2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

A two-stage working model strategy for network analysis under Hierarchical Exponential Random Graph Models

Ming Cao
Department of Biostatistics and Data Science
University of Texas Health Science Center at Houston
Houston, TX, USA
ming.cao@outlook.com

Yong Chen
Department of Biostatistics and Epidemiology
University of Pennsylvania
Philadelphia, PA, USA
ychen123@mail.med.upenn.edu

Kayo Fujimoto
Department of Health Promotion and Behavioral Sciences
University of Texas Health Science Center at Houston
Houston, TX, USA
kayo.fujimoto@uth.tmc.edu

Michael Schweinberger
Department of Statistics
Rice University
Houston, TX, USA
michael.schweinberger@rice.edu

Abstract—Social network data are complex and dependent data. At the macro-level, social networks often exhibit clustering in the sense that social networks consist of communities; and at the micro-level, social networks often exhibit complex network features such as transitivity within communities. Modeling real-world social networks requires modeling both the macro- and micro-level, but many existing models focus on one of them while neglecting the other. In recent work, [28] introduced a class of Exponential Random Graph Models (ERGMs) capturing community structure as well as micro-level features within communities. While attractive, existing approaches to estimating ERGMs with community structure are not scalable. We propose here a scalable two-stage strategy to estimate an important class of ERGMs with community structure, which induces transitivity within communities. At the first stage, we use an approximate model, called working model, to estimate the community structure. At the second stage, we use ERGMs with geometrically weighted dyadwise and edgewise shared partner terms to capture refined forms of transitivity within communities. We use simulations to demonstrate the performance of the two-stage strategy in terms of the estimated community structure. In addition, we show that the estimated ERGMs with geometrically weighted dyadwise and edgewise shared partner terms within communities outperform the working model in terms of goodness-of-fit. Last, but not least, we present an application to high-resolution human contact network data.

Index Terms—Social Networks, Hierarchical Exponential Random Graph Models, Latent Space Cluster Models, Multiphase Inference

I. Introduction

Social network data, which are data that can be represented by a graph consisting of edges between nodes, are popular in network science, engineering, computer science, statistics, and related fields [17]. The nodes may represent individuals or organizations, and each edge is a measure of the relation between a pair of nodes. For example, in the collaboration network of statisticians studied by [16], $Y_{i,j} = 1$ indicates that individual i collaborated with individual j and $Y_{i,j} = 0$ otherwise. The statistical analysis of social network data is concerned with the processes driving edge formation, and the dependencies among the edge formation processes.

Large social networks often exhibit both macro- and micro-level structure [32]: at the macro-level, nodes are often partitioned into communities; and at the micro-level, social networks often exhibit complex network features such as transitivity within communities (“a friend of a friend is my friend”).

To capture macro-level features of social networks such as community structure, Stochastic Block Models (SBMs) are popular [1, 21]. However, SBMs assume that all edges within communities are independent, which means that SBMs are incapable of capturing dependencies among edges. An example of a dependence among edges is transitivity (“a friend of a friend is my friend”), which implies that triples of edges are dependent, and therefore cannot be captured by SBMs.

To model micro-level features of social networks such as transitivity, Exponential Random Graph Models (ERGMs) are widely used [8, 19, 35]. The development of well-behaved ERGMs with geometrically weighted edgewise shared partner terms for capturing transitivity [33] along with advances in Markov Chain Monte Carlo (MCMC) methods [11, 14, 31] have led to widespread use of ERGMs [8, 19].

In recent work, [28] combined SBMs and ERGMs to capture both macro- and micro-level features of social networks. Called Hierarchical ERGMs (HERGMs), those models assume that the set of nodes is partitioned into

IEEE/ACM ASONAM 2018, August 28-31, 2018, Barcelona, Spain

communities, and that the micro-level features of communities are captured by community-specific ERGMs. However, the Bayesian approach to estimating HERGMs proposed by [28] is time-consuming and infeasible for social networks with more than 100 nodes [29], for at least two reasons. First, there are $\exp(K \log n)$ possible partitions of the set of nodes. Second, for each possible partition of the set of nodes, there are $\exp(\binom{n_k}{2} \log 2)$ possible subgraphs within community k, where n_k is the number of nodes in community k, $k = 1, \ldots, K$. Likelihood-based inference is time-consuming, because maximum likelihood and Bayesian approaches involve intractable sums over $\exp(K \log n)$ possible community structures and, for each possible community structure with K communities, there are $\exp(\binom{n_k}{2} \log 2)$ $(k = 1, \ldots, K)$ subgraphs within K communities.

This paper. We develop a scalable approach to estimating both the macro- and micro-level features of HERGMs. The approach is based on decoupling the estimation of the macro- and micro-level features of HERGMs, resulting in a two-stage estimation strategy. We apply the two-stage estimation strategy to an important class of ERGMs, able to induce transitivity within communities. At the first stage, we use an approximate model, called working model, to estimate the community structure. As a working model, we use latent class models with community structure, which capture community structure as well as transitivity within communities. At the second stage, community-specific ERGMs are estimated, using geometrically weighted dyadwise and edgewise shared partner terms to capture more refined forms of transitivity within communities. We use simulations to show the performance of the two-stage strategy in terms of the estimated community structure. In addition, we show that the estimated ERGMs with geometrically weighted dyadwise and edgewise shared partner terms within communities outperform the working model used in the first step in terms of goodness-of-fit. We demonstrate the usefulness of the two-stage estimation strategy by applying it to high-resolution human contact network data [23].

The rest of paper is organized as follows. In Sections III and IV, we review ERGMs and ERGMs with local dependence, respectively. In Section V, we propose a scalable two-stage estimation strategy for estimating ERGMs with local dependence. In Section VI, we use simulations to demonstrate the usefulness of the two-stage estimation strategy. In Section VII, we apply it to high-resolution human contact network data.

II. Notation

Throughout, we consider undirected graphs, but extensions to directed graphs and weighted graphs are straightforward. An undirected graph $G = (V, E)$ consists a set of nodes $V = \{1, \ldots, n\}$ and a set of edges $E \subseteq V \times V$. The edges of the undirected G can be represented by the $n \times n$ adjacency matrix Y with elements

$$Y_{i,j} = \begin{cases} 1 & \text{if there is an edge between nodes } i \text{ and } j \\ 0 & \text{otherwise.} \end{cases}$$

By convention, self-loops are excluded, so $Y_{i,i} = 0$, $i = 1, \ldots, n$. In addition, we assume that the set of nodes is partitioned into K subsets of nodes, labeled $1, \ldots, K$ and called communities. The memberships of nodes to communities are denoted by $Z = \{Z_i\}_{i=1}^n$, where $Z_i = k$ indicates that node i belongs to community $k \in \{1, \ldots, K\}$.

III. ERGMs

Exponential-family random graph models (ERGMs) [4, 10, 35] are models of random graphs and can be considered to be generalizations of the Bernoulli(π) random graphs of Erdös and Rényi [3]. ERGMs assume that the probability mass of graph Y has the exponential-family form

$$\mathbb{P}(Y = y) \;=\; \exp(\langle \theta, \, s(y) \rangle - \psi(\theta)), \qquad (1)$$

where $\langle \theta, \, s(y) \rangle$ denotes the inner product of a vector of network features $s(y)$ and a vector of weights θ, and $\psi(\theta)$ ensures that $\mathbb{P}(Y = y)$ sums to one. The special case of Bernoulli(π) random graphs is obtained when $s(y)$ is the number of edges and $\theta = \text{logit}(\pi)$ is the log odds of the probability of an edge.

In principle, network scientists are free to specify network features $s(y)$, which makes ERGMs appealing to network scientists. However, while ERGMs represent a rich modeling framework, ERGMs need to be applied with care, for at least two reasons. First, it is known that some specifications of ERGMs make more sense than others [2, 5, 27]. For example, using the number of edges and triangles as network features is problematic, because it assumes that, for each pair of nodes, every additional triangle has the same added value. Such model specifications are associated with model near-degeneracy [2, 5, 27]. To reduce model near-degeneracy, [33] proposed geometrically weighted terms, such as Geometrically Weighted Edgewise Shared Partner (GWESP) terms [12, 13]. ERGMs with GWESP terms imply that the added value of additional triangles decays at a geometric rate. Such models are more appealing in practice and mitigate model near-degeneracy [13, 27]. Second, when ERGMs are applied to large networks, the underlying dependence assumptions may be unwarranted: e.g., the classic ERGMs of [4] assume that each edge can depend on $2\,(n-2)$ other edges. In large social networks with millions or billions of nodes, it is not plausible that an edge depends on millions or billions of other edges.

IV. HERGMs

To alleviate the shortcomings of SBMs (which capture macro-, but not micro-level features) and ERGMs (which capture micro-, but not macro-level features), [28] introduced HERGMs with local dependence. The notion of local dependence was inspired by related notions of local dependence in physics, spatial statistics, and time series analysis, where the units of analysis interact with other units, but the interactions are local in the sense that each unit interacts with other units that are close to it in a well-defined sense: e.g., in two-dimensional Ising models in physics, particles are located on a two-dimensional lattice and each particle interacts with its 4 nearest neighbors on the lattice. In random graphs, it is likewise plausible that each edge interacts with a small number of other edges, rather than $2\,(n-2)$ other edges as the classic ERGMs of [4] assume. There are many possible definitions of the notion of local dependence. A simple definition was introduced by [28].

Definition: local dependence [28]. The dependence induced by a probability mass function $\mathbb{P}$ of Y is called local if there exists a partition of the set of nodes A into $K \geq 2$ non-empty communities $A_1, \ldots, A_K$, such that probability mass function of the whole graph decomposes into probability mass functions of within- and between-community subgraphs $Y_{k,l}$:

$$\begin{aligned}\mathbb{P}(Y = y) &= \prod_{k=1}^{K} \mathbb{P}(Y_{k,k} = y_{k,k}) \\ &\times \prod_{l=1}^{k-1} \mathbb{P}(Y_{k,l} = y_{k,l}, Y_{l,k} = y_{l,k}),\end{aligned} \tag{2}$$

where within-community probability mass functions induce dependence within subgraphs $Y_{k,k}$, whereas between-community probability mass functions induce independence between subgraphs.

In other words, local dependence breaks down the dependence of graph Y into within-community and between-community subgraphs, but leaves researchers the freedom to specify dependencies of interest within each community. [28] and [30] show that HERGMs with local dependence have desirable theoretical properties, such as being weakly projective and consistent.

In practice, HERGMs with local dependence can be specified by specifying conditional probability mass functions of the graph Y given the community memberships $Z = z$ of the form

$$\begin{aligned}\mathbb{P}(Y = y \,|\, Z = z) &= \prod_{k=1}^{K} \mathbb{P}(Y_{k,k} = y_{k,k} \,|\, Z = z) \\ &\times \prod_{l=1}^{k-1} \mathbb{P}(Y_{k,l} = y_{k,l} \,|\, Z = z),\end{aligned} \tag{3}$$

where the between-community probability mass functions are of the form

$$\mathbb{P}(Y_{k,l} = y_{k,l} \,|\, Z = z) = \prod_{i \in A_k, j \in A_l} \mathbb{P}(Y_{i,j} = y_{i,j} \,|\, Z = z),$$

and the within-community probability mass functions are of the form

$$\mathbb{P}(Y_{k,k} = y_{k,k} \,|\, Z = z) = \exp(\langle \theta_k,\, s_k(y_{k,k}) \rangle - \psi_k(\theta_k)),$$

where $s_k(y_{k,k})$ is a vector of network features for community k and θ_k is a vector of weights for community k. The function $\psi_k(\theta_k)$ ensures that $\mathbb{P}(Y_{k,k} = y_{k,k} \,|\, Z = z)$ sums to 1. To complete the model specification, a marginal probability mass function for the community memberships Z is required. In the simplest case, when the number of communities K is known, the community memberships of nodes can be assumed to be independent $\text{Multinomial}(\pi_1, \ldots, \pi_K)$ random variables. The case of unknown K is discussed in [28].

V. Two-stage estimation

While it is attractive that HERGMs capture both macro- and micro-level features of social networks, a fully Bayesian approach to estimating them is challenging, as pointed out in Section I. Here, we propose a scalable two-stage estimation strategy, tailored to HERGMs with transitivity within communities. We focus here on HERGMs with transitivity within communities, because transitivity is one of the most challenging, and important network phenomena [34].

In the first step, it estimates the macro-level structure: the community structure. In the second step, it estimates the micro-level structure: the parameters of ERGMs governing refined forms of transitivity within communities.

In short, we propose the following two-stage estimation strategy for estimating HERGMs with transitivity within communities:

- Stage 1: Estimate the community structure.
- Stage 2: Conditional on the estimated community structure, estimate community-specific ERGMs.

Note that success in Step 1 is critical to success in Step 2, because poor clustering results in poor community-specific ERGM inference. Therefore, the estimation of the communities needs to be based on a working model that is a good approximation of the HERGM under consideration. Since the HERGM captures both community structure and transitivity within communities, so should the working model.

There are at least two possible classes of working models: SBMs [1, 21] and latent space models with community structure [6]. SBMs capture community structure, but fail to capture transitivity within communities, as pointed out in Section I. In contrast, latent space models with community structure capture both community structure and transitivity within communities, and hence constitutes a useful class of working models. We therefore propose to use latent space models with community structure as working models.

In the following two subsections, we first describe latent space models with community structure for estimating communities in Step 1, and then discuss methods for estimating community-specific ERGMs in Step 2.

A. Step 1: estimating communities

To estimate communities in Step 2, we use an extension of latent space models with community structure [7, 9]. A comprehensive review of latent space models can be found in [26].

[9] proposed latent space models based on the assumption that each node i has a position Z_i in a latent d-dimensional Euclidean space, which may be interpreted as an unobserved social space. Conditional on the positions Z_i and Z_j of nodes i and j, edges $Y_{i,j}$ are independent. If there are no covariates, the conditional probability of an edge between nodes i and j takes the form

$$\operatorname{logit}\{\mathbb{P}(Y_{i,j} = 1 \mid Z_i, Z_j)\} = \alpha - \|Z_i - Z_j\|_2, \tag{4}$$

where α can be interpreted as the baseline propensity to form edges and $\|Z_i - Z_j\|_2$ denotes the Euclidean distance between the positions Z_i and Z_j of nodes i and j. Covariates can be incorporated [9].

[7] extended the latent space models of [9] by assuming that the set of nodes is partitioned into K communities, and that the position of each node is generated by a community-specific multivariate Gaussian. In other words, [7] assumed that the positions Z_i of nodes i are generated by a finite mixture of d-variate Gaussians:

$$Z_i \overset{\text{iid}}{\sim} \sum_{k=1}^{K} \lambda_k \operatorname{MVN}_d(\mu_k, \sigma_k^2 I_d), \tag{5}$$

where $\lambda_k > 0$ denotes the probability that a node belongs to the k-th community, and the probabilities λ_k satisfy $\sum_{k=1}^{K} \lambda_k = 1$. Here, μ_k are community-specific mean vectors, σ_k^2 are community-specific variances, and I_d denotes the identity matrix of order d.

Two important properties of so-called Latent Space Cluster Models (LSCMs) are that LSCMs can be used to assess who belongs to which community and that LSCMs induce a stochastic tendency toward transitivity. By the triangle inequality, if two nodes i and j are close to a third node k, then i and j are close to each other as well. Since the conditional probability of an edge between two nodes increases as the distance between the two nodes decreases, the existence of two edges between nodes i and k and nodes j and k implies that the edge between i and k exists with high probability as well. As a result, LSCMs induce a stochastic tendency toward transitivity.

Note that LSCMs do not model transitivity directly but induce transitivity indirectly by assuming that nodes are embedded in a latent d-dimensional Euclidean space. LSCMs are therefore not as rich models of transitivity as ERGMs, which allow to directly specify many models for many forms of transitivity, using geometrically weighted model terms and other sensible model terms. Nonetheless, LSCMs are useful approximations of HERGMs with transitivity within communities, because they capture both community structure and some form of transitivity within communities. As important, while estimating LSCMs is not trivial, there exist scalable estimation procedures for LSCMs. Two examples are the scalable methods of [22] and [25], which are both based on approximations of the likelihood function. We use here the Bayesian variational approach of [25] implemented in R package VBLPCM [24].

B. Step 2: estimating ERGMs

To estimate community-specific ERGMs in Step 2, we use the Monte Carlo maximum likelihood estimates proposed by [14], which are approximations of maximum likelihood estimates. For each community, the community-specific ERGM parameters are estimated by Monte Carlo maximum likelihood estimates. We use here the Monte Carlo maximum likelihood approach of [14] implemented in R packages ergm [15] and hergm [29]. A detailed description of the Monte Carlo maximum likelihood procedure can be found in [14], and some recent advances have been made by [11] and [18].

VI. Experiments

We use simulation results to demonstrate the proposed two-stage estimation strategy, with a focus on the following two performance aspects: first, perfect recovery of community structure, with a well-chosen working model and a moderate network size; and, second, a substantial improvement in goodness-of-fit of the estimated HERGM over the working model.

A. Simulation design

We consider HERGMs with conditional probability mass function

$$\mathbb{P}(Y = y \mid Z = z) \;=\; \binom{n_B}{y_B} p^{y_B} (1-p)^{n_B - y_B}$$

$$\prod_{k=1}^{K} \exp\big(\theta_{k,1} \operatorname{Edges}(y_{k,k}) + \theta_{k,2} \operatorname{GWDSP}(y_{k,k})$$

$$+ \theta_{k,3} \operatorname{GWESP}(y_{k,k}) - \psi(\theta_k)\big),$$

where y_B is the number of between-community edges and n_B is the number of possible between-community edges. Each community-specific ERGM has three terms, an edge term to capture the basic propensity to form edges within communities and a Geometrically Weighted Edgewise Shared Partner (GWESP) and a Geometrically Weighted Dyadwise Shared Partner (GWDSP) term to capture refined forms of closure within communities. The GWESP and GWDSP terms are weighted sums of the number of pairs of nodes with i shared partners (called dyadwise shared partners, DSP) and the number of connected pairs of nodes with i shared partners (called edgewise shared partners, ESP), respectively. The weights decay at a geometric rate, which implies that the added value of additional shared partners decays at a geometric

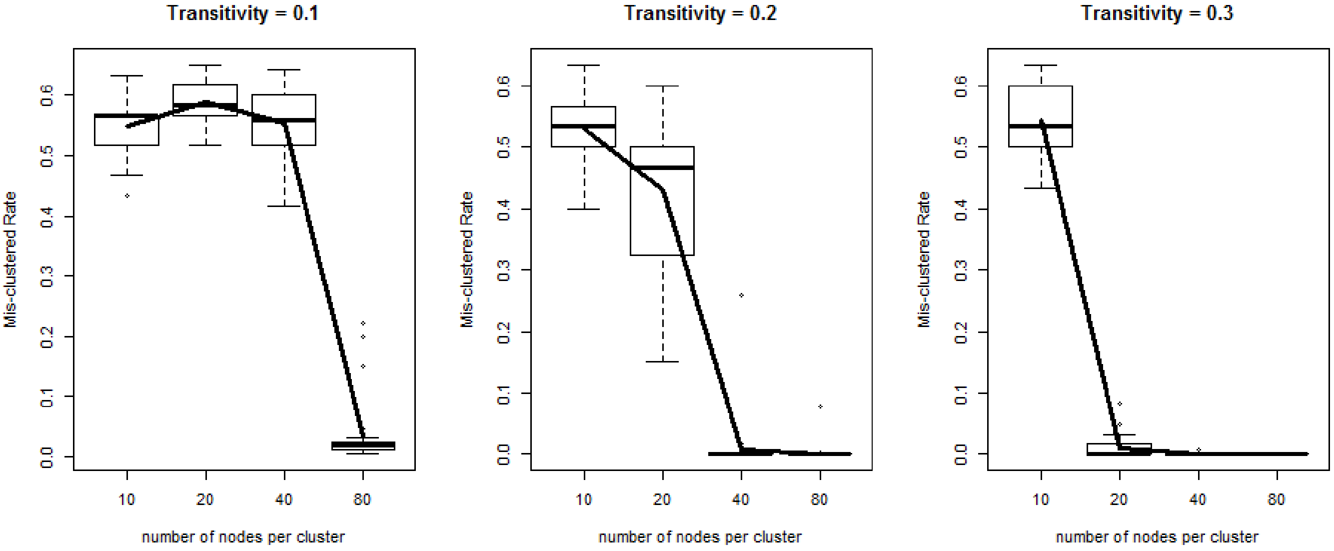


Fig. 1: The proportion of misclassified nodes decreases as the community sizes increase. From left to right, each column represents the strength of transitivity, ranging from weak (0.1), moderate (0.2), to strong (0.3). On the horizontal axis is community size (size of cluster), which is the same for all communities. On the vertical axis is the proportion of misclassified nodes (mis-clustered rate).

rate. A careful discussion of GWESP and GWDSP terms can be found in [12] and [13].

Networks with 3 communities of equal size are simulated from the HERGM specified in Equation (6), with between-community link probability $p = 0.05$ and within-community base density parameter $\theta_{k,1} = \text{logit}(0.05)$, $k = 1, \ldots, 6$. To assess the impact of community sizes and the strength of transitivity, we consider community sizes 10, 20, 40, and 80 and consider as strength of transitivity $\theta_{k,2}$ / $\theta_{k,3}$, ranging from weak (0.1), moderate (0.2), to strong (0.3). For each of the $3 \times 4 = 12$ settings, we simulated 100 networks and applied the two-stage estimation strategy to each of them. We assess the performance of the two-stage estimation procedure in Steps 1 and 2 below.

B. Stage 1 performance

To evaluate the performance of the two-stage estimation strategy in Step 1, we use the proportion of misclassified nodes to assess how well the community structure is estimated.

Figure 1 shows the proportion of misclassified nodes along with the community size and the strength of transitivity. Three observations are worth noting. First, the proportion of misclassified nodes decreases as the community sizes increase, regardless of the strength of transitivity. Second, almost-perfect recovery of the community structure is possible when the sizes of the communities is large (80), regardless of the strength of transitivity, or when the sizes of the communities is moderate (40) and the strength of transitivity is moderate (0.2) or large (0.3). Last, but not least, the proportion of misclassified nodes decreases faster when the transitivity is stronger.

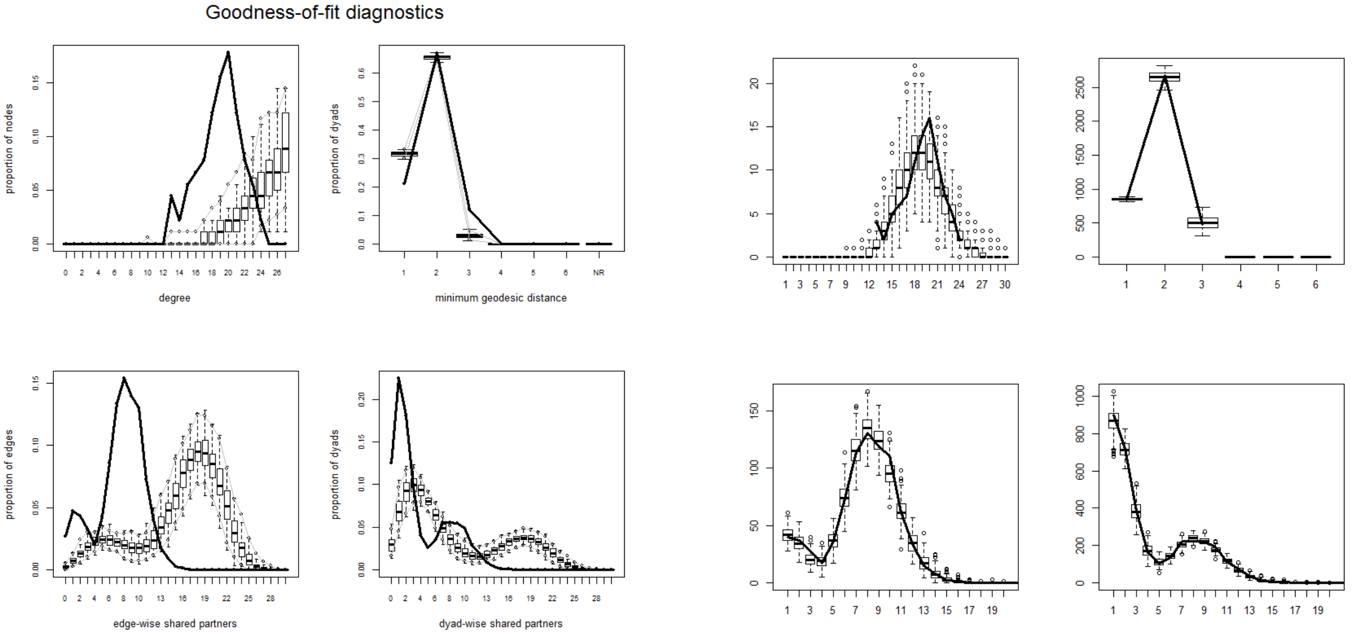
C. Stage 2 performance

To evaluate the performance of the two-stage estimation strategy in terms of goodness-of-fit and show that the HERGM estimated in Step 2 over the working model estimated in Step 1, we chose one simulated network at random from the subset of simulated networks for which the working model achieved perfect clustering. We then simulated two sets of networks, one from the working model estimated in Step 1 and one from the HERGM estimated in Step 2. We then assessed the goodness-of-fit in terms of common network features [13]: degrees, geodesic distances [20], and the numbers of edgewise shared partners (ESP) and dyadwise shared partners (DSP).

Figure 2 suggests that the working model fails to capture ESP and DSP, along with degrees. In contrast, the estimated HERGM captures all of those network features much better than the working model.

VII. Application to high-resolution human contact network data

We demonstrate the two-stage estimation strategy for HERGMs by using high-resolution human contact network data collected in a U.S. highschool [23], which can be downloaded from http://sing.stanford.edu/flu. Many infectious diseases, including Seasonal Influenza (flu), are transmitted via droplets during close proximity interactions (CPIs). Although it has long been known that



(a) Goodness-of-fit of the working model estimated in Step 1.

(b) Goodness-of-fit of the HERGM estimated in Step 2.

Fig. 2: Goodness-of-fit of the working model estimated in Step 1 (a) and the HERGM estimated in Step 2 (b) in terms of degrees, geodesic distances, ESP, and DSP.

Community k	Size of community k	Number of edges	$\widehat{\theta}_{k,1}$ (Edges)	$\widehat{\theta}_{k,2}$ (GWESP)	$\widehat{\theta}_{k,3}$ (GWDSP)
1	191	1843	-4.01 (-5.38,-3.16)	1.38 (1.2,1.61)	-0.12 (-0.16,-0.07)
2	88	622	-2.8 (-3.65,-2.2)	0.94 (0.78,1.1)	-0.14 (-0.17,-0.11)
3	76	375	-3.43 (-4.12,-2.65)	1.01 (0.87,1.15)	-0.11 (-0.16,-0.05)
4	163	859	-4.18 (-4.65,-3.67)	1.06 (0.96,1.18)	-0.06 (-0.09,-0.03)
5	91	340	-3.37 (-3.88,-2.9)	0.86 (0.76,0.97)	-0.08 (-0.12,-0.03)
6	106	510	-3.39 (-4.01,-2.78)	0.98 (0.85,1.1)	-0.1 (-0.14,-0.05)

TABLE I: Monte Carlo maximum likelihood estimates $\widehat{\theta}_{k,1}$, $\widehat{\theta}_{k,2}$, and $\widehat{\theta}_{k,3}$ of the parameters $\theta_{k,1}$, $\theta_{k,2}$, and $\theta_{k,3}$ of the community-specific ERGMs for communities $k = 1, \ldots, 6$, conditional on the estimated community structure with 6 communities.

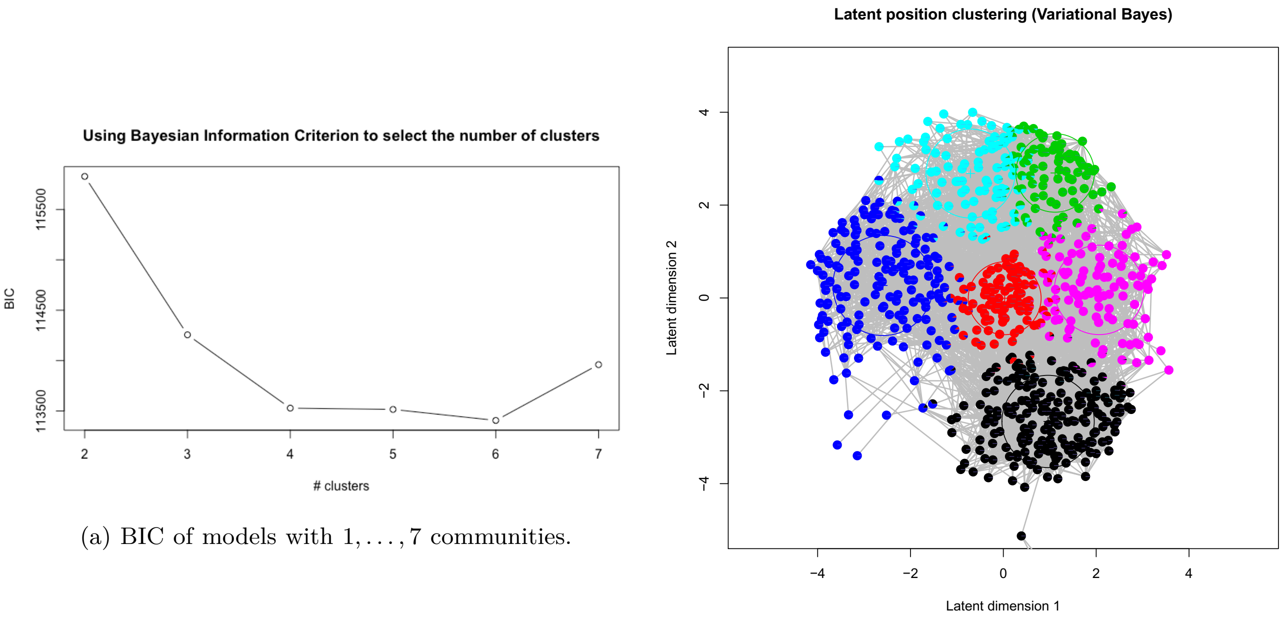
understanding how people interact is critical to containing the spread of infectious diseases, previous studies have been limited by the data collection design. Most of them have relied on surveys, which are prone to measurement errors. The advantage of [23]'s study is that CPIs were collected by wireless sensors, which are less prone to measurement errors. In particular, in [23]'s study students carried beacons, and whenever two students were sufficiently close to each other, the beacons of the students emitted signals. To construct an undirected graph, we ignore the directions of the signals and define $Y_{i,j} = 1$ if two students i and j were sufficiently close to each other for more than 10 minutes and $Y_{i,j} = 0$ otherwise. We take the largest component of the resulting network, containing 715 nodes and 7,005 edges. We are interested in HERGMs with transitivity within communities induced by GWESP and GWDSP terms, as specified in Equation (6). As a working model, we use two-dimensional LSCMs, as described in Section V-A.

The working model was estimated by the variational Bayes methods of [25] implemented in R package VBLPCM [24]. To select the number of communities, we used the Bayesian Information Criterion (BIC), as suggested by [7].

The BIC of models with $1, \ldots, 7$ communities is shown in Figure 3(a) and suggests that 6 communities is a reasonable choice. All of the following results are based on 6 communities.

Figure 3(b) shows the estimated positions of nodes in two-dimensional Euclidean space along with the estimated community structure with 6 communities. It can be seen that there are 6 distinct communities, as suggested by the BIC. In addition, the uncertainty about the community memberships of most nodes is low, as the pie plots are dominated by a single color.

Given the estimated community structure, we esti-



(a) BIC of models with $1, \ldots, 7$ communities.

(b) Estimated community structure with 6 communities.

Fig. 3: Figure (a) shows that the BIC is lowest when the number of communities is 6, suggesting that 6 communities is a reasonable choice. Figure (b) shows estimates of the positions of nodes in two-dimensional Euclidean space along with the 6 estimated communities.

mated the community-specific ERGMs. The Monte Carlo maximum likelihood estimates $\widehat{\theta}_{k,1}$, $\widehat{\theta}_{k,2}$, and $\widehat{\theta}_{k,3}$ of the parameters $\theta_{k,1}$, $\theta_{k,2}$, and $\theta_{k,3}$ of the community-specific ERGMs for communities $k = 1, \ldots, 6$ are shown in Table I. The estimates of the parameters $\theta_{k,2}$ of the GWESP terms are all large and positive, whereas the estimates of the parameters $\theta_{k,3}$ of the GWDSP terms are all small and negative. Taken together, these results suggest that, for each pair of students, dyadwise shared partners are discouraged, unless the two students are connected by an edge. If the two students are connected by an edge, additional shared partners have an added value, although the added value decays at a geometric rate. The inferred transitivity in the human contact network has implications in terms of epidemics, suggesting that infectious diseases can spread easily within the estimated communities.

VIII. Discussion

We introduced a scalable two-stage estimation strategy for estimating both the macro- and micro-level structure of social networks based on HERGMs and assessed its performance on simulated and real-world networks.

One limitation of the two-stage estimation strategy is that no universal working model exists. A working model has to be chosen on a case-by-case basis, and some working models may outperform others. For HERGMs with transitivity within communities, LSCMs are natural working models, but when HERGMs model other forms of local structure, other working models may be more appropriate.

Last, but not least, extensions to other forms of networks, including time-dependent networks as well as multilevel and multiplex networks, would be interesting. The basic ideas discussed here can be extended to such networks, although the practical implementation is non-trivial.

Acknowledgements

Ming Cao was supported by the UTHealth Innovation for Cancer Prevention Research Training Program Pre-doctoral Fellowship (Cancer Prevention and Research Institute of Texas award RP160015). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Cancer Prevention and Research Institute of Texas. Michael Schweinberger was partially supported by the National Science Foundation (NSF award DMS-1513644). The authors are grateful to Jonathan Stewart for helpful comments and suggestions.

References

[1] P. J. Bickel and A. Chen. A nonparametric view of network models and Newman-Girvan and other modularities. In Proceedings of the National Academy of Sciences, volume 106, pages 21068–21073, 2009.

[2] S. Chatterjee and P. Diaconis. Estimating and understanding exponential random graph models. The Annals of Statistics, 41:2428–2461, 2013.

[3] P. Erdős and A. Rényi. On the evolution of random graphs. Publications of the Mathematical Institute of the Hungarian Academy of Sciences, 5:17–61, 1960.

[4] O. Frank and D. Strauss. Markov graphs. Journal of the American Statistical Association, 81:832–842, 1986.

[5] M. S. Handcock. Statistical models for social networks: Inference and degeneracy. In R. Breiger, K. Carley, and P. Pattison, editors, Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers, pages 1–12. National Academies Press, Washington, D.C., 2003.

[6] M. S. Handcock, A. E. Raftery, and J. M. Tantrum. Model-based clustering for social networks. Journal of the Royal Statistical Society, Series A (with discussion), 170:301–354, 2007.

[7] M. S. Handcock, A. E. Raftery, and J. M. Tantrum. Model-based clustering for social networks. Journal of the Royal Statistical Society: Series A (Statistics in Society), 170(2):301–354, 2007.

[8] J. K. Harris. An Introduction to Exponential Random Graph Modeling. Sage, Thousand Oaks, California, 2013.

[9] P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. Journal of the American Statistical Association, 97:1090–1098, 2002.

[10] P. W. Holland and S. Leinhardt. An exponential family of probability distributions for directed graphs. Journal of the American Statistical Association, 76:33–65, 1981.

[11] R. M. Hummel, D. R. Hunter, and M. S. Handcock. Improving simulation-based algorithms for fitting ERGMs. Journal of Computational and Graphical Statistics, 21:920–939, 2012.

[12] D. R. Hunter. Curved exponential family models for social networks. Social Networks, 29:216–230, 2007.

[13] D. R. Hunter, S. M. Goodreau, and M. S. Handcock. Goodness of fit of social network models. Journal of the American Statistical Association, 103:248–258, 2008.

[14] D. R. Hunter and M. S. Handcock. Inference in curved exponential family models for networks. Journal of Computational and Graphical Statistics, 15:565–583, 2006.

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

[16] P. Ji and J. Jin. Coauthorship and citation networks for statisticians. The Annals of Applied Statistics, 10:1779–1812, 2016.

[17] E. D. Kolaczyk. Statistical Analysis of Network Data: Methods and Models. Springer-Verlag, New York, 2009.

[18] P. N. Krivitsky. Using contrastive divergence to seed Monte Carlo MLE for exponential-family random graph models. Computational Statistics and Data Analysis, 107:149–161, 2017.

[19] D. Lusher, J. Koskinen, and G. Robins. Exponential Random Graph Models for Social Networks. Cambridge University Press, Cambridge, UK, 2013.

[20] M. Newman, A.-L. Barabasi, and D. J. Watts. The structure and dynamics of networks. Princeton University Press, 2011.

[21] K. Nowicki and T. A. B. Snijders. Estimation and prediction for stochastic blockstructures. Journal of the American Statistical Association, 96:1077–1087, 2001.

[22] A. E. Raftery, X. Niu, P. D. Hoff, and K. Y. Yeung. Fast inference for the latent space network model using a case-control approximate likelihood. Journal of Computational and Graphical Statistics, 21:901–919, 2012.

[23] M. Salathé, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones. A high-resolution human contact network for infectious disease transmission. Proceedings of the National Academy of Sciences, 107(51):22020–22025, 2010.

[24] M. Salter-Townshend. VBLPCM: Variational Bayes Latent Position Cluster Model for networks, 2012. R package version 2.2.

[25] M. Salter-Townshend and T. B. Murphy. Variational Bayesian inference for the latent position cluster model for network data. Computational Statistics and Data Analysis, 57:661–671, 2013.

[26] M. Salter-Townshend, A. White, I. Gollini, and T. B. Murphy. Review of statistical network analysis: models, algorithms, and software. Statistical Analysis and Data Mining, 5:243–264, 2012.

[27] M. Schweinberger. Instability, sensitivity, and degeneracy of discrete exponential families. Journal of the American Statistical Association, 106(496):1361–1370, 2011.

[28] M. Schweinberger and M. S. Handcock. Local dependence in random graph models: characterization, properties and statistical inference. Journal of the Royal Statistical Society, Series B, 77:647–676, 2015.

[29] M. Schweinberger and P. Luna. HERGM: Hierarchical exponential-family random graph models. Journal of Statistical Software, 85:1–39, 2018.

[30] M. Schweinberger and J. Stewart. Finite-graph concentration and consistency results for random graphs with complex topological structures. 2017. Available at https://arxiv.org/abs/1702.01812.

[31] T. A. B. Snijders. Markov chain monte carlo estimation of exponential random graph models. Journal of Social Structure, 3(2):1–40, 2002.

[32] T. A. B. Snijders. Contribution to the discussion of Handcock, M.S., Raftery, A.E., and J.M. Tantrum, Model-based clustering for social networks. Journal

of the Royal Statistical Society, Series A, 170:322–324, 2007.

[33] T. A. B. Snijders, P. E. Pattison, G. L. Robins, and M. S. Handcock. New specifications for exponential random graph models. Sociological Methodology, 36:99–153, 2006.

[34] S. Wasserman and K. Faust. Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge, 1994.

[35] S. Wasserman and P. Pattison. Logit models and logistic regression for social networks: I. An introduction to Markov graphs and p^*. Psychometrika, 61:401–425, 1996.