Bernoulli **32**(1), 2026, 127–152
https://doi.org/10.3150/25-BEJ1852

Rates of convergence and normal approximations for estimators of local dependence random graph models

JONATHAN R. STEWART[1,a]

[1]*Department of Statistics, Florida State University, Tallahassee, FL, USA,* [a]*jrstewart@fsu.edu*

Local dependence random graph models are a class of block models for network data which allow for dependence among edges under a local dependence assumption defined around the block structure of the network. Since being introduced by Schweinberger and Handcock (*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** (2015) 647–676), research in the statistical network analysis and network science literatures have demonstrated the potential and utility of this class of models. In this work, we provide the first theory for estimation and inference which ensures consistent and valid inference of parameter vectors of local dependence random graph models. This is accomplished by deriving convergence rates of estimation and inference procedures for local dependence random graph models based on a single observation of the graph, allowing both the number of model parameters and the sizes of blocks to tend to infinity. First, we derive non-asymptotic bounds on the ℓ_2-error of maximum likelihood estimators with convergence rates, outlining conditions under which these rates are minimax optimal. Second, and more importantly, we derive non-asymptotic bounds on the error of the multivariate normal approximation. These theoretical results are the first to achieve both optimal rates of convergence and non-asymptotic bounds on the error of the multivariate normal approximation for parameter vectors of local dependence random graph models.

Keywords: Local dependence random graph model; minimax bounds; multivariate normal approximation; network data; statistical network analysis

1. Introduction

Local dependence random graph models, introduced by Schweinberger and Handcock (2015), are a class of statistical models for network data built around block structure, where a population of nodes $\mathscr{N}$, which we take without loss to be $\mathscr{N} \coloneqq \{1, \ldots, N\}$ ($N \geq 3$), is partitioned into $K \in \{1, 2, \ldots\}$ subsets $\mathscr{A}_1, \ldots, \mathscr{A}_K$ called blocks (also referred to as communities or subpopulations within the literature). The class owes its name to the fundamental assumption that dependence among edges is constrained to block-based subgraphs. We formally review local dependence random graph models in Section 1.1.

There are two key aspects to local dependence random graph models which help to explain the research interest received in both the statistical network analysis and network science literatures (Agneessens, Trincado-Munoz and Koskinen, 2024, Babkin et al., 2020, Dahbura et al., 2021, 2023, Mele, 2022, Schweinberger and Stewart, 2020, Stewart et al., 2019, Tolochko and Boomgaarden, 2024, Whetsell, Kroll and Dehart-Davis, 2021). First, block structure (or community structure) is a well-established structural phenomena relevant to many applications and networks encountered in our world (e.g., Holland, Laskey and Leinhardt, 1983, Newman and Girvan, 2004, Stewart et al., 2019). Second, local dependence random graph models possess desirable properties and behavior that circumvent early difficulties in constructing models of edge dependence, which include producing non-degenerate models of edge dependence (including transitivity) and consistency results for estimators (Schweinberger and Handcock, 2015, Schweinberger and Stewart, 2020).

Utilization of local dependence random graph models requires knowledge or estimates of both the block memberships of the nodes in the network, as well as the parameters of interest which determine the amount of probability mass placed on different configurations of the network. In practice, the parameter vectors of local dependence random graph models must always be estimated, whereas the block memberships of nodes can either be observed as part of the observation process (Schweinberger and Stewart, 2020, Stewart et al., 2019), or can be estimated (Babkin et al., 2020, Schweinberger, 2020). We will focus on the problem of estimating parameter vectors under the assumption that the block memberships of nodes have either been observed as part of the observation process or have been estimated.

In this work, we advance the literature on local dependence random graph models by providing the first statistical theory which elaborates conditions under which estimation and inference methodology based on a single observation of the graph can be expected to produce consistent and valid inference of parameter vectors of local dependence random graph models. The main results are non-asymptotic and cover settings where the number of model parameters and the sizes of the blocks tend to infinity. The main contributions of this work include:

1. Establishing the first non-asymptotic bounds on the ℓ_2-error of maximum likelihood estimators of parameters vectors of local dependence random graph models which hold with high probability,
2. Outlining conditions under which the rates of convergence implied by the bounds on the ℓ_2-error of maximum likelihood estimators are minimax optimal, and
3. Deriving the first non-asymptotic bound on the error of the multivariate normal approximation of a standardization of maximum likelihood estimators.

All results are stated in terms of interpretable quantities, allowing us to quantify the effect of key aspects of the statistical model and network structure upon convergence rates of the aforementioned errors. In so doing, we introduce the first principled approach to estimation and inference for local dependence random graph models by developing theoretical results which achieve both optimal rates of convergence and non-asymptotic bounds on the error of the multivariate normal approximation of maximum likelihood estimators.

1.1. Local dependence random graph models

We consider simple, undirected random graphs $\boldsymbol{X} \in \mathbb{X} \coloneqq \{0,1\}^{\binom{N}{2}}$ which are defined on the set of nodes $\mathscr{N} \coloneqq \{1,\ldots,N\}$ ($N \geq 3$). Edge variables between pairs of nodes $\{i,j\} \subset \mathscr{N}$ are given by

$$X_{i,j} \quad = \quad \begin{cases} 1 & \text{Nodes } i \text{ and } j \text{ are connected in the graph} \\ 0 & \text{Otherwise,} \end{cases}$$

assuming throughout that $X_{i,j} = X_{j,i}$ ($\{i,j\} \subset \mathscr{N}$) and $X_{i,i} = 0$ ($i \in \mathscr{N}$).

A *local dependence random graph* (Schweinberger and Handcock, 2015) is a random graph $\boldsymbol{X}$ where the set of nodes $\mathscr{N}$ is partitioned into K *blocks* $\mathscr{A}_1,\ldots,\mathscr{A}_K$ with probability distributions $\mathbb{P}$ of the form

$$\mathbb{P}(\boldsymbol{X}=\boldsymbol{x}) \quad = \quad \prod_{1\leq k\leq l\leq K} \mathbb{P}_{k,l}(\boldsymbol{X}_{k,l}=\boldsymbol{x}_{k,l}), \qquad \boldsymbol{x} \in \mathbb{X}, \tag{1}$$

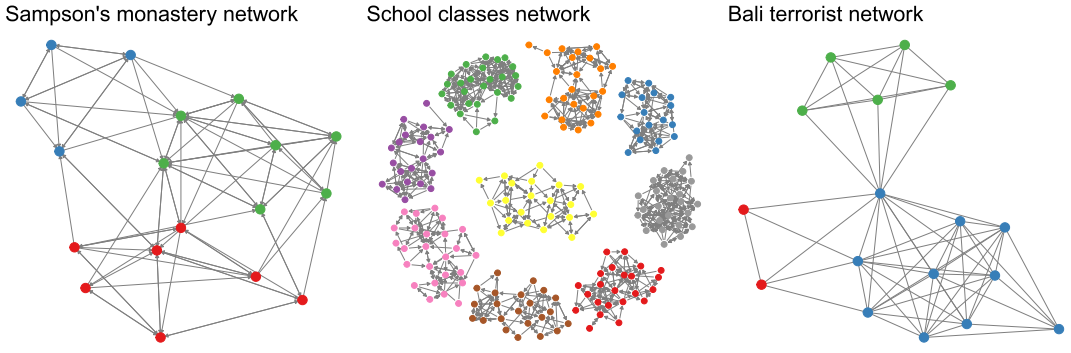

Figure 1. Three real data examples of networks for which local dependence random graph models would be applicable, including Sampson's monastery network, the school classes data set from Stewart et al. (2019), and the Bali terrorist network studied in Schweinberger and Handcock (2015). Node colors correspond to block memberships.

where the subgraphs $\boldsymbol{X}_{k,l}$ $(1 \leq k \leq l \leq K)$ are defined as follows:

$$\boldsymbol{X}_{k,l} \;\coloneqq\; \begin{cases} (X_{i,j})_{\{(i,j):\, i<j,\, i\in\mathscr{A}_k,\, j\in\mathscr{A}_k\}} \;\in\; \mathbb{X}_{k,k} \;\coloneqq\; \{0,1\}^{\binom{|\mathscr{A}_k|}{2}} & \text{if } k = l \\ (X_{i,j})_{\{(i,j):\, i\in\mathscr{A}_k,\, j\in\mathscr{A}_l\}} \;\in\; \mathbb{X}_{k,l} \;\coloneqq\; \{0,1\}^{|\mathscr{A}_k|\,|\mathscr{A}_l|} & \text{if } k \neq l. \end{cases}$$

We refer to the subgraphs $\boldsymbol{X}_{k,k}$ $(1 \leq k \leq K)$ as the *within-block subgraphs* and to the subgraphs $\boldsymbol{X}_{k,l}$ $(1 \leq k < l \leq K)$ as the *between-block subgraphs*. The probability distribution $\mathbb{P}_{k,l}$ is the marginal probability distribution of the subgraph $\boldsymbol{X}_{k,l}$ $(1 \leq k \leq l \leq K)$. A *local dependence random graph model* is any probability distribution $\mathbb{P}$ for $\boldsymbol{X}$ of the form (1). Figure 1 visualizes three networks which can be studied using local dependence random graph models. While the collection of block-based subgraphs $\boldsymbol{X}_{k,l}$ $(1 \leq k \leq l \leq K)$ are independent, edges within the same block-based subgraph can be dependent. The joint distribution $\mathbb{P}$ can be specified by specifying the marginal probability distributions $\mathbb{P}_{k,l}$ for the block-based subgraphs $\boldsymbol{X}_{k,l}$ $(1 \leq k \leq l \leq K)$.

It is worth noting that the block memberships are known in both Sampson's monastery network and the school classes network visualized in Figure 1, whereas the block memberships of the Bali terrorist network were estimated as in Schweinberger and Handcock (2015). When the block memberships correspond to tangible and observable quantities (e.g., school class memberships of students), data on the block memberships can be collected as part of the observation process. When this is not possible, the block memberships must be estimated, for example by using the two-step estimation methodology of Babkin et al. (2020), which estimates both the block memberships of nodes and the parameter vectors of local dependence random graph models.

Exponential families account for the most prevalent specifications of local dependence random graph models (e.g., Dahbura et al., 2021, Schubert and Brand, 2022, Schweinberger and Handcock, 2015, Stewart et al., 2019, Tolochko and Boomgaarden, 2024), indeed having been the statistical foundations for the class in the seminal work by Schweinberger and Handcock (2015). Moreover, exponential families provide a flexible statistical platform for constructing models of edge dependence in network data applications (Lusher, Koskinen and Robins, 2012, Schweinberger et al., 2020), and have been shown to be possess desirable statistical properties in local dependence random graph models, including the consistency of maximum likelihood estimators of canonical and curved exponential

families (Schweinberger and Stewart, 2020). An *exponential-family local dependence random graph model* can be specified via the marginal probability distributions of the block-based subgraphs:

$$\mathbb{P}_{k,l,\boldsymbol{\theta}_{k,l}}(\boldsymbol{X}_{k,l} = \boldsymbol{x}_{k,l}) \;=\; h_{k,l}(\boldsymbol{x}_{k,l}) \, \exp\left(\langle \boldsymbol{\theta}_{k,l}, \, s_{k,l}(\boldsymbol{x}_{k,l})\rangle - \psi_{k,l}(\boldsymbol{\theta}_{k,l})\right), \tag{2}$$

defined for each $\boldsymbol{x}_{k,l} \in \mathbb{X}_{k,l}$, where

- $s_{k,l} : \mathbb{X}_{k,l} \mapsto \mathbb{R}^{p_{k,l}}$ is a vector of sufficient statistics;
- $\boldsymbol{\theta}_{k,l} \in \mathbb{R}^{p_{k,l}}$ is the natural parameter vector;
- $h_{k,l} : \mathbb{X}_{k,l} \mapsto [0, \infty)$ is the reference function of the exponential family; and
- $\psi_{k,l}(\boldsymbol{\theta}_{k,l}) = \log \sum_{\boldsymbol{v} \in \mathbb{X}_{k,l}} h_{k,l}(\boldsymbol{v}) \, \exp(\langle \boldsymbol{\theta}_{k,l}, \, s_{k,l}(\boldsymbol{v})\rangle)$ is the log-normalizing constant.

It is straightforward to show that exponential family specifications of the marginal probability distributions of the within-block and between-block subgraphs will lead to a joint distribution which is also an exponential family.

A diverse range of models with the local dependence property in (1) can be constructed through different specifications of the sufficient statistics and reference functions. To allow for a general scope of well-structured models, we assume that the joint distributions of $\boldsymbol{X}$ take the form

$$\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{X} = \boldsymbol{x}) \;=\; \prod_{1 \leq k \leq l \leq K} \mathbb{P}_{k,l,\boldsymbol{\theta}_{k,l}}(\boldsymbol{X}_{k,l} = \boldsymbol{x}_{k,l}) \;=\; h(\boldsymbol{x}) \, \exp\left(\langle \boldsymbol{\theta}, \, s(\boldsymbol{x})\rangle - \psi(\boldsymbol{\theta})\right), \tag{3}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_W, \boldsymbol{\theta}_B) \in \mathbb{R}^{p+q}$ and $s(\boldsymbol{x}) = (s_W(\boldsymbol{x}_W), s_B(\boldsymbol{x}_B)) \in \mathbb{R}^{p+q}$, with the definitions

$$\boldsymbol{x}_W \coloneqq (\boldsymbol{x}_{1,1}, \ldots, \boldsymbol{x}_{K,K}), \qquad \boldsymbol{x}_B \coloneqq (\boldsymbol{x}_{1,2}, \ldots, \boldsymbol{x}_{1,K}, \boldsymbol{x}_{2,3}, \boldsymbol{x}_{2,4}, \ldots, \boldsymbol{x}_{K-1,K}),$$

$$h(\boldsymbol{x}) \coloneqq \prod_{1 \leq k \leq l \leq K} h_{k,l}(\boldsymbol{x}_{k,l}), \qquad \text{and} \qquad \psi(\boldsymbol{\theta}) \coloneqq \sum_{k=1}^{K} \psi_{k,k}(\boldsymbol{\theta}_W) + \sum_{1 \leq k < l \leq K} \psi_{k,l}(\boldsymbol{\theta}_B).$$

Throughout, we will assume that $p = \dim(\boldsymbol{\theta}_W)$ and $q = \dim(\boldsymbol{\theta}_B)$. The exponential family is then the set of probability distributions $\{\mathbb{P}_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \mathbb{R}^{p+q}\}$, where we note that the natural parameter space is equal to $\mathbb{R}^{p+q}$, a fact which follows trivially due to the fact that the support $\mathbb{X}$ of $\boldsymbol{X}$ is a finite set. We additionally assume throughout this work that the exponential family implied by (3) is minimal. While the product of the block-based subgraph distributions in (2) will form an exponential family, it may not be minimal, in which case we assume that the representation in (3) is the minimal representation of the exponential family obtained through reduction by sufficiency, reparameterization, and proper choice of reference measure; see Proposition 1.5 of Brown (1986). The assumption that an exponential family is minimal is not restrictive, as any non-minimal exponential family can be reduced to a minimal exponential family (Proposition 1.5, Brown, 1986).

We next provide examples of exponential-family local dependence random graph models in order to motivate the broad scope of this class of models, as well as to demonstrate how to construct local dependence random graph models. As the scope of possible models that can be constructed is large, we are unable to present a complete primer on the topic, and refer to works by Schweinberger and Handcock (2015), Stewart et al. (2019), and Schweinberger and Stewart (2020), for further information on and concrete examples of exponential-family local dependence random graph models.

1.2. Examples of exponential-family local dependence random graph models

1.2.1. Example 1: The stochastic block model

As a first example, we review the stochastic block model (Holland, Laskey and Leinhardt, 1983), which is a special case of a local dependence random graph model. The joint distribution for $\boldsymbol{X}$ is given by

$$\begin{aligned}\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{X}=\boldsymbol{x}) &\propto \left[\prod_{k=1}^{K}\prod_{i<j\,:\,i,j\in\mathscr{A}_k}\exp(\theta_{k,k}\,x_{i,j})\right]\left[\prod_{1\le k<l\le K}\prod_{(i,j)\in\mathscr{A}_k\times\mathscr{A}_l}\exp(\theta_{k,l}\,x_{i,j})\right] \\ &\propto \exp\left(\sum_{k=1}^{K}\theta_{k,k}\sum_{i<j\,:\,i,j\in\mathscr{A}_k}x_{i,j}+\sum_{1\le k<l\le K}\theta_{k,l}\sum_{(i,j)\in\mathscr{A}_k\times\mathscr{A}_l}x_{i,j}\right),\end{aligned} \tag{4}$$

where $\theta_{k,l}\in\mathbb{R}$ $(1\le k\le l\le K)$. The second line of (4) implies the minimal exponential family, where each block-based subgraph is a collection of independent and identically distributed Bernoulli random variables whose edge probability depends on the subgraph index (k,l) and the value of $\theta_{k,l}\in\mathbb{R}$.

1.2.2. Example 2: Transitivity in local dependence random graphs

The second example we present captures stochastic tendencies towards edge transitivity in networks, by including a sufficient statistic which models the stochastic tendency for an edge in the network to belong to a triangle. For this example, we consider joint distributions $\{\mathbb{P}_{\boldsymbol{\theta}}:\boldsymbol{\theta}\in\mathbb{R}^3\}$ for $\boldsymbol{X}$ of the form

$$\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{X}=\boldsymbol{x}) \propto \exp\big(\theta_1\,s_1(\boldsymbol{x})+\theta_2\,s_2(\boldsymbol{x})+\theta_3\,s_3(\boldsymbol{x})\big),$$

with natural parameters $(\theta_1,\theta_2,\theta_3)\in\mathbb{R}^3$, and where the sufficient statistics are given by

$$\begin{aligned}s_1(\boldsymbol{x}) &= \sum_{k=1}^{K}\sum_{i<j\,:\,i,j\in\mathscr{A}_k}x_{i,j} \\ s_2(\boldsymbol{x}) &= \sum_{k=1}^{K}\sum_{i<j\,:\,i,j\in\mathscr{A}_k}x_{i,j}\,\mathbb{1}\left(\sum_{h\in\mathscr{A}_k\setminus\{i,j\}}x_{i,h}\,x_{j,h}\ge 1\right) \\ s_3(\boldsymbol{x}) &= \sum_{1\le k<l\le K}\sum_{(i,j)\in\mathscr{A}_k\times\mathscr{A}_l}x_{i,j}.\end{aligned}$$

In words, $s_1(\boldsymbol{x})$ counts the number of edges in each of the within-block subgraphs $\boldsymbol{x}_{k,k}$ $(1\le k\le K)$, whereas $s_3(\boldsymbol{x})$ counts the number of edges in each of the between-block subgraphs $\boldsymbol{x}_{k,l}$ $(1\le k<l\le K)$. Neither of these statistics induce dependence, as when $\theta_2=0$, the joint distribution will factorize with respect to the edge variables in the graph, implying edges are independent.

The second sufficient statistic induces dependence among edge variables contained in the same within-block subgraph, noting that the form of the statistic in $s_2(\boldsymbol{x})$ ensures that distributions will not factorize with respect to the edge variables within the graph when $\theta_2\neq 0$. The second statistic $s_2(\boldsymbol{x})$ counts the number of edges $x_{i,j}$ between pairs of nodes i and j belonging to a common block $\mathscr{A}_k$ which are also mutually connected to at least one other node $h\in\mathscr{A}_k$ also belonging to the same common block, i.e., it counts the number of within-block edges which form at least one triangle within the respective block-based subgraph. We call such edges *transitive edges*.

Motivation for taking this approach to constructing models of edge dependence lies in foundational properties of exponential families. The mean-value parameter map of the exponential family is given by $\boldsymbol{\mu}(\boldsymbol{\theta}) \coloneqq \mathbb{E}_{\boldsymbol{\theta}}\left(s_1(\boldsymbol{X}), s_2(\boldsymbol{X}), s_3(\boldsymbol{X})\right)$ (p. 73–74, Brown, 1986), mapping the natural parameter space $\mathbb{R}^3$ to the interior of the convex hull of the image of $\mathbb{X}$ under the vector of sufficient statistics $s : \mathbb{X} \mapsto \mathbb{R}^3$:

$$\boldsymbol{\mu}(\boldsymbol{\theta}) \quad \in \quad \mathbb{M} \quad \coloneqq \quad \operatorname{int}\left(\operatorname{ConHull}\left(\left\{s(\boldsymbol{x}) \in \mathbb{R}^3 \,:\, \boldsymbol{x} \in \mathbb{X}\right\}\right)\right),$$

where $\operatorname{ConHull}(\mathscr{S})$ represents the convex hull of the set $\mathscr{S}$. Moreover, for a minimal exponential family (of which this example is), the map $\boldsymbol{\mu} : \mathbb{R}^3 \mapsto \mathbb{M}$ defines a homeomorphism between $\mathbb{R}^3$ and $\mathbb{M}$ (Theorem 3.6, Brown, 1986). This last point emphasizes a key modeling aspect of exponential-family local dependence random graph models, as for any point $\boldsymbol{u} \in \mathbb{M}$ parameterizing the expected values (mean values) of the sufficient statistics $(s_1(\boldsymbol{X}), s_2(\boldsymbol{X}), s_3(\boldsymbol{X}))$, we are guaranteed to be able to find a natural parameter vector $\boldsymbol{\theta} \in \mathbb{R}^3$ for which $\mathbb{E}_{\boldsymbol{\theta}}\left(s_1(\boldsymbol{X}), s_2(\boldsymbol{X}), s_3(\boldsymbol{X})\right) = \boldsymbol{u}$, allowing specified models to flexibly capture average tendencies of networks, including density, transitivity, and much more.

1.2.3. Example 3: Incorporating node and block heterogeneity into models

The third example shows how we are able to incorporate heterogeneous parameterizations for blocks, as well as for the stochastic propensities of different nodes to form edges, demonstrating ways in which the dimension of parameter vectors can grow in applications. For ease of presentation, we will build on Example 2 by extending the sufficient statistics which were specified in that example.

First, we will demonstrate how heterogeneity in node degrees can be incorporated into models. Suppose that nodes are divided into M non-overlapping groups or categories $\{1, \ldots, M\}$ which we represent as sets $\mathcal{G}_1, \ldots, \mathcal{G}_M$. Note that these are distinct from the blocks $\mathscr{A}_1, \ldots, \mathscr{A}_K$. In applications, these groups might be ranks in a department, gender, race, or any other categorical covariate which can be observed and treated as fixed. As such, each block may be comprised of different amount of nodes from each of the groups $\mathcal{G}_1, \ldots, \mathcal{G}_M$, an example of which is the school classes data set studied in Stewart et al. (2019), where each school class was comprised of different numbers of male and female students.

We replace $s_1(\boldsymbol{x})$ in Example 2 by multiple sufficient statistics:

$$s_m(\boldsymbol{x}) \quad = \quad \sum_{k=1}^{K} \sum_{i \in \mathcal{G}_m \cap \mathscr{A}_k} \sum_{j \in \mathscr{A}_k \setminus \{i\}} x_{i,j}, \qquad m \in \{1, \ldots, M\},$$

with natural parameters $(\theta_1, \ldots, \theta_M) \in \mathbb{R}^M$. A version of this statistic is implemented in the `R` package `ergm` under the name `nodefactor` (Krivitsky et al., 2023). In words, the model includes a sufficient statistic that, based on the value of the corresponding natural parameter, adjusts the baseline propensity for within-block edge formation involving nodes in that group. With no other sufficient statistics in the model, the log-odds of an edge would be given by

$$\log \frac{\mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = 1)}{\mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = 0)} \quad = \quad \theta_m + \theta_n, \qquad i \in \mathcal{G}_m \cap \mathscr{A}_k,\ j \in \mathcal{G}_n \cap \mathscr{A}_k, \qquad k \in \{1, \ldots, K\}.$$

This is reminiscent of the $p1$ model (Holland and Leinhardt, 1981) and the β-model (Chatterjee, Diaconis and Sly, 2011), in which each node is given its own distinct class.

We now show how heterogeneity can arise in the block-based subgraphs, by allowing different parameterizations for different blocks. The statistic $s_2(\boldsymbol{x})$ in Example 2 counts the number of transitive edges in each within-block subgraph $\boldsymbol{X}_{k,k}$ $(1 \le k \le K)$, using the same parameter for each within-block subgraph. It may be that different blocks display different tendencies towards transitivity. To make

this concrete, suppose that the individual blocks $\{1,\dots,K\}$ are partitioned into L groups or categories $\mathcal{H}_1,\dots,\mathcal{H}_L$. We then replace the sufficient statistic $s_2(\boldsymbol{x})$ in Example 2 by multiple statistics:

$$s_{M+l}(\boldsymbol{x}) \;=\; \sum_{k\in\mathcal{H}_l}\sum_{i<j\,:\,i,j\in\mathscr{A}_k} x_{i,j}\,\mathbb{1}\left(\sum_{h\in\mathscr{A}_k\setminus\{i,j\}} x_{i,h}\,x_{j,h}\;\geq\;1\right), \qquad l\in\{1,\dots,L\},$$

with natural parameters $(\theta_{M+1},\dots,\theta_{M+L})\in\mathbb{R}^L$. The complete model is given by

$$\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{X}=\boldsymbol{x}) \;\propto\; \exp\left(\sum_{t=1}^{M+L+1}\theta_t\,s_t(\boldsymbol{x})\right),$$

with natural parameter space $\mathbb{R}^{M+L+1}$, where the last sufficient statistic is equal to

$$s_{M+L+1}(\boldsymbol{x}) \;=\; \sum_{1\leq k<l\leq K}\;\sum_{(i,j)\in\mathscr{A}_k\times\mathscr{A}_l} x_{i,j}.$$

Example 3 helps to demonstrate how the number of model parameters can grow quickly in applications when significant generality, heterogeneity, or adaptability is needed to capture important aspects of the application. A version of Example 3 will be used in the simulation studies conducted in Section 3.

2. Theoretical guarantees

Our main theoretical results are presented in this section. We first review exponential family theory for local dependence random graph models in Section 2.1. Our consistency theory is then presented in Section 2.2. Section 2.2.1 derives rates of convergence in the ℓ_2-norm of maximum likelihood estimators, whereas Section 2.2.2 presents bounds on the minimax risk in the ℓ_2-norm which help to establish the minimax optimality (under mild conditions) of the upper bounds presented in Section 2.2.1. Lastly, but importantly, rates of convergence of the error of the multivariate normal approximation are obtained in Section 2.3, providing both non-asymptotic and asymptotic theory for multivariate normal approximations of maximum likelihood estimators of local dependence random graph models.

Due to space restrictions, all proofs are presented in the supplement (Stewart, 2026).

2.1. Preliminaries for exponential families

The log-likelihood of an exponential-family local dependence random graph model is

$$\ell(\boldsymbol{\theta},\boldsymbol{x}) \;\coloneqq\; \log\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{X}=\boldsymbol{x}) \;=\; \sum_{k=1}^{K}\ell_{k,k}(\boldsymbol{\theta}_W,\boldsymbol{x}_{k,k}) + \sum_{1\leq k<l\leq K}\ell_{k,l}(\boldsymbol{\theta}_B,\boldsymbol{x}_{k,l}),$$

where

$$\begin{aligned}\ell_{k,k}(\boldsymbol{\theta}_W,\boldsymbol{x}_{k,k}) &\coloneqq \langle\boldsymbol{\theta}_W,\,s_{k,k}(\boldsymbol{x}_{k,k})\rangle - \psi_{k,k}(\boldsymbol{\theta}_W) + \log h_{k,k}(\boldsymbol{x}_{k,k})\\ \ell_{k,l}(\boldsymbol{\theta}_B,\boldsymbol{x}_{k,l}) &\coloneqq \langle\boldsymbol{\theta}_B,\,s_{k,l}(\boldsymbol{x}_{k,l})\rangle - \psi_{k,l}(\boldsymbol{\theta}_B) + \log h_{k,l}(\boldsymbol{x}_{k,l}).\end{aligned}$$

The gradient $\nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}, \boldsymbol{x}) = (\nabla_{\boldsymbol{\theta}_W} \ell(\boldsymbol{\theta}, \boldsymbol{x}), \nabla_{\boldsymbol{\theta}_B} \ell(\boldsymbol{\theta}, \boldsymbol{x}))$ is given by

$$\begin{aligned} \nabla_{\boldsymbol{\theta}_W} \ell(\boldsymbol{\theta}, \boldsymbol{x}) &= \sum_{k=1}^{K} \left[s_{k,k}(\boldsymbol{x}_{k,k}) - \mathbb{E}_{k,k,\boldsymbol{\theta}_W} s_{k,k}(\boldsymbol{X}_{k,k}) \right] \\ \nabla_{\boldsymbol{\theta}_B} \ell(\boldsymbol{\theta}, \boldsymbol{x}) &= \sum_{1 \leq k < l \leq K} \left[s_{k,l}(\boldsymbol{x}_{k,l}) - \mathbb{E}_{k,l,\boldsymbol{\theta}_B} s_{k,l}(\boldsymbol{X}_{k,l}) \right], \end{aligned}$$

where $\mathbb{E}_{k,k,\boldsymbol{\theta}_W}$ and $\mathbb{E}_{k,l,\boldsymbol{\theta}_B}$ are the expectation operators with respect to the marginal probability distributions $\mathbb{P}_{k,k,\boldsymbol{\theta}_W}$ of $\boldsymbol{X}_{k,k}$ and $\mathbb{P}_{k,l,\boldsymbol{\theta}_B}$ of $\boldsymbol{X}_{k,l}$, respectively (Lemma 6.1, Stewart, 2026). We denote the set of maximum likelihood estimators for a given observation $\boldsymbol{x} \in \mathbb{X}$ by

$$\widehat{\boldsymbol{\Theta}} \equiv \widehat{\boldsymbol{\Theta}}(\boldsymbol{x}) \coloneqq \left\{ \boldsymbol{\theta}' \in \mathbb{R}^{p+q} : \ell(\boldsymbol{\theta}', \boldsymbol{x}) = \sup_{\boldsymbol{\theta} \in \mathbb{R}^{p+q}} \ell(\boldsymbol{\theta}, \boldsymbol{x}) \right\}.$$

For minimal and regular exponential families, the maximum likelihood estimator exists uniquely when it is exists, i.e., $|\widehat{\boldsymbol{\Theta}}| \in \{0, 1\}$ (Proposition 3.13, Sundberg, 2019). Regarding existence, the maximum likelihood estimator of natural parameter vectors of minimal exponential families exists when the sufficient statistic vector falls within the interior of the mean-value parameter space (Theorem 5.5., p. 148, Brown, 1986), in which case there exists a parameter vector $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^{p+q}$ for which $\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}}) = s(\boldsymbol{x})$ for a given observation $\boldsymbol{x} \in \mathbb{X}$ of the random graph $\boldsymbol{X}$, defining $\boldsymbol{\mu}(\boldsymbol{\theta}) \coloneqq \mathbb{E}_{\boldsymbol{\theta}}\, s(\boldsymbol{X})$ to be the mean-value parameter map (p. 73–74, Brown, 1986).

In practice, computing maximum likelihood estimators is not straightforward, as the log-normalizing constants are generally computationally intractable unless the marginal probability distributions of the block-based subgraphs $\boldsymbol{X}_{k,l}$ $(1 \leq k \leq l \leq K)$ are assumed to factorize further to reduce the computational burden of computing the normalizing constants, as both $\psi_{k,k}(\boldsymbol{\theta}_W)$ and $\psi_{k,l}(\boldsymbol{\theta}_B)$ involve summations of exponentially many terms in $\binom{|\mathscr{A}_k|}{2}$ and $|\mathscr{A}_k|\,|\mathscr{A}_l|$, respectively. It becomes infeasible to compute these summations in practice even for modest block sizes. The prevailing method for estimating exponential families of random graph models is Monte-Carlo maximum likelihood estimation (MCMLE) (Hunter and Handcock, 2006). The algorithm outlined in Hunter and Handcock (2006) applies directly to exponential-family local dependence random graph models (Schweinberger and Stewart, 2020, Stewart and Schweinberger, 2019, Stewart et al., 2019), and is used in the simulation studies conducted in Section 3 through the implementation in the `R` package `mlergm` (Stewart and Schweinberger, 2019).

We summarize the key aspects of MCMLE with exponential families of random graph models outlined in Hunter and Handcock (2006). The essential idea of MCMLE is to approximate intractable likelihood functions with stochastic approximations utilizing Markov Chain Monte Carlo (MCMC) methods. The crux of the methodology rests on a simple approximation of normalizing constants via importance sampling. To introduce the idea, let $\boldsymbol{\theta}_0 \in \mathbb{R}^{p+q}$ be a fixed parameter vector in the natural parameter space of an exponential-family local dependence random graph model. We can equivalently find maximum likelihood estimators $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}^\star$ by

$$\widehat{\boldsymbol{\theta}} = \operatorname*{arg\,max}_{\boldsymbol{\theta} \in \mathbb{R}^{p+q}} \left[\ell(\boldsymbol{\theta}, \boldsymbol{x}) - \ell(\boldsymbol{\theta}_0, \boldsymbol{x}) \right] = \operatorname*{arg\,max}_{\boldsymbol{\theta} \in \mathbb{R}^{p+q}} \left[\langle \boldsymbol{\theta} - \boldsymbol{\theta}_0,\, s(\boldsymbol{x}) \rangle - \log\left(\exp\left(\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0) \right) \right) \right].$$

In order to solve the above optimization problem, we need to be able to approximate the gradient corresponding to the above objective function, which is given by

$$\nabla_{\boldsymbol{\theta}} \left[\ell(\boldsymbol{\theta}, \boldsymbol{x}) - \ell(\boldsymbol{\theta}_0, \boldsymbol{x}) \right] = s(\boldsymbol{x}) - \nabla_{\boldsymbol{\theta}} \log\left(\exp\left(\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0) \right) \right). \tag{5}$$

The intractability of the normalizing constants in (5) makes direct computation infeasible, as discussed.

We approximate the term $\exp(\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0))$ via a change of measure argument:

$$\begin{aligned}
\exp(\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0)) &= \exp(-\psi(\boldsymbol{\theta}_0)) \sum_{\boldsymbol{x} \in \mathbb{X}} h(\boldsymbol{x}) \exp(\langle \boldsymbol{\theta}, s(\boldsymbol{x}) \rangle) \\
&= \exp(-\psi(\boldsymbol{\theta}_0)) \sum_{\boldsymbol{x} \in \mathbb{X}} h(\boldsymbol{x}) \exp(\langle \boldsymbol{\theta}, s(\boldsymbol{x}) \rangle) \frac{\exp(\langle \boldsymbol{\theta}_0, s(\boldsymbol{x}) \rangle)}{\exp(\langle \boldsymbol{\theta}_0, s(\boldsymbol{x}) \rangle)} \\
&= \mathbb{E}_{\boldsymbol{\theta}_0} \exp(\langle \boldsymbol{\theta} - \boldsymbol{\theta}_0, s(\boldsymbol{X}) \rangle),
\end{aligned}$$

where $\mathbb{E}_{\boldsymbol{\theta}_0}$ is the expectation operator corresponding to $\mathbb{P}_{\boldsymbol{\theta}_0}$. As a result, if we can approximate the expectation $\mathbb{E}_{\boldsymbol{\theta}_0} \exp(\langle \boldsymbol{\theta} - \boldsymbol{\theta}_0, s(\boldsymbol{X}) \rangle)$ via Monte Carlo methods, then we can approximate the ratio of normalizing constants $\exp(\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0))$. A key advantage of this approach lies in the fact that the expectation is taken with respect to a fixed distribution $\mathbb{P}_{\boldsymbol{\theta}_0}$. In general, we will not be able to sample directly from the distributions and will need to rely on MCMC sampling methods (see, e.g., Krivitsky et al., 2023, Snijders, 2002). Let $\widetilde{\boldsymbol{X}}_1, \ldots, \widetilde{\boldsymbol{X}}_n$ be an MCMC sample from $\mathbb{P}_{\boldsymbol{\theta}_0}$. Then, returning to (5), we have the following approximation:

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}} \left[\ell(\boldsymbol{\theta}, \boldsymbol{x}) - \ell(\boldsymbol{\theta}_0, \boldsymbol{x}) \right] &\approx s(\boldsymbol{x}) - \nabla_{\boldsymbol{\theta}} \log \left(\frac{1}{n} \sum_{i=1}^{n} \exp\left(\langle \boldsymbol{\theta} - \boldsymbol{\theta}_0, s(\widetilde{\boldsymbol{x}}_i) \rangle \right) \right) \\
&= s(\boldsymbol{x}) - \sum_{i=1}^{n} \left(\frac{\exp\left(\langle \boldsymbol{\theta} - \boldsymbol{\theta}_0, s(\widetilde{\boldsymbol{x}}_i) \rangle \right)}{\sum_{j=1}^{n} \exp\left(\langle \boldsymbol{\theta} - \boldsymbol{\theta}_0, s(\widetilde{\boldsymbol{x}}_j) \rangle \right)} \right) s(\widetilde{\boldsymbol{x}}_i).
\end{aligned}$$

Using this approximation, root finding algorithms—such as stochastic gradient descent or Fisher scoring algorithms—can be utilized to find the MCMLE approximation to the MLE; Hunter and Handcock (2006) outlines a stochastic Fisher scoring algorithm. The convergence of the MCMLE to the MLE depends on the convergence of the exact log-likelihood to the stochastic approximation (see, e.g., discussions in Geyer and Thompson, 1992), which will depend upon properties of the Markov chain utilized to generate sample networks. In the usual implementations (e.g., Krivitsky et al., 2023), these chains will be geometrically mixing toward the target sampling distribution and will provide good approximations provided sufficient computational resources have been expended. As a final point on the computational complexity, different model specifications, implementations of MCMC methodology, and block structures will have different mixing times and thus will require differing amounts of computational resources. With regards to scalability, access to parallel computing presents a significant opportunity to improve computation times by exploiting the independence of the block-based subgraphs to parallelize simulation; see discussions in Babkin et al. (2020), which analyzed networks with over 10,000 nodes utilizing parallel computing and an implementation of the stochastic Fisher scoring algorithm of Hunter and Handcock (2006) in the `R` package `mlergm` (Stewart and Schweinberger, 2019).

2.2. Convergence rates of maximum likelihood estimators

We derive non-asymptotic bounds on the ℓ_2-error of maximum likelihood estimators which hold with high probability. Our results extend those of Schweinberger and Stewart (2020), who derived consistency results for maximum likelihood estimators of canonical and curved exponential-family local dependence random graph models, but did not report rates of convergence. Additionally, Schweinberger and Stewart (2020) focused on estimation of only the within-block parameter vectors $\boldsymbol{\theta}_W$. In contrast, we establish consistency theory with rates of convergence for entire parameter vectors $(\boldsymbol{\theta}_W, \boldsymbol{\theta}_B)$ of exponential-family

local dependence random graph models, covering settings where the number of model parameters and sizes of blocks may tend to infinity, at appropriate rates. The consistency theory in this work is related to—but distinct from—the results in Stewart and Schweinberger (2020), who prove a general theorem for establishing consistency and rates of convergence of maximum likelihood and pseudolikelihood-based estimators of random graph models with dependent edges with respect to the ℓ_∞-norm under a more general weak dependence assumption. First, we focus specifically on local dependence random graph models and quantify rates of convergence in the ℓ_2-norm for this class of models and in terms of interpretable quantities related to local dependence random graphs, namely properties of the block structure, graph, and model. Second, our method of proof is fundamentally different from that of both Schweinberger and Stewart (2020) and Stewart and Schweinberger (2020), and the consistency theory in this work cannot be proved as a corollary to an existing result.

We outline some notational definitions and regularity assumptions for our theorems to follow, subsequently discussing each in turn. Let $\mathscr{B}_2(\boldsymbol{v}, r) \coloneqq \{\boldsymbol{v}' \in \mathbb{R}^{\dim(\boldsymbol{v})} : \|\boldsymbol{v}' - \boldsymbol{v}\|_2 < r\}$ be the open ℓ_2-ball with center $\boldsymbol{v}$ and radius $r > 0$ and denote by $\lambda_{\min}(\boldsymbol{A})$ and $\lambda_{\max}(\boldsymbol{A})$ the smallest and largest eigenvalues, respectively, of the matrix $\boldsymbol{A} \in \mathbb{R}^{d \times d}$. We write $a_N = O(b_N)$ when there exists a constant $C > 0$ and integer $N_0 \geq 1$ such that $a_N \leq C\, b_N$ for all $N \geq N_0$, and write $a_N = o(b_N)$ when there exists, for all $\delta > 0$, an integer $N_0(\delta) \geq 1$ such that $a_N \leq \delta\, b_N$ for all $N \geq N_0(\delta)$.

Assumption 1. Assume there exist $C_W > 0$ and $C_B > 0$, independent of N, p, and q, such that

$$\sup_{\boldsymbol{x}_{k,k} \in \mathbb{X}_{k,k}} \|s_{k,k}(\boldsymbol{x}_{k,k})\|_\infty \leq C_W \binom{|\mathscr{A}_k|}{2}, \qquad k \in \{1, \ldots, K\},$$

$$\sup_{\boldsymbol{x}_{k,l} \in \mathbb{X}_{k,l}} \|s_{k,l}(\boldsymbol{x}_{k,l})\|_\infty \leq C_B\, |\mathscr{A}_k|\, |\mathscr{A}_l|, \qquad \{k, l\} \subseteq \{1, \ldots, K\}.$$

Assumption 2. Assume there exists $\epsilon > 0$, independent of N, p, and q, such that

$$\widetilde{\lambda}^{\epsilon}_{\min,W} \coloneqq \inf_{\boldsymbol{\theta} \in \mathscr{B}_2(\boldsymbol{\theta}^\star, \epsilon)} \frac{\lambda_{\min}\left(-\mathbb{E}\, \nabla^2_{\boldsymbol{\theta}_W} \ell(\boldsymbol{\theta}, \boldsymbol{X})\right)}{K} > 0$$

$$\widetilde{\lambda}^{\epsilon}_{\min,B} \coloneqq \inf_{\boldsymbol{\theta} \in \mathscr{B}_2(\boldsymbol{\theta}^\star, \epsilon)} \frac{\lambda_{\min}\left(-\mathbb{E}\, \nabla^2_{\boldsymbol{\theta}_B} \ell(\boldsymbol{\theta}, \boldsymbol{X})\right)}{\binom{K}{2}} > 0.$$

Assumption 3. Define $A_{\text{avg}} \coloneqq K^{-1} \sum_{k=1}^{K} |\mathscr{A}_k|$ to be the average block size and

$$\widetilde{\lambda}^{\star}_{\max,W} \coloneqq \frac{\lambda_{\max}\left(-\mathbb{E}\, \nabla^2_{\boldsymbol{\theta}_W} \ell(\boldsymbol{\theta}^\star, \boldsymbol{X})\right)}{K} \quad \text{and} \quad \widetilde{\lambda}^{\star}_{\max,B} \coloneqq \frac{\lambda_{\max}\left(-\mathbb{E}\, \nabla^2_{\boldsymbol{\theta}_B} \ell(\boldsymbol{\theta}^\star, \boldsymbol{X})\right)}{\binom{K}{2}},$$

and assume that

$$\sqrt{A_{\text{avg}}}\, \frac{\sqrt{\widetilde{\lambda}^{\star}_{\max,W}}}{\widetilde{\lambda}^{\epsilon}_{\min,W}} = o\left(\sqrt{\frac{N}{p}}\right) \quad \text{and} \quad A_{\text{avg}}\, \frac{\sqrt{\widetilde{\lambda}^{\star}_{\max,B}}}{\widetilde{\lambda}^{\epsilon}_{\min,B}} = o\left(\sqrt{\frac{N^2}{q}}\right).$$

Assumption 4. Assume the largest block size $A_{\max} \coloneqq \max\{|\mathscr{A}_1|, \ldots, |\mathscr{A}_K|\}$ satisfies

$$A_{\max} \;\leq\; \min\left\{ \left(\frac{N\,\widetilde{\lambda}^{\star}_{\max,W}}{A_{\text{avg}}\,p^2} \right)^{1/4}, \quad \left(\frac{N^2\,\widetilde{\lambda}^{\star}_{\max,B}}{4\,A^2_{\text{avg}}\,q^2} \right)^{1/4} \right\}.$$

Remark 1 (Discussion of Assumption 1). We place a restriction on the scaling of the block-based sufficient statistic vectors with respect to the sizes of the blocks. The need for this arises out of a need to derive concentration inequalities for gradients of the log-likelihood, as well as a need to control third-order derivatives of the log-likelihood function in our method of proof for deriving bounds on the error of the multivariate normal approximation. The assumption is natural, as it essentially requires that the values of the sufficient statistics possess an upper-bound which is proportional to the number of edge variables in each of the respective block-based subgraphs. An example of interest is the transitive edge count statistic of a within-block subgraph $\boldsymbol{X}_{k,k}$, discussed also in in Section 1.2, given by

$$\sum_{i<j\,:\,i,j\in\mathscr{A}_k} x_{i,j}\,\mathbb{1}\left(\sum_{h\in\mathscr{A}_k\setminus\{i,j\}} x_{i,h}\,x_{j,h} \;\geq\; 1 \right) \;\leq\; \sum_{i<j\,:\,i,j\in\mathscr{A}_k} x_{i,j} \;\leq\; \binom{|\mathscr{A}_k|}{2},$$

which can be viewed as a special case of the geometrically-weighted edgewise shared partner statistic (Hunter and Handcock, 2006, Stewart et al., 2019). To further contextualize this assumption, it is helpful to note that Assumption 1 is related to the issue of instability of exponential-families of random graph models (Schweinberger, 2011). Maximal changes in the sufficient statistic vectors $s_{k,k}(\boldsymbol{x})$ $(1 \leq k \leq K)$ and $s_{k,l}(\boldsymbol{x})$ $(1 \leq k < l \leq K)$ due to changing the value of a single edge in $\boldsymbol{x}$ are defining characteristics of instability in exponential-family random graph models, in the sense of Schweinberger (2011). Assumption 1 implies limitations on the sensitivity of the sufficient statistic vectors to changes in the edges in the graph. Understanding this connection helps to explain why local dependence random graph models achieve statistical behavior and properties not achieved in early—but flawed—statistical models of edge dependence in network data (Chatterjee and Diaconis, 2013, Häggström and Jonasson, 1999, Jonasson, 1999, Schweinberger, 2011). Lastly, it is worth noting that Assumption 1 could be relaxed further, allowing for a larger upper bound. The result of this, however, would be looser upper-bounds on the ℓ_2-error and slower rates of convergence.

Remark 2 (Discussion of Assumption 2). Assumption 2 places a restriction on the scaling of the smallest eigenvalue of the joint Fisher information matrix by placing an assumption on the scaling of the smallest eigenvalue of the Fisher information matrices $-\mathbb{E}\,\nabla^2_{\boldsymbol{\theta}_W}\,\ell(\boldsymbol{\theta},\boldsymbol{X})$ and $-\mathbb{E}\,\nabla^2_{\boldsymbol{\theta}_B}\,\ell(\boldsymbol{\theta},\boldsymbol{X})$ corresponding to the within-block and between-block probability distributions, respectively, in a neighborhood $\mathscr{B}_2(\boldsymbol{\theta}^{\star},\epsilon)$ of the data-generating parameter vector $\boldsymbol{\theta}^{\star} = (\boldsymbol{\theta}^{\star}_W, \boldsymbol{\theta}^{\star}_B)$. The local dependence assumption and the assumption that the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_W, \boldsymbol{\theta}_B) \in \mathbb{R}^{p+q}$ partitions the within-block and between-block parameters implies that the joint Fisher information matrix $-\mathbb{E}\,\nabla^2_{\boldsymbol{\theta}}\,\ell(\boldsymbol{\theta},\boldsymbol{X})$ has the form

$$-\mathbb{E}\,\nabla^2_{\boldsymbol{\theta}}\,\ell(\boldsymbol{\theta},\boldsymbol{X}) \;=\; \begin{pmatrix} -\mathbb{E}\,\nabla^2_{\boldsymbol{\theta}_W}\,\ell(\boldsymbol{\theta},\boldsymbol{X}) & \mathbf{0}_{p,q} \\ \mathbf{0}_{q,p} & -\mathbb{E}\,\nabla^2_{\boldsymbol{\theta}_B}\,\ell(\boldsymbol{\theta},\boldsymbol{X}) \end{pmatrix},$$

where $\mathbf{0}_{m,n}$ is the $(m \times n)$-dimensional matrix of all zeros. Assumption 2 essentially assumes that the Fisher information matrices are invertible in a neighborhood of the data-generating parameter vector. Minimum eigenvalue restrictions of Fisher information matrices are standard in settings where the number of model parameters may tend to infinity (e.g., Janková and van de Geer, 2018, Portnoy, 1988,

Ravikumar, Wainwright and Lafferty, 2010). Notably, our assumption represents a restriction on what amounts to an average minimum eigenvalue (averaged over the block-based quantities in both the within-block and between-block cases). To understand why we have adopted this definition in our assumptions (relevant also to Assumption 3), instead of placing a restriction on the minimum eigenvalue of the Fisher information matrices corresponding to each block-based subgraph, observe through Weyl's inequality, the bound

$$\begin{aligned}\lambda_{\min}\left(-\mathbb{E}\,\nabla^2_{\boldsymbol{\theta}_W}\,\ell(\boldsymbol{\theta},\boldsymbol{X})\right) &\geq \sum_{k=1}^{K}\lambda_{\min}\left(-\mathbb{E}\,\nabla^2_{\boldsymbol{\theta}_W}\,\ell_{k,k}(\boldsymbol{\theta}_W,\boldsymbol{X}_{k,k})\right)\\ &\geq K\left(\min_{k\in\{1,\ldots,K\}}\lambda_{\min}\left(-\mathbb{E}\,\nabla^2_{\boldsymbol{\theta}_W}\,\ell_{k,k}(\boldsymbol{\theta}_W,\boldsymbol{X}_{k,k})\right)\right).\end{aligned}$$

If certain subgraphs do not contain any information about certain subsets of parameters, possibly due to heterogeneous parameterizations that allow different blocks to have different parameters, then it may be the case that $\lambda_{\min}(-\mathbb{E}\,\nabla^2_{\boldsymbol{\theta}_W}\,\ell_{k,k}(\boldsymbol{\theta}_W,\boldsymbol{X}_{k,k}))=0$ for some $k\in\{1,\ldots,K\}$ due to singularity. As a result and in order to cover more general settings and heterogeneous parameterizations, we place our minimum eigenvalue restriction on the scaling of the averaged smallest eigenvalue of joint Fisher information matrices.

Remark 3 (Discussion of Assumption 3). Assumption 3 places a regularity assumption on three key quantities, the average block size $A_{\text{avg}} \coloneqq K^{-1}\sum_{k=1}^{K}|\mathscr{A}_k|$, the average minimum eigenvalues of Fisher information matrices $\widetilde{\lambda}^{\epsilon}_{\min,W}$ and $\widetilde{\lambda}^{\epsilon}_{\min,B}$ in a neighborhood of the data-generating parameter vector $\boldsymbol{\theta}^{\star}=(\boldsymbol{\theta}^{\star}_W,\boldsymbol{\theta}^{\star}_B)$ (defined in Assumption 2), and the average maximum eigenvalues of Fisher information matrices $\widetilde{\lambda}^{\star}_{\max,W}$ and $\widetilde{\lambda}^{\star}_{\max,B}$ at the data-generating parameter vector. As will be seen in Theorem 2.1, Assumption 3 essentially outlines a scaling requirement of these three quantities (in their respective cases) which ensures consistent estimation under Theorem 2.1, in the sense that the upper bounds on the ℓ_2-error in Theorem 2.1 will tend to zero as the size of the network N tends to infinity. As such, Assumption 3 can be viewed as a minimal information criterion which requires that we obtain sufficient information about the parameter vector $(\boldsymbol{\theta}^{\star}_W,\boldsymbol{\theta}^{\star}_B)$ from an observation of the random graph $\boldsymbol{X}$.

Remark 4 (Discussion of Assumption 4). In our method of deriving concentration inequalities, we bound factors involving the influence of edge variables in the random graph by the size of the largest block size, noting that dependence is restricted to block-based subgraphs whose size is dominated by functions of the largest block size. Similar approaches have been taken in Schweinberger and Stewart (2020). Notably, Assumption 4 does not assume that the sizes of blocks are fixed and allows these quantities to grow without bound. However, this assumption places a restriction on how large blocks can be in order to ensure that the derived concentration inequalities are sufficiently sharp to facilitate the development of the statistical theory of this work.

2.2.1. Upper bounds on the ℓ_2-error of maximum likelihood estimators

The first theoretical result we present establishes upper-bounds on the ℓ_2-error of maximum likelihood estimators for exponential-family local dependence random graph models which hold with high probability, presented in Theorem 2.1. This paves the way for establishing bounds on rates of convergence of maximum likelihood estimators with respect to the ℓ_2-norm. We will address the question of optimal rates of convergence in Section 2.2.2, where we outline a set of sufficient conditions for which we prove the upper bounds in Theorem 2.1 are minimax optimal, in the sense that the upper bounds derived in Theorem 2.1 match (up to an unknown constant) the minimax rate of convergence.

Theorem 2.1. *Consider a minimal exponential-family local dependence random graph model satisfying Assumptions 1, 2, 3, and 4 and assume that* $p = \dim(\boldsymbol{\theta}^\star_W) \geq \log N$ *and* $q = \dim(\boldsymbol{\theta}^\star_B) \geq \log N$. *Then there exist constants* $C > 0$ *and* $N_0 \geq 3$, *independent of* N, p, *and* q, *such that, with probability at least* $1 - N^{-2}$, *the maximum likelihood estimator* $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}_W, \widehat{\boldsymbol{\theta}}_B) \in \mathbb{R}^{p+q}$ *exists, is unique, and satisfies*

$$\|\widehat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}^\star_W\|_2 \leq C \sqrt{A_{\text{avg}}} \frac{\sqrt{\widetilde{\lambda}^\star_{\max,W}}}{\widetilde{\lambda}^\epsilon_{\min,W}} \sqrt{\frac{p}{N}}$$

$$\|\widehat{\boldsymbol{\theta}}_B - \boldsymbol{\theta}^\star_B\|_2 \leq C A_{\text{avg}} \frac{\sqrt{\widetilde{\lambda}^\star_{\max,B}}}{\widetilde{\lambda}^\epsilon_{\min,B}} \sqrt{\frac{q}{N^2}},$$

for all integers $N \geq N_0$.

Theorem 2.1 provides the foundation for establishing convergence rates in the ℓ_2-norm of maximum likelihood estimators of exponential-family local dependence random graph models. The assumption that the exponential family is minimal ensures uniqueness of the maximum likelihood estimator when it exists (Proposition 3.13, Sundberg, 2019). Rates of convergence will depend on

- the dimensions of the parameters vectors $p = \dim(\boldsymbol{\theta}^\star_W)$ and $q = \dim(\boldsymbol{\theta}^\star_B)$;
- the ratios $\sqrt{\widetilde{\lambda}^\star_{\max,W}} \,/\, \widetilde{\lambda}^\epsilon_{\min,W}$ and $\sqrt{\widetilde{\lambda}^\star_{\max,B}} \,/\, \widetilde{\lambda}^\epsilon_{\min,B}$; and
- the average block size A_{avg},

with rates of convergence depending on the scaling of these quantities with respect to N. Theorem 2.1 additionally provides a set of sufficient conditions for the event that the maximum likelihood estimator exists to occur with high probability. Related to discussions in Section 1.2, the maximum likelihood estimator exists in the event $s(\boldsymbol{X}) \in \mathbb{M}$, recalling the definition of $\mathbb{M}$ from Section 1.2 as the mean-value parameter space of the exponential family. The assumptions of Theorem 2.1 ensure that the probability of the event $s(\boldsymbol{X}) \in \mathbb{M}$ occurs with high probability, provided the network size N is sufficiently large. This event essentially requires that the sufficient statistic not fall on the boundary of the convex hull of the image of $\mathbb{X}$ under the vector of sufficient statistics $s : \mathbb{X} \mapsto \mathbb{R}^{p+q}$, i.e., $\partial\,\mathbb{M}$. The probabilities for any fixed network size N, however, will depend on both properties of the network and the model specification. With regards to the latter, significant heterogeneity, such as in Example 3 in Section 1.2, can result in a higher-dimensional parameter space and therefore sufficient statistic vector, which can increase the chance of the sufficient statistic vector falling on the boundary $\partial\,\mathbb{M}$, in which event the maximum likelihood estimator will not exist.

We permit both $\widetilde{\lambda}^\star_{\max,W}$ and $\widetilde{\lambda}^\star_{\max,B}$ to scale faster than $\widetilde{\lambda}^\epsilon_{\min,W}$ and $\widetilde{\lambda}^\epsilon_{\min,B}$, respectively, provided consistency is still established (i.e., provided Assumption 3 is met). Within the context of exponential families of growing dimension in classical settings of a random sample of independent and identically distributed random vectors, Portnoy (1988) and Ghosal (2000) obtain similar convergence rates, in their respective settings. Notably, Theorem 2.1 of Portnoy (1988) arrives at a similar scaling requirement for the minimum and maximum eigenvalues of Fisher information matrices. A key difference is that both works place third order assumptions on the models (see the assumptions of Theorem 2.1 of Portnoy (1988), and Theorem 2.1 of Ghosal (2000)). We avoid the need for such assumptions through the method of proof of Theorem 2.1, but require a smoothness condition on minimum eigenvalues of Fisher information matrices, as Assumptions 2 and 3 restrict the scaling of maximum eigenvalues of the Fisher

information matrix at the data-generating parameter vector $\boldsymbol{\theta}^\star$ relative to minimum eigenvalues of the same within a neighborhood $\mathscr{B}_2(\boldsymbol{\theta}^\star, \epsilon)$ of $\boldsymbol{\theta}^\star$. If we assume additional regularity in the spectrum of the Fisher information matrices by assuming that

$$\widetilde{\lambda}^\star_{\min,W} \coloneqq \frac{\lambda_{\min}\left(-\mathbb{E}\,\nabla^2_{\boldsymbol{\theta}_W}\,\ell(\boldsymbol{\theta}^\star, \boldsymbol{X})\right)}{K} = O\left(\widetilde{\lambda}^\epsilon_{\min,W}\right)$$

$$\widetilde{\lambda}^\star_{\min,B} \coloneqq \frac{\lambda_{\min}\left(-\mathbb{E}\,\nabla^2_{\boldsymbol{\theta}_B}\,\ell(\boldsymbol{\theta}^\star, \boldsymbol{X})\right)}{\binom{K}{2}} = O\left(\widetilde{\lambda}^\epsilon_{\min,B}\right),$$

then we could prove a corollary to Theorem 2.1 which establishes the upper bounds

$$\|\widehat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}^\star_W\|_2 \leq C\sqrt{A_{\text{avg}}}\,\frac{\sqrt{\widetilde{\lambda}^\star_{\max,W}}}{\widetilde{\lambda}^\star_{\min,W}}\sqrt{\frac{p}{N}}$$

$$\|\widehat{\boldsymbol{\theta}}_B - \boldsymbol{\theta}^\star_B\|_2 \leq C\,A_{\text{avg}}\,\frac{\sqrt{\widetilde{\lambda}^\star_{\max,B}}}{\widetilde{\lambda}^\star_{\min,B}}\sqrt{\frac{q}{N^2}},$$

which are more analogous to the results of Portnoy (1988). Related to other works within the statistical network analysis, our consistency results and rates of convergence have key connections to theoretical results for the β-model, for example those obtained in Shao et al. (2021), which includes convergence rates for parameters of the β-model in the ℓ_2-norm, and also the ℓ_∞- and ℓ_1-norms.

As a final point, Theorem 2.1 assumes that the block memberships are known, i.e., the blocks $\mathscr{A}_1, \ldots, \mathscr{A}_K$ are observed or estimated without error. In many cases, the block memberships can be observed through the observation process (e.g., Schweinberger and Stewart, 2020, Stewart et al., 2019). However, in certain settings this may not be possible and the block memberships must be estimated (e.g., Babkin et al., 2020, Schweinberger, 2020). In both cases, the results of Theorem 2.1 can be regarded as the estimation error of an oracle estimate with perfect knowledge or estimation of the block structure of the network. The impact of imperfect block membership knowledge on theoretical guarantees (whether through a noisy observation or error in the estimation of block memberships of nodes) is an open question for future research.

2.2.2. *Minimax risk in the ℓ_2-norm and optimal rates of convergence*

We next turn to the question of whether the upper bounds on the ℓ_2-error established in Theorem 2.1 are optimal, in the sense that they match (up to an unknown constant) the rates of convergence of the minimax risk in the ℓ_2-norm.

We define the minimax risk with respect to the ℓ_2-norm to be

$$\begin{aligned}\mathcal{R}_{W,N} &\coloneqq \inf_{\widehat{\boldsymbol{\theta}}_W}\sup_{\boldsymbol{\theta}\in\mathbb{R}^{p+q}}\mathbb{E}_{\boldsymbol{\theta}}\,\|\widehat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_W\|_2 \\ \mathcal{R}_{B,N} &\coloneqq \inf_{\widehat{\boldsymbol{\theta}}_B}\sup_{\boldsymbol{\theta}\in\mathbb{R}^{p+q}}\mathbb{E}_{\boldsymbol{\theta}}\,\|\widehat{\boldsymbol{\theta}}_B - \boldsymbol{\theta}_B\|_2.\end{aligned} \tag{6}$$

The method by which we establish lower bounds to the minimax risk in the ℓ_2-norm requires placing an assumptions on the average value of the largest eigenvalues of Fisher information matrices, similar

to the roles of $\widetilde{\lambda}^{\star}_{\max,W}$ and $\widetilde{\lambda}^{\star}_{\max,B}$ in Theorem 2.1, extended now to a neighborhood $\mathscr{B}_2(\boldsymbol{\theta}^{\star}, \epsilon)$ of $\boldsymbol{\theta}^{\star}$. Fix $\epsilon > 0$, independent of N, p, and q, and define

$$\begin{aligned} \widetilde{\lambda}^{\epsilon}_{\max,W} &\coloneqq \sup_{\boldsymbol{\theta} \in \mathscr{B}_2(\boldsymbol{\theta}^{\star}, \epsilon)} \frac{\lambda_{\max}\left(-\mathbb{E}\,\nabla^2_{\boldsymbol{\theta}_W}\,\ell(\boldsymbol{\theta}, \boldsymbol{X})\right)}{K} \\ \widetilde{\lambda}^{\epsilon}_{\max,B} &\coloneqq \sup_{\boldsymbol{\theta} \in \mathscr{B}_2(\boldsymbol{\theta}^{\star}, \epsilon)} \frac{\lambda_{\max}\left(-\mathbb{E}\,\nabla^2_{\boldsymbol{\theta}_B}\,\ell(\boldsymbol{\theta}, \boldsymbol{X})\right)}{\binom{K}{2}}. \end{aligned} \tag{7}$$

We first establish lower bounds to the minimax risks $\mathcal{R}_{W,N}$ and $\mathcal{R}_{B,N}$ in Theorem 2.2, which enable us to outline sufficient conditions for the upper bounds on the ℓ_2-error presented in Theorem 2.1 to achieve (up to an unknown constant) the minimax rates of convergence; see Corollary 2.4. In the following results, it is helpful to recall that $\widetilde{\lambda}^{\epsilon}_{\min,W}$ and $\widetilde{\lambda}^{\epsilon}_{\min,B}$ are defined in Assumption 2, $\widetilde{\lambda}^{\star}_{\max,W}$ and $\widetilde{\lambda}^{\star}_{\max,B}$ are defined in Assumption 3, and $\widetilde{\lambda}^{\epsilon}_{\max,W}$ and $\widetilde{\lambda}^{\epsilon}_{\max,B}$ are defined in (7).

Theorem 2.2 (Lower bound to the minimax risk). *Consider an exponential-family local dependence random graph model satisfying Assumption 2. Then there exist constants $C_1 > 0$ and $C_2 > 0$, independent of N, p, and q, such that the minimax risks $\mathcal{R}_{W,N}$ and $\mathcal{R}_{B,N}$ defined in (6) satisfy*

$$\mathcal{R}_{W,N} \;\geq\; C_1 \sqrt{\frac{A_{\text{avg}}}{\widetilde{\lambda}^{\epsilon}_{\max,W}}}\,\sqrt{\frac{p}{N}} \;\geq\; C_1 \left(\frac{\widetilde{\lambda}^{\epsilon}_{\min,W}}{\widetilde{\lambda}^{\epsilon}_{\max,W}}\right) \frac{\sqrt{\widetilde{\lambda}^{\star}_{\max,W}}}{\widetilde{\lambda}^{\epsilon}_{\min,W}}\,\sqrt{A_{\text{avg}}}\,\sqrt{\frac{p}{N}}$$

$$\mathcal{R}_{B,N} \;\geq\; C_2 \frac{A_{\text{avg}}}{\sqrt{\widetilde{\lambda}^{\epsilon}_{\max,B}}}\,\sqrt{\frac{q}{N^2}} \;\geq\; C_2 \left(\frac{\widetilde{\lambda}^{\epsilon}_{\min,B}}{\widetilde{\lambda}^{\epsilon}_{\max,B}}\right) \frac{\sqrt{\widetilde{\lambda}^{\star}_{\max,B}}}{\widetilde{\lambda}^{\epsilon}_{\min,B}}\,A_{\text{avg}}\,\sqrt{\frac{q}{N^2}},$$

provided $p = \dim(\boldsymbol{\theta}^{\star}_W) = O(N\,\widetilde{\lambda}^{\epsilon}_{\max,W})$ *and* $q = \dim(\boldsymbol{\theta}^{\star}_B) = O(N^2\,\widetilde{\lambda}^{\epsilon}_{\max,B})$.

The role of Assumption 2 in Theorem 2.2 is to ensure that both $\widetilde{\lambda}^{\epsilon}_{\min,W}$ and $\widetilde{\lambda}^{\epsilon}_{\min,B}$ are bounded away from 0, ensuring all of the lower bounds are well defined, whereas we assume

$$p \;=\; \dim(\boldsymbol{\theta}^{\star}_W) \;=\; O(N\,\widetilde{\lambda}^{\epsilon}_{\max,W}) \qquad \text{and} \qquad q \;=\; \dim(\boldsymbol{\theta}^{\star}_B) \;=\; O(N^2\,\widetilde{\lambda}^{\epsilon}_{\max,B}) \tag{8}$$

in order to satisfy a technical condition in the proof of Theorem 2.2. Under the assumption that the maximum eigenvalues of Fisher information matrices are bounded away from 0, the condition in (8) requires that $p = O(N)$ and $q = O(N^2)$, which places a much less stringent restriction on the dimensions of parameters vectors when compared with Assumption 3. Two sets of lower bounds are presented in Theorem 2.2, with the first being the most sharp, but unhelpful in our pursuit of studying whether the rates of convergence implied in Theorem 2.1 are minimax optimal. The second, though looser, set of bounds approximately match the upper bounds on the ℓ_2-error established in Theorem 2.1. Indeed, this second set of bounds allows us to establish conditions for such minimax optimality in Corollary 2.4 which is presented below.

Note that the lower bound to the minimax risk presented in Theorem 2.2 considers $\boldsymbol{\theta}^{\star} \in \mathbb{R}^{p+q}$. The fact that the parameter space is unbounded introduces no complications when deriving lower bounds; however, when turning to the problem of deriving an upper bound to the minimax risk, an unbounded

parameter space presents new challenges. The following theorem obtains upper bounds on the minimax risk with respect to the ℓ_2-norm in a neighborhood of the data-generating parameter vector.

Theorem 2.3 (Upper bound to the minimax risk). *Under the assumptions of Theorem 2.1, there exist constants $C_1 > 0$, $C_2 > 0$, and $N_0 \geq 3$, independent of N, p, and q, such that the minimax risks restricted to a local neighborhood $\mathscr{B}_2(\boldsymbol{\theta}^\star, \epsilon)$ of a point $\boldsymbol{\theta}^\star \in \mathbb{R}^{p+q}$ satisfy, for all integers $N \geq N_0$,*

$$\inf_{\widehat{\boldsymbol{\theta}}_W} \sup_{\boldsymbol{\theta} \in \mathscr{B}_2(\boldsymbol{\theta}^\star, \epsilon)} \mathbb{E}_{\boldsymbol{\theta}} \|\widehat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_W\|_2 \leq C_1 \frac{\sqrt{\widetilde{\lambda}^\star_{\max,W}}}{\widetilde{\lambda}^\epsilon_{\min,W}} \sqrt{A_{\text{avg}}} \sqrt{\frac{p}{N}}$$

$$\inf_{\widehat{\boldsymbol{\theta}}_B} \sup_{\boldsymbol{\theta} \in \mathscr{B}_2(\boldsymbol{\theta}^\star, \epsilon)} \mathbb{E}_{\boldsymbol{\theta}} \|\widehat{\boldsymbol{\theta}}_B - \boldsymbol{\theta}_B\|_2 \leq C_2 \frac{\sqrt{\widetilde{\lambda}^\star_{\max,B}}}{\widetilde{\lambda}^\epsilon_{\min,B}} A_{\text{avg}} \sqrt{\frac{q}{N^2}},$$

where $\epsilon > 0$ is the same as in Assumptions 2 and 3 and in (7).

The final result of this section is concerned with outlining a set of sufficient conditions which allow us to establish the minimax optimality of Theorem 2.1.

Corollary 2.4. *Under the assumptions of Theorems 2.1 and 2.2, and the assumption that*

$$\widetilde{\lambda}^\epsilon_{\max,W} = O\left(\widetilde{\lambda}^\epsilon_{\min,W}\right) \quad \text{and} \quad \widetilde{\lambda}^\epsilon_{\max,B} = O\left(\widetilde{\lambda}^\epsilon_{\min,B}\right), \tag{9}$$

the maximum likelihood estimators $\widehat{\boldsymbol{\theta}}_W$ and $\widehat{\boldsymbol{\theta}}_B$ achieve the minimax rate of convergence, in the sense that the upper bounds on the ℓ_2-error of $\widehat{\boldsymbol{\theta}}_W$ and $\widehat{\boldsymbol{\theta}}_B$ presented in Theorem 2.1 match (up to an unknown constant which is independent of N, p, and q) the lower bounds to the minimax risks in Theorem 2.2.

If the exponential-family local dependence random graph model satisfies (9), then Corollary 2.4 establishes the minimax optimality of the rates of convergence for maximum likelihood estimators implied via Theorem 2.1. Such an assumption is common in the high-dimensional statistics literature (e.g., Janková and van de Geer, 2018, Ravikumar, Wainwright and Lafferty, 2010), where it is common to assume that minimum and maximum eigenvalues of Fisher information matrices corresponding to the sampling distribution are bounded away from 0 and from above, respectively. We can interpret condition (9) similarly, however applied to the joint Fisher information for the entire collection of random variables in the random graph (in contrast to the sampling distribution from which a random sample is generated) and in a neighborhood $\mathscr{B}_2(\boldsymbol{\theta}^\star, \epsilon)$ of the data-generating parameter vector $\boldsymbol{\theta}^\star$.

2.3. Convergence rates of the multivariate normal approximation

A key challenge to any statistical analysis of network data is finding rigorous justification for statistical inference methodology. The main contributing factor to this challenge lies in the fact that statistical analyses of network data are typically in the setting of a single collection of dependent random variables without the benefit of replication. In other words, any statistical inference will be based on a single observation of a collection of dependent binary random variables. It is common for inference of model parameters in exponential-family random graph models to utilize the normal approximation for carrying out inference about estimated coefficients (e.g., Krivitsky et al., 2023, Lusher, Koskinen and Robins,

2012, Stewart et al., 2019). Except in select cases, these inferences are performed without rigorous theoretical justification, owing to the difficulty of obtaining theoretical results establishing the validity of the normal approximation in scenarios with a single observations of a collection of dependent binary random variables.

The dependence structure of local dependence random graph models facilitates proof of rigorous theoretical results justifying the normal approximation for estimators, and in this section, we obtain rates of convergence of the multivariate normal approximation in scenarios of increasing model dimension. It is worth noting that our results imply the univariate normal approximation, as multiple univariate tests are frequently utilized in applications (e.g., Stewart et al., 2019). Similarly to our consistency results presented in Theorem 2.1, the quality of the multivariate normal approximation will depend on key quantities related to the block structure, graph, and model specification.

Throughout, $\boldsymbol{Z}_d$ will denote a d-dimensional multivariate normal random vector with mean vector $\boldsymbol{0}_d$ (the d-dimensional vector of all zeros) and covariance matrix $\boldsymbol{I}_d$ (the d-dimensional identity matrix). The probability distribution of $\boldsymbol{Z}_d$ is denoted by Φ_d.

In order to establish our multivariate normal approximation theory, we leverage a multivariate Berry-Esseen theorem provided in Raič (2019), together with a Taylor expansion of the log-likelihood equation. Utilizing properties of exponential families, we are able to derive non-asymptotic bounds on the error of the multivariate normal approximation for a standardization of the maximum likelihood estimator, providing the first results which elaborate conditions under which the normal approximation is expected to produce valid inferences in local dependence random graph models.

Theorem 2.5. *Consider a minimal exponential-family local dependence random graph model satisfying Assumptions 1, 2, 3, and 4 and assume that* $p = \dim(\boldsymbol{\theta}^\star_W) \geq \log N$ *and* $q = \dim(\boldsymbol{\theta}^\star_B) \geq \log N$. *Then there exist constants* $C_1 > 0$, $C_2 > 0$, *and* $N_0 \geq 3$, *independent of* N, p, *and* q, *and a random vector* $\boldsymbol{\Delta} \in \mathbb{R}^{p+q}$ *such that, for all integers* $N \geq N_0$ *and measurable convex sets* $C \subset \mathbb{R}^{p+q}$,

$$\left|\mathbb{P}(I(\boldsymbol{\theta}^\star)^{1/2}\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star) + \boldsymbol{\Delta} \in C) - \Phi_d(\boldsymbol{Z}_d \in C)\right|$$

$$\leq \; C_1\,(p+q)^{1/4}\,A_{\max}^7 \left[\sqrt{\frac{p^3}{(\widetilde{\lambda}^\star_{\min,W})^3\,N}} + \sqrt{\frac{q^3}{(\widetilde{\lambda}^\star_{\min,B})^3\,N^2}}\right],$$

where the random vector $\boldsymbol{\Delta}$ *satisfies*

$$\mathbb{P}\left(\|\boldsymbol{\Delta}\|_2 \leq C_2\,A_{\max}^6 \sqrt{A_{\text{avg}}\,\frac{(\widetilde{\lambda}^\star_{\max,W})^2}{(\widetilde{\lambda}^\epsilon_{\min,W})^5}\,\frac{p^5}{N} + A_{\text{avg}}^2\,\frac{(\widetilde{\lambda}^\star_{\max,B})^2}{(\widetilde{\lambda}^\epsilon_{\min,B})^5}\,\frac{q^5}{N^2}}\right) \;\geq\; 1 - \frac{1}{N^2}.$$

The standardization $I(\boldsymbol{\theta}^\star)^{1/2}\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star)$ is of a familiar form in multivariate normal approximation settings. The quantity $\boldsymbol{\Delta}$ can be interpreted as an error term or a random perturbation, arising due to a Taylor approximation. While our result is stated for $I(\boldsymbol{\theta}^\star)^{1/2}\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star) + \boldsymbol{\Delta}$, an important aspect of Theorem 2.5 lies in establishing that the random perturbation $\boldsymbol{\Delta}$ to $I(\boldsymbol{\theta}^\star)^{1/2}\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star)$ is small (in the ℓ_2-norm) with high probability, justifying basing inferences and derivations of confidence regions on $I(\boldsymbol{\theta}^\star)^{1/2}\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star)$ in applications. Indeed, under mild assumptions (which we state below), it is straightforward to establish that $\|\boldsymbol{\Delta}\|_2$ converges almost surely to 0 as $N \to \infty$.

A remark is in order regarding the term $(p+q)^{1/4}$ in the upper bound on the error of the multivariate normal approximation in Theorem 2.5. Current results on multivariate Berry-Esseen bounds involve terms which are functions of the dimension of the random vector (Raič, 2019). Here, the total dimension

of the random vector is $p+q$, as we are proving the joint multivariate normality of a standardization of the entire vector of maximum likelihood estimators $(\widehat{\boldsymbol{\theta}}_W, \widehat{\boldsymbol{\theta}}_B)$ which has dimension $p+q$. In other words, we are unable to separate the error into two terms which are functions of only quantities based on within-block and between-block quantities, as was done in our consistency theory in Section 2.2.

Typically, both $I_W(\boldsymbol{\theta}^\star_W)^{1/2}$ and $I_B(\boldsymbol{\theta}^\star_B)^{1/2}$ will be unknown, but can be approximated in practice. We can approximate both $I_W(\boldsymbol{\theta}^\star_W)$ and $I_B(\boldsymbol{\theta}^\star_B)$ through Monte-Carlo methods, as Fisher information matrices of canonical exponential families are the covariance matrices of the sufficient statistics. This is a common approach to estimating the Fisher information matrix in the exponential-family random graph model literature, owing to the fact that models are frequently estimated via Monte-Carlo maximum likelihood estimation, which already requires simulating sufficient statistic vectors (e.g., Hunter and Handcock, 2006, Krivitsky et al., 2023), and discussed in Section 2.1.

Under an additional regularity assumption, we can simplify the bounds presented in Theorem 2.5.

Assumption 5. Assume that there exist constants $L > 0$ and $U > 0$ such that

$$0 \;<\; L \;\leq\; \min\left\{\widetilde{\lambda}^{\epsilon}_{\min,W},\, \widetilde{\lambda}^{\epsilon}_{\min,B}\right\} \;\leq\; \max\left\{\widetilde{\lambda}^{\star}_{\max,W},\, \widetilde{\lambda}^{\star}_{\max,B}\right\} \;\leq\; U, \tag{10}$$

for all values of N, p, and q.

Assumption 5 is reminiscent of minimum and maximum eigenvalue restrictions in the high-dimensional statistics literature, where it is common to assume the minimum and maximum eigenvalues of Fisher information matrices are bounded away from 0 and from above, respectively (e.g., Janková and van de Geer, 2018, Ravikumar, Wainwright and Lafferty, 2010). Assumption 5 can be interpreted similarly, though applied to the averaged minimum and maximum eigenvalues of the joint Fisher information matrices; see also the discussions following Corollary 2.4.

Under Assumptions 1, 2, 3, 4, and 5, we may leverage Theorem 2.5 to establish, for all measurable convex sets $C \subset \mathbb{R}^{p+q}$, the new bound of

$$\left|\mathbb{P}(I(\boldsymbol{\theta}^\star)^{1/2}\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star) + \boldsymbol{\Delta} \in C) - \Phi_d(\mathbf{Z}_d \in C)\right| \;\leq\; C_1 (p+q)^{1/4} A_{\max}^7 \left[\sqrt{\frac{p^3}{N}} + \sqrt{\frac{q^3}{N^2}}\right],$$

where $\boldsymbol{\Delta}$ now satisfies

$$\mathbb{P}\left(\|\boldsymbol{\Delta}\|_2 \leq C_2\, A_{\max}^6 \sqrt{A_{\text{avg}}\,\frac{p^5}{N} + A_{\text{avg}}^2\,\frac{q^5}{N^2}}\right) \;\geq\; 1 - \frac{1}{N^2}.$$

In certain settings, it may be the case that properties of the network limit the sizes of the blocks, in which the size of the largest block $A_{\max}$ may be bounded for all network sizes. Under the additional assumption that the sizes of the blocks are bounded above, we can absorb the quantities involving $A_{\max}$ and A_{avg} into the constants $C_1 > 0$ and $C_2 > 0$ in the above bounds. This results in the following simple bounds on the error of the multivariate normal approximation:

$$\left|\mathbb{P}(I(\boldsymbol{\theta}^\star)^{1/2}\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star) + \boldsymbol{\Delta} \in C) - \Phi_d(\mathbf{Z}_d \in C)\right| \;\leq\; C_1\,(p+q)^{1/4} \left[\sqrt{\frac{p^3}{N}} + \sqrt{\frac{q^3}{N^2}}\right],$$

where $\boldsymbol{\Delta}$ will then satisfy

$$\mathbb{P}\left(\|\boldsymbol{\Delta}\|_2 \leq C_2 \sqrt{\frac{p^5}{N} + \frac{q^5}{N^2}} \right) \geq 1 - \frac{1}{N^2},$$

for all measurable convex sets $C \subset \mathbb{R}^{p+q}$. Note, in the above results, that the probability bounds approach 1 sufficiently fast, allowing us to establish, through the Borel–Cantelli lemma, that $\|\boldsymbol{\Delta}\|_2$ converges $\mathbb{P}$-almost surely to 0 as $N \to \infty$, provided the upper bounds on $\|\boldsymbol{\Delta}\|_2$ tend to 0 as $N \to \infty$.

Finally, to deliver a simple and easily interpretable result for statistical inference, we prove a corollary to Theorem 2.5 establishing the asymptotic multivariate normality of maximum likelihood estimators.

Corollary 2.6. *Under the assumptions of Theorem 2.5, Assumption 5, and assuming*

$$\lim_{N\to\infty} \max\left\{ A_{\max}^6 \sqrt{A_{\text{avg}} \frac{p^5}{N} + A_{\text{avg}}^2 \frac{q^5}{N^2}},\ (p+q)^{1/4} A_{\max}^7 \left[\sqrt{\frac{p^3}{N}} + \sqrt{\frac{q^3}{N^2}} \right] \right\} = 0,$$

we have the distributional limit $I(\boldsymbol{\theta}^\star)^{1/2} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star) \xrightarrow{D} \mathbf{Z}_{p+q}$ *as* $N \to \infty$.

Corollary 2.6 can be proved directly by observing that the assumptions of the corollary ensure the error bounds in Theorem 2.5 converge to 0 in the limit as $N \to \infty$. As a result of Corollary 2.6, standard procedures for constructing confidence regions, univariate confidence intervals, and performing statistical hypothesis tests for significance of parameters are justified using the asymptotic approximation of the variance-covariance matrix $I(\boldsymbol{\theta}^\star) = \mathbb{V}\, s(\boldsymbol{X})$, which we discuss above. When the sizes of the blocks are bounded as above, the essential condition for asymptotic multivariate normality becomes

$$\lim_{N\to\infty} \sqrt{\frac{p^5}{N} + \frac{q^5}{N^2}} = 0,$$

restricting the maximum growth with N of the dimensions of the parameters vectors $p = \dim(\boldsymbol{\theta}_W)$ and $q = \dim(\boldsymbol{\theta}_B)$, suggesting that both $p = \dim(\boldsymbol{\theta}_W) = o(N^{1/5})$ and $q = \dim(\boldsymbol{\theta}_B) = o(N^{2/5})$ must hold in our theory for the error of the multivariate normal approximation to vanish in the limit as $N \to \infty$.

Up to now, we required knowledge of the Fisher information matrix $I(\boldsymbol{\theta}^\star) = \mathbb{V}\, s(\boldsymbol{X})$. We end the section with a result concerning the estimation of this term for practical implementation. We define

$$\widetilde{I}_W^\star \coloneqq \frac{\mathbb{E}[-\nabla_{\boldsymbol{\theta}_W}^2 \ell(\boldsymbol{\theta}^\star, \boldsymbol{X})]}{K} \quad \text{and} \quad \widetilde{I}_B^\star \coloneqq \frac{\mathbb{E}[-\nabla_{\boldsymbol{\theta}_B}^2 \ell(\boldsymbol{\theta}^\star, \boldsymbol{X})]}{\binom{K}{2}}.$$

Natural estimators for each are given by

$$\widehat{I}_W \coloneqq \frac{1}{K} \sum_{k=1}^{K} \left(s_{k,k}(\boldsymbol{X}_{k,k}) - \overline{s}_W(\boldsymbol{X}_W) \right) \left(s_{k,k}(\boldsymbol{X}_{k,k}) - \overline{s}_W(\boldsymbol{X}_W) \right)^\top$$

$$\widehat{I}_B \coloneqq \frac{1}{\binom{K}{2}} \sum_{1 \leq k < l \leq K} \left(s_{k,l}(\boldsymbol{X}_{k,l}) - \overline{s}_B(\boldsymbol{X}_B) \right) \left(s_{k,l}(\boldsymbol{X}_{k,l}) - \overline{s}_B(\boldsymbol{X}_B) \right)^\top$$

with the definition

$$\overline{s}_W(\boldsymbol{X}_W) \;\coloneqq\; \frac{1}{K}\sum_{k=1}^{K} s_{k,k}(\boldsymbol{X}_{k,k}) \qquad \text{and} \qquad \overline{s}_B(\boldsymbol{X}_B) \;\coloneqq\; \frac{1}{\binom{K}{2}}\sum_{1\le k<l\le K} s_{k,l}(\boldsymbol{X}_{k,l}).$$

The following theorem establishes bounds on the error $|||\widehat{I}_W - \widetilde{I}^{\star}_W|||_2$ and $|||\widehat{I}_B - \widetilde{I}^{\star}_B|||_2$ which hold with high probability, where $|||\cdot|||_2$ denotes the spectral matrix norm.

Theorem 2.7. *Under the assumptions of Theorem 2.5, the events*

$$|||\widehat{I}_W - \widetilde{I}^{\star}_W|||_2 \;\le\; C\,A_{\max}^2\sqrt{A_{\text{avg}}\,\widetilde{\lambda}^{\star}_{\max,W}}\left(\sqrt{\frac{p\,\log(p)}{N}}+\sqrt{\frac{p^2}{N}}\right)$$

$$|||\widehat{I}_B - \widetilde{I}^{\star}_B|||_2 \;\le\; C\,A_{\max}^2\,A_{\text{avg}}\sqrt{\widetilde{\lambda}^{\star}_{\max,B}}\left(\sqrt{\frac{q\,\log(q)}{N^2}}+\sqrt{\frac{q^2}{N^2}}\right)$$

jointly occur with probability at least $1-4N^{-2}$.

The conclusions of Theorem 2.7 reiterate the conclusions of our previous theoretical results, that if certain quantities related to the sizes of blocks and properties of models through the spectral properties of Fisher information matrices are sufficiently well-behaved, and the dimensions of the parameter vectors do not grow too quickly with N, then accurate estimation and valid inferences of parameter vectors of local dependence random graph models will be obtained with high probability.

3. Simulation results

3.1. Simulation study 1: Convergence rates of maximum likelihood estimators

Simulation study 1 demonstrates that the rate of growth of the dimension of parameter vectors plays a key role in the finite sample performance. We consider three cases in a setting which controls certain aspects of the graph. Throughout this study, we assume that the sizes of the blocks are all fixed at 50, i.e., $|\mathscr{A}_k| = 50$ for all $k \in \{1,\ldots,K\}$. In order to vary the size of the network N, we vary the number of blocks $K \in \{1,5,10,15,20\}$, which results in networks of size $N \in \{50,250,500,750,1000\}$. We focus on a special case of Example 3 from Section 1.2, by assuming that each node $i \in \mathscr{N}$ is assigned to a group $\mathscr{G}_1,\ldots,\mathscr{G}_M$ ($M \ge 2$). The specific form of this model is then given by

$$\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{X}=\boldsymbol{x}) \;\propto\; \exp\left(\sum_{m=1}^{M}\theta_m\,s_m(\boldsymbol{x})+\theta_{m+1}\,s_{m+1}(\boldsymbol{x})\right),$$

where

$$s_m(\boldsymbol{x}) \;=\; \sum_{k=1}^{K}\sum_{i\in\mathscr{A}_k\cap\mathscr{G}_m}\sum_{j\in\mathscr{A}_k\setminus\{i\}} x_{i,j}, \qquad m\in\{1,\ldots,M\},$$

and

$$s_{m+1}(\boldsymbol{x}) \;=\; \sum_{k=1}^{K}\sum_{i<j\,:\,i\in\mathscr{A}_k,j\in\mathscr{A}_k} x_{i,j}\,\mathbb{1}\left(\sum_{h\in\mathscr{A}_k\setminus\{i,j\}} x_{i,h}\,x_{j,h}\;\ge\;1\right).$$

For this simulation study we will focus on the within-block parameter vector in order to easily compare the trade-off between the dimension of the parameter vector p and the size of the network N. We can then assume that $X_{i,j} = 0$ with probability one for all $\{i, j\} \subset \mathcal{N}$ belonging to distinct blocks, i.e., the between-block subgraphs $\boldsymbol{X}_{k,l}$ $(1 \leq k < l \leq K)$ are empty subgraphs with probability one.

We consider three cases:

- Case 1: $M = 3$, in which case $p = 4$ for all $N \in \{50, 250, 500, 750, 1000\}$.
- Case 2: $M = \lceil N^{2/5} \rceil$, in which case $p \in \{6, 11, 14, 16, 17\}$ depending on the size of the network.
- Case 3: $M = \lceil \sqrt{N} \rceil$, in which case $p \in \{9, 17, 24, 29, 33\}$ depending on the size of the network.

For each case and network size $N \in \{50, 250, 500, 750, 1000\}$, we simulate 500 networks from $\mathbb{P}_{\boldsymbol{\theta}}$ where $\theta_{M+1} = .5$ and $(\theta_1, \ldots, \theta_M) \overset{iid}{\sim} \text{Unif}(-1.5, -.5)$. The value of θ_{M+1} ensures there is a reasonably strong tendency towards transitivity in the network, and the values of $(\theta_1, \ldots, \theta_M)$ result in networks with plausible densities. The results of the Simulation study 1 are summarized in Figure 2.

The finite sample performance of this study suggests, as would be expected based on the results of Theorem 2.1, that the rate at which the ℓ_2-error converges to 0 is fastest in Case 1 for which the model dimension is fixed, and slowest in Case 3 for which the model dimension is on the order of $\sqrt{N}$. We compute a predicted error bound based on Theorem 2.1 by estimating the constant terms, which in this simulation study include the average block sizes A_{avg} and the largest block size $A_{\max}$, as well as the terms quantifying averaged eigenvalues of the Fisher information matrices. This can be accomplished by estimating constants for each network size by

$$\widehat{C}_N \;\coloneqq\; Q_{N,.95} \Big/ \sqrt{\frac{p}{N}}, \qquad N \in \{50, 250, 500, 750, 1000\},$$

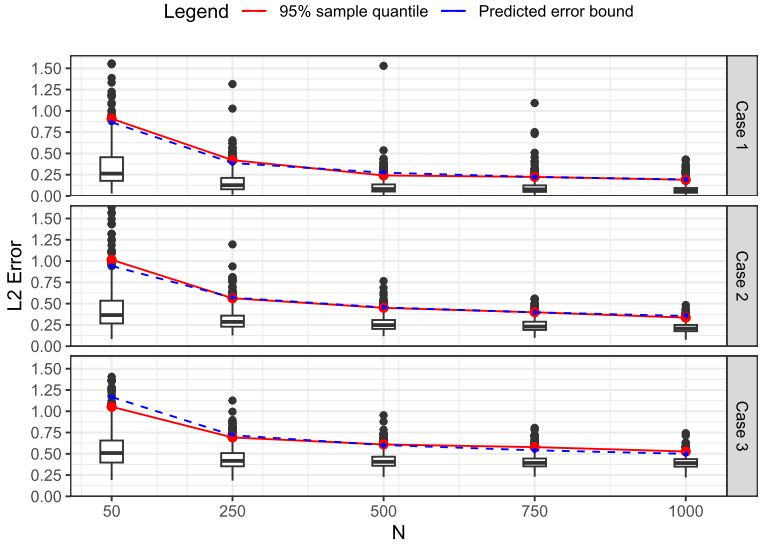

Figure 2. The results of Simulation study 1, which demonstrates the trade-off in finite sample performance of maximum likelihood estimators based on the number of model parameters and size of the network. Each boxplot for each combination of case and network size is based on 500 replications. Boxplots display the empirical distribution of the ℓ_2-error, whereas the red lines track the 95% sample quantiles and the blue dashed lines track the error bounds predicted by Theorem 2.1.

where $Q_{N,.95}$ is the 95% sample quantile of the ℓ_2-errors of the maximum likelihood estimators based on the 500 replications, and then using the estimate

$$\widehat{C} \;\coloneqq\; \frac{1}{5} \sum_{N \in \{50,250,500,750,1000\}} \widehat{C}_N$$

to obtain an overall estimate of the constant term. The predicted error bounds are then defined as

$$\widetilde{E}_N \;\coloneqq\; \widehat{C}\sqrt{\frac{p}{N}}, \qquad N \in \{50, 250, 500, 750, 1000\}.$$

The dashed blue lines track the values of $\widetilde{E}_N$ in Figure 2, whereas the red lines track $Q_{N,.95}$.

Notably, the predicted error bound closely matches the 95% sample quantile of the simulated ℓ_2-errors. Theorem 2.1 establishes a bound which should hold with high probability, provided N is sufficiently large. Figure 2 demonstrates that the predicted error bounds most closely match the realized 95% sample quantile of the simulated ℓ_2-errors for larger network sizes. It is also worth noting that an additional source of variation here may be due to the fact that the constant term is not actually constant in the network size, as the quantities $\widetilde{\lambda}^{\star}_{\max,W}$ and $\widetilde{\lambda}^{\epsilon}_{\min,W}$ may depend on N. With that said, though, the simulation reveals close agreement with the predicted error bounds.

3.2. Simulation study 2: Error of the normal approximation

The second simulation study we conduct explores the error of the normal approximation, leveraging results in Theorem 2.5. We consider the same probability distribution as in Simulation study 1, in the following two cases:

- Case 1: Fixed parameter dimension $p = 5$ with $M = 4$ categories of each node group and networks of size $N \in \{250, 500, 750, 1000\}$.
- Case 2: Growing parameter dimension $p = 2K$ with $M = 2K - 1$ categories of each node group, where there are 50 nodes per block and the number of blocks vary over $K \in \{5, 10, 15, 20\}$, resulting in networks of size $N \in \{250, 500, 750, 1000\}$.

We generate 500 replications in each case, simulating networks from the same probability distributions as in Simulation study 1 and in the same manner.

We study the quality of the normal approximation by constructing confidence intervals for the transitive edge parameter and Quantile-Quantile plots for the standardized maximum likelihood estimator of the transitive edge parameter. Our results demonstrate the empirical Type I error in the former matches the theoretical Type I error, with the Quantile-Quantile plots not revealing significant departure from normality. For each case, we constructed 95% confidence intervals and computed the empirical Type I error control. Letting θ_{m+1} and $\widehat{\theta}_{m+1}$ denote the transitive edge parameter and the maximum likelihood estimator of the transitive edge parameter, we leverage Theorem 2.5 to construct confidence intervals:

$$\mathbb{P}\left(\theta^{\star}_{m+1} \in \left[\widehat{\theta}_{m+1} - q_{1-\alpha/2}\sqrt{[\boldsymbol{S}^{-1}]_{m+1,m+1}},\ \widehat{\theta}_{m+1} + q_{1-\alpha/2}\sqrt{[\boldsymbol{S}^{-1}]_{m+1,m+1}}\right]\right) \approx 1-\alpha, \quad \alpha \in (0,1),$$

where $q_{1-\alpha/2}$ denotes the $(1-\alpha/2)$%-quantile of the univariate standard normal distribution and $\boldsymbol{S}$ denotes the sample variance-covariance matrix obtained by sampling sufficient statistics through MCMC methods; see the discussions in Section 2.1. For Case 1, the empirical coverage was (.96, .95, .95, .96) corresponding to network sizes of (250, 500, 750, 1000), and for Case 2, the same was (.96, .95, .95, .96) corresponding to network sizes of (250, 500, 750, 1000). The Quantile-Quantile plots for each case across the different network sizes are presented in Figure 3.

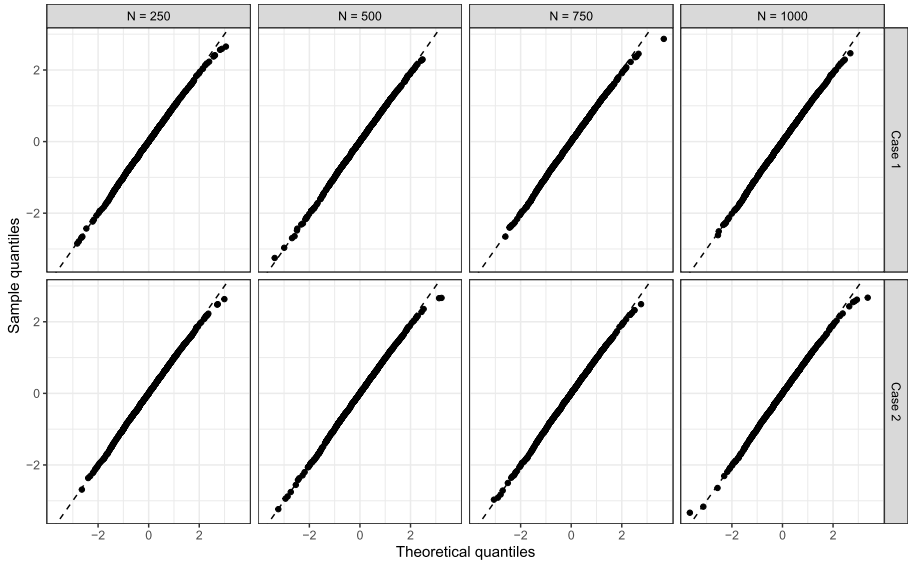

Figure 3. Quantile-Quantile plots showing the results of Simulation Study 2. The sample quantiles of the standardized maximum likelihood estimates of the transitive edge parameter are plotted against the theoretical quantiles based on the standard normal approximation in each of the two cases studied in Simulation study 2 across networks of size $N \in \{250, 500, 750, 1000\}$.

4. Conclusions

In this work, we have proved the first rigorous theory for both estimation and statistical inference of local dependence random graph models. We have established minimax optimal rates of convergence in the ℓ_2-norm of maximum likelihood estimators of exponential-family local dependence random graph models, accompanying these results with finite-sample error bounds on the multivariate normal approximation of a standardization of maximum likelihood estimators. Notably, our results allow for both the number of parameters and the sizes of blocks to grow unbounded with the size of the network.

Our consistency and normal approximation theory are non-asymptotic, although we have stated helpful asymptotic results along the way, which enable us to understand how key aspects of the model (through the spectrum of Fisher information matrices and the dimension of parameter vectors) and properties of the network (through the number and sizes of blocks and nodes) impact rates of convergence for both the statistical error (in the ℓ_2-norm) and the multivariate normal approximation. Our results cover general settings and heterogeneous parameterizations, as exemplified in the examples in Section 1.2 and our simulation studies in Section 3, which allow our results to cover a broad scope.

Results were derived under the assumption that we have perfect knowledge of the block memberships of nodes in the network. This may be reasonable in certain settings where we can observe the block memberships of nodes, but might be violated in other settings where we obtain imperfect observations of the block memberships of nodes, whether through a noisy observation process or error in the estimates of the block memberships. The effect of imperfect knowledge of the block memberships of nodes on the aforementioned errors and convergence rates is an open question.

Acknowledgments

The author is grateful to two anonymous reviewers, an associate editor, the editor, and to Dr. Michael Schweinberger, for constructive comments and suggestions that have greatly improved the paper.

Funding

The author was supported by NSF award SES-2345043 and the Department of Defense Test Resource Management Center under contracts FA8075-18-D-0002 and FA8075-21-F-0074.

Supplementary Material

Supplement to "Rates of convergence and normal approximations for estimators of local dependence random graph models (DOI: 10.3150/25-BEJ1852SUPP; .pdf). The supplementary material contains the proofs for all results in Section 2, as well as additional technical results and the proofs of these technical results used to prove the main results in Section 2.

References

Agneessens, F., Trincado-Munoz, F.J. and Koskinen, J. (2024). Network formation in organizational settings: Exploring the importance of local social processes and team-level contextual variables in small groups using Bayesian hierarchical ERGMs. *Soc. Netw.* **77** 104–117. https://doi.org/10.1016/j.socnet.2022.07.001

Babkin, S., Stewart, J.R., Long, X. and Schweinberger, M. (2020). Large-scale estimation of random graph models with local dependence. *Comput. Statist. Data Anal.* **152** 107029. MR4130894 https://doi.org/10.1016/j.csda.2020.107029

Brown, L.D. (1986). *Fundamentals of Statistical Exponential Families: With Applications in Statistical Decision Theory*. Institute of Mathematical Statistics. MR0882001 https://doi.org/10.1214/lnms/1215466757

Chatterjee, S. and Diaconis, P. (2013). Estimating and understanding exponential random graph models. *Ann. Statist.* **41** 2428–2461. MR3127871 https://doi.org/10.1214/13-AOS1155

Chatterjee, S., Diaconis, P. and Sly, A. (2011). Random graphs with a given degree sequence. *Ann. Appl. Probab.* **21** 1400–1435. MR2857452 https://doi.org/10.1214/10-AAP728

Dahbura, J.N.M., Komatsu, S., Nishida, T. and Mele, A. (2021). A structural model of business card exchange networks. arXiv preprint, arXiv:2105.12704. https://doi.org/10.48550/arXiv.2105.12704

Dahbura, J.N.M., Komatsu, S., Nishida, T. and Mele, A. (2023). Homophily and community structure at scale: An application to a large professional network. *AEA Pap. Proc.* **113** 156–160. https://doi.org/10.1257/pandp.20231094

Geyer, C.J. and Thompson, E.A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **54** 657–683. MR1185217 https://doi.org/10.1111/j.2517-6161.1992.tb01443.x

Ghosal, S. (2000). Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity. *J. Multivariate Anal.* **74** 49–68. MR1790613 https://doi.org/10.1006/jmva.1999.1874

Häggström, O. and Jonasson, J. (1999). Phase transition in the random triangle model. *J. Appl. Probab.* **36** 1101–1115. MR1742153 https://doi.org/10.1239/jap/1032374758

Holland, P.W., Laskey, K.B. and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. MR0718088 https://doi.org/10.1016/0378-8733(83)90021-7

Holland, P.W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *J. Amer. Statist. Assoc.* **76** 33–50. MR0608176 https://doi.org/10.1080/01621459.1981.10477598

Hunter, D.R. and Handcock, M.S. (2006). Inference in curved exponential family models for networks. *J. Comput. Graph. Statist.* **15** 565–583. MR2291264 https://doi.org/10.1198/106186006X133069

Janková, J. and van de Geer, S. (2018). Inference in high-dimensional graphical models. In *Handbook of Graphical Models* (M. Maathuis, M. Drton, S. Lauritzen and M. Wainwright, eds.) 325–350. CRC Press. MR3888003 https://doi.org/10.1080/01621459.2020.1801279

Jonasson, J. (1999). The random triangle model. *J. Appl. Probab.* **36** 852–867. MR1737058 https://doi.org/10.1239/jap/1032374639

Krivitsky, P.N., Hunter, D.R., Morris, M. and Klumb, C. (2023). ergm 4: New features for analyzing exponential-family random graph models. *J. Stat. Softw.* **105** 1–44. https://doi.org/10.18637/jss.v105.i06

Lusher, D., Koskinen, J. and Robins, G. (2012). *Exponential Random Graph Models for Social Networks*. Cambridge University Press. https://doi.org/10.1017/CBO9780511894701

Mele, A. (2022). A structural model of homophily and clustering in social networks. *J. Bus. Econom. Statist.* **40** 1377–1389. MR4439296 https://doi.org/10.1080/07350015.2021.1930013

Newman, M.E. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E* **69** 026113. MR1975193 https://doi.org/10.1103/PhysRevE.69.026113

Portnoy, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.* **16** 356–366. MR0924876 https://doi.org/10.1214/aos/1176350710

Raič, M. (2019). A multivariate Berry-Esseen theorem with explicit constants. *Bernoulli* **25** 2824–2853. MR4003566 https://doi.org/10.3150/18-BEJ1072

Ravikumar, P., Wainwright, M.J. and Lafferty, J.D. (2010). High-dimensional Ising model selection using ℓ_1-regularized logistic regression. *Ann. Statist.* **38** 1287–1319. MR2662343 https://doi.org/10.1214/09-AOS691

Schubert, D. and Brand, A. (2022). "Whom should I talk to?": Role prescription and hierarchy building in supervised living groups. *Soc. Incl.* **10** 295–306. https://doi.org/10.17645/si.v10i3.5406

Schweinberger, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *J. Amer. Statist. Assoc.* **106** 1361–1370. MR2896841 https://doi.org/10.1198/jasa.2011.tm10747

Schweinberger, M. (2020). Consistent structure estimation of exponential-family random graph models with block structure. *Bernoulli* **26** 1205–1233. MR4058365 https://doi.org/10.3150/19-BEJ1153

Schweinberger, M. and Handcock, M.S. (2015). Local dependence in random graph models: Characterization, properties and statistical inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 647–676. MR3351449 https://doi.org/10.1111/rssb.12081

Schweinberger, M. and Stewart, J. (2020). Concentration and consistency results for canonical and curved exponential-family models of random graphs. *Ann. Statist.* **48** 374–396. MR4065166 https://doi.org/10.1214/19-AOS1810

Schweinberger, M., Krivitsky, P.N., Butts, C.T. and Stewart, J.R. (2020). Exponential-family models of random graphs: Inference in finite, super and infinite population scenarios. *Statist. Sci.* **35** 627–662. MR4175389 https://doi.org/10.1214/19-STS743

Shao, M., Zhang, Y., Wang, Q., Zhang, Y., Luo, J. and Yan, T. (2021). L-2 Regularized maximum likelihood for β-model in large and sparse networks. arXiv preprint, arXiv:2110.11856. https://doi.org/10.48550/arXiv.2110.11856

Snijders, T.A. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *J. Soc. Struct.* **3** 1–40.

Stewart, J.R. (2026). Supplement to "Rates of convergence and normal approximations for estimators of local dependence random graph models." https://doi.org/10.3150/25-BEJ1852SUPP

Stewart, J.R. and Schweinberger, M. (2019). R package mlergm. Multilevel Exponential-Family Random Graph Models. The Comprehensive R Archive Network.

Stewart, J.R. and Schweinberger, M. (2020). Pseudo-likelihood-based M-estimation of random graphs with dependent edges and parameter vectors of increasing dimension. arXiv preprint, arXiv:2012.07167. https://doi.org/10.48550/arXiv.2012.07167

Stewart, J., Schweinberger, M., Bojanowski, M. and Morris, M. (2019). Multilevel network data facilitate statistical inference for curved ERGMs with geometrically weighted terms. *Soc. Netw.* **59** 98–119. https://doi.org/10.1016/j.socnet.2018.11.003

Sundberg, R. (2019). *Statistical Modelling by Exponential Families*. Cambridge University Press. MR3969949 https://doi.org/10.1017/9781108604574

Tolochko, P. and Boomgaarden, H.G. (2024). Same but different: A comparison of estimation approaches for exponential random graph models for multiple networks. *Soc. Netw.* **76** 1–11. https://doi.org/10.1016/j.socnet.2023.05.003

Whetsell, T.A., Kroll, A. and Dehart-Davis, L. (2021). Formal hierarchies and informal networks: How organizational structure shapes information search in local government. *J. Public Adm. Res. Theory* **31** 653–669. https://doi.org/10.1093/jopart/muab003

Received January 2024 and revised January 2025