boldsymbol{y}_{i,j}!)^{3.33}$, rather than Poisson's $(\boldsymbol{y}_{i,j}!)^1$. Highly negative CMP coefficients may also be interpreted as the model being an overfit.

There is a highly significant negative coefficient on the baseline propensity for ties. An interpretation for this is that, from the point of view of a single dyad, the probability of a given pair of actors having a tie is deflated, but if they do have a tie, it is likely to be across multiple social contexts. Both faction leaders appear to have greater overall propensities to interact than the other club members, and, interestingly, they appear to have similar effect sizes to each other.

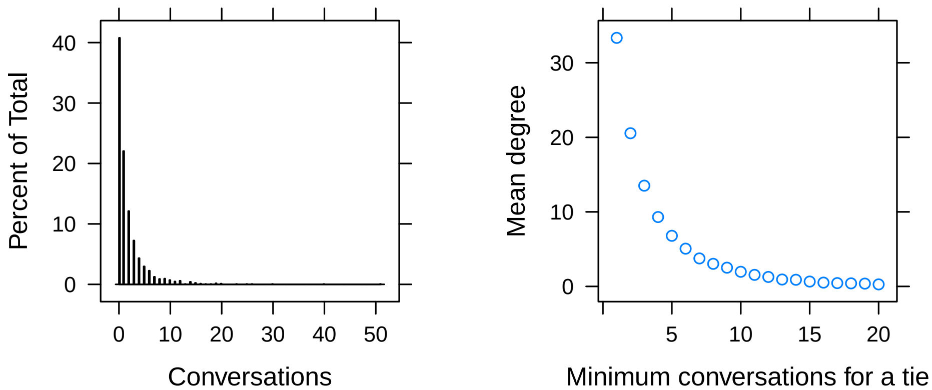
Taken separately, the faction similarity effect is highly statistically significant and positive, indicating a positive faction cohesion. The transitivity effect is significant by the Wald test, but a Monte Carlo test gives its one-sided (since only positive transitivity is of interest) P-value as 0.11. Put together, the transitivity loses any potential significance. (Notably, the estimated correlation between their parameter estimates is -0.34.) This suggests that they are explaining the same aspect of the network structure, but that faction allegiance is the much stronger explanation. Though factions may themselves be endogenous due to influence through social relations or, as Zachary concludes, the two processes reinforced each other over time, at observation time, faction allegiance explains network structure better than transitivity.

6.2 Example 2: Interactions in a fraternity

In a series of studies in the 1970s, Bernard et al. (1979–1980) assessed accuracy of retrospective sociometric surveys in a number of settings, including a college fraternity whose 58 occupants had all lived there for at least three months. To record the true amounts of interaction, for several days, unobtrusive observers were sent to periodically walk through the fraternity to note students engaged in conversation. Obtaining these network data from Batagelj and Mrvar (2006), we model these observed pairwise interaction counts.

The raw distribution of counts, given in Figure 5a, appears to be strongly overdispersed relative to Poisson, and, indeed, relative to the geometric distribution: the mean of counts is 2.0, while their standard deviation (not variance) is 3.4. At least some of this is due to actor heterogeneity: the square root of the within-actor variance of the counts is 3.1. Excluding extreme observations (all values over 30) does not make a qualitative difference. (The statistics are 1.9, 3.0, and 2.8, respectively.) Nor does there appear a natural place to threshold the counts to produce a binary network. (See Figure 5b.) We thus model the baseline shape of the distribution of counts using the following terms:

baseline propensity to have ties: number of dyads with nonzero value;
baseline intensity of interactions: sum of dyad values; and



(a) Count distribution (b) Effect of thresholding

Figure 5: Conversation count summaries for Bernard and Killworth's fraternity network

underdispersion: the statistic (8).
(We have also attempted to use CMP (via (6)) but found the process to be unstable due to the greater-than-geometric level of dispersion.)

Little was recorded about the social roles of the fraternity members, so we consider the effects of endogenous social forces:
actor heterogeneity: the undirected version of (12);
transitivity of intensities: the statistic (13).
Faust (2007), in particular, found that in many empirical networks, much of the apparent triadic effects are accounted for by variations in degree distribution and other lower-order effects. Thus, we consider four models: baseline shape only (B), baseline with heterogeneity (BH), baseline with transitivity (BT), and all terms (BHT), to explore this concept in a valued setting.

We report the model fits in Table 2. MCMC diagnostics, described by Goodreau et al. (2008a), show adequate mixing and unimodal distributions of sufficient statistics, and networks simulated from these fits have, on average, statistics equal to the observed sufficient statistics. The baseline dyadwise distribution terms are difficult to interpret, but the highly negative coefficient on underdispersion suggests a a strong degree of overdispersion, as expected. Some of this overdispersion appears to be absorbed by modeling actor heterogeneity, however. There are indications a high degree of heterogeneity in

Table 2: Results from fitting the models to Bernard and Killsworth's fraternity network

Term	Estimates (Std. Errors) B	BH	BT	BHT
Ties	**5.60** (0.21)	**4.96** (0.17)	**6.24** (0.21)	**4.98** (0.17)
Intensity	**3.65** (0.05)	**3.13** (0.06)	**3.40** (0.07)	**3.12** (0.06)
Underdispersion	**−9.71** (0.22)	**−8.23** (0.20)	**−10.52** (0.22)	**−8.26** (0.19)
Heterogeneity		**1.48** (0.06)		**1.46** (0.07)
Transitivity			**0.46** (0.05)	0.03 (0.04)

Coefficients statistically significant at $\alpha = 0.05$ are **bolded**.
Standard errors incorporate the uncertainty introduced by approximating of the likelihood using MCMC (Hunter and Handcock, 2006).

individuals' interactions, over and above that expected for even the overdispersed baseline distribution. (Monte Carlo P-val. < 0.001 based on 10,000 draws.)

Without accounting for actor heterogeneity (i.e., Model BT), there appears to be a strong transitivity effect — a friend of a friend is a friend — and the Monte Carlo test confirms this with a similar P-value. However, if actor heterogeneity is accounted for, the transitivity effects vanish (simulated one-sided P-val. $= 0.43$), suggesting that the underlying social process is better explained by a relatively small number of highly social individuals whose ties to each other and to (less social) third parties create excess transitive ties for the overall amount of interaction observed. At the same time, if, instead of using (13) as the test statistic, we use a less conservative statistic of the form (14) with $v_{\text{2-path}}(x_1, x_2) = \sqrt{x_1 x_2}$ (geometric mean), $v_{\text{combine}}(x_1, \ldots, x_{n-2}) = \sum_{k=1}^{n-2} x_k$, and $v_{\text{affect}}(x_1, x_2) = \sqrt{x_1 x_2}$, the effect's significance seems to increase (one-sided P-val. $= 0.07$). However, when we attempted to fit the model with this effect, the process exhibited the degeneracy-like bimodality. This suggests that there is a trade-off between stability and power to detect subtle effects.

7 Discussion

We have generalized the exponential-family random graph models to networks whose relationships are unbounded counts, explored the issues that arise when generalizing, and proposed ways to model several common network features for count data. We demonstrated our development by analyzing two very different networks to examine the interaction of friend-of-a-friend effects with homophily and individual heterogeneity.

This paper focused on modeling counts. More generally, one can express a valued ERGM by replacing the set of possible dyad values $\mathbb{N}_0$ by a more general set $\mathbb{S}$ and replacing $h(\boldsymbol{y})$ with a more general σ-finite measure space $(\mathcal{Y}, \mathsf{Y}, P_h)$ with reference measure P_h, then postulating a probability measure $P_{\boldsymbol{\theta};P_h,\boldsymbol{\eta},\boldsymbol{g}}$ with Radon-Nikodym derivative of $P_{\boldsymbol{\theta};P_h,\boldsymbol{\eta},\boldsymbol{g}}$ with respect to P_h,

$$\frac{dP_{\boldsymbol{\theta};P_h,\boldsymbol{\eta},\boldsymbol{g}}}{dP_h}(\boldsymbol{y}) = \frac{\exp\left(\boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \boldsymbol{g}(\boldsymbol{y})\right)}{\kappa_{P_h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{\theta})},$$

(Barndorff-Nielsen, 1978, pp. 115–116; Brown, 1986, pp. 1–2) with the normalizing constant

$$\kappa_{P_h,\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{\theta}) = \int_{\mathcal{Y}} \exp\left(\boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \boldsymbol{g}(\boldsymbol{y})\right) dP_h(\boldsymbol{y}).$$

For binary and count data, and discrete data in general, P_h could be specified as a function relative to the counting measure, while for continuous data, it could be defined with respect to the Lebesgue measure. Still, as with count data, the shape of this function would need to be specified.

Other scenarios might call for more complex specifications of the reference measure. Some network data, such as measurements of duration of conversation (Wyatt et al., 2010) and international trade volumes (Westveld and Hoff, 2011) are continuous measurements except for having a positive probability of two actors not conversing at all or two countries having no measured trade. Westveld and Hoff use a normal distribution to model the log-transformed trade volume, imputing $0 = \log(1)$ for 0 observed trade volumes (all nonzero trade volumes being greater than 1 unit), and they note this issue and address it by pointing out that in their (latent-variable) model, an impact of such an outlier would be contained. Valued ERGMs may provide a more principled approach by specifying a semicontinuous P_h, such as one that puts a mass of 1/2 on 0 and 1/2 on Lebesgue measure on $(0, \infty)$.

We have also focused on data that do not impose any constraints on the sample space: $\mathcal{Y} \equiv \mathbb{S}^{\mathbb{Y}}$. But, some types of network data, such as those where each actor ranks the others (Newcomb, 1961) may be viewed as imposing a more complex constraint on sample spaces: setting $\mathbb{S} = \{1..n-1\}$ and constraining $\mathcal{Y}$ to ensure that each ego assigns a unique rank to each alter gives a sample space of permutations that could, with a counting measure, serve as the reference measure for an ERGM on rank data. These, and other applications are a subject for ongoing and future work.

This paper focuses on models for cross-sectional networks, where a single snapshot of relationship states or relationships aggregated over a time period are observed. For longitudinal data, comprising multiple snapshots of networks over the same actors over time, binary ERGMs have been used as a basis for discrete-time models for network tie evolution by Robins and Pattison (2001), Wyatt et al. (2009; 2010), Hanneke, Fu, and Xing (2010), and Krivitsky and Handcock (2010). Valued ERGMs can be directly applied to the discrete temporal ERGMs of Hanneke et al. (2010) although their adaptation to the work of Krivitsky and Handcock (2010) may be less straightforward, especially if the benefits to interpretability of the separable models are to be retained.

In practice, networks are not always observed completely. Handcock and Gile (2010) develop an approach to ERGM inference for partially observed or sampled binary networks. It would be natural to extend this approach to valued networks and valued ERGMs.

Some methods for assessing a network model's fit, particularly MCMC diagnostics (Goodreau et al., 2008a) can be used with little or no modification. Others, like the goodness-of-fit methods of Hunter, Goodreau, and Handcock (2008a) may require development of characteristics meaningful for valued networks. It may also be possible to extend the stability criteria of Schweinberger (2011) to models with infinite sample spaces.

Acknowledgments

The author thanks Mark S. Handcock for helpful discussions and comments on early drafts; Stephen E. Fienberg for his feedback on this manuscript; and Michael Schweinberger, David R. Hunter, Tom A. B. Snijders, and Xiaoyue Niu for their comments and advice. This research was supported by Portuguese Foundation for Science and Technology Ciência 2009 Program, ONR award N000140811015, and NIH award 1R01HD068395-01.

References

Ole E. Barndorff-Nielsen. *Information and Exponential Families in Statistical Theory*. John Wiley & Sons, Inc., New York, 1978. ISBN 0471995452. 7, 12, 31

Vladimir Batagelj and Andrej Mrvar. Pajek datasets. Available at `http://vlado.fmf.uni-lj.si/pub/networks/data/`, 2006. URL `http://vlado.fmf.uni-lj.si/pub/networks/data/`. 2, 14, 28

H. Russell Bernard, Peter D. Killworth, and Lee Sailer. Informant accuracy in social network data IV: A comparison of clique-level structure in behavioral and cognitive network data. *Social Networks*, 2(3):191–218, 1979–1980. ISSN 0378-8733. doi:10.1016/0378-8733(79)90014-5. 2, 28

Julian Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B*, 36: 192–236, 1974. ISSN 0035-9246. 2

Lawrence D. Brown. *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*, volume 9 of *Lecture Notes — Monograph Series*. Institute of Mathematical Statistics, Hayward, California, 1986. ISBN 0-940600-10-2. URL `http://projecteuclid.org/euclid.lnms/1215466757`. 7, 31, 40, 42

Jana Diesner and Kathleen M. Carley. Exploration of communication networks from the Enron email corpus. In *Proceedings of Workshop on Link Analysis, Counterterrorism and Security, SIAM International Conference on Data Mining 2005*, pages 21–23, 2005. 2

Katherine Faust. Very local structure in social networks. *Sociological Methodology*, 37(1):209–256, December 2007. ISSN 1467-9531. doi:10.1111/j.1467-9531.2007.00179.x. 29

Ove Frank and David Strauss. Markov graphs. *Journal of the American Statistical Association*, 81(395):832–842, 1986. ISSN 0162-1459. 2, 22

Linton C. Freeman and S. C. Freeman. A semi-visible college: Structural effects of seven months of EIES participation by a social networks community. In M. M. Henderson and M. J. McNaughton, editors, *Electronic Communication: Technology and Impacts*, volume 52 of *AAAS Symposium*, pages 77–85, Washington, D.C., 1980. American Association for Advancement of Science. 2

Charles J. Geyer. Likelihood inference for spatial point processes. In Ole E. Barndorff-Nielsen, Wilfrid S. Kendall, and Marie-Colete N. M. van Lieshout, editors, *Stochastic Geometry: Likelihood and Computation*, volume 80 of *Monographs on Statistics and Applied Probability*, pages 79–141. Chapman & Hall/CRC Press, Boca Raton, Florida, 1999. ISBN 0-8493-0396-6. 10

Charles J. Geyer and Elizabeth A. Thompson. Constrained Monte Carlo maximum likelihood for dependent data (with discussion). *Journal of the Royal Statistical Society. Series B*, 54(3):657–699, 1992. ISSN 0035-9246. 9

Ana Goldenberg, Alice X. Zheng, Stephen E. Fienberg, and Edoardo M. Airoldi. A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2(2):129–233, 2009. URL `http://arxiv.org/abs/0912.5410v1`. 1

Steven M. Goodreau, Mark S. Handcock, David R. Hunter, Carter T. Butts, and Martina Morris. A `statnet` tutorial. *Journal of Statistical Software*, 24(9):1–26, May 2008a. ISSN 1548-7660. URL `http://www.jstatsoft.org/v24/i09`. 12, 27, 29, 32

Steven M. Goodreau, James Kitts, and Martina Morris. Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks. *Demography*, 45(1):103–125, February 2008b. ISSN 0070-3370. 1, 4, 25

Mark S. Handcock. Assessing degeneracy in statistical models of social networks. Working Paper 39, Center for Statistics and the Social Sciences, University of Washington, Seattle, WA, December 2003. URL `http://www.csss.washington.edu/Papers/`. 11

Mark S. Handcock. Statistical exponential-family models for signed networks. Unpublished manuscript, 2006. 3

Mark S. Handcock and Krista J. Gile. Modeling social networks from sampled data. *Annals of Applied Statistics*, 4(1):5–25, 2010. ISSN 1932-6157. doi:10.1214/08-AOAS221. 32

Mark S. Handcock, David R. Hunter, Carter T. Butts, Steven M. Goodreau, Pavel N. Krivitsky, and Martina Morris. *`ergm`: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks*. Seattle, WA, 2010. URL `http://CRAN.R-project.org/package=ergm`. Version 2.2-6. Project home page at `http://statnetproject.org`. 10

Steve Hanneke, Wenjie Fu, and Eric P. Xing. Discrete temporal models of social networks. *Electronic Journal of Statistics*, 4:585–605, 2010. ISSN 1935-7524. doi:10.1214/09-EJS548. 32

Kathleen M. Harris, F. Florey, Joyce Tabor, Peter S. Bearman, J. Jones, and J. Richard Udry. The national longitudinal study of adolescent health: Research design. Technical report, University of North Carolina, 2003. URL `http://www.cpc.unc.edu/projects/addhealth/design/`. 2, 4

Peter D. Hoff. Bilinear mixed effects models for dyadic data. *Journal of the American Statistical Association*, 100(469):286–295, 2005. ISSN 0162-1459. 3, 15, 22

Paul W. Holland and Samuel Leinhardt. An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76(373):33–65, 1981. ISSN 0162-1459. 2, 18, 22

David R. Hunter and Mark S. Handcock. Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15(3):565–583, 2006. ISSN 1061-8600. 4, 5, 8, 9, 10, 22, 23, 27, 30

David R. Hunter, Steven M. Goodreau, and Mark S. Handcock. Goodness of fit for social network models. *Journal of the American Statistical Association*, 103(481):248–258, March 2008a. ISSN 0162-1459. doi:10.1198/016214507000000446. 32

David R. Hunter, Mark S. Handcock, Carter T. Butts, Steven M. Goodreau, and Martina Morris. `ergm`: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, 24(3):1–29, May 2008b. ISSN 1548-7660. URL `http://www.jstatsoft.org/v24/i03`. 5, 7, 10

Frank P. Kelly and Brian D. Ripley. A note on Strauss's model for clustering. *Biometrika*, 63(2):357–360, 1976. ISSN 0006-3444. 17, 19

Pavel N. Krivitsky and Mark S. Handcock. A separable model for dynamic networks. *Under review*, November 2010. URL `http://arxiv.org/abs/1011.1937`. 32

Pavel N. Krivitsky, Mark S. Handcock, Adrian E. Raftery, and Peter D. Hoff. Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks*, 31 (3):204–213, July 2009. ISSN 0378-8733. doi:10.1016/j.socnet.2009.04.001. 3, 14, 15, 22

Pavel N. Krivitsky, Mark S. Handcock, and Martina Morris. Adjusting for network size and composition effects in exponential-family random graph models. *Statistical Methodology*, 8(4):319–339, July 2011. ISSN 1572-3127. doi:10.1016/j.stamet.2011.01.005. 4, 5, 7, 15

Diane Lambert. Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1):1–14, 1992. ISSN 0040-1706. 15

Emmanuel Lazega and Philippa E. Pattison. Multiplexity, generalized exchange and cooperation in organizations: a case study. *Social Networks*, 21(1):67–90, 1999. ISSN 0378-8733. doi:10.1016/S0378-8733(99)00002-7. 1

Mahendra Mariadassou, Stéphane Robin, and Corinne Vacher. Uncovering latent structure in valued graphs: A variational approach. *Annals of Applied Statistics*, 4(2):715–742, 2010. ISSN 1932-6157. doi:10.1214/10-AOAS361. 3, 22

Peter McCullagh and John A. Nelder. *Generalized Linear Models*, volume 37 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, second edition, August 1989. ISBN 0-412-31760-5. 16, 17

Martina Morris and Mirjam Kretzschmar. Concurrent partnerships and the spread of HIV. *AIDS*, 11(5):641–648, April 1997. 1

Martina Morris, Mark S. Handcock, and David R. Hunter. Specification of exponential-family random graph models: Terms and computational aspects. *Journal of Statistical Software*, 24(4):1–24, May 2008. ISSN 1548-7660. URL `http://www.jstatsoft.org/v24/i04`. 2, 14, 19, 40

Theodore M. Newcomb. *The Acquaintance Process*. Holt, Rinehart, Winston, New York, 1961. 1, 2, 32

Philippa Pattison and Stanley Wasserman. Logit models and logistic regressions for social networks: II. Multivariate relations. *British Journal of Mathematical and Statistical Psychology*, 52(2):169–193, November 1999. ISSN 0007-1102. 3

Kenneth E. Read. Cultures of the central highlands, New Guinea. *Southwestern Journal of Anthropology*, 10(1):1–43, Spring 1954. 2

Alessandro Rinaldo, Stephen E. Fienberg, and Yi Zhou. On the geometry of discrete exponential families with application to exponential random graph models. *Electronic Journal of Statistics*, 3:446–484, 2009. ISSN 1935-7524. doi:10.1214/08-EJS350. 3, 11

Herbert Robbins and Sutton Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407, September 1951. ISSN 00034851. 10

Garry Robins and Philippa Pattison. Random graph models for temporal processes in social networks. *Journal of Mathematical Sociology*, 25(1): 5–41, 2001. 32

Garry Robins, Philippa Pattison, and Stanley S. Wasserman. Logit models and logistic regressions for social networks: III. Valued relations. *Psychometrika*, 64(3):371–394, 1999. ISSN 0033-3123. 3

Samuel F. Sampson. *A Novitiate in a Period of Change: An Experimental and Case Study of Social Relationships*. Ph.D. thesis (university micofilm, no 69-5775), Department of Sociology, Cornell University, Ithaca, New York, 1968. 2

Michael Schweinberger. Instability, sensitivity, and degeneracy of discrete exponential families. *Journal of the American Statistical Association*, 0 (0):1–10, 2011. doi:10.1198/jasa.2011.tm10747. 5, 11, 12, 23, 32

Galit Shmueli, Thomas P. Minka, Joseph B. Kadane, Sharad Borle, and Peter Boatwright. A useful distribution for fitting discrete data: Revival of the Conway–Maxwell–Poisson distribution. *Journal of the Royal Statistical Society: Series C*, 54(1):127 –142, January 2005. ISSN 1467-9876. doi:10.1111/j.1467-9876.2005.00474.x. 13, 16, 25, 40

Tom A. B. Snijders. Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2), 2002. 10

Tom A. B. Snijders, Philippa E. Pattison, Garry L. Robins, and Mark S. Handcock. New specifications for exponential random graph models. *Sociological Methodology*, 36(1):99–153, 2006. 5, 11, 22

Tom A. B. Snijders, Gerhard G. van de Bunt, and Christian E. G. Steglich. Introduction to stochastic actor-based models for network dynamics. *Social Networks*, 32(1):44–60, 2010. ISSN 0378-8733. doi:10.1016/j.socnet.2009.02.004. 22, 23

David Strauss and Michael Ikeda. Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85(409):204–212, 1990. ISSN 0162-1459. 6

Andrew C. Thomas and Joseph K. Blitzstein. Valued ties tell fewer lies: Why not to dichotomize network edges with thresholds. January 2011. URL `http://arxiv.org/abs/1101.0788`. 2

Marijtje A. J. van Duijn, Tom A. B. Snijders, and Bonne J. H. Zijlstra. p_2: a random effects model with covariates for directed graphs. *Statistica Neerlandica*, 58(2):234–254, 2004. 22

Michael D. Ward and Peter D. Hoff. Persistent patterns of international commerce. *Journal of Peace Research*, 44(2):157, 2007. 1

Anton H. Westveld and Peter D. Hoff. A mixed effects model for longitudinal relational and network data, with applications to international trade and conflict. *Annals of Applied Statistics*, 5(2A):843–872, September 2011. doi:10.1214/10-AOAS403. 2, 31

Danny Wyatt, Tanzeem Choudhury, and Jeff Bilmes. Dynamic multi-valued network models for predicting face-to-face conversations. In *NIPS-09 workshop on Analyzing Networks and Learning with Graphs*, Whistler, British

Columbia, Canada, December 2009. Neural Information Processing Systems (NIPS). 2, 3, 32

Danny Wyatt, Tanzeem Choudhury, and Jeff Blimes. Discovering long range properties of social networks with multi-valued time-inhomogeneous models. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10)*, Atlanta, Georgia, USA, July 2010. Association for the Advancement of Artificial Intelligence. 3, 23, 31, 32

Wayne W. Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33(4):452–473, 1977. ISSN 0091-7710. 2, 24, 25, 26, 28

A A sampling algorithm for a Poisson-reference ERGM

We use a Metropolis-Hastings sampling algorithm (Algorithm 1) to sample from a Poisson-reference ERGM, using a Poisson kernel with its mode at the present value of a dyad and, occasionally (with a specified probability π_0), proposing a jump directly to 0. Because, as we discuss in Section 5.2.2, counts of interactions are often zero-inflated relative to Poisson, setting $\pi_0 > 0$ can be used to speed-up mixing. For highly overdispersed distributions, a Poisson kernel may be trivially replaced by a geometric or even negative-binomial kernel.

This algorithm selects the dyad on which the jump is to be proposed at random. A possible improvement to this algorithm would be to adapt to it the tie-no-tie (TNT) proposal (Morris et al., 2008), which optimizes sampling in sparse (zero-inflated) networks by focusing on dyads which have a nonzero values.

B Non-steepness of the Conway–Maxwell–Poisson family

Expressed in its exponential-family canonical form, a random variable X with the Conway–Maxwell–Poisson distribution has the pmf

$$\Pr_{\boldsymbol{\theta};\boldsymbol{\eta},\boldsymbol{g}}(X = x) = \frac{\exp\left(\boldsymbol{\theta}_1 x + \boldsymbol{\theta}_2 \log(x!)\right)}{\kappa_{\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{\theta})},\ x \in \mathbb{N}_0$$

with the normalizing constant

$$\kappa_{\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{\theta}) = \sum_{x'=0}^{\infty} \frac{\exp\left(\boldsymbol{\theta}_1 x' + \boldsymbol{\theta}_2 \log(x'!)\right)}{\kappa_{\boldsymbol{\eta},\boldsymbol{g}}(\boldsymbol{\theta})}.$$

Theorem B.1. *The Conway–Maxwell–Poisson family is not regular.*

Proof. The natural parameter space of CMP is

$$\boldsymbol{\Theta}_{\mathrm{N}} = \{\boldsymbol{\theta}' \in \mathbb{R}^2 : \boldsymbol{\theta}_2 < 0 \vee (\boldsymbol{\theta}_2 = 0 \wedge \boldsymbol{\theta}_1 < 0)\}$$

(Shmueli et al., 2005). Due to the boundary at $\boldsymbol{\theta}_2 = 0$, $\boldsymbol{\Theta}_{\mathrm{N}}$ is not an open set, and hence the family is not regular (Brown, 1986, p. 2). □

Algorithm 1 Sampling from a Poisson-reference ERGM with no constraints, optimized for zero-inflated distributions

Let:

RandomChoose(A) return a random element of a set A

Uniform(a, b) return a random draw from the Uniform(a, b) distribution

Poisson$_{\neq y}(\lambda)$ return a random draw from the Poisson(λ) distribution, conditional on not drawing y

$p(y^*; y) = \frac{\exp\left(-\left(y+\frac{1}{2}\right)\right)\left(y+\frac{1}{2}\right)^{y^*}/y^*!}{1-\exp\left(-\left(y+\frac{1}{2}\right)\right)\left(y+\frac{1}{2}\right)^{y}/y!}$, the pmf of a Poisson$_{\neq y}(y + \frac{1}{2})$ draw

Input: $\boldsymbol{y}^{(0)} \in \mathcal{Y}$, T sufficiently large, $\mathbb{Y}$, $\boldsymbol{g}$, $\boldsymbol{\eta}$, $\pi_0 \in [0, 1)$

Output: a draw from the specified Poisson-reference ERGM

for $t \leftarrow 1..T$ **do**
  $(i, j) \leftarrow$ RandomChoose$(\mathbb{Y})$ {Select a dyad at random.}
  if $\boldsymbol{y}_{i,j} \neq 0 \wedge$ Uniform$(0, 1) < \pi_0$ **then**
    $y^* \leftarrow 0$ {Propose a jump to 0 with probability π_0.}
  else
    $y^* \leftarrow$ Poisson$_{\neq \boldsymbol{y}_{i,j}^{(t-1)}}\left(\boldsymbol{y}_{i,j}^{(t-1)}\right)$ {Propose a jump to a new value.}
  $$q \leftarrow \begin{cases} \frac{\pi_0+(1-\pi_0)p(0;y^*)}{p(y^*;0)} & \boldsymbol{y}_{i,j}^{(t-1)} = 0 \\ \frac{p(\boldsymbol{y}_{i,j}^{(t-1)};0)}{\pi_0+(1-\pi_0)p(0;\boldsymbol{y}_{i,j}^{(t-1)})} & \boldsymbol{y}_{i,j}^{(t-1)} \neq 0 \wedge y^* = 0 \\ \frac{\cancel{(1-\pi_0)}p(\boldsymbol{y}_{i,j}^{(t-1)};y^*)}{\cancel{(1-\pi_0)}p(y^*;\boldsymbol{y}_{i,j}^{(t-1)})} & \text{otherwise} \end{cases}$$
  $$r \leftarrow q \times \frac{\boldsymbol{y}_{i,j}^{(t-1)}!}{y^*!} \times \exp\left(\boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \boldsymbol{\Delta}_{i,j}^{\boldsymbol{y}_{i,j}^{(t-1)} \to y^*}\left(\boldsymbol{y}^{(t-1)}\right)\right)$$
  if Uniform$(0, 1) < r$ **then**
    $\boldsymbol{y}^{(t)} \leftarrow \boldsymbol{y}_{(i,j)=y^*}^{(t-1)}$ {Accept the proposal.}
  else
    $\boldsymbol{y}^{(t)} \leftarrow \boldsymbol{y}^{(t-1)}$ {Reject the proposal.}
return $\boldsymbol{y}^{(T)}$

Theorem B.2. *The Conway–Maxwell–Poisson family is not steep.*

Proof. A necessary and sufficient condition for a non-regular exponential family to be steep is that

$$\forall_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_{\mathrm{N}} \setminus \boldsymbol{\Theta}_{\mathrm{N}}^{\circ}} \mathrm{E}_{\boldsymbol{\theta};\boldsymbol{\eta},\boldsymbol{g}}(\|\boldsymbol{g}(X)\|) = \infty,$$

where $\boldsymbol{\Theta}_{\mathrm{N}}^{\circ}$ is the open interior of $\boldsymbol{\Theta}_{\mathrm{N}}$, and their set difference is thus the non-open boundary of the natural parameter space that is contained within it. (Brown, 1986, Proposition 3.3, p. 72) For CMP, this boundary

$$\boldsymbol{\Theta}_{\mathrm{N}} \setminus \boldsymbol{\Theta}_{\mathrm{N}}^{\circ} = \{\boldsymbol{\theta}' \in \mathbb{R}^2 : \boldsymbol{\theta}_2 = 0 \wedge \boldsymbol{\theta}_1 < 0\}.$$

There, $X \sim \mathrm{Geometric}(p = 1 - \exp(\boldsymbol{\theta}_1))$. Noting that $X \geq 0$ a.s., $\log(X!) \geq 0$ a.s., and $\log(x!) \leq (x+1)\log\left(\frac{x+1}{e}\right) + 1$,

$$\begin{aligned}
\mathrm{E}_{\boldsymbol{\theta};\boldsymbol{\eta},\boldsymbol{g}}(\|\boldsymbol{g}(X)\|) &= \mathrm{E}_{\mathrm{Geometric}(p=1-\exp(\boldsymbol{\theta}_1))}(\|[X, \log(X!)]^\top\|) \\
&\leq \mathrm{E}_{\mathrm{Geometric}(p=1-\exp(\boldsymbol{\theta}_1))}\left(X + \log(X!)\right) \\
&\leq \mathrm{E}_{\mathrm{Geometric}(p=1-\exp(\boldsymbol{\theta}_1))}\left(X + (X+1)\log\left(\frac{X+1}{e}\right) + 1\right) \\
&\leq \mathrm{E}_{\mathrm{Geometric}(p=1-\exp(\boldsymbol{\theta}_1))}\left(X + (X+1)^2 + 1\right) \\
&< \infty,
\end{aligned}$$

since the first and second moments of the geometric distribution are finite. Therefore, CMP is not steep. $\square$

Because the non-steep boundary corresponds to the most dispersed distribution that CMP can represent, maximum likelihood estimator properties for data which are highly overdispersed are not guaranteed.