On the Question of Effective Sample Size in Network Modeling: An Asymptotic Inquiry

Pavel N. Krivitsky and Eric D. Kolaczyk

Abstract. The modeling and analysis of networks and network data has seen an explosion of interest in recent years and represents an exciting direction for potential growth in statistics. Despite the already substantial amount of work done in this area to date by researchers from various disciplines, however, there remain many questions of a decidedly foundational nature—natural analogues of standard questions already posed and addressed in more classical areas of statistics—that have yet to even be posed, much less addressed. Here we raise and consider one such question in connection with network modeling. Specifically, we ask, "Given an observed network, what is the sample size?" Using simple, illustrative examples from the class of exponential random graph models, we show that the answer to this question can very much depend on basic properties of the networks expected under the model, as the number of vertices n_V in the network grows. In particular, adopting the (asymptotic) scaling of the variance of the maximum likelihood parameter estimates as a notion of effective sample size (n_{eff}), we show that when modeling the overall propensity to have ties and the propensity to reciprocate ties, whether the networks are sparse or not under the model (i.e., having a constant or an increasing number of ties per vertex, respectively) is sufficient to yield an order of magnitude difference in n_{eff}, from $O(n_V)$ to $O(n_V^2)$. In addition, we report simulation study results that suggest similar properties for models for triadic (friend-of-a-friend) effects. We then explore some practical implications of this result, using both simulation and data on food-sharing from Lamalera, Indonesia.

Key words and phrases: Asymptotic normality, consistency, mutuality, triadic closure, exponential-family random graph model, maximum likelihood.

1. INTRODUCTION

Since roughly the mid-1990s, the study of networks has increased dramatically. Researchers from across

Pavel N. Krivitsky is Lecturer, School of Mathematics and Applied Statistics and National Institute for Applied Statistics Research Australia (NIASRA), University of Wollongong, NSW 2522, Australia (e-mail: pavel@uow.edu.au; URL: http://www.krivitsky.net/research/). Eric D. Kolaczyk is Professor, Department of Mathematics and Statistics, Boston University, Boston, Massachusetts 02215, USA (e-mail: kolaczyk@bu.edu; URL: http://math.bu.edu/people/kolaczyk/).

the sciences—including biology, bioinformatics, computer science, economics, engineering, mathematics, physics, sociology and statistics—are more and more involved with the collection and statistical analysis of data associated with networks. As a result, statistical methods and models are being developed in this area at a furious pace, with contributions coming from a wide spectrum of disciplines. See, for example, the work of Jackson (2008), Kolaczyk (2009) and Newman (2010) for recent overviews from the perspective of economics, statistics and statistical physics, respectively.

A cross-sectional network is typically represented mathematically by a graph, say, $G = (V, E)$, where

V is a set of n_V vertices (commonly written $V = \{1, \ldots, n_V\}$) and E is a set of $|E|$ ties [represented as vertex pairs $(u, v) \in E$]. Ties can be either directed [wherein (u, v) is distinct from (v, u)] or undirected. Prominent examples of networks represented in this fashion include the World Wide Web graph (with vertices representing web pages and directed ties representing hyperlinks pointing from one page to another), protein–protein interaction networks in biology (with vertices representing proteins and undirected ties representing an affinity for two proteins to bind physically) and friendship networks (with vertices representing people and ties representing friendship nominations in a social survey).

A great deal of attention in the literature has been focused on the natural problem of modeling networks—of the presence and absence of their ties in particular. There is by now a wide variety of network models that have been proposed, ranging from models of largely mathematical interest to models designed to be fit statistically to data. See, for example, the sources cited above or, for a shorter treatment, the review paper by Airoldi et al. (2009). The derivation and study of network models is a unique endeavor, due to a number of factors. First, the defining aspect of networks is their relational nature, and hence the task is effectively one of modeling complex dependencies among the vertices. Second, quite often there is no convenient space associated with the network, and so the type of distance and geometry that can be exploited in modeling other dependent phenomena, like time series and spatial processes are not, generally, available when modeling networks. Finally, network problems frequently are quite large, involving hundreds if not thousands or hundreds of thousands of vertices and their ties. Since a network of n_V vertices can in principle have on the order of $O(n_V^2)$ ties, in network modeling and analysis—particularly statistical analysis of network data—the sheer magnitude of the network can be a critical factor in this area.

Suppose that we observe a network, in the form of a directed graph $G = (V, E)$, where V is a set of $n_V = |V|$ vertices and E is a set of ordered vertex pairs, indicating ties. We will focus on graphs with no self-loops: $(u, u) \notin E$ for any $u \in V$. Alternatively, we may think of G in terms of its $n_V \times n_V$ adjacency matrix Y, where $Y_{ij} = 1$, if $(i, j) \in E$, and 0, otherwise, with $Y_{ii} \equiv 0$. What is our sample size in this setting? At the opening workshop of the recent Program on Complex Networks, held in August of 2010 at the Statistical and Applied Mathematical Sciences Institute (SAMSI), in North Carolina, USA, this question in fact evoked three different responses:

(1) it is the number of unique entries in Y, that is, $n_V(n_V - 1)$;
(2) it is the number of vertices, that is, n_V; or
(3) it is the number of networks, that is, one.

Which answer is correct? And, why should it matter?

Despite the already vast literature on network modeling, to the best of our knowledge this question has yet to be formally posed much less answered. Closest to doing so are, perhaps, Frank and Snijders (1994) and Snijders and Borgatti (1999), who offer some discussion of this issue in the context of jackknife and bootstrap estimation of variance in network contexts. That this should be so is particularly curious given that the analogous questions have been asked and answered in other areas involving dependent data. In particular, the notion of an *effective sample size* has been found to be useful in various contexts involving dependent data, including survey sampling, time series analysis, spatial analysis and even genetic case–control studies (Thiébaux and Zwiers, 1984; Yang et al., 2011). Given a sample of size n in such contexts, an effective sample size—say, n_{eff}—typically is defined in connection with the variance of an estimator of interest. An understanding of n_{eff}, as a function of n, can help lend important insight into a variety of fundamental and interrelated concerns, including the precision with which inference can be done, the amount of information contributed by the data toward learning a parameter(s) and, more practically, the resources needed for data collection.

For example, in survey sampling, where nontrivial dependencies can arise through the use of complex sampling designs, n_{eff} generally is taken to be the sample size necessary under simple random sampling with replacement to obtain a variance equal to that resulting from the actual design used (e.g., Lavrakas, 2008). Alternatively, consider a simple AR(1) time series model, where $(X_t - \mu) = \phi(X_{t-1} - \mu) + Z_t$, for $|\phi| < 1$ and Z_t independent and identically distributed normal random variables, with mean zero and variance σ^2. For a sample of size n, the sample mean $\bar{X}_n$, the natural and unbiased estimator of μ, has a variance that behaves asymptotically in n like $\sigma^2/[n(1 - \phi)^2]$. Contrasting this expression with σ^2/n, corresponding to the case of independent and identically distributed X_t (i.e., equivalent to the case where $\phi \equiv 0$), the value $n_{\text{eff}} = n(1 - \phi)^2$ is sometimes interpreted as an effective sample size.

In these and similar contexts, it is often possible to show that whereas nominally the relevant (asymptotic) variance scales inversely with the sample size n, under dependency a different scaling obtains, reflecting a combination of (a) the nominal sample size n and, importantly, (b) the dependency structure in the data. Since networks are defined by relational data and, hence, consist of random variables that are inherently dependent, it seems not unreasonable to hope that we might similarly gain insight into the above question "What is the sample size?" in a network setting, with the corresponding n_{eff} expected to be some function of the number of vertices n_V, modified by characteristics of the network structure itself.

Following a similar practice in these other fields, therefore, we will interpret the scaling of the asymptotic variances of maximum likelihood estimates in a network model as an effective sample size. In this paper we provide some initial insight into the question of what is the effective sample size in network modeling, focusing on the impact of what is arguably the most fundamental of network characteristics—sparsity. A now commonly acknowledged characteristic of real-world networks is that the actual number of ties tends to scale much more like the number of vertices [i.e., $O(n_V)$] than the number of potential ties [i.e., $O(n_V^2)$]. Here we demonstrate that two very different regimes of asymptotics, corresponding to responses 1 and 2 above, obtain for maximum likelihood estimates in the context of a simple case of the popular exponential random graph models, under nonsparse and sparse variants of the models. Response 3 suggests no meaningful asymptotics other than via independent replication. These may arise in some unexpected settings, such as with discrete-time Markov models for evolution of networks over time (Hanneke, Fu and Xing, 2010; Krivitsky and Handcock, 2014, e.g.); however, we do not explore this direction here.

We will also show that the notion of regime of asymptotics relates to the notion of consistency, as it applies to networks. Krivitsky, Handcock and Morris (2011) showed, informally, that their offset model was consistent, in the sense that if the network's asymptotic regime agreed with the model, the coefficients of the nonoffset terms would converge to some asymptotic value. Although the results of Shalizi and Rinaldo (2013) suggest that consistency may be meaningless for linear ERGMs with nontrivial dependence structure, our results, both theoretical and simulated, suggest that offsets that control the asymptotic regime of the network model can produce consistency-like properties.

As a technical aside, we note that the exponential random graph models we consider here are only relatively simple versions of those commonly used in practice. We choose to work with these models because (i) they are amenable to relatively standard tools in producing the theoretical results we require, while, nevertheless, (ii) they are sufficient in allowing us to highlight in a straightforward and illustrative manner our key finding—that the question of effective sample size in network settings can in fact be expected to be nontrivial and that the answer in general is likely to be subtle, depending substantially on basic model assumptions. That such insight may be obtained already for the simplest models in this class not only speaks to the fundamental nature of our results, but also appears to be fortunate, in that it would appear that theoretical analysis of the key quantity involved in our calculations becomes decidedly more delicate when even moderately more sophisticated models are considered. We provide further comments in this direction at the end of this paper.

The rest of this paper is organized as follows. Some background and definitions are provided in Section 2. Our main results are presented in Section 3, first for the case where ties arise as independent coin flips; second, for the case in which flips corresponding to ties to and from a given pair of vertices are dependent; and, third, for the case of triadic (friend-of-a-friend) effects, which we study via simulation. We then illustrate some practical implications of our results through a simulation study in Section 4, exploring coverage of confidence intervals associated with our asymptotic arguments, and through application to food-sharing networks in Section 5, where we examine the extent to which real-world data can be found to support nonsparse versus sparse variants of our models. Finally, some additional discussion may be found in Section 6.

2. BACKGROUND

There are many models for networks. [See Kolaczyk (2009), Chapter 6, or the review paper by Airoldi et al. (2009).] The class of exponential random graph models has a history going back roughly 30 years and is particularly popular with practitioners in social network analysis. This class of models specifies that the distribution of the adjacency matrix Y follow an exponential family form, that is, $p_\theta(Y = y) \propto \exp(\theta^\top g(y))$, for

vectors θ of parameters and $g(\cdot)$ of sufficient statistics. However, despite this seemingly appealing feature, work in the last 10 years has shown that exponential random graph models must be handled with some care, as both their theoretical properties and computational tractability can be rather sensitive to model specification. See Robins et al. (2007), for example, and Chatterjee and Diaconis (2013) for a more theoretical treatment.

Here we concern ourselves only with certain examples of the simplest type of exponential random graph models, wherein the dyads (Y_{ij}, Y_{ji}) and $(Y_{k\ell}, Y_{\ell k})$ are assumed independent, for $(i, j) \neq (k, \ell)$, and identically distributed. These independent dyad models arguably have the smallest amount of dependency to still be interesting as network models. A variant of the models introduced by Holland and Leinhardt (1981), they are in fact too simple to be appropriate for modeling in most situations of practical interest. However, they are ideal for our purposes, as they allow us to quickly obtain nontrivial insight into the question of effective sample size in network modeling using relatively standard tools and arguments.

Outside of Section 3.3, the models we consider are all variations of the form

$$p_{\alpha,\beta}(Y = y) = \prod_{i<j} \frac{\exp\{\alpha(y_{ij} + y_{ji}) + \beta y_{ij} y_{ji}\}}{1 + 2e^{\alpha} + e^{2\alpha+\beta}}$$

(2.1)
$$= \left(\frac{1}{1 + 2e^{\alpha} + e^{2\alpha+\beta}}\right)^{\binom{n_V}{2}}$$
$$\cdot \exp\{\alpha s(y) + \beta m(y)\},$$

with sufficient statistics

$$(2.2) \quad s(y) \equiv \sum_{i<j}(y_{ij} + y_{ji}) \quad \text{and} \quad m(y) \equiv \sum_{i<j} y_{ij} y_{ji},$$

a so-called *Bernoulli model with reciprocity*. The parameter α governs the propensity of pairs of vertices i and j to form a tie (i, j), and the parameter β governs the tendency toward reciprocity, forming a tie (j, i) that reciprocates (i, j). This model can be motivated from the independence and homogeneity assumptions given above by an argument analogous to that of Frank and Strauss (1986) using the Hammersley–Clifford Theorem (Besag, 1974), with dependence graph being $D = \{\{(i, j), (j, i)\} : (i, j) \in V^2 \wedge i < j\}$, the set of cliques of D being $\{\{(i, j)\} : (i, j) \in V^2 \wedge i \neq j\} \cup \{\{(i, j), (j, i)\} : (i, j) \in V^2 \wedge i < j\}$, and simplifying for homogeneity.

Of interest will be both this general model and the restricted model $p_\alpha \equiv p_{\alpha,0}$, wherein $\beta = 0$ and there is

no reciprocity, and not just dyads, but individual potential ties within dyads are independent. We will refer to this latter model simply as the *Bernoulli model*. Realizations of networks from this model without and with reciprocity [holding expected tie count $s(y)$ fixed] are given in Figure 1(a) and (b), respectively.

Importantly, in both the Bernoulli model and the Bernoulli model with reciprocity, we will examine the question of effective sample size under both the original model parameterization and a reparameterisation in which parameter(s) are shifted by a value $\log n_V$. Krivitsky, Handcock and Morris (2011) introduced such shifts in an undirected context as a way of adjusting models like (2.1) for network size such that realizations with fixed α and β would produce network distributions with asymptotically constant expected mean degree, $E_{\alpha,\beta}[2s(Y)/n_V]$, for varying n_V. That is, a configuration (α, β) that would produce a typical $n_V = 100$ realization like that in Figure 1(a) would produce an $n_V = 200$ realization like that in Figure 1(d). The model's baseline asymptotic behavior is to have a constant expected density, $E_{\alpha,\beta}[2s(Y)/\{n_V(n_V - 1)\}]$, such that a parameter configuration that would produce a network like 1(a) for $n_V = 100$ would produce a network like 1(c) for $n_V = 200$.

In a directed context, "degree" of a given vertex i is ambiguous, as it can refer to the number of ties that vertex makes to others ($\sum_{j \neq i} Y_{ij}$, "outdegree"), the number of ties others make to that vertex ($\sum_{j \neq i} Y_{ji}$, "indegree"), the number of others to whom that vertex has at least one connection of either type [$\sum_{j \neq i} \max(Y_{ij}, Y_{ji})$], and the number of connections that vertex has [$\sum_{j \neq i}(Y_{ij} + Y_{ji})$]. In this work, we use either of the first two. Then, "mean degree" of Y is $s(Y)/n_V$, with mean outdegree and mean indegree trivially equal; and density is $s(Y)/\{n_V(n_V - 1)\}$.

Motivated by similar concerns, we use the presence or absence of such shifts to produce two different types of asymptotic behavior in our network model classes, corresponding to sparse (asymptotically finite mean degree) and nonsparse (asymptotically infinite mean degree) networks, respectively. Because it is widely recognized that most large real-world networks are sparse networks, this distinction is critical and, as we show below, it has fundamental implications on effective sample size and consistency.

3. MAIN RESULTS

3.1 Bernoulli Model

We first present our results for the Bernoulli model. Let p_α denote the model $p_{\alpha,0}$, as defined above, and
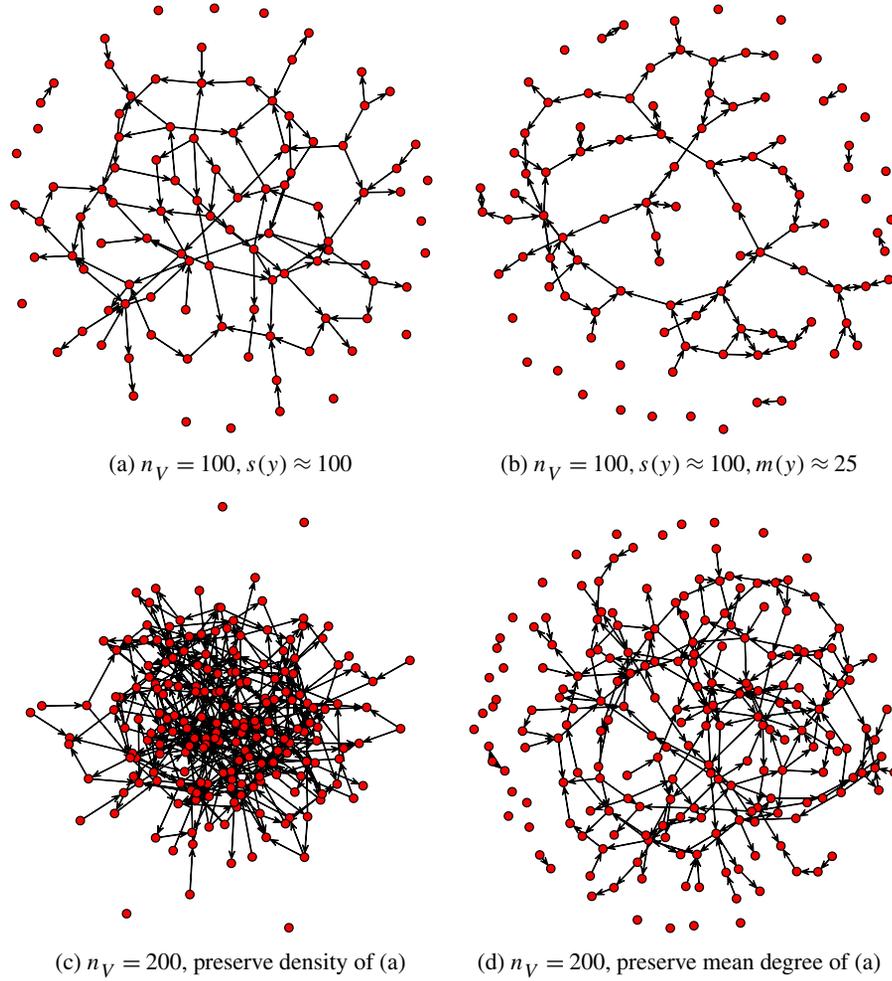
(a) $n_V = 100, s(y) \approx 100$ (b) $n_V = 100, s(y) \approx 100, m(y) \approx 25$

(c) $n_V = 200$, preserve density of (a)   (d) $n_V = 200$, preserve mean degree of (a)

FIG. 1. *Sampled networks drawn from four configurations of* (2.1). (a) *shows a realization from a model with expected mean degree* 1 *on* 100 *vertices, and no reciprocity effect.* (b) *shows a realization from a model with the same network size and mean degree as* (a), *but with reciprocity parameter β set such that the expected number of mutual ties is* 25. (c) *is a realization of the model from* (a), *scaled to* 200 *vertices, preserving density; while* (d) *preserves mean degree.*

let $p_\alpha^\dagger$ denote the same model, but under the mapping $\alpha \mapsto \alpha - \log n_V$ of the density parameter. Then, it is easy to show that under p_α the mean vertex in- and out-degree tends to infinity and the network density stays at $\text{logit}^{-1}(\alpha)$ as $n_V \to \infty$, while under $p_\alpha^\dagger$, the mean degree tends to e^α while the density tends to zero. In fact, the limiting in- and out-degree distributions tend to a Poisson law with the stated mean.

From the perspective of traditional random graph theory, the offset model of Krivitsky, Handcock and Morris (2011) is asymptotically equivalent to the standard formulation of an Erdős–Rényi random graph, in which the probability of a tie scales like e^α/n_V. Alternatively, from the perspective of social network theory, it is useful to examine the log-odds that $Y_{ij} = 1$, conditional on the status of all other potential ties. Defining $Y_{[-ij]}$ to be the network Y with tie (i, j) removed if

present, this can be expressed as

$$\text{logit } p(Y_{ij} = 1 | Y_{[-ij]} = y_{[-ij]})$$
$$\equiv \log \frac{p(Y_{ij} = 1 | Y_{[-ij]} = y_{[-ij]})}{p(Y_{ij} = 0 | Y_{[-ij]} = y_{[-ij]})}.$$

This quantity goes from being a constant value α under $p = p_\alpha$ to a value $\alpha - \log n_V$ under $p_\alpha^\dagger$. This reflects the intuition that as long as there is a cost associated with forming and maintaining a network tie, an individual will be able to maintain ties with a shrinking fraction of the network as the network grows, with the average number of maintained ties being unaffected by the growth of the network beyond a certain point (Krivitsky, Handcock and Morris, 2011).

Given the observation of a network Y randomly generated with respect to either of these models, initial

insight into the effective sample size can be obtained by studying the asymptotic behavior of the Fisher information, which we denote $\mathcal{I}(\alpha)$ and $\mathcal{I}^\dagger(\alpha)$ under p_α and $p_\alpha^\dagger$, respectively. Straightforward calculation shows that while

$$\mathcal{I}(\alpha) = \binom{n_V}{2} \frac{2e^\alpha}{(1 + e^\alpha)^2},$$

in contrast,

$$\mathcal{I}^\dagger(\alpha) = \binom{n_V}{2} \frac{2e^\alpha / n_V}{(1 + e^\alpha / n_V)^2} \approx n_V e^\alpha.$$

So $\mathcal{I}(\alpha) = O(n_V^2)$, while $\mathcal{I}(\alpha)^\dagger = O(n_V)$, a difference by an order of magnitude.

The implications of this difference are immediately apparent when we consider the asymptotic behavior of the maximum likelihood estimates of α under the two models.

THEOREM 3.1. *Let $\hat{\alpha}$ and $\hat{\alpha}^\dagger$ denote the maximum likelihood estimates of the parameter α_0 under models p_{α_0} and $p_{\alpha_0}^\dagger$, respectively, where $\alpha_0 \in [\alpha_{\min}, \alpha_{\max}]$, for finite $\alpha_{\min}, \alpha_{\max}$. Then under the model p_{α_0}, the estimator $\hat{\alpha}$ is $\binom{n_V}{2}^{1/2}$-consistent for α_0, and*

$$\binom{n_V}{2}^{1/2} (\hat{\alpha} - \alpha_0) \to N\left(0, \left\{ \frac{2e^{\alpha_0}}{(1 + e^{\alpha_0})^2} \right\}^{-1}\right),$$

while under the model $p_{\alpha_0}^\dagger$, the estimator $\hat{\alpha}^\dagger$ is $n_V^{1/2}$-consistent for α_0, and

$$\sqrt{n_V}(\hat{\alpha}^\dagger - \alpha_0) \to N(0, e^{-\alpha_0}).$$

The proof of these results uses largely standard techniques for asymptotics of estimating equations, but with a few interesting twists. Note that, for fixed n_V, the dyads (Y_{ij}, Y_{ji}) constitute $n_V(n_V - 1)/2$ independent and identically distributed bivariate random variables under both p_α and $p_\alpha^\dagger$. Consistency of the estimators in both cases can be argued by verifying, for example, the conditions of Theorem 5.9 of van der Vaart (2000) for consistency of estimating equations. Similarly, the proof of asymptotic normality of the estimators can be based on the usual technique of a Taylor series expansion of the log-likelihood and, due to the fact that we have assumed an exponential family distribution, the asymptotic normality of the sufficient statistic $s(y)$ in (2.2); however, in the case of the sparse model $p_\alpha^\dagger$, the dyads $\{(Y_{ij}, Y_{ji})\}_{i<j}$ follow a different distribution for each n_V, and therefore an array-based central limit theorem is required to show the asymptotic normality of $s(y)$. But since increasing the number of vertices from, say, $n_V - 1$ to n_V, as $n_V \to \infty$,

increases the number of dyads in our model by $n_V - 1$, a standard triangular array central limit theorem is not appropriate here. Rather, a double array central limit theorem is needed, such as Theorem 7.1.2 of Chung (2001). A full derivation is provided in the supplemental article (Krivitsky and Kolaczyk, 2014).

3.2 Bernoulli Model with Reciprocity

From Theorem 3.1 we see that the effective sample size n_{eff} in this context can be either on the order of n_V or of n_V^2, depending on the scaling of the assumed model, that is, on whether the model is sparse or not. From a nonnetwork perspective, these results can be largely anticipated by the rescaling involved, in that the transformation $\alpha \mapsto \alpha - \log n_V$ induces a rescaling of the expected number of ties by n_V^{-1}. Now, however, consider the full Bernoulli model with reciprocity, $p_{\alpha,\beta}$, defined in (2.1). Even with just two parameters the situation becomes notably more subtle.

Let $\mathcal{I}(\alpha, \beta)$ be the 2×2 Fisher information matrix under this model. Then calculations (not shown) completely analogous to those required for our previous results show that $\mathcal{I}(\alpha, \beta) = O(n_V^2)$ and, similarly, asymptotic properties of the maximum likelihood estimate of (α, β) analogous to those for p_α hold.

Let us focus then on sparse versions of $p_{\alpha,\beta}$. The offset used previously, that is, mapping α to $\alpha - \log n_V$, is not, by itself, satisfactory. Call the resulting model $p_{\alpha,\beta}^\dagger$. Standard arguments show that the limiting in- and out-degree distributions under this model will be Poisson with mean parameter e^α. On the other hand, the expected number of *reciprocated* out-ties a vertex has, $E_{\alpha,\beta}^\dagger[2m(Y)/n_V]$, behaves like $e^{2\alpha+\beta}/n_V$, and therefore tends to zero as $n_V \to \infty$. Thus, β plays no role in the limiting behavior of the model, and, indeed, reciprocity vanishes. This fact can also be understood through examination of the Fisher information matrix, say, $\mathcal{I}^\dagger(\alpha, \beta)$, in that direct calculation shows that

$$\mathcal{I}^\dagger(\alpha, \beta) = \begin{bmatrix} O(n_V) & O(1) \\ O(1) & O(1) \end{bmatrix}.$$

That is, only the information on α grows with the network. Under $p_{\alpha,\beta}^\dagger$, only the affinity parameter α can be inferred in a reliable manner.

However, the same intuition that suggests that, as the network becomes larger, a given vertex i will have an opportunity for contact with a smaller and smaller fraction of it also suggests that if there is a preexisting relationship in the form of a tie from j to i, such an opportunity likely exists for a tie from i to j regardless of how large the network may be. This, as well as

direct examination of the exact expression for the information matrix $\mathcal{I}^{\dagger}(\alpha, \beta)$, suggests that the $-\log n_V$ penalty on tie log-probability should not apply to reciprocating ties, which may be implemented by mapping $\beta \mapsto \beta + \log n_V$. Call this model, in which $p_{\alpha,\beta}^{\dagger}$ is augmented with this additional offset for β, the model $p_{\alpha,\beta}^{\ddagger}$. The corresponding conditional log-odds of a tie now have the form

$$\operatorname{logit} p_{\alpha,\beta}^{\ddagger}(Y_{ij} = 1 | Y_{[-ij]} = y_{[-ij]})$$
$$= \begin{cases} \alpha - \log n_V, & \text{if } y_{ji} = 0, \\ \alpha + \beta, & \text{if } y_{ji} = 1, \end{cases}$$

which exactly captures the intuition described.

It can be shown that under $p_{\alpha,\beta}^{\ddagger}$ we have $\mathcal{I}^{\ddagger}(\alpha, \beta) = O(n_V)$, indicating that information on both parameters grows at the same rate in n_V. It can also be shown that the limiting in- and out-degree distribution is now Poisson with mean parameter $e^{\alpha} + e^{2\alpha+\beta}$, and that $\mathrm{E}_{\alpha,\beta}^{\ddagger}[2m(Y)/n_V]$ tends to $e^{2\alpha+\beta}$. So, both parameters play a role in the limiting behavior of the model and the additional offset induces an asymptotically constant expected per-vertex reciprocity in addition to asymptotically constant expected mean degree.

Finally, we have the following analogue of Theorem 3.1.

THEOREM 3.2. *Let $(\hat{\alpha}^{\ddagger}, \hat{\beta}^{\ddagger})$ denote the maximum likelihood estimate of the parameter (α_0, β_0) under the model $p_{\alpha_0,\beta_0}^{\ddagger}$, where $(\alpha_0, \beta_0) \in [\alpha_{\min}, \alpha_{\max}] \cdot [\beta_{\min}, \beta_{\max}]$, for finite $\alpha_{\min}, \alpha_{\max}, \beta_{\min}, \beta_{\max}$. Then $(\hat{\alpha}^{\ddagger}, \hat{\beta}^{\ddagger})$ is $n_V^{1/2}$-consistent for (α_0, β_0), and*

$$\sqrt{n_V} \begin{pmatrix} \hat{\alpha}^{\ddagger} - \alpha_0 \\ \hat{\beta}^{\ddagger} - \beta_0 \end{pmatrix}$$
$$\rightarrow N \left(0, e^{-\alpha_0} \begin{bmatrix} 1 & -2 \\ -2 & 4 + 2e^{-\alpha_0 - \beta_0} \end{bmatrix} \right).$$

Proof of this theorem, using arguments directly analogous to those of Theorem 3.1, may be found in the supplemental article (Krivitsky and Kolaczyk, 2014). From the theorem we see that under the sparse model $p_{\alpha,\beta}^{\ddagger}$, as under $p_{\alpha}^{\dagger}$, the effective sample size n_{eff} is n_V.

3.3 Triadic Effects

Although there has been some work on obtaining closed-form asymptotics for ERGMs with triadic—friend-of-a-friend—effects (Chatterjee, Diaconis and Sly, 2011) or showing that they might not exist (Shalizi and Rinaldo, 2013), these results do not appear to be directly applicable to the per-capita asymptotic regimes

that we consider in this work. Therefore, in this section, we use simulation in an attempt to extend the intuition developed in Section 3.2—that reciprocating ties should not be "penalized" for the network size—to these triadic effects. For the sake of simplicity, we will consider undirected networks only.

A tie between i and k and a tie between k and j—that is, that i knows k and k knows j—should create a preexisting relationship between i and j. That is, k can "introduce" i and j regardless of how large the network is otherwise. Thus, given $i - k - j$ relationships, a potential relationship between i and j should not be penalized for network size (though $i - k - j$ themselves are); and more such two-paths (i.e., $i - k' - j$) should have no further effect on this penalty. This suggests an offset on the statistic called the *transitive ties* (Snijders, van de Bunt and Steglich, 2010, equation 8) or, equivalently, Geometrically-Weighted Edgewise Shared Partners (GWESP) (Morris, Handcock and Hunter, 2008) with parameter α fixed at 0, that is,

$$(3.1) \qquad t(y) = \sum_{i<j} y_{ij} \max_k(y_{ik}, y_{jk}).$$

Unlike the more familiar count of the number of triangles ($\sum_{i<j<k} y_{ij} y_{ik} y_{jk}$), $t(y)$ only considers whether a two-path between i and j exists, not how many of them there are. [This also makes it far less prone to ERGM degeneracy (Schweinberger, 2011).]

Consider the following model, with tie count and transitive tie count (3.1):

$$p_{\alpha,\gamma}(Y = y)$$
$$(3.2) \qquad \propto \exp\{-\log(n_V)(\alpha^{\star} s(y) - \gamma^{\star} t(y)) + \alpha_0 s(y) + \gamma_0 t(y)\}.$$

As with $p_{\alpha}^{\dagger}$, the coefficient on s is penalized by network size, in the form of $\log(n_V)\alpha^{\star}$, with $\alpha^{\star}$ being 1 in $p_{\alpha}^{\dagger}$. However, the penalty is then partially negated by increasing the coefficient on t by $\log(n_V)\gamma^{\star}$. This means that, on a sparse network,

$$\operatorname{logit} p_{\alpha,\gamma}(Y_{ij} = 1 | Y_{[-ij]} = y_{[-ij]})$$
$$\approx \begin{cases} \alpha_0 - \alpha^{\star} \log n_V, & \text{if } \neg \exists_{k \neq i,j} y_{ik} y_{kj} = 1, \\ \alpha_0 + 3\gamma_0 - (\alpha^{\star} - 3\gamma^{\star}) \log n_V, \\ \qquad\qquad\qquad \text{if } \exists_{k \neq i,j} y_{ik} y_{kj} = 1. \end{cases}$$

This approximation holds because on an otherwise empty network having ties (i, k) and (k, j), adding a tie (i, j) creates not one but three transitive ties, by making all three of the ties transitive, leading to the coefficient of 3 on the γ's. However, as the network becomes

more dense, this ceases to hold exactly, because (i, k) and/or (k, j) may already be transitive when (i, j) is added, so only two or one transitive tie might be created.

Therefore, on a sufficiently sparse network (i.e., sufficiently large n_V for a given mean degree), in order to cancel the network size penalty for a tie (i, j) but retain it for (i, k) and (k, j), $\gamma^\star = \alpha^\star/3$. With $\alpha^\star = 1$ per the same reasoning as before, this means that our heuristic suggests that $\gamma^\star \approx 1/3$. We verify this empirically as follows. Define $t'(y)$—per-capita transitive ties—as

$$t'(y) = \frac{1}{n_V} \sum_i \sum_{j \neq i} y_{ij} \max_{k \neq i, j} (y_{ik} y_{jk}).$$

In other words, for each vertex i, the number of its neighbors who have ties to at least one other neighbor of i is counted, and the resulting measures averaged over all vertices in the network. It can be shown easily that $t'(y) \equiv 2t(y)/n_V$.

We can then ask if there exist constant values of $\alpha^\star$ and $\gamma^\star$ that produce stable mean degree $(2s(y)/n_V)$ and stable per-capita transitive ties $(t'(y))$. We constructed a series of 40 networks, sized from 100 to 12,000, logarithmically spaced, for each of three configurations of mean degree, 6, 9 and 12; and two levels of per-capita transitivity for each: $1/2$ of the mean degree and $1/4$ of the mean degree. (This was done because per-capita transitivity cannot exceed the mean degree.) For each combination of n_V, $2s/n_V$ and $t'/(2s/n_V)$, we used simulated annealing to construct a network y with these statistics, and then fit an ERGM $p_{\alpha, \gamma}$ (without offsets) to it to obtain point estimates for what is effectively $(-\log(n_V)\alpha^\star + \alpha_0, \log(n_V)\gamma^\star + \gamma_0)$. The calculations were performed using the `ergm` package (Handcock et al., 2014; Hunter et al., 2008) for the R computing environment (R Core Team, 2013).

We show the results in Figure 2. Our intuition seems to be confirmed, to the extent that our network sizes are sufficiently large to confirm or disconfirm it. The trend in both parameter estimates appears to become more linear (in $\log n_V$) as n_V increases, suggesting that unique $\alpha^\star$ and $\gamma^\star$ exist. For $\hat{\alpha}$, the asymptotic slope (i.e., $-\alpha^\star$) is very close to -1 regardless of the mean degree and the amount of transitivity, and for $\hat{\alpha}$, the slope (i.e., $\gamma^\star$) decreases as $\log n_V$ increases, though it does not quite obtain the exact value of $1/3$ for the network sizes considered. [Considering only $n_V >$ 5,000, $2s(y)/n_V = 6$, and $t'(y)/(2s(y)/n_V) = 1/4$—the fastest-converging configuration—gave the slope of 0.35.]

Notably, even though given a particular value of the sufficient statistic $(s(y), t(y))$, the natural parameters (α, γ) would be determined exactly, we have to use Monte Carlo MLE (Hunter and Handcock, 2006) to estimate them, so there is some noise in the point estimates.

Overall, it appears that the coefficients of sparser networks with weaker transitivity tend to approach linearity faster. Thus, we performed a follow-up simulation study, this one with mean degree 2, transitivity proportion $1/8$ and 40 values of n_V between 10,000 and 40,000, logarithmically spaced.

Based on all of the values of n_V considered, $\hat{\alpha}^\star = 1.00037$ [95% CI: $(1.00029, 1.00044)$] and $\hat{\gamma}^\star = 0.3377$ [95% CI: $(0.3369, 0.3386)$], closer to the theoretical values of 1 and $1/3$ than the smaller network sizes. The confidence intervals do not include the theoretical values, but we would not expect the asymptotic values to be attained for any finite network size. Indeed, there is evidence of nonlinearity in that range [P-value of predictor $\log(n_V)^2$ term is <0.0001 for the $\alpha^\star$ response and 0.04 for the $\gamma^\star$ response, with negative coefficient for both]. Furthermore, fitting only the 20 data points with $n_V > 20,000$ produces $(\hat{\alpha}^\star, \hat{\gamma}^\star) = (1.000072, 0.3347)$, and fitting only the 10 data points with $n_V > 29,000$, $(\hat{\alpha}^\star, \hat{\gamma}^\star) = (1.00030, 0.3334)$.

This very strongly suggests meaningful and interpretable asymptotic behavior for triadic closure ERGM terms as well. In particular, the asymptotic linearity with a known coefficient suggests a form of consistency for "intercepts" α_0 and γ_0, as it is they that control the asymptotic mean degree and per-vertex amount of triadic closure in (3.2).

To relate this to the notion of effective sample size n_{eff} used earlier, defined through the scaling of the information matrix $\mathcal{I}(\alpha, \gamma)$, we simulated the sufficient statistics from the above-described fits. For an exponential family, the variance–covariance matrix of sufficient statistics under the MLE approximates the information matrix (Hunter and Handcock, 2006, equation 3.5, e.g.). We find that the entries of $\hat{\mathcal{I}}(\hat{\alpha}, \hat{\gamma})/n_V = \text{Var}_{\hat{\alpha}, \hat{\gamma}}([s(Y), t(Y)])/n_V$ do not exhibit any trend at all as a function of n_V, for fixed mean degree and per-vertex transitivity. [In particular, for a linear trend, P-values are 0.31, 0.49 and 0.41 for $\text{Var}(s(Y))/n_V$, $\text{Var}(t(Y))/n_V$ and $\text{Cov}(s(Y), t(Y))/n_V$, respectively. Exploratory plots do not show any pattern, except for greater variability in estimates of variance for higher n_V.] This strongly suggests that the asymptotics of the model (3.2) have an effective sample size n_{eff} of n_V as well.
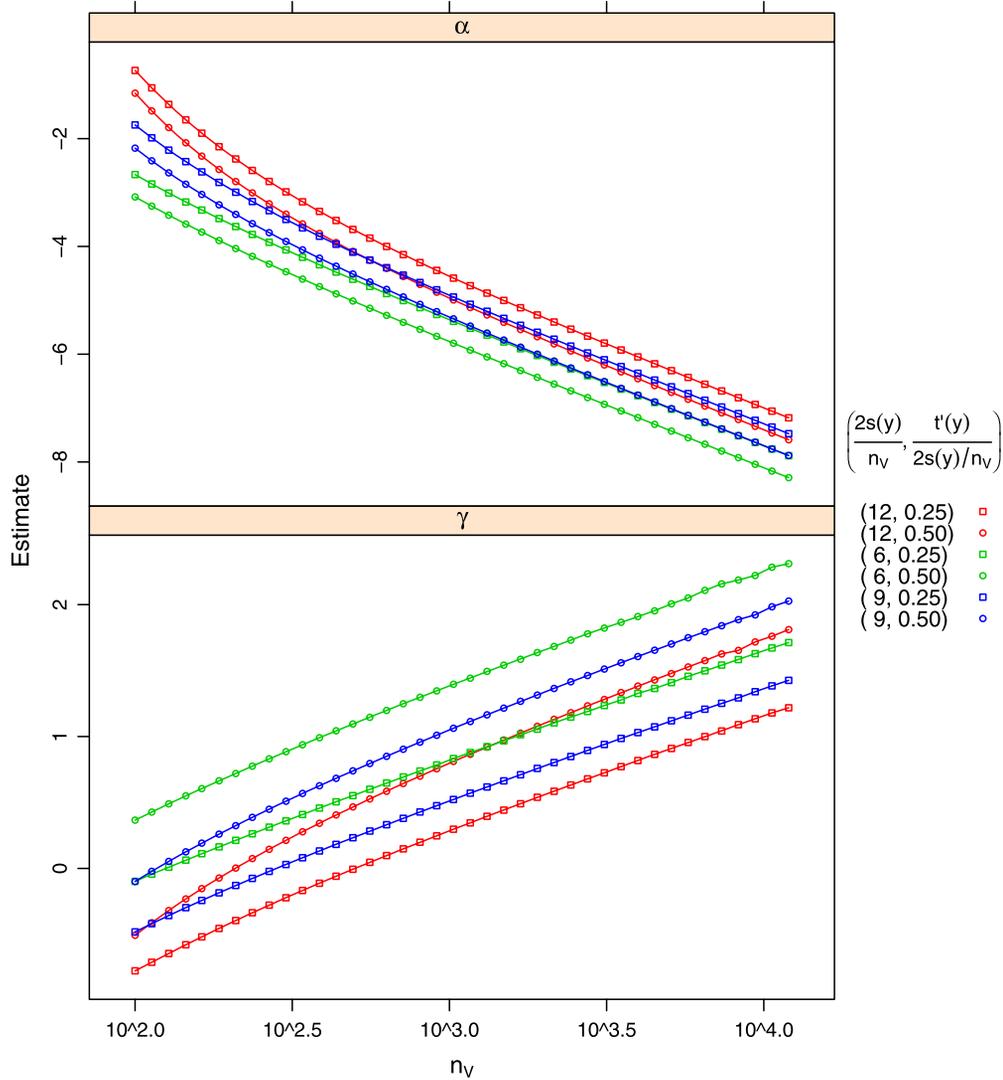
FIG. 2. *Maximum likelihood estimates from fitting $p_{\alpha,\gamma}$ to networks with a variety of sizes, densities (distinguished by color) and levels of transitivity (distinguished by plotting symbol). Note that the horizontal axis is plotted on the logarithmic scale.*

4. COVERAGE OF WALD CONFIDENCE INTERVALS

Our asymptotic arguments in Section 3 were developed primarily for the purpose of establishing the scaling associated with the asymptotic variance, so as to provide insight into the question of effective sample size—our main focus here. However, the asymptotically normal distributions we have derived are of no little independent interest themselves, as they serve as a foundation for doing formal inference on the model parameters in practice. By way of illustration, here we explore their use for constructing confidence intervals, particularly those based on Theorem 3.2: under a model $p_{\alpha,\beta}^{\ddagger}$, the Wald confi-

dence intervals using plug-in estimators for the standard errors are $\hat{\alpha}^{\ddagger} \pm z_{(1-\mathrm{CL})/2}^{*}\sqrt{e^{-\hat{\alpha}^{\ddagger}}/n_V}$ for α and $\hat{\beta}^{\ddagger} \pm z_{(1-\mathrm{CL})/2}^{*}\sqrt{e^{-\hat{\alpha}^{\ddagger}}(4+2e^{-\hat{\alpha}^{\ddagger}-\hat{\beta}^{\ddagger}})/n_V}$ for β.

Because our asymptotics are in $n_V = |V|$, we examine a variety of network sizes. The desired asymptotic properties of the network are expressed in terms of the per-capita mean value parameters—$\mathrm{E}_{\alpha,\beta}^{\ddagger}[s(Y)/n_V]$ and $\mathrm{E}_{\alpha,\beta}^{\ddagger}[2m(Y)/n_V]$. We study two configurations:

(1) $(\mathrm{E}_{\alpha,\beta}^{\ddagger}[s(Y)/n_V], \mathrm{E}_{\alpha,\beta}^{\ddagger}[m(Y)/n_V]) = (1, 0.25)$
and

(2) $(\mathrm{E}_{\alpha,\beta}^{\ddagger}[s(Y)/n_V], \mathrm{E}_{\alpha,\beta}^{\ddagger}[m(Y)/n_V]) = (1, 0.40)$.

In other words, the expected mean outdegree is set to 1, and expected numbers of *out-ties* that are reciprocated are $0.25 \cdot 2 = 0.5$ and $0.40 \cdot 2 = 0.8$ per vertex, respectively. These represent two levels of mutuality, though note that even (1) represents substantial mutuality, especially for larger networks.

For each $n_V = 10, 15, 20, \ldots, 200$, we estimate the natural parameters of the model $p^{\ddagger}_{\alpha,\beta}$ corresponding to the desired mean value parameters, and then simulate 100,000 networks from each configuration, evaluating the MLE and constructing a Wald confidence interval at each level of the customary 80%, 90%, 95% and 99%, for α and for β (individually), checking the coverage.

For some of the smaller sample sizes, the simulated network statistics for some realizations were not in the interior of their convex hull (Barndorff-Nielsen, 1978, Theorem 9.13, page 151). That is, their values were the maximal or minimal possible: $s(y) = 0$ (empty graph), $s(y) = n_V(n_V - 1)$ (complete graph), $m(y) = 0$ (no ties reciprocated), and/or $m(y) = s(y)/2$ (every extant tie reciprocated). For those, the MLE did not exist. [For (1), the fraction was 8.2% for $n_V = 10$ and none of the 100,000 realizations had no MLE for $n_V \geq 55$. For (2), it was 14.2% for $n_V = 10$ and none of the realizations had no MLE for $n_V \geq 65$.]

Our results are conditional on the MLE existing. From the frequentist perspective, one might argue that if the MLE did not exist for a real data set, we would not have reported that type of confidence interval, so it should be excluded from the simulation as well.

We report coverages for selected network sizes in Table 1 and provide a visualization in Figure 3. Overall, the 80% coverage appears to be varied—and not very conservative—while higher levels of confidence appear to be more consistently conservative, particularly for estimates of β. Coverage for α appears to oscillate as a function of network size. This is particularly noticeable for the lower confidence levels and stronger mutuality (2). Tendency of a confidence interval for a binomial proportion to oscillate around the nominal level is a known phenomenon (Brown, Cai and DasGupta, 2001, 2002, and others), though it is interesting to note that it appears to be more prominent for the density, rather than mutuality, parameter and that it appears to be stronger for stronger mutuality.

5. EXAMPLE: FOOD-SHARING NETWORKS IN LAMALERA

While the results of Section 3 are important in establishing how closely the question of effective sample size in network modeling is tied to the structural property of (non)sparseness expected of the networks modeled, there remains the important practical question of establishing in applications just which model (i.e., sparse or nonsparse) is most appropriate. While a full and detailed study of this question is beyond the scope of this work, we present here an initial exploration.

Note that, in exploring this question, we face a problem similar to that pointed out by Krivitsky, Handcock and Morris (2011): it requires a collection of closed networks of a variety of sizes yet substantively similar social structure. Furthermore, our results are limited to modeling density and reciprocity, so the networks should be well approximated by this model. Here, we use data collected by Nolin (2010), in which each of 317 households in Lamalera, Indonesia was asked to list the households to whom they have given and households from whom they have received food in the preceding season. Lamalera is split, administratively, into two villages, which are further subdivided into wards, and then into neighborhoods. Nolin (2010) fit several ERGMs to the network, finding that distance between households had a significant effect on the propensity to share, as did kinship between members of the households involved. Nolin also found a significant positive mutuality effect.

In our study, we make use of the geographic effect by constructing a series of 24 overlapping subnetworks, consisting of Lamalera itself, its 2 constituent villages, 6 wards and 15 neighborhoods, with network sizes ranging from 12 to 317. We then fit the baseline model $p_{\alpha,\beta}$ to each network. If $p_{\alpha,\beta}$ is the most realistic asymptotic regime for these data, we would expect estimates $\hat{\alpha}$ and $\hat{\beta}$ to have no relationship to $\log n_V$ for the corresponding network. If $p^{\dagger}_{\alpha,\beta}$ is the most realistic, we would expect no relationship between $\log n_V$ and $\hat{\beta}$, but an approximately linear relationship with $\hat{\alpha}$, with slope around -1. Last, if $p^{\ddagger}_{\alpha,\beta}$ is the most realistic, we would expect the slope of the relationship between $\log n_V$ and $\hat{\alpha}$ to be around -1 and between $\log n_V$ and $\hat{\beta}$ to be around $+1$.

The estimated coefficients and the slopes are given in Figure 4. The results are suggestive. The relationship between $\hat{\alpha}$ and $\log n_V$ is clearly negative, while the relationship between $\hat{\beta}$ and $\log n_V$ is clearly positive, and the magnitudes of both slopes are closer to 1 than to 0 (although both are far from equaling 1). Overlap between the subnetworks induces dependence among the coefficients, so it is not possible to formally

Simulated Theorem 3.2 confidence interval coverage levels for selected network sizes and two levels of reciprocity: lower (1) and higher (2)

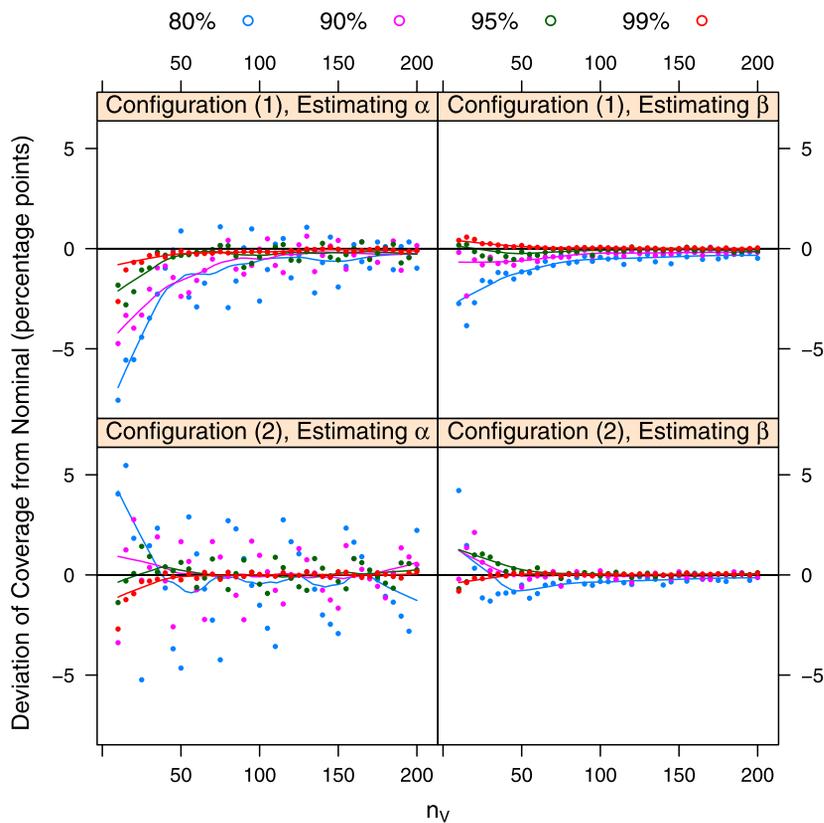|		Coverage							
		80.0%		**90.0%**		**95.0%**		**99.0%**	
	n_V	α	β	α	β	α	β	α	β
(1)	10	72.4%	77.3%	85.3%	89.8%	93.2%	95.2%	96.4%	99.4%
	20	74.5%	77.3%	86.0%	89.4%	92.9%	94.9%	98.3%	99.5%
	50	80.9%	78.8%	87.6%	89.4%	94.7%	94.8%	98.9%	99.2%
	100	77.4%	79.6%	90.0%	90.0%	94.6%	94.9%	98.9%	99.1%
	200	79.0%	79.5%	90.1%	89.8%	94.9%	94.9%	98.9%	99.0%
(2)	10	84.0%	84.2%	86.6%	89.8%	93.6%	94.3%	96.3%	98.2%
	20	81.8%	80.3%	92.8%	92.1%	95.1%	96.0%	98.1%	98.8%
	50	75.3%	79.5%	91.7%	89.4%	95.6%	95.1%	98.8%	99.0%
	100	78.5%	79.7%	91.0%	90.2%	94.5%	94.9%	99.0%	99.1%
	200	82.2%	79.9%	90.5%	89.9%	95.3%	95.1%	99.2%	99.1%

FIG. 3.   *Scatterplot of differences between simulated coverage and nominal coverage for the two configurations studied, as a function of network size n_V. Color denotes the nominal coverage levels, and smoothing lines have been added. Note that the differences are differences in percentage points* (simulated % − nominal %), *not percent differences* ($\frac{\text{simulated \% − nominal \%}}{\text{nominal \%}} \cdot 100\%$).
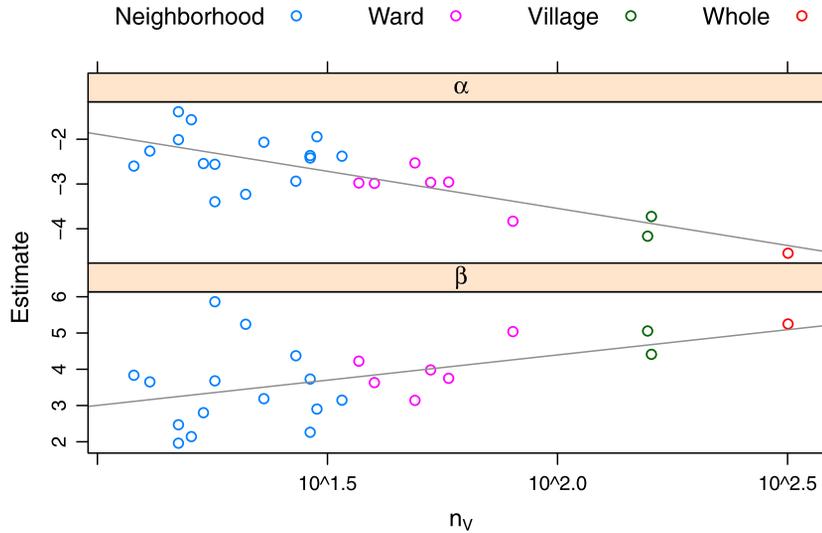
FIG. 4.   *Maximum likelihood estimates from fitting $p_{\alpha,\beta}$ to each subdivision of the Lamalera food-sharing network. Note that the horizontal axis is plotted on the logarithmic scale. Colors indicate subdivision type. The least-squares coefficients from regressing $\hat{\alpha}$ and $\hat{\beta}$ on $\log n_V$ are -0.72 and $+0.60$, respectively.*

test or estimate how significant this difference is. Nevertheless, the preponderance of evidence is that $p_{\alpha,\beta}^{\ddagger}$ is the best of the three considered. That is, a sparse model that does not enforce sparsity on reciprocating ties appears to be preferable here.

A possible explanation for why the magnitudes of the slopes are substantially less than 1 is that both the argument of Krivitsky, Handcock and Morris (2011) and our argument in Section 3.2 rely on the assumption that the network is closed—no relationships of interest are to or from vertices outside of the network observed—or, at least, that the stable mean degree and per-capita reciprocity are for the ties within it. However, while there is likely to be very little food sharing out of or into Lamalera, and relatively little between the two villages it comprises (7% of all food-sharing ties in the network are between villages), there is more sharing between the wards (28% are between wards), and even more between neighborhoods (44%). Thus, the closed-network assumption is violated. (The respective between-subdivision percentages for reciprocated ties are 6%, 22% and 39%.) When each of the subdivisions of the network is considered in isolation, these ties are lost, so the smaller subdivisions appear, to the model, to have smaller mean degree and per-capita mutuality. (See Figure 5.) This, in turn, means that smaller subdivisions have a decreased $\hat{\alpha}$ (increasing the slope for it in Figure 4) and, because mutual ties suffer less of this "attrition" than ties do overall, the $\hat{\beta}$, after adjusting for the decreased $\hat{\alpha}$, is increased for smaller networks, thus

reducing the slope for $\hat{\beta}$ in Figure 4. It is not unlikely that this pattern will hold in any network with an unobserved spatial structure, whose subnetworks of interest are contiguous regions in this space.

6. DISCUSSION

Unlike conventional data, network data typically do not have an unambiguous notion of sample size. The theoretical developments and the examples we have presented show that the effective sample size n_{eff} associated with a network depends strongly on the model assumed for how the network scales. In particular, in the case of reciprocity, whether or not the model for scaling takes into account the notion of the preexisting relationship affects whether reciprocity is even meaningful for large networks. In the case of triadic effects, a similar notion—along with the intuition that as the network size changes each individual's view of triadic closure should not—implies a specific scaling regime which, in turn, implies a specific notion of the effective sample size.

The models we study here are relatively simple examples of network models. However, with reciprocity, our work includes an important aspect that already allows us a glimpse beyond the treatments of, say, Chatterjee, Diaconis and Sly (2011) and Rinaldo, Petrović and Fienberg (2013), for so-called beta models, where the dependency induced here by reciprocity is absent. In addition, the results for reciprocity suggest that the effective modeling of triadic (e.g., friend
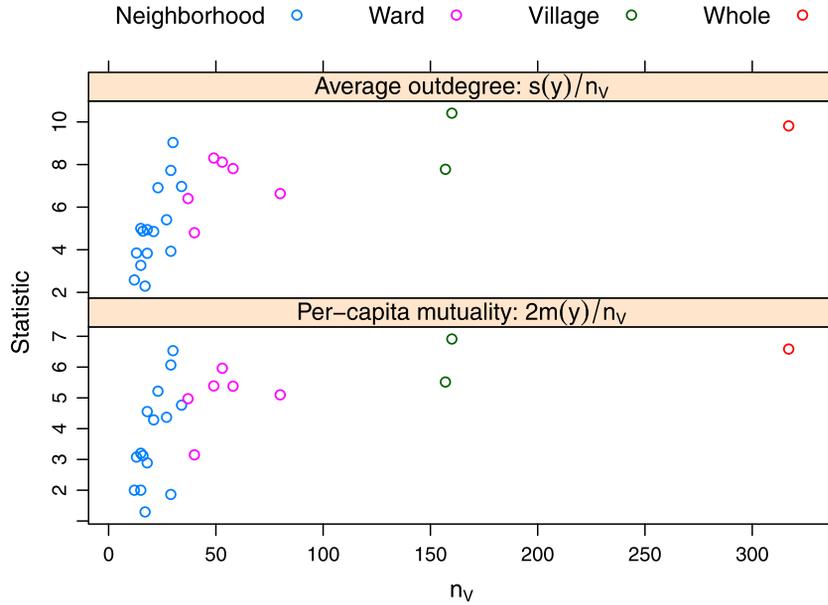
FIG. 5.   *Per-capita network statistics as a function of n_V. Colors indicate subdivision type. Note that the larger subdivisions have more within-subdivision ties.*

of a friend of a friend) effects in a manner indexed to network size requires a more complex treatment yet. However, our simulation shows, perhaps somewhat surprisingly, that if triadic closure is considered on a per-capita basis, effective sample size ultimately behaves similarly to the way it does in the simpler cases.

We note that asymptotic theory supporting methods for the construction of confidence intervals for network parameters is only beginning to emerge. The most traction appears to have been gained in the context of stochastic block models (e.g., Bickel and Chen, 2009; Choi, Wolfe and Airoldi, 2012; Celisse, Daudin and Pierre, 2012 and Rohe, Chatterjee and Yu, 2011), although progress is beginning to be had with exponential random graph models as well (e.g., Chatterjee, Diaconis and Sly, 2011; Chatterjee and Diaconis, 2013 and Rinaldo, Petrović and Fienberg, 2013). Most of these works present consistency results for maximum likelihood and related estimators, with the exception of Bickel and Chen (2009), which also includes results on asymptotic normality of estimators. Our work contributes to this important but nascent area with both our theoretical developments and our simulation studies. In particular, the asymptotic regime of $p_{\alpha,\gamma}$ is one that neither appears to become degenerate nor approaches Erdős–Rényi.

The lack of an established understanding of the distributional properties of parameter estimates in commonly used network models is particularly unfortunate

given that a number of software packages now allow for the easy computation of such estimates. For example, packages for computing estimates of parameters in fairly general formulations of exponential random graph models routinely report both estimates and, ostensibly, standard errors, where the latter are based on standard arguments for exponential families. Unfortunately, practitioners do not always seem to be aware that the use of these standard errors for constructing normal-theory confidence intervals and tests is lacking fully formal justification. From that perspective, our work appears to be one of the first to begin laying the necessary theoretical foundation to justify practical confidence interval procedures in exponential random graph models. See Haberman (1981) for another contribution in this direction, proposed as part of the discussion of the original paper of Holland and Leinhardt (1981).

In order to successfully build upon our work, and extend our results to more sophisticated instances of exponential random graph models, certain technical challenges must be overcome. First, we note that our notion of effective sample size is tied directly to the asymptotic behavior of the Fisher information matrix of our model [denoted $\mathcal{I}(\theta)$ in the supplemental article (Krivitsky and Kolaczyk, 2014)]. Given that exponential random graph models are, by definition, of exponential family form, this information matrix is in principle given by the matrix of partial second derivatives of the cumulant generating function [denoted ψ in

the supplemental article, so that $\mathcal{I} = \partial^2 \psi(\theta)/\partial\theta\,\partial\theta^\top$].
Due to the use of independent dyads in our theoretical
work (i.e., our models are variations on Bernoulli models), the corresponding likelihoods factor over dyads,
and hence the information matrices are simply proportional to powers of n_V (i.e., linear or quadratic). This
is in analogy to the canonical setting of independent
and identically distributed observations. However, in
more general settings beyond the case of independent
dyads—including even the models with triadic effects
we studied in simulation—the likelihood cannot be expected to factor in such a simple manner. Hence, the
analysis of the Fisher information promises to be decidedly more subtle. In fact, there appears to be almost no work to date studying this matrix in any detail. To the best of our knowledge, the only such work
is the recent manuscript by Pu et al. (2013), introducing a deterministic approach to approximating this matrix (stochastic approximations may, of course, be produced using MCMC) based on a lower bound of the cumulant generating function. This bound, however, has
only an implicit representation.

Second, in the case of more general exponential random graph models than those studied here, there will
be a need for a correspondingly more sophisticated
central limit theorem, in order to produce results on
asymptotic normality analogous to those we present
for the simpler models we study. Even for our models,
the tool we used was somewhat nonstandard, in that
we required a double-array central limit theorem. The
more general case will require a central limit theorem
capable of handling the nontrivial global dependencies
induced by effects even as seemingly simple as triadic
closure or the like. Progress on the first point above is a
likely prerequisite to understanding the nature of these
dependencies sufficiently well to know just what sort
of central limit theorem is required.

Finally, there is, as always with exponential random
graph models, the issue of instability and degeneracy
that must be kept in mind (e.g., Handcock, 2003 and
Chatterjee, Diaconis and Sly, 2011). It has been discovered only relatively recently that substantial care
must be taken in specifying network effects in exponential random graph models. Without such care, it
is possible to produce models for which the corresponding distributions turn out to be near-degenerate
and, in turn, the estimation of parameters highly unstable. Schweinberger (2011) has recently shed important light on this issue, showing that instability
and degeneracy are related to the scaling of the linear
term in exponential family distributions generally and,

more specifically, in exponential random graph models. These scaling results can be expected to have implications on the role that scaling necessarily plays in
the types of calculations we have presented here.

SUPPLEMENTARY MATERIAL

**Supplement to "On the Question of Effective
Sample Size in Network Modeling: An Asymptotic
Inquiry"** (DOI: 10.1214/14-STS502SUPP; .pdf). This
document contains proofs of the results reported in the
body of the article.

REFERENCES

AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P.
(2009). A survey of statistical network models. *Found. Trends
Mach. Learn.* **2** 129–233.

BARNDORFF-NIELSEN, O. (1978). *Information and Exponential
Families in Statistical Theory*. Wiley, Chichester. MR0489333

BESAG, J. (1974). Spatial interaction and the statistical analysis of
lattice systems. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **36** 192–
236. MR0373208

BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of
network models and Newman–Girvan and other modularities.
Proc. Natl. Acad. Sci. USA **106** 21068–21073.

BROWN, L. D., CAI, T. T. and DASGUPTA, A. (2001). Interval
estimation for a binomial proportion. *Statist. Sci.* **16** 101–133.
MR1861069

BROWN, L. D., CAI, T. T. and DASGUPTA, A. (2002). Confidence
intervals for a binomial proportion and asymptotic expansions.
Ann. Statist. **30** 160–201. MR1892660

CELISSE, A., DAUDIN, J.-J. and PIERRE, L. (2012). Consistency of maximum-likelihood and variational estimators in
the stochastic block model. *Electron. J. Stat.* **6** 1847–1899.
MR2988467

CHATTERJEE, S. and DIACONIS, P. (2013). Estimating and understanding exponential random graph models. *Ann. Statist.* **41**
2428–2461. MR3127871

CHATTERJEE, S., DIACONIS, P. and SLY, A. (2011). Random
graphs with a given degree sequence. *Ann. Appl. Probab.* **21**
1400–1435. MR2857452

CHOI, D. S., WOLFE, P. J. and AIROLDI, E. M. (2012). Stochastic
blockmodels with a growing number of classes. *Biometrika* **99**
273–284. MR2931253

CHUNG, K. L. (2001). *A Course in Probability Theory*, 3rd ed. Academic Press, Inc., San Diego, CA. MR1796326

FRANK, O. and SNIJDERS, T. A. B. (1994). Estimating the size of hidden populations using snowball sampling. *J. Official Statistics* **10** 53–67.

FRANK, O. and STRAUSS, D. (1986). Markov graphs. *J. Amer. Statist. Assoc.* **81** 832–842. MR0860518

HABERMAN, S. J. (1981). An exponential family of probability distributions for directed graphs: Comment. *J. Amer. Statist. Assoc.* **76** 60–61.

HANDCOCK, M. S. (2003). Assessing degeneracy in statistical models of social networks. Technical Report No. 39, Center for Statistics and the Social Sciences, Univ. Washington, Seattle, WA.

HANDCOCK, M. S., HUNTER, D. R., BUTTS, C. T., GOODREAU, S. M., KRIVITSKY, P. N. and MORRIS, M. (2014). ergm: Fit, simulate and diagnose exponential-family models for networks. The Statnet project. Available at http://www.statnet.org. R package version 3.1.2.

HANNEKE, S., FU, W. and XING, E. P. (2010). Discrete temporal models of social networks. *Electron. J. Stat.* **4** 585–605. MR2660534

HOLLAND, P. W. and LEINHARDT, S. (1981b). An exponential family of probability distributions for directed graphs. *J. Amer. Statist. Assoc.* **76** 33–65. MR0608176

HUNTER, D. R. and HANDCOCK, M. S. (2006). Inference in curved exponential family models for networks. *J. Comput. Graph. Statist.* **15** 565–583. MR2291264

HUNTER, D. R., HANDCOCK, M. S., BUTTS, C. T., GOODREAU, S. M. and MORRIS, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *J. Stat. Softw.* **24** 1–29.

JACKSON, M. O. (2008). *Social and Economic Networks*. Princeton Univ. Press, Princeton, NJ. MR2435744

KOLACZYK, E. D. (2009). *Statistical Analysis of Network Data. Methods and Models*. Springer, New York. MR2724362

KRIVITSKY, P. N. and HANDCOCK, M. S. (2014). A separable model for dynamic networks. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **76** 29–46. MR3153932

KRIVITSKY, P. N., HANDCOCK, M. S. and MORRIS, M. (2011). Adjusting for network size and composition effects in exponential-family random graph models. *Stat. Methodol.* **8** 319–339. MR2800354

KRIVITSKY, P. N. and KOLACZYK, E. D. (2014). Supplement to "On the question of effective sample size in network modeling: An asymptotic inquiry". DOI:10.1214/14-STS502SUPP.

LAVRAKAS, P. J. (2008). *Encyclopedia of Survey Research Methods*. SAGE Publications, Thousand Oaks, CA.

MORRIS, M., HANDCOCK, M. S. and HUNTER, D. R. (2008). Specification of exponential-family random graph models: Terms and computational aspects. *J. Stat. Softw.* **24** 1–24.

NEWMAN, M. E. J. (2010). *Networks. An Introduction*. Oxford Univ. Press, Oxford. MR2676073

NOLIN, D. A. (2010). Food-sharing networks in Lamalera, Indonesia: Reciprocity, kinship, and distance. *Hum. Nat.* **21** 243–268.

PU, W., CHOI, J., AMIR, E. and ESPELAGE, D. L. (2013). Learning exponential random graph models. Unpublished manuscript. Available at https://www.ideals.illinois.edu/handle/2142/45098.

R CORE TEAM (2013). R: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

RINALDO, A., PETROVIĆ, S. and FIENBERG, S. E. (2013). Maximum likelihood estimation in the β-model. *Ann. Statist.* **41** 1085–1110. MR3113804

ROBINS, G., SNIJDERS, T. A. B., WANG, P., HANDCOCK, M. S. and PATTISON, P. (2007). Recent developments in exponential random graph (p^*) models for social networks. *Soc. Networks* **29** 192–215.

ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. MR2893856

SCHWEINBERGER, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *J. Amer. Statist. Assoc.* **106** 1361–1370. MR2896841

SHALIZI, C. R. and RINALDO, A. (2013). Consistency under sampling of exponential random graph models. *Ann. Statist.* **41** 508–535. MR3099112

SNIJDERS, T. A. B. and BORGATTI, S. P. (1999). Non-parametric standard errors and tests for network statistics. *Connections* **22** 161–170.

SNIJDERS, T. A. B., VAN DE BUNT, G. G. and STEGLICH, C. E. G. (2010). Introduction to stochastic actor-based models for network dynamics. *Soc. Networks* **32** 44–60.

THIÉBAUX, H. J. and ZWIERS, F. W. (1984). The interpretation and estimation of effective sample size. *J. Climate Appl. Meteor.* **23** 800–811.

VAN DER VAART, A. W. (2000). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge.

YANG, Y., REMMERS, E. F., OGUNWOLE, C. B., KASTNER, D. L., GREGERSEN, P. K. and LI, W. (2011). Effective sample size: Quick estimation of the effect of related samples in genetic case–control association analyses. *Comput. Biol. Chem.* **35** 40–49. MR2815360