

Local Dependence in Exponential Random Network Models

Abstract

Graph representations are used across disciplines for the analysis and visualization of relational data. Exponential random graph models allow for a general method of modelling the underlying stochastic process that has generated the observed data conditional on observer attributes of the vertices, or nodes. Recent developments in ERGMs have introduced the notions of local dependence and the exponential random network model, or ERNM. Local dependence enforces the assumption of independence between edges that connect nodes that are, in some sense, “far apart.” This is formalized by the introduction of a neighborhood structure on the graph: a partition of the vertices with the property that edges between two neighborhoods are stochastically independent of all other edge variables. This independence allows for the proof of a desirable consistency condition and a central limit theorem for statistics of the graph. The random network model allows for the joint modelling of both the graph and random attributes of the vertices. This has useful applications in network analysis, as it allows researchers to make inferences about how graphical features affect vertices and vice versa. This thesis combines these two developments to show the original result that ERNMs with the local dependence property have the same useful consistency property and that a similar central limit theorem also holds.

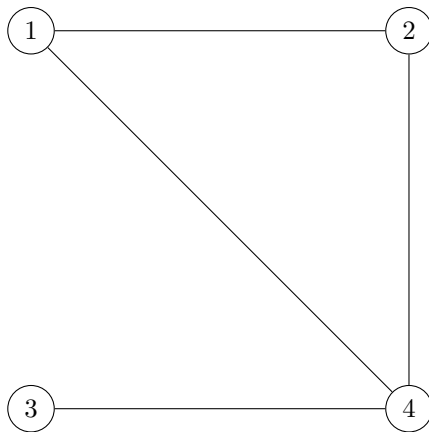


Figure 1: An undirected graph on four vertices.

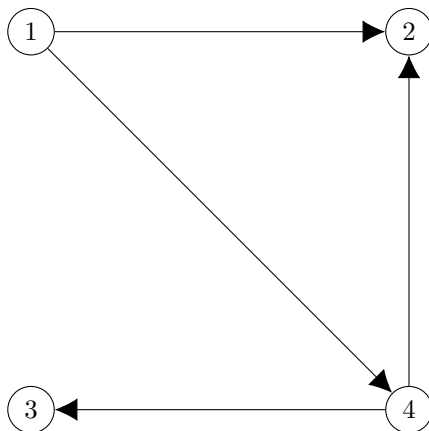


Figure 2: A directed graph on four vertices

1 Introduction

1.1 Random graphs

The mathematical object we call a graph is a set of vertices and edges, so for a graph G , we write $G = (V, E)$ where V is the set of vertices and E are ordered or unordered pairs (i, j) for $i, j \in V$. These objects can be represented visually by pictures like Figure 1. If the pairs in E are ordered, then we say that G is directed (or sometimes that G is a digraph), otherwise we say that G is undirected. The example in Figure 1 is undirected; when drawing directed graph, we can indicate the direction of the the edge (i, j) by drawing an arrow from node i to node j , as in Figure 2. If we only consider the presence or absence of an edge between vertices, then we say that the graph G is unweighted. In a weighted graph, each edge is assigned a numeric value, called the weight. The case of an unweighted graph is equivalent a weighted graph where each edge weight $w_{ij} \in \{0, 1\}$.

Generally, the most convenient way to represent a graph is as an adjacency matrix. For a graph with n vertices, the adjacency matrix is the $n \times n$ matrix with the weight of edge (i, j) in the i, j position. Note that if a graph G is undirected, then its adjacency matrix is symmetric and if G is an unweighted graph, then all its entries are either 1 or 0. For example, the graph in Figure 1 has adjacency matrix

$$\begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}. \tag{1}$$

Here, we have adopted the convention that vertices do not have an edge with themselves, so the diagonal elements of this matrix are all zero. Sometimes we will also represent a graph as an edge list, which is simply the set of pairs in E with their weights.

If we assume that a graph G is generated by some kind of stochastic process, then we consider the whole graph as a random variable. This is equivalent to considering the random matrix Y , where Y is the adjacency matrix of G . We can consider the n^2 random variables Y_{ij} for $i, j \in \{1, \dots, n\}$, taking values in the set of possible weights for the random graph in consideration. For example, if G is an undirected binary random graph, then this can be simplified to the $n^2/2 - n$ Bernoulli random variables $Y_{(i,j)}$ for $1 \leq i < j \leq n$. By breaking the random graph into these simpler parts, we can begin to apply familiar statistical ideas to these complex objects.

1.2 Network modeling

Graphs are a convenient way to represent relational data, or data that contain information about a relationship between actors. This brings us to the area of network analysis, where random graphs are used to model many different kinds of relationships. For example, in sociology, networks are commonly used to model relationships among people, like marriages within a group of people. Political scientists use random graphs to represent things like the network of trade agreements between nations. Finally, in epidemiology, random networks can model an underlying contact network, say of people who shake hands or otherwise come into contact in a way that could spread a disease, over which a transmissible infection spreads. This allows the researcher to estimate how the medical features of the disease and the sociological features of the population interact to promote or inhibit the spread of disease. As an example of this, see Groendyke, Welch, & Hunter (2012), where, among other things, they examine if closing schools would have had an effect on the spread of an 1861 measles outbreak.

To formalize these notions, we begin with a random graph Y on n vertices (called nodes in the field of network analysis) considered as a random matrix with sample space \mathcal{Y} , where \mathcal{Y} is the set of all possible graphs on n vertices. There may be some restrictions on \mathcal{Y} , depending on the application being considered. Usually, we prohibit edges that connect a vertex to itself by forcing the diagonal entries of Y to be 0, as in (1). However, specific contexts may require nonobvious constraints put in place by the researcher; for example, when using a network to model the romantic relationships within a group of people, the researcher must exclude networks with ties between nodes that do not have the appropriate gender, depending on the respective preferences of each pair of nodes.

This thesis explores two classes of network models: the exponential random graph model (ERGM) and the exponential random network model (ERNM). These models are quite similar in that they both model the structural features of the network, such as node degree and triangle formation, as a function of nodal covariates, say age or gender. However, the ERNM also models the effect of network features on the nodes. For example, Fellows & Handcock (2012) show that teens with friends who smoke or drink are more likely to smoke or drink themselves.

The notion of local dependence was introduced by Schweinberger & Handcock (2015) in the context of ERGMs. Here we extend this idea to ERNMs and provide new proofs of their two main theorems in this context. The first theorem shows that the stochastic process of sampling smaller subnetworks from a large network satisfies a desirable consistency condition. This means that as we sample larger and larger subnetworks, the model parameter estimates become better and better approximations of the true parameter values. Shalizi & Rinaldo (2013) showed that this is not the case for most ERGMs. The second theorem shows that certain types of statistics of locally dependent random networks have an asymptotically normal distribution, which is also desirable from the point of view of making inferences based on these models.

2 Exponential random graph models

The first model of interest to this thesis is the exponential random graph model (ERGM), introduced by S. Wasserman & Pattison (1996). Broadly, given a network, this model allows us to specify structural features that are of interest and test in a principled way if there are more of those kinds of features present than would be expected if the network were formed by chance. This allows the researcher to draw conclusions about the latent process that generated the network.

This model can be expressed by the equation

$$P(Y = y|\theta) = \frac{e^{\theta \cdot g(y)}}{C(\theta, \mathcal{Y})}, \quad (2)$$

where $g(y)$ is a vector of statistics that depend on the network y and θ is a vector of parameters. The function $C(\theta, \mathcal{Y})$ is a normalizing constant. We may easily extend the model to include nodal covariates (like demographic information) by introducing the fixed $n \times p$ matrix x which is used to calculate some of the components of g .

Note that the constant in the denominator depends both on the support of the random variable Y , which is fixed and—in principle—known, and the unknown parameter θ . In general, this constant will have the form

$$C(\theta, \mathcal{Y}) = \sum_{y \in \mathcal{Y}} e^{\theta \cdot g(y)}, \quad (3)$$

which is impossible to compute in all but the most trivial cases, as the sum is over all possible networks. For the simplest undirected binary network with n nodes, we have $|\mathcal{Y}| = 2^{n(n-1)/2}$, so for 10 nodes, the constant C will be a sum of 2^{45} terms. As each term requires a non-trivial calculation, we are unable to calculate the value of C directly for a network of any reasonable size.

With the model in place, we can begin to estimate the parameter θ via maximum likelihood estimation. To do that, we define the likelihood function, which takes the parameter space into the reals. Our goal is to find the value of θ that maximizes this function given the observed data. We call this θ the maximum likelihood estimator, or MLE. Intuitively, we are looking for the value of θ that makes the network we have observed most likely according to our model. In our case this function is

$$L(\theta|y) = \frac{e^{\theta \cdot g(y)}}{C(\theta, \mathcal{Y})}. \quad (4)$$

Note that C also varies with θ .

The fact that this term varies with θ means that a naive optimization algorithm will be prohibitively slow. Exploring the parameter space almost always involves evaluating the function at many different points. For example, the common gradient descent algorithm involves evaluating the function many times in the neighborhood of a point to approximate the gradient. This then allows the algorithm to choose a new point where the function is larger. This process then repeats. Even when given relatively simple functions, this algorithm can require hundreds of function calls. In our case, that is not feasible, as we cannot even evaluate our function once.

Difficulties aside, this model will be the jumping off point for our exposition of the properties of the exponential random network model.

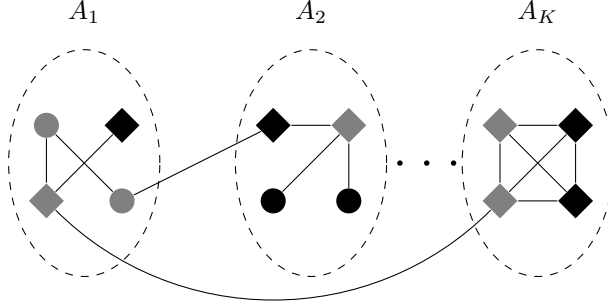


Figure 3: A locally dependent random network with neighborhoods A_1, A_2, \dots, A_K and two binary node attributes, represented as gray or black and circle or diamond.

3 Locally dependent exponential random network models

To begin, we define a random network, developed by Fellows & Handcock (2012). By way of motivation, note that in the ERGM the nodal variates are fixed and are included in the model as explanatory variables in making inferences about network structure. Furthermore, there is a class of models that we do not discuss here that consider the network as a fixed explanatory variable in modeling (random) nodal attributes. It is not difficult to come up with situations where a researcher would like to *jointly* model both the network and the node attributes. Thus we define a class of networks in which both the network structure and the attributes of the individual nodes are modeled as random quantities.

Definition 3.1 (Random network). Let \mathcal{N} be a countable collection of nodes (which we take to be a subset of \mathbb{N}). Let Y be the random graph on the nodes \mathcal{N} with support \mathcal{Y} . Then for each element $n \in \mathcal{N}$, let there be a corresponding random vector of node attributes $X_n \in \mathbb{R}^q$ and collect them into the $n \times q$ random matrix X with support \mathcal{X} . The *random network* is the random variable $Z = (Y, X)$ with support $\mathcal{Z} = \mathcal{Y} \times \mathcal{X}$.

Now we wish to model these objects, so we follow the ERGM and turn to the exponential family (Fellows & Handcock, 2012). We write

$$P(Y = y, X = x|\eta) \propto e^{\eta \cdot g(y,x)}. \quad (5)$$

This looks very similar to the ERGM, but note the explicit dependence on the quantity x . More concretely, we can include terms that depend *only* on x , which would have no place in an ERGM. We can further express the difference of the two models by rewriting the left hand side of 5 as

$$P(X = x, Y = y|\eta) = P(Y = y|X = x, \eta)P(X = x|\eta),$$

where the first term on the right hand side is the ERGM and the second term is

$$P(X = x|\eta) = \frac{C(\mathcal{Z}, \eta, x)}{C(\mathcal{Z}, \eta)},$$

where

$$C(\mathcal{Z}, \eta, x) = \int_{\{(v,u) \in \mathcal{Z}: u=x\}} P(X = x|\eta).$$

Roughly, this is the proportion of the total sample space \mathcal{Z} that is possible with x fixed. This is not, in general, equal to one, so the ERNM is not equal to the ERGM (Fellows & Handcock, 2012).

3.1 Definitions and notation

We will consistently refer to a set of nodes, A_k , as the k -th neighborhood, with an uppercase K representing the total number of neighborhoods and a lowercase k representing a specific neighborhood. The variable \mathcal{N} will refer to the domain of a random network, usually the union of a collection of neighborhoods. Nodes within the network will be indexed by the variables i and j , with $Z_{ij} = (\{Y_{ij}, X_i, X_j\})$, where Y_{ij} is referring to the edge between nodes i and j , and X_i and X_j refer to the random vectors of node attributes. Abstracting this further, i and j will also refer to tuples of nodes, so we will write $\vec{i} = (i_1, i_2, \dots, i_q) \in \mathcal{N} \times \mathcal{N} \times \dots \times \mathcal{N}$. The variables Z and Y will also often carry a subscript of W or B (for example Y_{Bij}) which emphasizes that the edge from i to j is within or between neighborhoods, respectively. Finally, for lack of a better notation, the indicator function $\mathbb{I}_B(i, j)$ (where B is for between) is one if $i \in A_l$ and $j \in A_p$ where $l \neq p$, and zero otherwise.

Definition 3.2 (Local dependence property). Extending the definition in Schweinberger & Handcock (2015), a random network model satisfies *the local dependence property* if there is a partition of the node set \mathcal{N} into neighborhoods A_1, A_2, \dots, A_K for $K \geq 2$ such that the network variables Z_{ij} are dependent when $i, j \in A_\ell$ for some ℓ and independent otherwise. We also require that nodal attributes depend only on the attributes of nodes within the same neighborhood. Thus, the probability measure can be written as

$$P(Z \in \mathbf{Z}) = \prod_{k=1}^K \left[P_{kk}(Z_{kk} \in \mathbf{Z}_{kk}) \prod_{\ell=1}^{k-1} P_{k\ell}(Z_{k\ell} \in \mathbf{Z}_{k\ell}, Z_{\ell k} \in \mathbf{Z}_{\ell k}) \right], \quad (6)$$

where Z_{mn} is the subnetwork consisting of the random graph ties from nodes in A_m to those in A_n and the appropriate node variables and \mathbf{Z}_{mn} is a subset of the sample space of Z_{mn} . Furthermore, the measures P_{kk} can induce dependence between dyads while the measures P_{kl} induce independence.

Definition 3.3 (Sparsity). Also from Schweinberger & Handcock (2015), we say a locally dependent random network is δ -sparse if there is some $\delta > 0$ and some $C > 0$ such that

$$\mathbb{E}(|Y_{Bij}|^p) \leq Cn^{-\delta}, \quad (p = 1, 2) \quad (7)$$

where $n = |\mathcal{N}|$ and Y_{Bij} signifies the tie between neighborhoods from node $i \in A_l$ to node $j \in A_m$ where $l \neq m$.

3.2 Preliminary theorems

In proving our theorems, we will make use of several other central limit theorems, all of which can be found in Billingsley (1995). The first is the Lindeberg-Feller central limit theorem for triangular arrays. The second is Lyapounov's condition, which gives a convenient way to show that the Lindeberg-Feller theorem holds. Finally, we make use of a central limit theorem for dependent random variables. For the sake of brevity, in this section we state each of these without proof.

Theorem 3.1 (Billingsley, 1995 Theorem 27.2). *For each n take X_{n1}, \dots, X_{nr_n} , independent with $\mathbb{E}(X_{ns}) = 0$ for all n and s (where no generality is lost in this assumption). Then we have $\sigma_{ns}^2 = \text{Var}(X_{ns}) = \mathbb{E}(X_{ns}^2)$. Next, set $s_s^2 = \sum_{s=1}^{r_n} \sigma_{ns}^2$. Now set*

$$S_n = X_{n1} + \dots + X_{nr_n}.$$

If the Lindeberg condition,

$$\lim_{n \rightarrow \infty} \sum_{s=1}^{r_n} \frac{1}{s_n^2} \int_{|X_{ns}| \geq \epsilon s_n} X_{ns}^2 = 0 \quad (8)$$

holds for all $\epsilon > 0$, then $S_n \xrightarrow{d} N(0, 1)$.

Theorem 3.2 (Billingsley, 1995 Theorem 27.3). *Let S_n be as before. Then if Lyapounov's condition,*

$$\lim_{n \rightarrow \infty} \sum_{s=1}^{r_n} \frac{1}{s_n^{2+\delta}} \mathbb{E} \left(|X_{ns}|^{s+\delta} \right) = 0 \quad (9)$$

holds for some $\delta > 0$, then the Lindeberg condition also holds. Therefore $S_n \xrightarrow{d} N(0, 1)$.

Theorem 3.3 (Billingsley, 1995 Theorem 27.4). *Suppose that X_1, X_2, \dots is stationary and α -mixing with $\alpha_n = O(n^{-5})$ and that $\mathbb{E}(X_n) = 0$ and $\mathbb{E}(X_n^{12}) < \infty$. Note that the condition on α is stronger than what we require. Our X_n will be M -dependent, meaning that each X_n is independent of all X_m where $|n - m| > M$. It is true that an M -dependent sequence is α -mixing for constant α . Then, if $S_n = X_1 + \dots + X_n$, we have*

$$\frac{\text{Var}(S_n)}{n} \rightarrow \sigma^2. \quad (10)$$

Then, if $\sigma > 0$, we have $S_n \xrightarrow{d} N(0, 1)$.

The final theorem is Slutsky's theorem, a classic result of asymptotic theory in statistics.

Theorem 3.4 (Wasserman, 2004 Theorem 5.5). *Let X_n, X, Y_n be random variables and let c be a constant. Then, if $X_n \xrightarrow{d} X$ and $Y \xrightarrow{p} c$ we have $X_n + Y_n \xrightarrow{d} X + c$ and $X_n Y_n \xrightarrow{d} cX$.*

3.3 Consistency under sampling

With these in place, we attempt to extend a result about locally dependent ERGMs proven by Schweinberger & Handcock (2015) to locally dependent ERNMs. In short, this theorem states that the parameters estimated by modeling a small sample of a larger network can be generalized to the overall network. It was shown by Shalizi & Rinaldo (2013) that most useful formulations of ERGMs do not form projective exponential families in the sense that the distribution of a subgraph cannot be, in general, recovered by marginalizing the distribution of a larger graph with respect to the edge variables not included in the smaller graph. Hence, we are unable to generalize parameter estimates from the subnetwork to the total network.

To show that locally dependent ERNMs do form a projective family, let \mathbb{A} be a collection of sets \mathcal{A} , where each \mathcal{A} is a finite collection of neighborhoods. Also, allow the set \mathbb{A} to be an ideal, so that if $\mathcal{A} \in \mathbb{A}$, every subset of \mathcal{A} is also in \mathbb{A} and if $\mathcal{B} \in \mathbb{A}$, then $\mathcal{A} \cup \mathcal{B} \in \mathbb{A}$. If $\mathcal{A} \subset \mathcal{B}$, think of passing from the set \mathcal{A} to the set \mathcal{B} as taking a larger sample of the (possibly infinite) set of neighborhoods in the larger network. Then let $\{\mathcal{P}_{\mathcal{A}, \theta}\}_{\mathcal{A} \in \mathbb{A}}$ be the collection of ERNMs with parameter θ indexed by the sets in \mathbb{A} . For each $\mathcal{A} \in \mathbb{A}$, let $\mathcal{P}_{\mathcal{A}, \Theta} = \{\mathcal{P}_{\mathcal{A}, \theta}\}_{\theta \in \Theta}$ be the collection of ERNMs on the neighborhoods in \mathcal{A} with parameter $\theta \in \Theta$ where $\Theta \subset \mathbb{R}^p$ is open. Assume that each distribution in $\mathcal{P}_{\mathcal{A}, \Theta}$ has the same support $\mathcal{Z}_{\mathcal{A}}$ and that $\mathcal{A} \subset \mathcal{B}$ if and only if $\mathcal{Z}_{\mathcal{B}} = \mathcal{Z}_{\mathcal{A}} \times \mathcal{Z}_{\mathcal{B} \setminus \mathcal{A}}$. Then, the exponential family $\{\mathcal{P}_{\mathcal{A}, \Theta}\}_{\mathcal{A} \in \mathbb{A}}$ is projective in the sense of Shalizi & Rinaldo (2013 Definition 1) precisely when Theorem 3.6 holds.

This follows from a specific case of the general definition given by Shalizi & Rinaldo (2013). There, for every pair \mathcal{A} and \mathcal{B} with $\mathcal{A} \subset \mathcal{B}$, they define the natural projection mapping $\pi_{\mathcal{B} \rightarrow \mathcal{A}} : \mathcal{Z}_{\mathcal{B}} \rightarrow \mathcal{Z}_{\mathcal{A}}$. Informally, this mapping projects the set $\mathcal{Z}_{\mathcal{B}}$ down to $\mathcal{Z}_{\mathcal{A}}$ by simply removing the extra data. For example if $\mathcal{B} = \{A_1, A_2\}$ and $\mathcal{A} = \{A_1\}$ as in Figure 3, then the mapping $\pi_{\mathcal{B} \rightarrow \mathcal{A}}$ is shown in Figure 4.

This is desirable because Shalizi & Rinaldo (2013) have demonstrated the following theorem.

Theorem 3.5 (Shalizi & Rinaldo, 2013 Theorem 3). *If the exponential model family $\{\mathcal{P}_{\mathcal{A}, \Theta}\}_{\mathcal{A} \in \mathbb{A}}$ is projective and the log of the normalizing constant can be written as*

$$\begin{aligned} \log(C(\theta, \mathcal{Z})) &= \log \left(\int_{\mathcal{Z}} e^{\theta \cdot g(z)} dz \right) \\ &= r(|\mathcal{Z}|) a(\theta), \end{aligned} \quad (11)$$

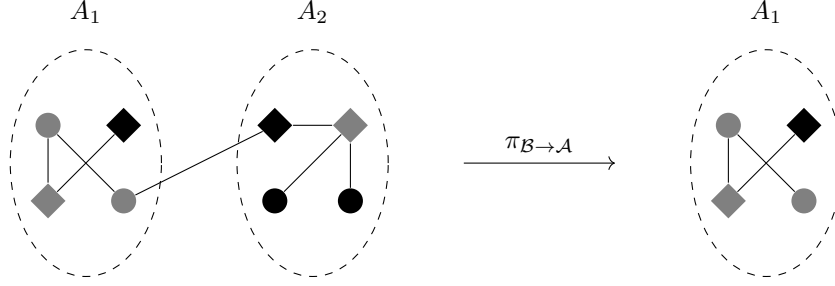


Figure 4: The projection mapping from $\mathcal{B} = \{A_1, A_2\}$ to $\mathcal{A} = \{A_1\}$.

where r is a positive, monotone increasing function of some positive measure on \mathcal{Z} and a is a differentiable function of θ , then the maximum likelihood estimator exists and is strongly consistent, meaning that the MLE, $\hat{\theta} \xrightarrow{a.s.} \theta$, where θ is the unknown parameter being estimated.

This is trivially achieved by setting $r = 1$ for all values of $|\mathcal{Z}|$ and setting $a(\theta) = \log(C(\theta, \mathcal{Z}))$. We have differentiability of a with respect to θ by a result from multivariable calculus that follows from Fubini's theorem. From a practical perspective, this means that a researcher using this model can assume that parameters estimated from samples of a large network are increasingly good approximations for the true parameter values as the sample size increases.

Theorem 3.6. Let A_1, A_2, \dots be a sequence of neighborhoods and define the sequence $\{\mathcal{N}_K\} = \bigcup_{i=1}^K A_i$. Then let Z_1, Z_2, \dots be the sequence of locally dependent random networks on the \mathcal{N}_K . For each Z_K , there is the corresponding set of neighborhoods \mathcal{A}_K . Let P_K be a generic probability distribution from the family $\{P_{K\theta}\}_{\theta \in \Theta}$. Let the network

$$\pi_{\mathcal{A}_{K+1} \rightarrow \mathcal{A}_K}(Z_{K+1}) = Z_{K+1 \setminus K},$$

with corresponding distribution $P_{K+1 \setminus K}$. Then

$$P_K(Z_K \in \mathbf{Z}_K) = P_{K+1}(Z_K \in \mathbf{Z}_K, Z_{K+1 \setminus K} \in \mathcal{Z}_{K+1 \setminus K}), \quad (12)$$

my_dev where \mathcal{Z}_K is the sample space of the distribution P_K and $\mathbf{Z}_K \subset \mathcal{Z}_K$. This is a specific case of the definition of projectibility for a general exponential family given by Shalizi & Rinaldo (2013).

Proof. This follows from the definition of local dependence, in much the same way as the proof for ERGMs by Schweinberger & Handcock (2015) does. We have

$$\begin{aligned} P_{K+1}(Z_K \in \mathbf{Z}_K, Z_{K+1 \setminus K} \in \mathcal{Z}_{K+1 \setminus K}) &= P_{K+1}(Z_K \in \mathbf{Z}_K)P_{K+1 \setminus K}(Z_{K+1 \setminus K} \in \mathcal{Z}_{K+1 \setminus K}) \\ &= P_K(Z_K \in \mathbf{Z}_K)(1) \\ &= P_K(Z_K \in \mathbf{Z}_K), \end{aligned}$$

where the measure becomes P_k from the product definition of a locally dependent random network. \square

3.4 Asymptotic normality of statistics

In this section we will prove that certain classes of statistics of locally dependent random networks are asymptotically normally distributed as the number of neighborhoods tends to infinity. The statistics we consider can be classified into three types: first, statistics which depend only on the graph structure; second, statistics that depend on both the graph and the nodal variates; and third, statistics that depend only on the

nodal variates. The first class of statistics has already been considered by Schweinberger & Handcock (2015), but we will reproduce the proof here, as the second proof is very similar. The third class of statistics becomes normal in the limit by a central limit theorem for M dependent random variables in Billingsley (1995).

Before we begin to explicitly define each of these classes, we clarify the notation that will be used. A general statistic will be a function

$$S : \mathcal{N}^d \rightarrow \mathbb{R},$$

where \mathcal{N}^d is the d -fold Cartesian product of the set of nodes, \mathcal{N} , with itself:

$$\mathcal{N}^d = \underbrace{\mathcal{N} \times \cdots \times \mathcal{N}}_{d \text{ times}}.$$

Additionally, the statistic will often carry a subscript K , indicating that the statistic is of the random network with K neighborhoods.

Formally, as explained in Schweinberger & Handcock (2015), the first class of statistics contains those that have the form

$$S_K = \sum_{i \in \mathcal{N}^d} S_{Ki},$$

where

$$S_{Ki} = \prod_{l,p \in i} Y_{lp},$$

a product q of edge variables that captures the interaction desired. We will also make use of the set A_k^d , which is a similar cartesian product. When we write $i \in A_k^d$, we mean the every component of the d -tuple i is an element of A_k . Furthermore, by a catachrestic abuse of notation, we will write $l, p \in i$ to mean that l and p are vertices contained in the d -tuple i . Now we are ready to prove the first case of the theorem.

Theorem 3.7. *Let A_1, A_2, \dots, A_K be a sequence of neighborhoods of size at most M and form the sequence of domains $\mathcal{N}_K = \bigcup_{k=1}^K A_k$. Then let Z_1, Z_2, \dots, Z_K be the sequence of unweighted random networks on the \mathcal{N}_K . Then, let the statistic $S_K : \mathcal{N}_K^d \rightarrow \mathbb{R}$ be given. Furthermore, assume the statistic depends only on the graph variables of the Z_K . We also assume that the Z_K satisfy the local dependence property and that they are δ -sparse, for some $\delta > d$. Finally, we require that $\text{Var}(W_K^*) \rightarrow \infty$, where W_K^* is defined in (14). Then*

$$\frac{S_K - \mathbb{E}(S_K)}{\sqrt{\text{Var}(S_K)}} \xrightarrow[K \rightarrow \infty]{d} N(0, 1). \quad (13)$$

Proof. As the networks Z_K are unweighted, all edge variables $Y_{ij} \in \{0, 1\}$. Let $\mu_{ij} = \mathbb{E}(Y_{ij})$. Then define $V_{ij} = Y_{ij} - \mu_{ij}$. Therefore, without loss of generality, we may work with V_{ij} , which has the convenient property that $\mathbb{E}(V_{ij}) = 0$. This means that we can similarly shift our statistics of interest, S_K . Therefore, call $S_K^* = S_K - \mathbb{E}(S_K)$, so that $\mathbb{E}(S_K^*) = 0$.

Note that we can write

$$S_K^* = W_K^* + B_K^*,$$

with

$$W_K^* = \sum_{k=1}^K W_{K,k}^* = \sum_{k=1}^K \sum_{i \in \mathcal{N}_K^d} \mathbb{I}(i \in A_k^d) S_{Ki}^* \quad (14)$$

and

$$B_K^* = \sum_{i \in \mathcal{N}_K^d} \mathbb{I}_B(i) S_{Ki}^*, \quad (15)$$

where the indicator functions restrict the sums to within the k -th neighborhood and between neighborhoods of the graph, respectively. Specifically, $\mathbb{I}_B(i) = 1$ when the d -tuple of nodes i contains nodes from different neighborhoods, or exactly when $\mathbb{I}(i \in A_k^d) = 0$ for all neighborhoods k . By splitting the statistic into the within and between neighborhood portions, we are able to make use of the independence relation between edges that connect neighborhoods. We also have $\mathbb{E}(W_K^*) = 0$ and $\mathbb{E}(B_K^*) = 0$, as each quantity is a sum of random variables with mean zero.

The idea of this proof is to gain control over the variances of B_K^* and all the elements of the sequence W_{Kk}^* . We can then show that B_K^* is converging in probability to zero and that the triangular array W_K^* satisfies Lyapounov's condition, and is thus asymptotically normal. Finally, Slutsky's theorem allows us to extend the result to S_K^* .

To bound the variance of B_K^* , note that

$$\text{Var}(B_K^*) = \sum_{i \in \mathcal{N}_K^d} \sum_{j \in \mathcal{N}_K^d} \mathbb{I}_B(i) \mathbb{I}_B(j) \text{Cov}(S_{Ki}^*, S_{Kj}^*).$$

Despite independence, some of these covariances may be nonzero if the two terms of the statistic both involve the same edge. For example, in Figure 3, a statistic that counted the number of edges between gray nodes plus the number of edges between diamond shaped nodes would have a nonzero covariance term because of the edge between the two nodes that are both gray and diamond shaped. To show that, in the limit, these covariances vanish, we need only concern ourselves with the nonzero terms in the sum. That is, only those terms where $\mathbb{I}_B(i) \mathbb{I}_B(j) = 1$. This happens exactly when both S_{Ki}^* and S_{Kj}^* involve a between neighborhood edge variable. So, note that we have

$$\begin{aligned} \text{Cov}(S_{Ki}^*, S_{Kj}^*) &= \mathbb{E}(S_{Ki}^* S_{Kj}^*) - \mathbb{E}(S_{Ki}^*) \mathbb{E}(S_{Kj}^*) \\ &= \mathbb{E}(S_{Ki}^* S_{Kj}^*), \end{aligned} \quad (16)$$

as the expectation of each term is zero. Next we take $Y_{l_1 l_2}$ to be one of the (possibly many) between neighborhood edge variables in this product (such that $\mathbb{I}_B(i) = 1$ where i is any tuple containing l_1 and l_2) and $V_{l_1 l_2}$ to be the recentered random variable corresponding to $Y_{l_1 l_2}$. Then,

$$\begin{aligned} \text{Cov}(S_{Ki}^*, S_{Kj}^*) &= \mathbb{E} \left(\prod_{m,n \in i} V_{mn} \prod_{m,n \in j} V_{mn} \right) \\ &= \mathbb{E} \left(V_{l_1 l_2}^p \prod_{m,n \in (i \cup j) \setminus \{l_1, l_2\}} V_{mn} \right), \quad (p = 1, 2) \end{aligned}$$

where we must consider the case where $p = 1$ to account for the covariance of S_{Ki}^* and S_{Kj}^* when $i \neq j$ and the case where $p = 2$ to account for the variance of S_{Ki}^* , which is computed in the case where $i = j$. So, if $p = 1$, then

$$\begin{aligned} \mathbb{E} \left(V_{l_1 l_2} \prod_{m,n \in (i \cup j) \setminus \{l_1, l_2\}} V_{mn} \right) &= \mathbb{E}(V_{l_1 l_2}) \mathbb{E} \left(\prod_{m,n \in (i \cup j) \setminus \{l_1, l_2\}} V_{mn} \right) \\ &= 0, \end{aligned}$$

by the local dependence property and the assumption that $\mathbb{E}(V_{l_1 l_2}) = 0$. The local dependence property allows us to factor out the expectation of $V_{l_1 l_2}$, as this edge is between neighborhoods, and therefore independent of every other edge in the graph. Now, if we have $p = 2$, then, by sparsity and the fact that the product below is at most 1,

$$\mathbb{E}(V_{l_1 l_2}^2) \mathbb{E} \left(\prod_{m,n \in (i \cup j) \setminus \{l_1, l_2\}} V_{mn} \right) \leq DCn^{-\delta},$$

where D is a constant that bounds the expectation above. There exists such a constant because each of the V_{mn} are bounded by definition, so a product of them is bounded. So as K grows large, the between neighborhood covariances all become asymptotically negligible. Therefore, we can conclude that

$$\text{Var}(B_K^*) = \sum_{i \in \mathcal{N}_K^d} \sum_{j \in \mathcal{N}_K^d} \mathbb{I}_B(i) \mathbb{I}_B(j) \text{Cov}(S_{K_i}^*, S_{K_j}^*) \leq DCn^{2d(-\delta)}. \quad (17)$$

So we have

$$\text{Var} B_K^* \rightarrow 0. \quad (18)$$

Then, for all $\epsilon > 0$, Chebyshev's inequality gives us

$$\lim_{K \rightarrow \infty} P(|B_K^*| > \epsilon) \leq \lim_{K \rightarrow \infty} \frac{1}{\epsilon^2} \text{Var}(B_K^*) = 0, \quad (19)$$

so

$$B_K^* \xrightarrow{P} 0. \quad (20)$$

Next, we bound the within neighborhood covariances, as we also have

$$\text{Var}(W_{Kk}^*) = \sum_{i \in \mathcal{N}_K^d} \sum_{j \in \mathcal{N}_K^d} \mathbb{I}(i, j \in A_k) \text{Cov}(S_{K_i}^*, S_{K_j}^*).$$

As the covariance forms an inner product on the space of square integrable random variables, the Cauchy-Schwarz inequality gives us

$$\mathbb{I}(i, j \in A_k) |\text{Cov}(S_{K_i}^*, S_{K_j}^*)| \leq \mathbb{I}(i, j \in A_k) \sqrt{\text{Var}(S_{K_i}^*)} \sqrt{\text{Var}(S_{K_j}^*)}. \quad (21)$$

Then, as each $S_{K_i}^*$ has expectation zero, we know that

$$\text{Var}(S_{K_i}^*) = \mathbb{E}(S_{K_i}^{*2}) - \mathbb{E}(S_{K_i}^*)^2 = \mathbb{E}(S_{K_i}^{*2}). \quad (22)$$

As $S_{K_i}^{*2} \leq 1$, we know $\text{Var}(S_{K_i}^*) \leq 1$ for all tuples i , so we have the bound

$$\mathbb{I}(i, j \in A_k) |\text{Cov}(S_{Ki}^*, S_{Kj}^*)| \leq \mathbb{I}(i, j \in A_k). \quad (23)$$

Now all that remains is to apply the Lindeberg-Feller central limit theorem to the double sequence $W_K^* = \sum_{k=1}^K W_{K,k}^*$. To that end, first note that, as each neighborhood contains at most a finite number of nodes, M , we can show that

$$\begin{aligned} \text{Var}(W_{K,k}^*) &= \sum_{i \in \mathcal{N}_K^d} \sum_{j \in \mathcal{N}_K^d} \mathbb{I}(i, j \in A_k^d) \text{Cov}(S_{Ki}^*, S_{Kj}^*) \\ &= \sum_{i \in \mathcal{N}_K^d} \sum_{j \in \mathcal{N}_K^d} \mathbb{I}(i, j \in A_k^d) \mathbb{E}(S_{Ki}^* S_{Kj}^*) \\ &\leq \sum_{i \in \mathcal{N}_K^d} \sum_{j \in \mathcal{N}_K^d} 1 \\ &\leq M^{2d}. \end{aligned} \quad (24)$$

Now we prove that Lyapounov's condition (25) holds for the constant in the exponent $\delta = 2$. So

$$\begin{aligned} \lim_{K \rightarrow \infty} \sum_{k=1}^K \frac{1}{\text{Var}(W_K^*)^2} \mathbb{E}(|W_{K,k}^*|^4) &= \lim_{K \rightarrow \infty} \frac{1}{\text{Var}(W_K^*)^2} \sum_{k=1}^K \mathbb{E}(W_{K,k}^{*2}) \mathbb{E}(W_{K,k}^{*2}) \\ &\leq \lim_{K \rightarrow \infty} \frac{M^{2d}}{\text{Var}(W_K^*)^2} \sum_{k=1}^K \mathbb{E}(W_{K,k}^*)^2 \\ &= \lim_{K \rightarrow \infty} \frac{M^{2d}}{\text{Var}(W_K^*)^2} \sum_{k=1}^K \text{Var}(W_{K,k}^*) \\ &= \lim_{K \rightarrow \infty} \frac{M^{2d}}{\text{Var}(W_K^*)} \\ &= 0. \end{aligned} \quad (25)$$

where $\text{Var}(W_K^*)$ tends to infinity by assumption. Therefore, Lyapounov's condition holds, and so by the Lindeberg-Feller central limit theorem, we have,

$$\frac{W_K^*}{\sqrt{\text{Var}(W_K^*)}} \xrightarrow{K \rightarrow \infty} N(0, 1). \quad (26)$$

Slutsky's theorem (3.4) gives the final result for $S_K^* = W_K^* + B_K^*$. Then we have

$$\frac{S_K^*}{\sqrt{\text{Var}(S_K^*)}} \xrightarrow{K \rightarrow \infty} N(0, 1), \quad (27)$$

as desired. \square

The second class of statistics are those that depend on both the graph and the nodal variates. These have a very similar form as the statistics previously considered. Now we require that

$$S_K = \sum_{i \in \mathcal{N}^d} S_{Ki},$$

where

$$S_{Ki} = \prod_{l,p \in i} Y_{lp} h(X_l, X_p),$$

a product with at most q terms.

Theorem 3.8. *If S_K is a statistic depending on both the random graph and the random nodal attributes, the sequence of random networks are as before, and the function h is uniformly bounded in the sense that, for all l and m , there is some B such that*

$$P(|h(X_l, X_m)|^p > B) = 0, \quad (p = 1, 2)$$

then we also have

$$S_K \xrightarrow[K \rightarrow \infty]{d} N(0, 1).$$

Proof. This proof is very similar to the proof of Theorem 3.7. We write

$$S_K = WK + BK,$$

exactly as before, incorporating the function h into each S_{Ki} as we did above. Then the binary nature of the graph and the uniform boundedness of h allow us to once again recenter the Y_{ij} , meaning that we will work with $V_{ij}h(X_i, X_j) = Y_{ij}h(X_i, X_j) - \mu_{ij}$. We also have $\mathbb{E}(V_{ij}h(X_i, X_j)) = 0$, so $\mathbb{E}(S_{Ki}^*) = 0$ as well.

For the between neighborhood covariances, we once again choose $V_{l_1 l_2}$, a between neighborhood network variable. Then we once again write

$$\begin{aligned} \text{Cov}(S_{Ki}^*, S_{Kj}^*) &= \mathbb{E} \left(\prod_{m,n \in i} V_{mn} h(X_m, X_n) \prod_{m,n \in j} V_{mn} h(X_m, X_n) \right) \\ &= \mathbb{E} \left((V_{l_1 l_2} h(X_{l_1}, X_{l_2}))^p \prod_{m,n \in (i \cup j) \setminus \{l_1, l_2\}} V_{mn} h(X_m, X_n) \right), \quad (p = 1, 2) \\ &= \mathbb{E}((V_{l_1 l_2} h(X_{l_1}, X_{l_2}))^p) \mathbb{E} \left(\prod_{m,n \in (i \cup j) \setminus \{l_1, l_2\}} V_{mn} h(X_m, X_n) \right), \end{aligned}$$

by the local dependence property. Then, when $p = 1$, we have $\mathbb{E}(V_{l_1 l_2} h(X_{l_1}, X_{l_2})) = 0$ by assumption, so the covariance is identically zero. When $p = 2$ we have

$$\mathbb{E}((V_{l_1 l_2} h(X_{l_1}, X_{l_2}))^2) \leq Cn^{-\delta}$$

by sparsity and

$$\mathbb{E} \left(\prod_{m,n \in (i \cup j) \setminus \{l_1, l_2\}} V_{mn} h(X_m, X_n) \right) \leq (DB)^{2q-2}$$

almost surely by uniform boundedness and the fact this product has at most $2q - 2$ terms. This follows from the fact that h is bounded by B and that V_{mn} is bounded by some constant D , by definition. So

$$\text{Cov}(S_{K_i}^*, S_{K_j}^*) \leq B^{2q-2} C n^{-\delta},$$

which tends to zero as K grows large. So, again by Chebyshev's inequality, we have

$$B_K^* \xrightarrow{K \rightarrow \infty} 0.$$

Next we bound the within neighborhood covariances. Now with each $|S_{K_i}^*| \leq B^q$, we have

$$\mathbb{I}(i, j \in A_k) |\text{Cov}(S_{K_i}^*, S_{K_j}^*)| \leq \mathbb{I}(i, j \in A_k) B^{2q}. \quad (28)$$

Now, we show that Lyapounov's condition (25) holds for the same $\delta = 2$. Once again note that each neighborhood has at most M nodes, so

$$\begin{aligned} \text{Var}(W_{K,k}^*) &= \sum_{i \in \mathcal{N}_K^d} \sum_{j \in \mathcal{N}_K^d} \mathbb{I}(i, j \in A_k) \text{Cov}(S_{K_i}^*, S_{K_j}^*) \\ &\leq \sum_{i \in \mathcal{N}_K^d} \sum_{j \in \mathcal{N}_K^d} \mathbb{I}(i, j \in A_k) B^{2q} \\ &\leq M^{2d} B^{2q}. \end{aligned}$$

Then Lyapounov's condition is

$$\begin{aligned} \lim_{K \rightarrow \infty} \sum_{k=1}^K \frac{1}{\text{Var}(W_K^*)^2} \mathbb{E}(|W_{K,k}^*|^4) &= \lim_{K \rightarrow \infty} \frac{1}{\text{Var}(W_K^*)^2} \sum_{k=1}^K \mathbb{E}(W_{K,k}^{*2}) \mathbb{E}(W_{K,k}^{*2}) \\ &\leq \lim_{K \rightarrow \infty} \frac{M^{2d} B^{2q}}{\text{Var}(W_K^*)^2} \sum_{k=1}^K \mathbb{E}(W_{K,k}^{*2}) \\ &= \lim_{K \rightarrow \infty} \frac{M^{2d} B^{2q}}{\text{Var}(W_K^*)^2} \sum_{k=1}^K \text{Var}(W_{K,k}^*) \\ &= \lim_{K \rightarrow \infty} \frac{M^{2d} B^{2q}}{\text{Var}(W_K^*)} \\ &= 0. \end{aligned} \quad (29)$$

Therefore, by the Lindeberg-Feller central limit theorem and Slutsky's theorem, we have

$$\frac{S_K^*}{\sqrt{\text{Var}(S_K^*)}} \xrightarrow{K \rightarrow \infty} N(0, 1). \quad \square$$

Finally, the last class of statistic is that which depends only on the nodal variates. This result follows directly from a central limit theorem for M -dependent random variables, which can be found in Billingsley (1995, p. 364). Establishing this theorem requires that we assume that the statistic in question depends only on a single variable across nodes. Therefore we assume that the statistic depends only on a single nodal covariate.

Theorem 3.9. *Take the sequence Z_K as before, and let X_K be the vector of nodal variates for each Z_K . Call each entry of this vector X_{K_i} , the variate corresponding to node i . Furthermore, we assume that $\mathbb{E}(X_{K_i}^{12}) < \infty$, $\mathbb{E}(X_{K_i} = 0)$. Then,*

$$\lim_{K \rightarrow \infty} \frac{\text{Var}(\sum_{i=1}^n X_{Ki})}{n} = \sigma^2,$$

where $n = |\mathcal{N}|$. Furthermore, if $\sigma > 0$, then

$$\frac{\sum_{i=1}^n X_{Ki}}{\sqrt{n \text{Var}(\sum_{i=1}^n X_{Ki})}} \xrightarrow{K \rightarrow \infty} N(0, 1).$$

Proof. Two random variables X_{Kl} and X_{Kp} are dependent if and only if l and p are in the same neighborhood. Without loss of generality, assume that the neighborhoods are such that all nodes within a neighborhood are indexed by consecutive integers. Then let $M = \limsup |A_K|$. Then the sequence X_{Ki} is M -dependent, so the result follows by application of Theorem 27.4 in Billingsley (1995). \square

In practice, the hypothesis that the twelfth moment exists is satisfied for most reasonable distributional assumptions about nodal covariates. Furthermore, the assumption that all nodal variates have expectation zero can easily be satisfied by recentering the observed data. Finally, the delta method gives us an asymptotically normal distribution for a differentiable statistic of the nodal variate. The univariate nature of the statistic is a fundamental limitation of this approach, however I am unable to find an analogous multidimensional central limit theorem that would allow us to establish the asymptotic normality of a statistic of multiple nodal variates.

4 Discussion and examples

4.1 Examples of convergence

In this section we simulate locally dependent random networks and calculate a statistic from each class discussed in Chapter 3. These networks grow larger and larger throughout the simulation, with each having n vertices and $n/10$ neighborhoods, to explore the asymptotic properties of our statistics. The nodes have two attributes. The first is a group, coded as 0 or 1, that effects the network formation. Edges are more likely to form between vertices in the same group. The second is a random attribute, also coded as 0 or 1. For each vertex, the value of this attribute depends on the attributes of the other vertices that it is connected to. The attribute can be thought of as a contagious infection or behavior that spreads along edges of the graph. The C++ and R code used to generate these networks is shown in section ???. This sort of complex relationship between the vertices and the edges is exactly the kind of generative process that ERNMs hope to capture.

Figure 5 shows an illustration of a simulated network with 100 vertices. The coloring shows the neighborhood membership of each vertex, while the presence or absence of fill indicates the value of the simulated attribute. Note the clustering of the neighborhoods and the attribute values. This is exactly the kind of joint clustering the locally dependent ERNM models. Furthermore, this also makes very intuitive sense as a realistic network structure. For example, suppose the attribute we are considering is smoking. This network shows that people who are friends with smokers are more likely to smoke themselves. This was the analysis done by Fellows & Handcock (2012) when introducing the ERNM.

Adding local dependence to this model is what allows us to show central limit theorems, however. In Figures 6 and 7, we can see the desired convergence towards a normal distribution that is guaranteed by the theorems proved in Chapter 3. In Figure 7, the quantile-quantile plots become very linear when the number of nodes reaches 200 and the number of neighborhoods reaches 20. All of the skewness has disappeared by this point and the tails of the distributions are almost perfectly normal. Finally, note that the slope is becoming shallower as the number of vertices grows. This corresponds to the shrinking of the statistic's variance, which is exactly what we hope to see as we grow the sample size.

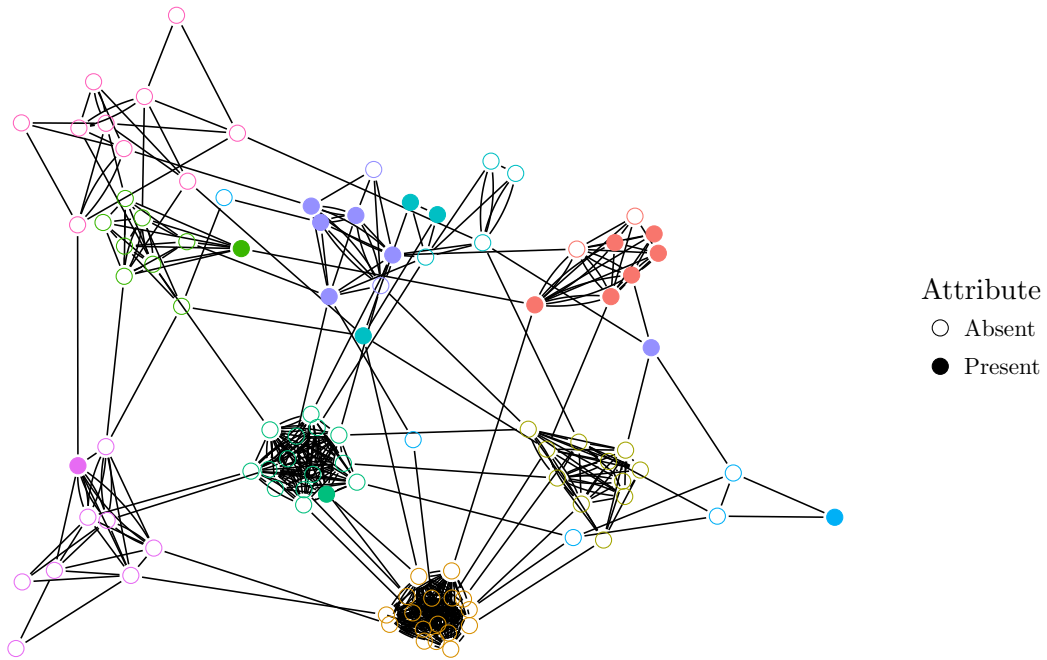


Figure 5: A simulated network with 100 vertices, colored according to neighborhood. The vertex fill indicates the presence or absence of the simulated attribute.

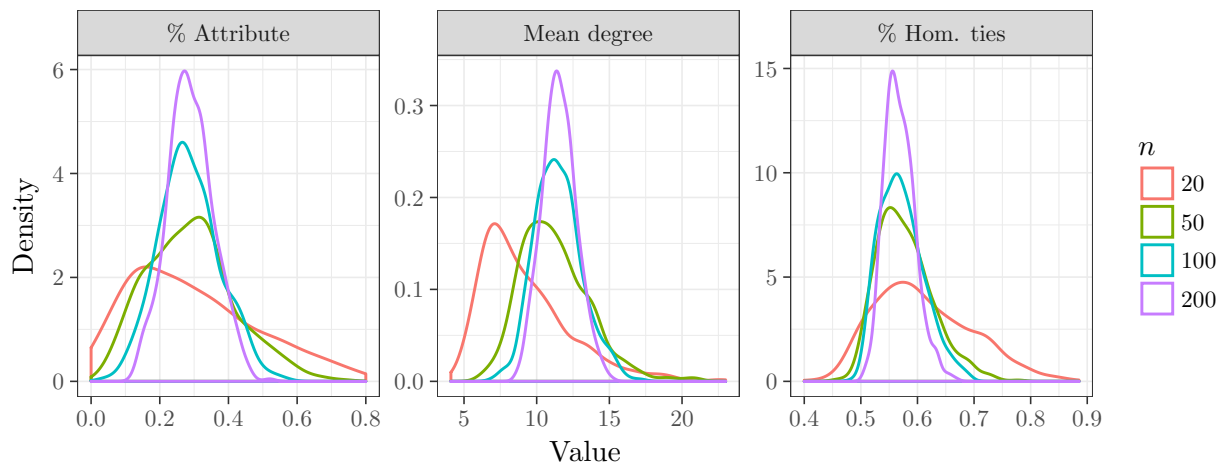


Figure 6: Density plots of simulated statistics of locally dependent random networks.

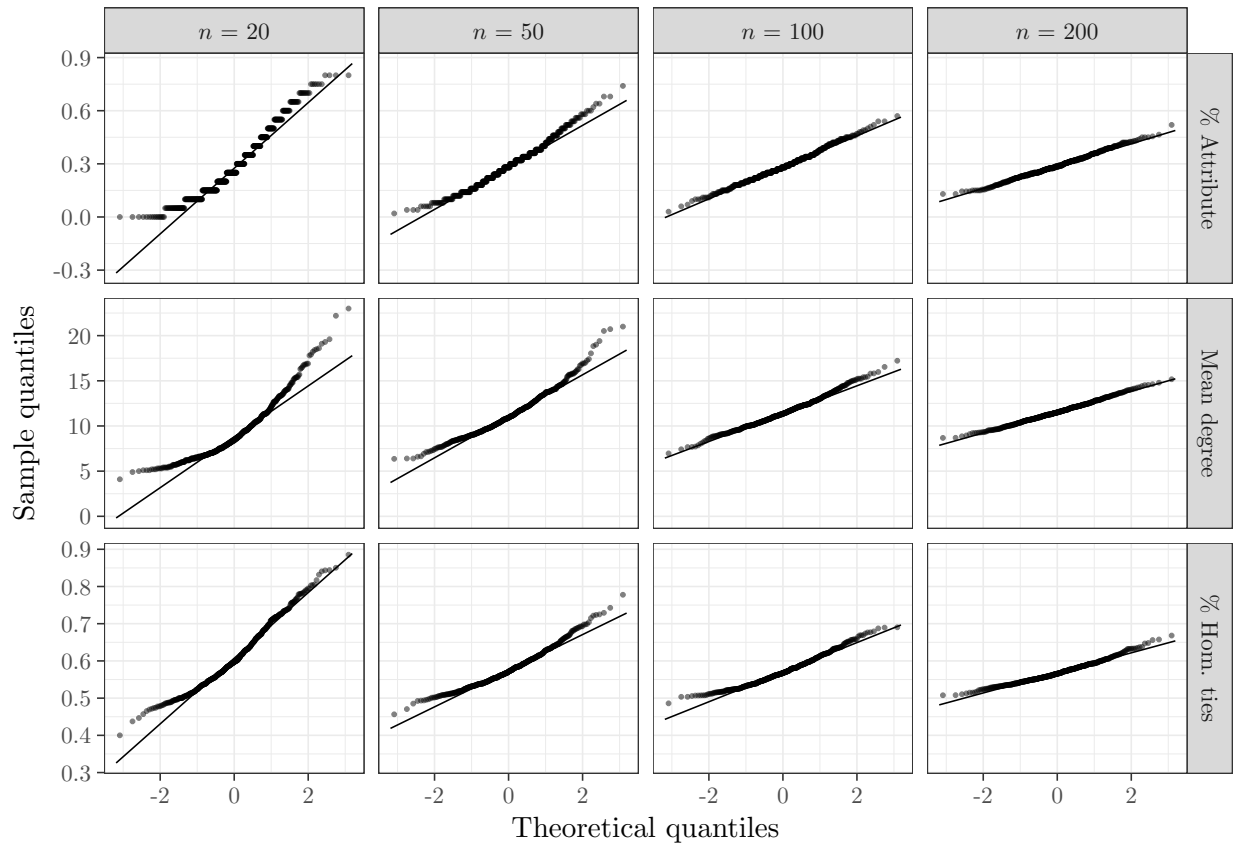


Figure 7: Normal Q-Q plots for simulated statistics of locally dependent random networks.

Table 1: Standard error estimates using the jackknife at the vertex level.

Statistic	Simulated standard error	Jackknife standard error
% Attribute	0.0666614	0.0203652
Mean degree	1.1684929	0.7662253
% Hom. ties	0.0269267	0.0184758

Table 2: Standard error estimates using the jackknife at the neighborhood level.

Statistic	Simulated standard error	Jackknife standard error
% Attribute	0.0666614	0.0677752
Mean degree	1.1684929	1.4487743
% Hom. ties	0.0269267	0.0372798

4.2 Standard error estimation

Given the asymptotic normality, all that remains to make these theorems useful for inference is to find a method to estimate the standard error of a statistic from a single observation. To that end, here we simulate a single network with 200 vertices by the same procedure as above and attempt to recover the standard errors of the statistics. We can compare these estimates to the standard errors of the empirical distributions we found in our previous simulation.

There are several approaches to approximating the standard errors. The first is to follow the vertex-level jackknife procedure outlined by Snijders & Borgatti (1999). This method is a modification of the standard jackknife estimator of standard deviation. At each iteration of the procedure, we remove one vertex from the network and recalculate the statistics of interest. Then we estimate the standard error with the equation

$$\widehat{\text{s.e.}} = \sqrt{\frac{n-2}{2n} \sum_{i=1}^n (g_{-i} - \bar{g})^2}, \quad (30)$$

where g is the quantity whose standard error we are approximating and g_{-i} is g calculated with vertex i removed. This differs from the standard jackknife estimator in the multiplicative constant. Heuristically, this constant accounts for the fact that we are seeing much less variation in the jackknife networks than we would expect to see in the true distribution. This follows from the fact that the variance of a statistic of the edge variables is (approximately) inversely proportional to the number of edge variables, $n(n-1)$, not the number of vertices, n (Snijders & Borgatti, 1999). We can see in Table 1 that this procedure underestimates the variance of our statistics.

This is most likely because of the dependence between vertices in the process that generated the network.

The second method is a standard jackknife, but leverages the neighborhood structure and the local dependence of the network. We do this by applying the jackknife at the level of the neighborhoods. That is to say, we remove each neighborhood in turn and recompute the statistic and then compute the jackknife estimator

$$\widehat{\text{s.e.}} = \sqrt{\frac{m-1}{m} \sum_{i=1}^m (g_{-i} - \bar{g})^2}, \quad (31)$$

where m is the number of neighborhoods. This process gives much more appropriate estimates, shown in Table 2.

4.3 Discussion of future work

Further development of locally dependent ERNMs for modelling complex social processes will have major impacts in the social sciences. The most pressing issue is the creation of software that allows for estimating parameters in these models. Code for fitting locally dependent ERGMs was developed by Schweinberger & Handcock (2015) and a Markov chain Monte Carlo algorithm for ERNMs was implemented by Fellows & Handcock (2012). However, it still remains to combine these two developments to work with locally dependent ERNMs.

Further mathematical work is also required to extend Theorem 3.9 to statistics involving more than one vertex attribute. This proof requires a multivariate analogue of the central limit theorem for M -dependent random variables. As my adviser told me, where there is a univariate central limit theorem, there is invariably a corresponding multidimensional theorem. However, we are unable to find a proof of this, so that result must be the topic of future work.

Finally, an investigation into the properties of standard error estimates like those produced above is much needed. Network sample sizes tend to be relatively small, so being forced to aggregate at the neighborhood level is a nontrivial issue. Furthermore, the jackknife procedure is most likely not making full use of all the information available within the observed network. However, a bootstrap procedure is, as far as my research shows, undefined for network data. Difficulty comes from being unable to sample with replacement. If one were to sample a network's vertices or edges with replacement, most likely duplicate edges would appear in the resampled networks. It is unclear how duplicates should be handled when calculating network statistics, so that sort of procedure is difficult to define.

References

- Billingsley, P. (1995). *Probability and measure* (3. ed., authorized reprint). New Delhi: Wiley India.
- Caimo, A., & Friel, N. (2014). Bergm: Bayesian exponential random graphs in R. *Journal of Statistical Software*, *61*(2). Retrieved from <http://www.jstatsoft.org/v61/i02/>
- Eddelbuettel, D. (2013). *Seamless R and C++ integration with rcpp*. New York: Springer.
- Eddelbuettel, D., & Francois, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, (40, 8).
- Fellows, I., & Handcock, M. S. (2012). Exponential-family random network models. *ArXiv Preprint ArXiv:1208.0121*.
- Groendyke, C., Welch, D., & Hunter, D. R. (2012). A network-based analysis of the 1861 Haggeloch measles data. *Biometrics*, *68*(3), 755–765. <http://doi.org/10.1111/j.1541-0420.2012.01748.x>
- Ismay, C. (2017). *thesisdown: An updated R Markdown thesis template using the bookdown package*.
- Robinson, D. (2016). *Broom: Convert statistical analysis objects into tidy data frames*. Retrieved from <https://CRAN.R-project.org/package=broom>
- Schweinberger, M., & Handcock, M. S. (2015). Local dependence in random graph models: Characterization, properties and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *77*(3), 647–676.
- Shalizi, C. R., & Rinaldo, A. (2013). Consistency under sampling of exponential random graph models. *The Annals of Statistics*, *41*(2), 508–535. <http://doi.org/10.1214/12-AOS1044>
- Snijders, T. A. B., & Borgatti, S. P. (1999). Non-parametric standard errors and tests for network statistics. *Connections*, *22*(2), 61–70.
- Wasserman, L. (2004). *All of statistics*. New York: Springer-Verlag.
- Wasserman, S., & Pattison, P. (1996). Logit models and logistic regressions for social networks: I. an introduction to markov graphs and p*. *Psychometrika*, *61*(3), 401–425.
- Wickham, H. (2016a). *Purrr: Functional programming tools*. Retrieved from <https://CRAN.R-project.org/package=purrr>
- Wickham, H. (2016b). *Tidyverse: Easily install and load 'tidyverse' packages*. Retrieved from <https://CRAN.R-project.org/package=tidyverse>
- Xie, Y. (2016). knitr: A general-purpose package for dynamic report generation in R (Version R package version 1.15.19).