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

JONATHAN R. STEWART

Department of Statistics
Florida State University
117 N Woodward Ave
Tallahassee, FL 32306-4330
E-mail: jrstewart@fsu.edu

Abstract 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 [33], 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 statistical disclaimers which provide conditions under which estimation and inference procedures can be expected to provide accurate and valid inferences. This is accomplished by deriving convergence rates of 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 the first non-asymptotic bounds on the ℓ_2-error of maximum likelihood estimators, along with convergence rates. Second, and more importantly, we derive the first non-asymptotic bounds on the error of the multivariate normal approximation. In so doing, we introduce the first principled approach to providing statistical disclaimers through quantifying the uncertainty about statistical conclusions based on data.

Keywords: statistical network analysis, network data, local dependence random graph model, multivariate normal approximation.

1. Introduction

Local dependence random graph models, introduced by Schweinberger and Handcock [33], are a class of statistical models for network data built around block structure, where a population of nodes $\mathcal{N}$, which we take without loss to be $\mathcal{N} \coloneqq \{1, \ldots, N\}$ ($N \geq 3$), is partitioned into $K \geq 1$ subsets $\mathcal{A}_1, \ldots, \mathcal{A}_K$ 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 that help explain the research interest received in both the statistical network analysis and network science literatures [38, 34, 2, 42, 24, 8, 1, 9, 40]. First, block structure (or community structure) is a well-established structural phenomena relevant to many applications and networks encountered in our world [e.g., 11, 25, 38]. 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 [33, 34].

In this work, we advance the literature on local dependence random graph models by providing the first statistical disclaimers which elaborate conditions under which estimation and inference procedures based on a single observation of the graph can be expected to produce accurate and valid inferences of local dependence random graph models. All 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 local dependence random graph models which hold with high probability, and
2. Deriving the first non-asymptotic bound on the error of the multivariate normal approximation of a standardization of the maximum likelihood estimator.

All results are stated in terms of interpretable quantities, allowing us to quantify the effect key aspects of the statistical model and network structure hold with respect to the convergence rates of the aforementioned errors. In so doing, we introduce the first principled approach to providing statistical disclaimers through quantifying the uncertainty about statistical conclusions based on data.

1.1. Local dependence random graph models

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

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

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

A *local dependence random graph* [33] is a random graph $\boldsymbol{X}$ where the set of nodes $\mathcal{N}$ is partitioned into K *blocks* $\mathcal{A}_1, \ldots, \mathcal{A}_K$ with joint probability distributions $\mathbb{P}$ of the form

$$\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) \;=\; \prod_{1 \leq k \leq l \leq K} \mathbb{P}_{k,l}(\boldsymbol{X}_{k,l} = \boldsymbol{x}_{k,l}), \qquad \boldsymbol{x} \in \mathbb{X}, \tag{1.1}$$

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

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

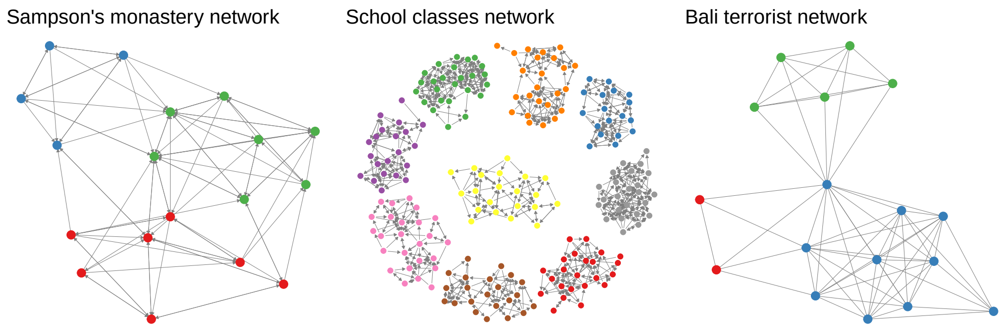

Figure 1. Three real data examples of networks for which local dependence random graph models would be applicable, including Sampson's monastery network, the school classes data set from Stewart et al. [38], and the Bali terrorist network studied in Schweinberger and Handcock [33]. Node colors correspond to block memberships.

We refer to the subgraphs $\boldsymbol{X}_{k,k}$ $(k = 1, \ldots, K)$ 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 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 (1.1). Figure 1 visualizes three networks which can be studied using local dependence random graph models. While the 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 $\boldsymbol{X}_{k,l}$ $(1 \leq k \leq l \leq K)$.

Exponential families account for the most prevalent specifications of local dependence random graph models [e.g., 38, 8, 42, 31], indeed having been the statistical foundations for the class in the seminal work by Schweinberger and Handcock [33]. Moreover, exponential families provide a flexible statistical platform for constructing models of edge dependence in network data applications [23, 35], 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 [34]. An *exponential-family local dependence random graph model* can be specified via the marginal probability distributions of the block-based subgraphs:

$$\mathbb{P}_{k,l,\boldsymbol{\theta}_{k,l}}(\boldsymbol{X}_{k,l} = \boldsymbol{v}) \quad = \quad h_{k,l}(\boldsymbol{v}) \; \exp\left(\langle \boldsymbol{\theta}_{k,l},\, s_{k,l}(\boldsymbol{v})\rangle - \psi_{k,l}(\boldsymbol{\theta}_{k,l})\right), \qquad \boldsymbol{v} \in \mathbb{X}_{k,l}, \qquad (1.2)$$

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 can be constructed with the local dependence property in (1.1) through different definitions of the sufficient statistics and reference functions. It is convenient to assume that the sufficient statistic vectors are homogeneous across the within-block subgraph and the between-block subgraph models, in the sense that the sufficient statistics essentially quantify the same effects in either case. In this approach, the natural parameter vectors are likewise assumed to be homogeneous, or in the case of size-based adjustments to the parameterizations, admit an affine transformation of a homogeneous parameter vector; see Krivitsky, Handcock and Morris [17], Stewart et al. [38], and Schweinberger and Stewart [34] for examples and further discussions of size-based parameterizations, with the latter two focusing specifically on local dependence random graphs. Throughout, we assume that $\boldsymbol{\theta}_{k,k} = \boldsymbol{\theta}_W$ for some $\boldsymbol{\theta}_W \in \mathbb{R}^p$ and all $k = 1, \dots, K$, and that $\boldsymbol{\theta}_{k,l} = \boldsymbol{\theta}_B$ for some $\boldsymbol{\theta}_B \in \mathbb{R}^q$ and all $1 \leq k < l \leq K$. As a result, we can write the joint distribution of $\boldsymbol{X}$ as

$$\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{X} = \boldsymbol{x}) \;\; = \;\; h(\boldsymbol{x}) \, \exp\left(\langle \boldsymbol{\theta}, \, s(\boldsymbol{x}) \rangle - \psi(\boldsymbol{\theta})\right), \tag{1.3}$$

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

$$s_W(\boldsymbol{x}) \;\; \coloneqq \;\; \sum_{k=1}^{K} s_{k,k}(\boldsymbol{x}_{k,k}) \quad \text{and} \quad s_B(\boldsymbol{x}) \;\; \coloneqq \;\; \sum_{1 \leq k < l \leq K} s_{k,l}(\boldsymbol{x}_{k,l}),$$

$$h(\boldsymbol{x}) \;\; \coloneqq \;\; \prod_{1 \leq k \leq l \leq K} h_{k,l}(\boldsymbol{x}_{k,l}),$$

and

$$\psi(\boldsymbol{\theta}) \;\; \coloneqq \;\; \sum_{k=1}^{K} \psi_{k,k}(\boldsymbol{\theta}_W) + \sum_{1 \leq k < l \leq K} \psi_{k,l}(\boldsymbol{\theta}_B).$$

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.

As the scope of possible models that can be constructed is large, we refer readers to works by Schweinberger and Handcock [33], Stewart et al. [38], and Schweinberger and Stewart [34], for concrete examples of exponential-family local dependence random graph models. We proceed assuming familiarity with specifications of this class in the literature.

2. Rates of convergence

In this section, we present our main results in Theorems 1 and 2, which establish rates of convergence for our consistency theory and normality theory, respectively. Together, these two theorems represent the first statistical disclaimers for estimation and inference of local dependence random graph models which elaborate conditions under which accurate estimation and valid inferences can be expected to be obtained.

We first review exponential family theory for local dependence random graph models in Section 2.1. We then turn to establishing rates of convergence of maximum likelihood and pseudolikelihood estimators in Section 2.2, and rates of convergence of the error of the multivariate normal approximation in Section 2.3.

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

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

which follows from Lemma 8 in the supplementary materials, where $\mathbb{E}_{k,k,\boldsymbol{\theta}_W}$ is the expectation operator with respect to the marginal distribution $\mathbb{P}_{k,k,\boldsymbol{\theta}_W}$ of $\boldsymbol{X}_{k,k}$ and $\mathbb{E}_{k,l,\boldsymbol{\theta}_B}$ is the expectation operator with respect to the marginal distribution $\mathbb{P}_{k,l,\boldsymbol{\theta}_B}$ of $\boldsymbol{X}_{k,l}$. 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 of [39]).

2.2. Convergence rates of the maximum likelihood estimator

We derive non-asymptotic bounds on the ℓ_2-error of the maximum likelihood estimator which hold with high probability. Our results extend the results of Schweinberger and Stewart [34], 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 [34] focused on estimation of only the parameter vectors of within-block probability distributions. In contrast, we establish consistency theory with rates of convergence for entire parameter vectors of exponential-family local dependence random graph models in this work, covering settings where the number of model parameters $p + q$ and the sizes of blocks $|\mathcal{A}_1|, \dots, |\mathcal{A}_K|$ 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 [37], 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 in terms of interpretable quantities related to local dependence random graphs, e.g., properties of the block structure, graph, and model. Second, our method of proof is fundamentally different from that of both Schweinberger and Stewart [34] and Stewart and Schweinberger [37], and as such the consistency theory in this work cannot be proved as a corollary to an existing result. As a final point of contrast, we establish the first non-asymptotic bound on the error of the multivariate normal approximation of a standardization of the maximum likelihood estimator, which in turn provides the first rigorous derivation of a statistical inference procedure for estimators of local dependence random graph models. Our normality theory is presented in Section 2.3.

The following notational definitions and two regularity assumptions will be utilized in Theorem 1. Define $A_{\max} \coloneqq \max\{|\mathcal{A}_1|, \dots, |\mathcal{A}_K|\}$ to be the size of the largest block and $\mathcal{B}_2(\boldsymbol{v}, r) \coloneqq \{\boldsymbol{v}' \in \mathbb{R}^{\dim(\boldsymbol{v})} : \|\boldsymbol{v}' - \boldsymbol{v}\|_2 < r\}$ to be the open ball with respect to the ℓ_2-norm in $\mathbb{R}^{\dim(\boldsymbol{v})}$ with center $\boldsymbol{v}$ and radius $r > 0$. Let $d_H : \mathcal{Y}^m \times \mathcal{Y}^m \mapsto \{0, 1, \dots, m\}$ be the Hamming distance which is defined by $d_H(\boldsymbol{y}, \boldsymbol{y}') = \sum_{i=1}^m \mathbb{1}(y_i \neq y_i')$ for all $(\boldsymbol{y}, \boldsymbol{y}') \in \mathcal{Y}^m \times \mathcal{Y}^m$, and $\lambda_{\min}(\boldsymbol{A})$ the smallest eigenvalue of the matrix $\boldsymbol{A}$.

Assumption A.1. Assume there exist constants $\epsilon^\star > 0$, $\xi_{W,\epsilon^\star} > 0$, and $\xi_{B,\epsilon^\star} > 0$, independent of N, p, and q, such that

$$\begin{aligned} \inf_{k \in \{1,\dots,K\}} \inf_{\boldsymbol{\theta}_W \in \mathcal{B}_2(\boldsymbol{\theta}_W^\star, \epsilon^\star)} \lambda_{\min}(-\mathbb{E}_{k,k} \nabla^2_{\boldsymbol{\theta}_W} \ell_{k,k}(\boldsymbol{\theta}_W, \boldsymbol{X}_{k,k})) &\geq \xi_{W,\epsilon^\star} > 0 \\ \inf_{\{k,l\} \subseteq \{1,\dots,K\}} \inf_{\boldsymbol{\theta}_B \in \mathcal{B}_2(\boldsymbol{\theta}_B^\star, \epsilon^\star)} \lambda_{\min}(-\mathbb{E}_{k,l} \nabla^2_{\boldsymbol{\theta}_B} \ell_{k,l}(\boldsymbol{\theta}_B, \boldsymbol{X}_{k,l})) &\geq \xi_{B,\epsilon^\star} > 0. \end{aligned}$$

Assumption A.2. Assume there exist constants $L_W \in (0, \infty)$ and $L_B \in (0, \infty)$, independent of N, p, and q, such that, for all $k \in \{1, \dots, K\}$ and $\{k, l\} \subseteq \{1, \dots, K\}$,

$$\begin{aligned} \sup_{(\boldsymbol{v}, \boldsymbol{v}') \in \mathbb{X}_{k,k} \times \mathbb{X}_{k,k} \,:\, d_H(\boldsymbol{v}, \boldsymbol{v}') = 1} \|s_{k,k}(\boldsymbol{v}) - s_{k,k}(\boldsymbol{v}')\|_\infty &\leq L_W \binom{|\mathcal{A}_k|}{2} \\ \sup_{(\boldsymbol{v}, \boldsymbol{v}') \in \mathbb{X}_{k,l} \times \mathbb{X}_{k,l} \,:\, d_H(\boldsymbol{v}, \boldsymbol{v}') = 1} \|s_{k,l}(\boldsymbol{v}) - s_{k,l}(\boldsymbol{v}')\|_\infty &\leq L_B \, |\mathcal{A}_k| \, |\mathcal{A}_l|. \end{aligned}$$

Assumption (A.1) requires that the smallest eigenvalue of negative expected Hessians of estimating functions in a neighborhood of the data-generating parameter vector $\boldsymbol{\theta}^\star$ are bounded away from 0. When the estimating function is the log-likelihood, this corresponds to Fisher information matrices. Minimum eigenvalue restrictions of Fisher information matrices are standard in settings where the number of model parameters may tend to infinity [e.g., 27, 29, 14]. Our theory relates deviations of the gradient of the log-likelihood $\nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}, \boldsymbol{X})$ from its expectation $\mathbb{E}\, \nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}, \boldsymbol{X})$ to deviations of maximum likelihood estimators $\widehat{\boldsymbol{\theta}} \in \widehat{\boldsymbol{\Theta}}$ from the data-generating parameter vector $\boldsymbol{\theta}^\star \in \mathbb{R}^{p+q}$. As such, we require concentration inequalities for the both $\|\nabla_{\boldsymbol{\theta}_W}\, \ell(\boldsymbol{\theta}, \boldsymbol{x}) - \mathbb{E}\, \nabla_{\boldsymbol{\theta}_W}\, \ell(\boldsymbol{\theta}, \boldsymbol{X})\|_2$ and $\|\nabla_{\boldsymbol{\theta}_B}\, \ell(\boldsymbol{\theta}, \boldsymbol{x}) - \mathbb{E}\, \nabla_{\boldsymbol{\theta}_B}\, \ell(\boldsymbol{\theta}, \boldsymbol{X})\|_2$ Concentration inequalities for functions of dependent random variables with countable support often quantify bounds on probabilities of deviations of functions from expected values in terms of the sensitivity of the function to changes in values of its arguments [e.g., 6, 16]. To this end, Assumption (A.2) assumes the block-based sufficient statistics are Lipschitz with respect to the Hamming distance, where $L_W \binom{|\mathcal{A}_k|}{2} > 0$ and $L_B\, |\mathcal{A}_k|\, |\mathcal{A}_l| > 0$ are the respective Lipschitz coefficients. As a result, the block-based sufficient statistics $s_{k,l}(\boldsymbol{x})$ $(1 \leq k \leq l \leq K)$ are allowed to scale with the number of edge variables in the block, which we will allow to grow with N.

Theorem 1. *Consider a minimal exponential-family local dependence random graph model satisfying both Assumptions (A.1) and (A.2). Then there exist constants $C_1 > 0$, $C_2 > 0$, and $N_0 \geq 3$, independent of N, p, and q, such that, for all $N \geq N_0$ and with probability at least $1 - 2\, N^{-2}$, the maximum likelihood estimator $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}_W, \widehat{\boldsymbol{\theta}}_B)$ exists uniquely and satisfies*

$$\begin{aligned} \|\widehat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_W^\star\|_2 &\leq C_1\, A_{\max}^6 \sqrt{\frac{p\, \log N}{N}} \\ \|\widehat{\boldsymbol{\theta}}_B - \boldsymbol{\theta}_B^\star\|_2 &\leq C_2\, A_{\max}^6 \sqrt{\frac{q\, \log N}{N^2}}, \end{aligned}$$

provided each upper bound tends to 0 as $N \to \infty$.

Theorem 1 is proved in Section 3.1. Theorem 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. Observe that the assumption that the exponential family is minimal is not restrictive, as any non-minimal exponential family can be reduced to a minimal exponential family (Proposition 1.5 of [3]). The rates of convergence in Theorem 1 depend on the number of model parameters $p + q$ as well as the size of the largest block $A_{\max}$, and the scaling of these quantities with respect to the number of nodes in the network N. We discuss the results of Theorem 1 further in Section 2.4.

We lastly discuss the requirement that the exhibited upper bounds on the ℓ_2-error tend to 0 in the limit as $N \to \infty$. Recall that Assumption (A.1) is a local assumption about the smallest eigenvalue of the negative expected Hessians of the log-likelihood. The proof of Theorem 1 establishes that the event $\widehat{\boldsymbol{\Theta}} \subseteq \mathcal{B}_2(\boldsymbol{\theta}^\star, \epsilon^\star)$ occurs with probability at least $1 - 2\,N^{-2}$, ensuring that it is legitimate to restrict focus to the subset of the parameter space $\mathcal{B}_2(\boldsymbol{\theta}^\star, \epsilon^\star)$. The assumption that the exhibited upper bounds on the ℓ_2-error tend to 0 in the limit as $N \to \infty$ are central to establishing this fact. Note that if this assumption were not satisfied, Theorem 1 would not establish consistency in the ℓ_2-norm. As a result, the assumptions of Theorem 1 ensure that we consider sequences of random graphs for which our theory will establish consistency in the ℓ_2-norm.

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., 18, 23, 38]. 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., 38]. Similarly to our consistency results presented in Theorem 1, the quality of the multivariate normal approximation will depend on key quantities related to the block structure, graph, and model specification.

Throughout, $\boldsymbol{Z}$ will denote a d-dimensional multivariate normal random vector with mean vector $\mathbf{0}_d$ (the d-dimensional vector of all zeros) and covariance matrix $\boldsymbol{I}_d$ (the d-dimensional identity matrix) and Φ will denote the corresponding probability distribution. We establish our multivariate normal approximation theory in Theorem 2 under essentially the same assumptions as our consistency theory in Theorem 1, with the addition of a single assumption regarding the scaling of the sufficient statistics of the exponential family.

Assumption A.3. Assume that there exist constants $C_W > 0$ and $C_B > 0$, independent of N, p, and q, such that

$$\begin{aligned} \sup_{k \in \{1,\dots,K\}} \quad \sup_{\boldsymbol{x}_{k,k} \in \mathbb{X}_{k,k}} \quad \|s_{k,k}(\boldsymbol{x}_{k,k})\|_\infty &\leq C_W \binom{|\mathcal{A}_k|}{2} \\ \sup_{\{k,l\} \subseteq \{1,\dots,K\}} \quad \sup_{\boldsymbol{x}_{k,l} \in \mathbb{X}_{k,l}} \quad \|s_{k,l}(\boldsymbol{x}_{k,l})\|_\infty &\leq C_B\, |\mathcal{A}_k|\, |\mathcal{A}_l|. \end{aligned}$$

Assumption (A.3) places 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 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 $\boldsymbol{X}_{k,k}$, which is given by

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

which can be viewed as a special case of the geometrically-weighted edgewise shared partner statistic [12, 38]. To further contextualize this assumption, it is helpful to note that Assumption (A.3), and also in fact Assumption (A.2), is related to the issue of instability of exponential-families of random graph models [32]. The maximal changes in the sufficient statistic vectors $s_{k,k}(\boldsymbol{x})$ $(k = 1, \ldots, K)$ 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 [32]. Both Assumptions (A.2) and (A.3) control 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—models of edge dependence in network data [13, 15, 32, 5].

In order to establish our multivariate normal approximation theory, we leverage a multivariate Berry-Essen theorem provided in Raič [28] 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 explain when the normal approximation is expected to produce valid inferences in local dependence random graph models. We prove Theorem 2 in Section 3.2.

Theorem 2. *Under the assumptions of Theorem 1 and Assumption (A.3), there exist constants $C_1 > 0$, $C_2 > 0$, and $N_0 \geq 3$, independent of N, p, and q, and a random vector $\widetilde{\boldsymbol{R}} \in \mathbb{R}^{p+q}$ such that, for all $N > N_0$ and measurable convex sets $\mathcal{A} \subset \mathbb{R}^{p+q}$,*

$$|\mathbb{Q}(\mathcal{A}) - \Phi(\mathcal{A})| \;\;\leq\;\; C_1 \sqrt{\frac{(p+q)^{7/2} \, A_{\max}^{26}}{N}},$$

where $\mathbb{Q}$ is the probability distribution of $I(\boldsymbol{\theta}^\star)^{1/2} \, (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star) - \widetilde{\boldsymbol{R}}$ and

$$\mathbb{P}\left(\|\widetilde{\boldsymbol{R}}\|_2 \;\leq\; C_2 \, A_{\max}^{22} \, \log N \, \sqrt{\frac{p^5}{N} + \frac{q^5}{N^2}} \right) \;\;\geq\;\; 1 - 2\, N^{-2}.$$

The standardization $I(\boldsymbol{\theta}^\star)^{1/2}\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star)$ is of a familiar form in multivariate normal approximation settings. While our result is stated for $I(\boldsymbol{\theta}^\star)^{1/2}\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star) + \widetilde{\boldsymbol{R}}$, a key aspect of Theorem 2 lies in establishing that the remainder term $\widetilde{\boldsymbol{R}}$ is small (in the ℓ_2-norm) with high probability, justifying basing inferences on $I(\boldsymbol{\theta}^\star)^{1/2}\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star)$ in applications. Typically, the quantities $I_W(\boldsymbol{\theta}_W^\star)^{1/2}$ and $I_B(\boldsymbol{\theta}_B^\star)^{1/2}$ will be unknown, but can be approximated. Both $I_W(\boldsymbol{\theta}_W^\star)$ and $I_B(\boldsymbol{\theta}_B^\star)$ can be approximated 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 requires simulating sufficient statistic vectors of the exponential family [e.g., 12, 18].

2.4. Implications of Theorems 1 and 2

We end the section with a discussion comparing the results of Theorems 1 and 2. For simplicity, assume that $p = q$, or less restrictively that $p \propto q$, where we may absorb the scaling factor into the constants of our results. In this setting, Theorem 1 establishes consistency of the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}^\star$ in the ℓ_2-norm provided

$$\lim_{N\to\infty} \sqrt{\frac{p\, A_{\max}^{12}\, \log N}{N}} \;=\; 0,$$

whereas our multivariate normal approximation theory in Theorem 2 requires

$$\lim_{N\to\infty} \sqrt{\frac{p^5\, A_{\max}^{44}\, (\log N)^2}{N}} \;=\; 0.$$

The latter limit places more stringent requirements on the growth rate of both $A_{\max}$ and p. To obtain the results of both Theorems 1 and 2, the following suffices:

- The size of the largest block $A_{\max}$ satisfies (ignoring logarithmic terms):

$$A_{\max} \;=\; o\left(\frac{N^{1/45}}{p^{1/9}}\right).$$

- The model dimension p (for both the within- and between-block probability distributions) has a conservative growth rate of at most (ignoring logarithmic terms):

$$p \;=\; o\left(\frac{N^{1/5}}{A_{\max}^{9}}\right).$$

In the above, we bounded $A_{\max}^{44}$ by $A_{\max}^{45}$ for ease of presentation. We illustrate the potential scaling of $A_{\max}$ and p with N further with an example. Assume that $p \propto N^{1/10}$. In this case, $A_{\max}$ must satisfy $A_{\max} = o(N^{1/90})$. While our results are applicable to scenarios where the sizes of blocks must grow slowly with the size of the network N, it has been observed that larger networks do not possess significantly larger blocks, with empirical evidence suggesting many communities are not larger than 100 nodes (see work by Leskovec et al. [21] and the discussion of Rohe, Chatterjee and Yu [30]). Our results therefore cover scenarios where the sizes of the blocks $|\mathcal{A}_1|, \ldots, |\mathcal{A}_K|$ and the dimension of models $p+q$ may grow unbounded as $N \to \infty$, provided quantities grow at appropriate rates.

Finally, we note that our results are framed in terms of the ℓ_2-norm, which places certain restrictions on the scaling of the model dimension. Notably, this would exclude models where the number of model parameters grows at a rate of N, namely the β-model [4, 37]. However, it is worth observing that in the case of the β-model, consistency results have only been obtained with respect to the ℓ_∞-norm (see the discussions following Corollary 1 of Stewart and Schweinberger [37]), and the best known normality results are for a fixed subset of the estimated parameters [e.g., 7], or in some modern cases for a growing but smaller subset of the estimated parameters [36]. In contrast, we are interested here in establishing the multivariate normal approximation for entire vectors of estimated parameters, and not solely for a proper subset of estimated parameters.

3. Proofs of Theorems

We prove Theorem 1 in Section 3.1 and prove Theorem 2 in Section 3.2. Auxiliary results used to prove each theorem are provided in the supplementary materials.

3.1. Proof of Theorem 1

Our method of proof utilizes a general M-estimation argument. For ease of presentation, we first introduce the general argument and then apply various technical results presented as lemmas in the supplementary material to obtain the results of Theorem 1.

General M-estimation framework for rates of convergence. Consider a random estimating function $m : \mathbb{R}^d \times \mathbb{X} \mapsto \mathbb{R}$ and define $M(\boldsymbol{\theta}) \coloneqq \mathbb{E}\, m(\boldsymbol{\theta}, \boldsymbol{X})$ for $\boldsymbol{\theta} \in \mathbb{R}^d$. We make the following assumptions concerning $m(\boldsymbol{\theta}, \boldsymbol{x})$ and $M(\boldsymbol{\theta})$:

1. Assume that $m(\boldsymbol{\theta}, \boldsymbol{x})$ is concave in $\boldsymbol{\theta} \in \mathbb{R}^d$ and continuously differentiable at all $\boldsymbol{\theta} \in \mathbb{R}^d$ and for all $\boldsymbol{x} \in \mathbb{X}$.
2. Assume that $M(\boldsymbol{\theta})$ is strictly concave in $\boldsymbol{\theta} \in \mathbb{R}^d$ and that $\boldsymbol{\theta}^\star \in \mathbb{R}^d$ is the unique global maximizer of $M(\boldsymbol{\theta})$.
3. Assume that $M(\boldsymbol{\theta})$ is twice continuously differentiable and that the negative Hessian $\boldsymbol{H}(\boldsymbol{\theta}) \coloneqq -\nabla^2_{\boldsymbol{\theta}}\, M(\boldsymbol{\theta})$ of $M(\boldsymbol{\theta})$ is positive definite for all $\boldsymbol{\theta} \in \mathcal{B}_2(\boldsymbol{\theta}^\star, \epsilon^\star)$.

When $m(\boldsymbol{\theta}, \boldsymbol{x})$ is the log-likelihood corresponding to a minimal exponential family, standard exponential family theory establishes that the above conditions (1) and (2) hold (e.g., Proposition 3.10 of [39]). As a result, $\nabla_{\boldsymbol{\theta}} M(\boldsymbol{\theta}^\star) = \mathbf{0}_d$, where $\mathbf{0}_d$ is the d-dimensional zero vector. By Theorem 6.3.4 of Ortega and Rheinboldt [26], if the event

$$\inf_{\boldsymbol{\theta}\in\partial\mathcal{B}_2(\boldsymbol{\theta}^\star,\,\epsilon)} \langle(\boldsymbol{\theta}-\boldsymbol{\theta}^\star),\ \nabla_{\boldsymbol{\theta}}\, m(\boldsymbol{\theta},\boldsymbol{X})\rangle \quad\geq\quad 0 \tag{3.1}$$

occurs, where $\partial\mathcal{B}_2(\boldsymbol{\theta}^\star,\,\epsilon)$ denotes the boundary of $\mathcal{B}_2(\boldsymbol{\theta}^\star,\,\epsilon) \coloneqq \{\boldsymbol{\theta}\in\mathbb{R}^d : \|\boldsymbol{\theta}-\boldsymbol{\theta}^\star\|_2 < \epsilon\}$, then a root of $\nabla_{\boldsymbol{\theta}}\, m(\boldsymbol{\theta},\boldsymbol{X})$ exists in $\overline{\mathcal{B}_2}(\boldsymbol{\theta}^\star,\epsilon) \coloneqq \mathcal{B}_2(\boldsymbol{\theta}^\star,\,\epsilon)\cup\partial\mathcal{B}_2(\boldsymbol{\theta}^\star,\,\epsilon)$, in which case a global maximizer $\boldsymbol{\theta}_0 = \arg\max_{\boldsymbol{\theta}\in\mathbb{R}^d}\, m(\boldsymbol{\theta},\boldsymbol{X})$ exists and satisfies $\|\boldsymbol{\theta}_0-\boldsymbol{\theta}^\star\|_2 \leq \epsilon$.

The key to our approach lies in demonstrating that condition (3.1) holds with high probability for a chosen $\epsilon \in (0,\epsilon^\star)$ which helps to establish rates of convergence of estimators. In order to do so, we leverage the multivariate mean value theorem to establish that there exists, for each $\boldsymbol{\theta}\in\partial\mathcal{B}_2(\boldsymbol{\theta}^\star,\,\epsilon)$, a parameter vector

$$\dot{\boldsymbol{\theta}} \quad=\quad t\,\boldsymbol{\theta} + (1-t)\,\boldsymbol{\theta}^\star \quad\in\quad \mathcal{B}_2(\boldsymbol{\theta}^\star,\,\epsilon) \quad\subset\quad \mathcal{B}_2(\boldsymbol{\theta}^\star,\,\epsilon^\star), \qquad \text{for some } t\in(0,1),$$

such that

$$\begin{aligned}\langle(\boldsymbol{\theta}-\boldsymbol{\theta}^\star),\ \nabla_{\boldsymbol{\theta}}\, M(\boldsymbol{\theta})\rangle \quad&=\quad \langle(\boldsymbol{\theta}-\boldsymbol{\theta}^\star),\ \nabla_{\boldsymbol{\theta}}\, M(\boldsymbol{\theta}^\star)\rangle + \langle(\boldsymbol{\theta}-\boldsymbol{\theta}^\star),\ \boldsymbol{H}(\dot{\boldsymbol{\theta}})\,(\boldsymbol{\theta}-\boldsymbol{\theta}^\star)\rangle\\ &=\quad \langle(\boldsymbol{\theta}-\boldsymbol{\theta}^\star),\ \boldsymbol{H}(\dot{\boldsymbol{\theta}})\,(\boldsymbol{\theta}-\boldsymbol{\theta}^\star)\rangle,\end{aligned}$$

recalling that $\nabla_{\boldsymbol{\theta}}\, M(\boldsymbol{\theta}^\star) = \mathbf{0}_d$. Observe that

$$\begin{aligned}\langle(\boldsymbol{\theta}-\boldsymbol{\theta}^\star),\ \boldsymbol{H}(\dot{\boldsymbol{\theta}})\,(\boldsymbol{\theta}-\boldsymbol{\theta}^\star)\rangle \quad&=\quad \frac{\langle(\boldsymbol{\theta}-\boldsymbol{\theta}^\star),\ \boldsymbol{H}(\dot{\boldsymbol{\theta}})\,(\boldsymbol{\theta}-\boldsymbol{\theta}^\star)\rangle}{\langle\boldsymbol{\theta}-\boldsymbol{\theta}^\star,\,\boldsymbol{\theta}-\boldsymbol{\theta}^\star\rangle}\,\|\boldsymbol{\theta}-\boldsymbol{\theta}^\star\|_2^2\\ &\geq\quad \lambda_{\min}(\boldsymbol{H}(\dot{\boldsymbol{\theta}}))\,\epsilon^2,\end{aligned}$$

recognizing that the Rayleigh quotient of $\boldsymbol{H}(\dot{\boldsymbol{\theta}})$ is bounded below by the smallest eigenvalue $\lambda_{\min}(\boldsymbol{H}(\dot{\boldsymbol{\theta}}))$ of $\boldsymbol{H}(\dot{\boldsymbol{\theta}})$ and that $\|\boldsymbol{\theta}-\boldsymbol{\theta}^\star\|_2 = \epsilon$ for all $\boldsymbol{\theta}\in\partial\mathcal{B}_2(\boldsymbol{\theta}^\star,\epsilon)$. As a result,

$$\inf_{\boldsymbol{\theta}\in\partial\mathcal{B}_2(\boldsymbol{\theta}^\star,\epsilon)} \langle(\boldsymbol{\theta}-\boldsymbol{\theta}^\star),\ \nabla_{\boldsymbol{\theta}}\, M(\boldsymbol{\theta})\rangle \quad\geq\quad \inf_{\boldsymbol{\theta}\in\mathcal{B}_2(\boldsymbol{\theta}^\star,\epsilon^\star)} \lambda_{\min}(\boldsymbol{H}(\boldsymbol{\theta}))\,\epsilon^2, \tag{3.2}$$

as $\partial\mathcal{B}_2(\boldsymbol{\theta}^\star,\epsilon)\subset\mathcal{B}_2(\boldsymbol{\theta}^\star,\epsilon^\star)$. Assumption (A.2) will ensure that

$$\inf_{\boldsymbol{\theta}\in\mathcal{B}_2(\boldsymbol{\theta}^\star,\epsilon^\star)} \lambda_{\min}(\boldsymbol{H}(\boldsymbol{\theta})) \quad>\quad 0. \tag{3.3}$$

As a result of (3.2) and (3.3), the event

$$\sup_{\boldsymbol{\theta}\in\mathcal{B}_2(\boldsymbol{\theta}^\star,\epsilon^\star)} |\langle(\boldsymbol{\theta}-\boldsymbol{\theta}^\star),\ \nabla_{\boldsymbol{\theta}}\, m(\boldsymbol{\theta},\boldsymbol{X}) - \nabla_{\boldsymbol{\theta}}\, M(\boldsymbol{\theta})\rangle| \leq \inf_{\boldsymbol{\theta}\in\mathcal{B}_2(\boldsymbol{\theta}^\star,\epsilon^\star)} \lambda_{\min}(\boldsymbol{H}(\boldsymbol{\theta}))\,\epsilon^2 \tag{3.4}$$

implies the event (3.1). Thus, demonstrating that event (3.4) occurs with probability at least $1-2\,N^{-2}$ demonstrates that event (3.1) occurs with probability at least $1-2\,N^{-2}$.

Rates of convergence for maximum likelihood estimators. The log-likelihood equation of an exponential-family local dependence random graph model has the form

$$\ell(\boldsymbol{\theta}, \boldsymbol{x}) \quad = \quad \sum_{k=1}^{K} \ell_{k,k}(\boldsymbol{\theta}_W, \boldsymbol{x}_{k,k}) + \sum_{1 \leq k < l \leq K} \ell_{k,l}(\boldsymbol{\theta}_B, \boldsymbol{x}_{k,l}),$$

which implies that the maximum likelihood estimator $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}_W, \widehat{\boldsymbol{\theta}}_B)$ is given by

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_W &= \operatorname*{arg\,max}_{\boldsymbol{\theta}_W \in \mathbb{R}^p} \sum_{k=1}^{K} \ell_{k,k}(\boldsymbol{\theta}_W, \boldsymbol{x}_{k,k}) \\ \widehat{\boldsymbol{\theta}}_B &= \operatorname*{arg\,max}_{\boldsymbol{\theta}_B \in \mathbb{R}^q} \sum_{1 \leq k < l \leq K} \ell_{k,l}(\boldsymbol{\theta}_B, \boldsymbol{x}_{k,l}), \end{aligned} \tag{3.5}$$

owing to the fact that the subgraphs $\boldsymbol{X}_{k,l}$ $(1 \leq k \leq l \leq K)$ are independent and that the parameter vectors $\boldsymbol{\theta}_W$ and $\boldsymbol{\theta}_B$ partition the parameters in $\boldsymbol{\theta}$. Hence, each optimizer in (3.5) can be found separately and independently. Define

$$\begin{aligned} m_W(\boldsymbol{\theta}_W, \boldsymbol{x}_W) &\coloneqq \sum_{k=1}^{K} \ell_{k,k}(\boldsymbol{\theta}_W, \boldsymbol{x}_{k,k}) \\ m_B(\boldsymbol{\theta}_B, \boldsymbol{x}_B) &\coloneqq \sum_{1 \leq k < l \leq K} \ell_{k,l}(\boldsymbol{\theta}_B, \boldsymbol{x}_{k,l}), \end{aligned}$$

$M_W(\boldsymbol{\theta}_W) \coloneqq \mathbb{E}\, m_W(\boldsymbol{\theta}_W, \boldsymbol{X}_W)$, and $M_B(\boldsymbol{\theta}_B) \coloneqq \mathbb{E}\, m_B(\boldsymbol{\theta}_B, \boldsymbol{X}_B)$, where

$$\boldsymbol{X}_W \coloneqq (\boldsymbol{X}_{k,k} : 1 \leq k \leq K) \qquad \text{and} \qquad \boldsymbol{X}_B \coloneqq (\boldsymbol{X}_{k,l} : 1 \leq k < l \leq K).$$

Due to the above considerations,

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

where $\mathbf{0}_{d,r}$ is the $(d \times r)$-dimensional matrix of all zeros, and where

$$\boldsymbol{H}_W(\boldsymbol{\theta}_W) \coloneqq \sum_{k=1}^{K} \boldsymbol{H}_{k,k}(\boldsymbol{\theta}_W) \qquad \text{and} \qquad \boldsymbol{H}_B(\boldsymbol{\theta}_B) \coloneqq \sum_{1 \leq k < l \leq K} \boldsymbol{H}_{k,l}(\boldsymbol{\theta}_B),$$

defining

$$\begin{aligned} \boldsymbol{H}_{k,k}(\boldsymbol{\theta}_W) &\coloneqq -\mathbb{E}\, \nabla^2_{\boldsymbol{\theta}_W} \ell_{k,k}(\boldsymbol{\theta}_W, \boldsymbol{X}_{k,k}), && \text{for all } k = 1, \ldots, K \\ \boldsymbol{H}_{k,l}(\boldsymbol{\theta}_B) &\coloneqq -\mathbb{E}\, \nabla^2_{\boldsymbol{\theta}_B} \ell_{k,l}(\boldsymbol{\theta}_B, \boldsymbol{X}_{k,l}), && \text{for all } 1 \leq k < l \leq K. \end{aligned}$$

Note that the interchange of differentiation and integration in this setting is trivial as the expectations are finite sums.

We demonstrate that event (3.4) occurs with probability at least $1 - 2\,N^{-2}$ for the within-block and between-block cases separately. For all $\boldsymbol{\theta} \in \mathcal{B}_2(\boldsymbol{\theta}^\star, \epsilon^\star)$, an application of Weyl's inequality and Assumption (A.1) shows that

$$\begin{aligned} \lambda_{\min}(\boldsymbol{H}_W(\boldsymbol{\theta}_W)) &\geq \sum_{k=1}^{K} \lambda_{\min}(\boldsymbol{H}_{k,k}(\boldsymbol{\theta}_W)) \geq \xi_{W,\epsilon^\star}\, K \\ \lambda_{\min}(\boldsymbol{H}_B(\boldsymbol{\theta}_B)) &\geq \sum_{1 \leq k < l \leq K} \lambda_{\min}(\boldsymbol{H}_{k,l}(\boldsymbol{\theta}_B)) \geq \xi_{B,\epsilon^\star} \binom{K}{2}. \end{aligned} \tag{3.6}$$

Let $\epsilon_W \in (0, \epsilon^\star/\sqrt{2})$ and $\epsilon_B \in (0, \epsilon^\star/\sqrt{2})$ and assume that $\boldsymbol{\theta}_W \in \partial\mathcal{B}_2(\boldsymbol{\theta}_W^\star, \epsilon_W)$ and $\boldsymbol{\theta}_B \in \partial\mathcal{B}_2(\boldsymbol{\theta}_B^\star, \epsilon_B)$. By construction, $(\boldsymbol{\theta}_W, \boldsymbol{\theta}_B) \in \mathcal{B}_2(\boldsymbol{\theta}^\star, \epsilon^\star)$ and event (3.4) becomes

$$\begin{aligned} |\langle \boldsymbol{\theta}_W - \boldsymbol{\theta}_W^\star,\, \nabla_{\boldsymbol{\theta}_W} m_W(\boldsymbol{\theta}_W, \boldsymbol{X}_W) - \nabla_{\boldsymbol{\theta}_W} M_W(\boldsymbol{\theta}_W) \rangle| &\leq \epsilon_W^2\, \xi_{W,\epsilon^\star}\, K \\ |\langle \boldsymbol{\theta}_B - \boldsymbol{\theta}_B^\star,\, \nabla_{\boldsymbol{\theta}_B} m_B(\boldsymbol{\theta}_B, \boldsymbol{X}_B) - \nabla_{\boldsymbol{\theta}_B} M_B(\boldsymbol{\theta}_B) \rangle| &\leq \epsilon_B^2\, \xi_{B,\epsilon^\star} \binom{K}{2}, \end{aligned}$$

using the lower bound in (3.6). By the Cauchy-Schwarz inequality,

$$\begin{aligned} & |\langle \boldsymbol{\theta}_W - \boldsymbol{\theta}_W^\star,\, \nabla_{\boldsymbol{\theta}_W} m_W(\boldsymbol{\theta}_W, \boldsymbol{X}_W) - \nabla_{\boldsymbol{\theta}_W} M_W(\boldsymbol{\theta}_W) \rangle| \\ \leq\;& \|\boldsymbol{\theta}_W - \boldsymbol{\theta}_W^\star\|_2\, \|\nabla_{\boldsymbol{\theta}_W} m_W(\boldsymbol{\theta}_W, \boldsymbol{X}_W) - \nabla_{\boldsymbol{\theta}_W} M_W(\boldsymbol{\theta}_W)\|_2 \\ =\;& \epsilon_W\, \|\nabla_{\boldsymbol{\theta}_W} m_W(\boldsymbol{\theta}_W, \boldsymbol{X}_W) - \nabla_{\boldsymbol{\theta}_W} M_W(\boldsymbol{\theta}_W)\|_2 \\ \leq\;& \epsilon_W\, \sqrt{p}\, \|\nabla_{\boldsymbol{\theta}_W} m_W(\boldsymbol{\theta}_W, \boldsymbol{X}_W) - \nabla_{\boldsymbol{\theta}_W} M_W(\boldsymbol{\theta}_W)\|_\infty, \end{aligned}$$

additionally using the inequality $\|\boldsymbol{v}\|_2 \leq \sqrt{d}\, \|\boldsymbol{v}\|_\infty$ ($\boldsymbol{v} \in \mathbb{R}^d$). Similarly,

$$\begin{aligned} & |\langle \boldsymbol{\theta}_B - \boldsymbol{\theta}_B^\star,\, \nabla_{\boldsymbol{\theta}_B} m_B(\boldsymbol{\theta}_B, \boldsymbol{X}_B) - \nabla_{\boldsymbol{\theta}_B} M_B(\boldsymbol{\theta}_B) \rangle| \\ \leq\;& \epsilon_B\, \sqrt{q}\, \|\nabla_{\boldsymbol{\theta}_B} m_B(\boldsymbol{\theta}_B, \boldsymbol{X}_B) - \nabla_{\boldsymbol{\theta}_B} M_B(\boldsymbol{\theta}_B)\|_\infty. \end{aligned}$$

It suffices to demonstrate, for all $\boldsymbol{\theta}_W \in \partial\mathcal{B}_2(\boldsymbol{\theta}_W^\star, \epsilon_W)$ and $\boldsymbol{\theta}_B \in \partial\mathcal{B}_2(\boldsymbol{\theta}_B^\star, \epsilon_B)$, that events

$$\begin{aligned} \|\nabla_{\boldsymbol{\theta}_W} m_W(\boldsymbol{\theta}_W, \boldsymbol{X}_W) - \nabla_{\boldsymbol{\theta}_W} M_W(\boldsymbol{\theta}_W)\|_\infty &\leq \frac{\epsilon_W\, \xi_{W,\epsilon^\star}}{\sqrt{p}}\, K \\ \|\nabla_{\boldsymbol{\theta}_B} m_B(\boldsymbol{\theta}_B, \boldsymbol{X}_B) - \nabla_{\boldsymbol{\theta}_B} M_B(\boldsymbol{\theta}_B)\|_\infty &\leq \frac{\epsilon_B\, \xi_{B,\epsilon^\star}}{\sqrt{q}} \binom{K}{2} \end{aligned} \tag{3.7}$$

occur with probability at least $1 - 2\,N^{-2}$. Define, for all $t > 0$, the events

$$\begin{aligned} \mathcal{D}_W(t) &:= \left\{ \boldsymbol{x} \in \mathbb{X} : \sup_{\boldsymbol{\theta}_W \in \partial\mathcal{B}_2(\boldsymbol{\theta}_W^\star, \epsilon_W)} \|\nabla_{\boldsymbol{\theta}_W} m_W(\boldsymbol{\theta}_W, \boldsymbol{X}_W) - \nabla_{\boldsymbol{\theta}_W} M_W(\boldsymbol{\theta}_W)\|_\infty \geq t \right\} \\ \mathcal{D}_B(t) &:= \left\{ \boldsymbol{x} \in \mathbb{X} : \sup_{\boldsymbol{\theta}_B \in \partial\mathcal{B}_2(\boldsymbol{\theta}_B^\star, \epsilon_B)} \|\nabla_{\boldsymbol{\theta}_B} m_B(\boldsymbol{\theta}_B, \boldsymbol{X}_B) - \nabla_{\boldsymbol{\theta}_B} M_B(\boldsymbol{\theta}_B)\|_\infty \geq t \right\}. \end{aligned}$$

Applying Lemma 1 in the supplementary material, for all $t > 0$,

$$\begin{aligned}
\mathbb{P}\left(\mathcal{D}_W\left(\frac{\epsilon_W\,\xi_{W,\epsilon^\star}}{\sqrt{p}}\,K\right)\right) &\leq 2\,\exp\left(-\frac{2\,\epsilon_W^2\,\xi_{W,\epsilon^\star}^2\,N}{p\,L_W^2\,A_{\max}^{12}} + \log\,p\right)\\
\mathbb{P}\left(\mathcal{D}_B\left(\frac{\epsilon_B\,\xi_{B,\epsilon^\star}}{\sqrt{q}}\binom{K}{2}\right)\right) &\leq 2\,\exp\left(-\frac{2\,\epsilon_B^2\,\xi_{B,\epsilon^\star}^2\,N^2}{q\,L_B^2\,A_{\max}^{12}} + \log\,q\right),
\end{aligned}$$

using the inequality $K \geq N\,/\,A_{\max}$. As a result, choosing

$$\begin{aligned}
\epsilon_W &\coloneqq \left(\frac{L_W\,\sqrt{3}}{\xi_{W,\epsilon^\star}\sqrt{2}}\right) A_{\max}^6 \sqrt{\frac{p\,\log N}{N}}\\
\epsilon_B &\coloneqq \left(\frac{L_B\,\sqrt{3}}{\xi_{B,\epsilon^\star}\sqrt{2}}\right) A_{\max}^6 \sqrt{\frac{q\,\log N}{N^2}}
\end{aligned} \tag{3.8}$$

and assuming $\max\{p, q\} \leq N$ establishes the bound

$$\begin{aligned}
\mathbb{P}\left(\mathcal{D}_W\left(\frac{\epsilon_W\,\xi_{W,\epsilon^\star}}{\sqrt{p}}\,K\right)\right) &\leq 2\,N^{-2}\\
\mathbb{P}\left(\mathcal{D}_B\left(\frac{\epsilon_B\,\xi_{B,\epsilon^\star}}{\sqrt{q}}\binom{K}{2}\right)\right) &\leq 2\,N^{-2}.
\end{aligned}$$

As a result, event (3.7) occurs with probability at least $1 - 2\,N^{-2}$, implying there exist constants $C_1 \coloneqq (L_W/\xi_{W,\epsilon^\star})\sqrt{3/2} > 0$ and $C_2 \coloneqq (L_B/\xi_{B,\epsilon^\star})\sqrt{3/2} > 0$, independent of N, p, and q, such that, with probability at least $1 - 2\,N^{-2}$, the maximum likelihood estimator $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}_W, \widehat{\boldsymbol{\theta}}_B)$ exists uniquely and satisfies

$$\begin{aligned}
\|\widehat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_W^\star\|_2 &\leq C_1\,A_{\max}^6 \sqrt{\frac{p\,\log N}{N}}\\
\|\widehat{\boldsymbol{\theta}}_B - \boldsymbol{\theta}_B^\star\|_2 &\leq C_2\,A_{\max}^6 \sqrt{\frac{q\,\log N}{N^2}}.
\end{aligned}$$

Uniqueness follows from the assumption that the exponential-family local dependence random graph model is minimal along with Proposition 3.13 of Sundberg [39].

Finally, we show the restriction to $\mathcal{B}_2(\boldsymbol{\theta}^\star, \epsilon^\star)$ to be legitimate. Having assumed that

$$\begin{aligned}
\lim_{N\to\infty}\, A_{\max}^6 \sqrt{\frac{p\,\log N}{N}} &= 0\\
\lim_{N\to\infty}\, A_{\max}^6 \sqrt{\frac{q\,\log N}{N^2}} &= 0,
\end{aligned}$$

there exists an $N_0 \geq 3$ such that, for all $N \geq N_0$, $\max\{\epsilon_W, \epsilon_B\} \leq \epsilon^\star/\sqrt{2}$. Thus, for all $N \geq N_0$ and with probability at least $1 - 2\,N^{-2}$, the unique vector $\widehat{\boldsymbol{\theta}} \in \widehat{\boldsymbol{\Theta}}$ satisfies

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star\|_2 \;=\; \sqrt{\|\widehat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_W^\star\|_2^2 + \|\widehat{\boldsymbol{\theta}}_B - \boldsymbol{\theta}_B^\star\|_2^2} \;\leq\; \sqrt{\epsilon_W^2 + \epsilon_B^2} \;<\; \epsilon^\star,$$

which implies, for all $N \geq N_0$, that

$$\mathbb{P}(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star\|_2 \leq \epsilon^\star) \quad \geq \quad 1 - 2\, N^{-2},$$

justifying the restriction to the subset of the parameter space $\mathcal{B}_2(\boldsymbol{\theta}^\star, \epsilon^\star) \subset \mathbb{R}^{p+q}$. □

3.2. Proof of Theorem 2

For ease of presentation, we first present a general argument for bounding the error of the multivariate normal approximation, and then show how it can be applied to maximum likelihood estimators of exponential-families of local dependence random graph models in order to establish the desired result.

Bounding the error of the multivariate normal approximation. Consider a general estimating function $m : \mathbb{R}^d \times \mathbb{X} \mapsto \mathbb{R}$ which admits the following form:

$$m(\boldsymbol{\theta}, \boldsymbol{x}) \quad = \quad \sum_{k=1}^{K} m_{k,k}(\boldsymbol{\theta}_W, \boldsymbol{x}_{k,k}) + \sum_{1 \leq k < l \leq K} m_{k,l}(\boldsymbol{\theta}_B, \boldsymbol{x}_{k,l}),$$

and assume that $m(\boldsymbol{\theta}, \boldsymbol{x})$ is thrice continuously differentiable in elements of $\boldsymbol{\theta} \in \mathbb{R}^{p+q}$. By assumption, the subgraphs $\boldsymbol{X}_{k,l}$ $(1 \leq k \leq l \leq K)$ are mutually independent, implying that for a fixed $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the collection of random variables $m_{k,k}(\boldsymbol{\theta}_W, \boldsymbol{X}_{k,k})$ $(k = 1, \ldots, K)$ and $m_{k,l}(\boldsymbol{\theta}_B, \boldsymbol{X}_{k,l})$ $(1 \leq k < l \leq K)$ are likewise mutually independent. Observe that

$$\nabla_{\boldsymbol{\theta}}\, m(\boldsymbol{\theta}, \boldsymbol{x}) \quad = \quad \sum_{k=1}^{K} \nabla_{\boldsymbol{\theta}}\, m_{k,k}(\boldsymbol{\theta}_W, \boldsymbol{x}_{k,k}) + \sum_{1 \leq k < l \leq K} \nabla_{\boldsymbol{\theta}}\, m_{k,l}(\boldsymbol{\theta}_B, \boldsymbol{x}_{k,l}), \tag{3.9}$$

noting that the collection of random vectors $\nabla_{\boldsymbol{\theta}}\, m_{k,k}(\boldsymbol{\theta}_W, \boldsymbol{X}_{k,k})$ $(k = 1, \ldots, K)$ and $\nabla_{\boldsymbol{\theta}}\, m_{k,l}(\boldsymbol{\theta}_B, \boldsymbol{X}_{k,l})$ $(1 \leq k < l \leq K)$ are also mutually independent. Assume that

$$\begin{aligned} \mathbb{E}\, \nabla_{\boldsymbol{\theta}}\, m_{k,k}(\boldsymbol{\theta}_W^\star, \boldsymbol{X}_{k,k}) \quad &= \quad \boldsymbol{0}_{p+q}, \qquad k = 1, \ldots, K \\ \mathbb{E}\, \nabla_{\boldsymbol{\theta}}\, m_{k,l}(\boldsymbol{\theta}_B^\star, \boldsymbol{X}_{k,l}) \quad &= \quad \boldsymbol{0}_{p+q}, \qquad 1 \leq k < l \leq K. \end{aligned} \tag{3.10}$$

As a result, $\mathbb{E}\, \nabla_{\boldsymbol{\theta}}\, m(\boldsymbol{\theta}^\star, \boldsymbol{X}) = \boldsymbol{0}_{p+q}$,

Let $\boldsymbol{\theta} \in \mathbb{R}^{p+q}$ and $\boldsymbol{x} \in \mathbb{X}$ be fixed. By a multivariate Taylor expansion,

$$\nabla_{\boldsymbol{\theta}}\, m(\boldsymbol{\theta}, \boldsymbol{X}) \quad = \quad \nabla_{\boldsymbol{\theta}}\, m(\boldsymbol{\theta}^\star, \boldsymbol{X}) + \nabla_{\boldsymbol{\theta}}^2\, m(\boldsymbol{\theta}^\star, \boldsymbol{X})\, (\boldsymbol{\theta} - \boldsymbol{\theta}^\star) + \boldsymbol{R}, \tag{3.11}$$

where $\boldsymbol{R} \in \mathbb{R}^{p+q}$ is a vector of remainders in the Lagrange form. The remainder terms R_i $(i = 1, \ldots, p+q)$ are given by

$$\begin{aligned} R_i \quad = \quad & \sum_{j=1}^{p+q} \frac{1}{2} \left[\frac{\partial^2}{\partial\, \theta_j^2} \left[\nabla_{\boldsymbol{\theta}}\, m(\dot{\boldsymbol{\theta}}^{(i)}, \boldsymbol{X}) \right]_i \right] (\theta_j - \theta_j^\star)^2 \\ & + \sum_{1 \leq j < r \leq p+q} \frac{1}{2} \left[\frac{\partial^2}{\partial\, \theta_j\, \partial\, \theta_r} \left[\nabla_{\boldsymbol{\theta}}\, m(\dot{\boldsymbol{\theta}}^{(i)}, \boldsymbol{X}) \right]_i \right] (\theta_j - \theta_j^\star)\, (\theta_r - \theta_r^\star), \end{aligned} \tag{3.12}$$

where $\dot{\boldsymbol{\theta}}^{(i)} = t_i\, \boldsymbol{\theta} + (1 - t_i)\, \boldsymbol{\theta}^\star$ (for some $t_i \in [0,1]$, $i = 1, \ldots, p+q$). Assume that $\boldsymbol{C} \coloneqq \mathbb{V}\, \nabla_{\boldsymbol{\theta}}\, m(\boldsymbol{\theta}^\star, \boldsymbol{X})$ is non-singular and that $\nabla_{\boldsymbol{\theta}}\, m(\boldsymbol{\theta}, \boldsymbol{x})$ has a root given by $\boldsymbol{\theta}_0 \in \mathbb{R}^{p+q}$. Taking $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, we re-arrange (3.11) with the observation $\boldsymbol{X} = \boldsymbol{x}$ in order to obtain

$$\nabla_{\boldsymbol{\theta}}\, m(\boldsymbol{\theta}^\star, \boldsymbol{x}) \quad = \quad \nabla^2_{\boldsymbol{\theta}}\, m(\boldsymbol{\theta}^\star, \boldsymbol{x})\, (\boldsymbol{\theta}^\star - \boldsymbol{\theta}_0) - \boldsymbol{R}.$$

From the form of (3.9), $\nabla_{\boldsymbol{\theta}}\, m(\boldsymbol{\theta}^\star, \boldsymbol{x})$ is a sum of independent random vectors. Define

$$\boldsymbol{Y}_{k,l} \quad \coloneqq \quad \begin{cases} \boldsymbol{C}^{-1/2}\;\; \nabla_{\boldsymbol{\theta}}\, m_{k,k}(\boldsymbol{\theta}^\star_W, \boldsymbol{X}_{k,k}), & \text{if } k = l \\[2mm] \boldsymbol{C}^{-1/2}\;\; \nabla_{\boldsymbol{\theta}}\, m_{k,l}(\boldsymbol{\theta}^\star_B, \boldsymbol{X}_{k,l}), & \text{if } k \neq l \end{cases},$$

and $\boldsymbol{S} \coloneqq \sum_{1 \leq k \leq l \leq K} Y_{k,l}$. Denote the distribution of $\boldsymbol{S}$ by $\mathbb{P}_{\boldsymbol{S}}$. Observe that, by (3.10), $\mathbb{E}\, \boldsymbol{S} = \boldsymbol{0}_{p+q}$ and that, by the definition of $\boldsymbol{C}$, $\mathbb{V}\, \boldsymbol{S} = \boldsymbol{I}_{p+q}$. Applying Lemma 3 in the supplementary materials, for all measurable convex sets $\mathcal{A} \subset \mathbb{R}^{p+q}$,

$$\begin{aligned} |\mathbb{P}_{\boldsymbol{S}}(\boldsymbol{S} \in \mathcal{A}) - \Phi(\boldsymbol{Z} \in \mathcal{A})| \quad &\leq \quad (42\, (p+q)^{1/4} + 16) \sum_{1 \leq k \leq l \leq K} \mathbb{E}\, \|\boldsymbol{Y}_{k,l}\|_2^3 \\ &\leq \quad 58\, (p+q)^{1/4} \sum_{1 \leq k \leq l \leq K} \mathbb{E}\, \|\boldsymbol{Y}_{k,l}\|_2^3. \end{aligned}$$

Normality results for $\boldsymbol{S}$ can be extended to a standardization of $(\boldsymbol{\theta}^\star - \boldsymbol{\theta})$ via:

$$\boldsymbol{S} \quad \overset{D}{=} \quad \boldsymbol{C}^{-1/2} \left[\nabla^2_{\boldsymbol{\theta}}\, m(\boldsymbol{\theta}^\star, \boldsymbol{x})\, (\boldsymbol{\theta}^\star - \boldsymbol{\theta}_0) - \boldsymbol{R}\right]. \tag{3.13}$$

Multivariate normal approximation for maximum likelihood estimators. Define

$$\begin{aligned} m_{k,k}(\boldsymbol{\theta}_W, \boldsymbol{x}_{k,k}) \quad &\coloneqq \quad \ell_{k,k}(\boldsymbol{\theta}_W, \boldsymbol{x}_{k,k}), \qquad k = 1, \ldots, K, \\ m_{k,l}(\boldsymbol{\theta}_B, \boldsymbol{x}_{k,l}) \quad &\coloneqq \quad \ell_{k,l}(\boldsymbol{\theta}_B, \boldsymbol{x}_{k,l}), \qquad 1 \leq k < l \leq K. \end{aligned}$$

We verify that the assumptions placed on $m(\boldsymbol{\theta}, \boldsymbol{x})$ in the general argument presented above are met in the case of maximum likelihood estimation.

By Lemma 8 in the supplementary material,

$$\begin{aligned} \nabla_{\boldsymbol{\theta}}\, m_{k,k}(\boldsymbol{\theta}_W, \boldsymbol{x}_{k,k}) \quad &= \quad (s_{k,k}(\boldsymbol{x}_{k,k}) - \mathbb{E}_{\boldsymbol{\theta}}\, s_{k,k}(\boldsymbol{X}_{k,k}),\, \boldsymbol{0}_q), \\ \nabla_{\boldsymbol{\theta}}\, m_{k,l}(\boldsymbol{\theta}_B, \boldsymbol{x}_{k,l}) \quad &= \quad (\boldsymbol{0}_p,\, s_{k,l}(\boldsymbol{x}_{k,l}) - \mathbb{E}_{\boldsymbol{\theta}}\, s_{k,l}(\boldsymbol{X}_{k,l})), \end{aligned}$$

noting that $\boldsymbol{\theta} = (\boldsymbol{\theta}_W, \boldsymbol{\theta}_B) \in \mathbb{R}^{p+q}$, implying $\nabla_{\boldsymbol{\theta}}\, m(\boldsymbol{\theta}, \boldsymbol{x}) = s(\boldsymbol{x}) - \mathbb{E}_{\boldsymbol{\theta}}\, s(\boldsymbol{x})$. Observe that

$$\begin{aligned} \mathbb{E}\, [\nabla_{\boldsymbol{\theta}}\, m_{k,k}(\boldsymbol{\theta}^\star_W, \boldsymbol{X}_{k,k})] \quad &= \quad \boldsymbol{0}_{p+q}, \qquad k = 1, \ldots, K, \\ \mathbb{E}\, [\nabla_{\boldsymbol{\theta}}\, m_{k,l}(\boldsymbol{\theta}^\star_B, \boldsymbol{X}_{k,l})] \quad &= \quad \boldsymbol{0}_{p+q}, \qquad 1 \leq k < l \leq K, \end{aligned}$$

implying $\mathbb{E} \nabla_{\boldsymbol{\theta}}\, m(\boldsymbol{\theta}^\star, \boldsymbol{X}) = \boldsymbol{0}_{p+q}$. Lemma 8 additionally establishes that

$$\nabla^2_{\boldsymbol{\theta}}\, m(\boldsymbol{\theta}^\star, \boldsymbol{x}) \quad = \quad \mathbb{V}\, s(\boldsymbol{X}) \quad = \quad \mathbb{V}\, (s(\boldsymbol{X}) - \mathbb{E}\, s(\boldsymbol{X})) \quad = \quad \mathbb{V}\, \nabla_{\boldsymbol{\theta}}\, m(\boldsymbol{\theta}^\star, \boldsymbol{X}),$$

implying $\boldsymbol{C} = \mathbb{V}\, s(\boldsymbol{X}) = \nabla_{\boldsymbol{\theta}}^2\, m(\boldsymbol{\theta}, \boldsymbol{x}) = -\mathbb{E}\, \nabla_{\boldsymbol{\theta}}^2\, \ell(\boldsymbol{\theta}, \boldsymbol{X})$, non-singular for all $\boldsymbol{\theta} \in \mathcal{B}_2(\boldsymbol{\theta}^\star, \epsilon^\star)$ by Assumption (A.1). Restricting to $\boldsymbol{\theta} \in \mathcal{B}_2(\boldsymbol{\theta}^\star \epsilon^\star)$, we have verified all conditions placed on $m(\boldsymbol{\theta}, \boldsymbol{x})$ in the general argument outlined above. In this case, (3.13) reduces to

$$\boldsymbol{S} \;\; \stackrel{D}{=} \;\; I(\boldsymbol{\theta}^\star)^{1/2}\, (\boldsymbol{\theta}^\star - \boldsymbol{\theta}) - I(\boldsymbol{\theta}^\star)^{-1/2}\, \boldsymbol{R},$$

where $I(\boldsymbol{\theta}^\star) = \mathbb{V}\, s(\boldsymbol{X})$ is the Fisher information matrix evaluated at the data-generating parameter vector $\boldsymbol{\theta}^\star \in \mathbb{R}^{p+q}$. The local dependence assumption and the partitioning of $s(\boldsymbol{X}) = (s_W(\boldsymbol{X}_W), s_B(\boldsymbol{X}_B))$ imply that

$$I(\boldsymbol{\theta}^\star) \;\; = \;\; \begin{pmatrix} I_W(\boldsymbol{\theta}_W^\star) & \mathbf{0}_{p,q} \\ \mathbf{0}_{q,p} & I_B(\boldsymbol{\theta}_B^\star) \end{pmatrix},$$

where $\mathbf{0}_{d,r}$ is the $(d \times r)$-dimensional matrix consisting of all zeros, where

$$\begin{aligned} I_W(\boldsymbol{\theta}_W^\star) \;\; &:= \;\; \sum_{k=1}^{K} -\mathbb{E}\, \nabla_{\boldsymbol{\theta}_W}^2\, \ell_{k,k}(\boldsymbol{\theta}_W^\star, \boldsymbol{X}_{k,k}) \\ I_B(\boldsymbol{\theta}_B^\star) \;\; &:= \;\; \sum_{1 \leq k < l \leq K} -\mathbb{E}\, \nabla_{\boldsymbol{\theta}_B}^2\, \ell_{k,l}(\boldsymbol{\theta}_B^\star, \boldsymbol{X}_{k,l}). \end{aligned}$$

The proof is completed by establishing the following two additional results.

I. **Convergence rate of the multivariate normal approximation**

We establish the convergence rate of the multivariate normal approximation by bounding $\sum_{1 \leq k \leq l \leq K} \mathbb{E}\, \|\boldsymbol{Y}_{k,l}\|_2^3$. In order to do so, we bound each term:

$$\begin{aligned} \|\boldsymbol{Y}_{k,k}\|_2 \;\; &= \;\; \|I_W(\boldsymbol{\theta}_W^\star)^{-1/2}\, [s_{k,k}(\boldsymbol{x}_{k,k}) - \mathbb{E}_{\boldsymbol{\theta}}\, s_{k,k}(\boldsymbol{X}_{k,k})]\, \|_2 \\ &\leq \;\; \|I_W(\boldsymbol{\theta}_W^\star)^{-1/2}\|_2\, \|s_{k,k}(\boldsymbol{x}_{k,k}) - \mathbb{E}_{\boldsymbol{\theta}}\, s_{k,k}(\boldsymbol{X}_{k,k})\|_2 \\ &\leq \;\; \frac{\|s_{k,k}(\boldsymbol{x}_{k,k}) - \mathbb{E}_{\boldsymbol{\theta}}\, s_{k,k}(\boldsymbol{X}_{k,k})\|_2}{\sqrt{\xi_{W,\epsilon^\star}\, K}}, \end{aligned}$$

using the bound

$$\sup_{k \in \{1,\ldots,K\}} \frac{1}{\lambda_{\min}(-\mathbb{E}\, \nabla_{\boldsymbol{\theta}_W}^2\, \ell_{k,k}(\boldsymbol{\theta}_W, \boldsymbol{X}_{k,k}))\, K} \;\; \leq \;\; \frac{1}{\xi_{W,\epsilon^\star}\, K},$$

which follows from Weyl's inequality and Assumption (A.1):

$$\lambda_{\min}(I_W(\boldsymbol{\theta}_W^\star)) \;\; \geq \;\; \sum_{k=1}^{K} \lambda_{\min}(-\mathbb{E}\, \nabla_{\boldsymbol{\theta}_W}^2\, \ell_{k,k}(\boldsymbol{\theta}_W^\star, \boldsymbol{X})) \;\; \geq \;\; \xi_{W,\epsilon^\star}\, K,$$

noting that $I_W(\boldsymbol{\theta}_W^\star) = \sum_{k=1}^K -\mathbb{E}\, \nabla_{\boldsymbol{\theta}_W}^2\, \ell_{k,k}(\boldsymbol{\theta}_W^\star, \boldsymbol{X})$. By the definition of L_W,

$$\begin{aligned}
\|s_{k,k}(\boldsymbol{x}_{k,k}) - \mathbb{E}_{\boldsymbol{\theta}}\, s_{k,k}(\boldsymbol{X}_{k,k})\|_\infty &\leq \sup_{\boldsymbol{x}_{k,k}' \in \mathbb{X}_{k,k}} \|s_{k,k}(\boldsymbol{x}_{k,k}) - s_{k,k}(\boldsymbol{x}_{k,k}')\|_\infty \\
&\leq \sup_{\boldsymbol{x}_{k,k}' \in \mathbb{X}_{k,k}} L_W \binom{|\mathcal{A}_k|}{2} d_H(\boldsymbol{x}_{k,k}, \boldsymbol{x}_{k,k}') \\
&\leq L_W \binom{|\mathcal{A}_K|}{2}^2,
\end{aligned}$$

which in turn implies the inequality

$$\begin{aligned}
\|s_{k,k}(\boldsymbol{x}_{k,k}) - \mathbb{E}_{\boldsymbol{\theta}}\, s_{k,k}(\boldsymbol{X}_{k,k})\|_2 &\leq \sqrt{p}\, \|s_{k,k}(\boldsymbol{x}_{k,k}) - \mathbb{E}_{\boldsymbol{\theta}}\, s_{k,k}(\boldsymbol{X}_{k,k})\|_\infty \\
&\leq \sqrt{p}\, L_W \binom{|\mathcal{A}_k|}{2}^2 \leq \frac{\sqrt{p}\, L_W\, A_{\max}^4}{4}.
\end{aligned}$$

Collecting the above bounds,

$$\|\boldsymbol{Y}_{k,k}\|_2^3 \leq \left(\frac{\sqrt{p}\, L_W\, A_{\max}^4}{4\sqrt{\xi_{W,\epsilon^\star}\, K}}\right)^3 = \frac{p^{3/2}\, L_W^3\, A_{\max}^{12}}{64\, \xi_{W,\epsilon^\star}^{3/2}\, K^{3/2}},$$

which implies

$$\sum_{k=1}^K \mathbb{E}\, \|\boldsymbol{Y}_{k,k}\|_2^3 \leq \frac{p^{3/2}\, L_W^3\, A_{\max}^{12}}{64\, \xi_{W,\epsilon^\star}^{3/2}\, K^{1/2}}.$$

A similar argument will reveal, by the definition of L_B,

$$\begin{aligned}
\|s_{k,l}(\boldsymbol{x}_{k,l}) - \mathbb{E}_{\boldsymbol{\theta}}\, s_{k,l}(\boldsymbol{X}_{k,l})\|_\infty &\leq \sup_{\boldsymbol{x}_{k,l}' \in \mathbb{X}_{k,l}} \|s_{k,l}(\boldsymbol{x}_{k,l}) - s_{k,l}(\boldsymbol{x}_{k,l}')\|_\infty \\
&\leq \sup_{\boldsymbol{x}_{k,l}' \in \mathbb{X}_{k,l}} L_B\, |\mathcal{A}_k|\, |\mathcal{A}_l|\, d_H(\boldsymbol{x}_{k,l}, \boldsymbol{x}_{k,l}') \\
&\leq L_B\, (|\mathcal{A}_k|\, |\mathcal{A}_l|)^2,
\end{aligned}$$

for $1 \leq k < l \leq K$, which will instead yield the bound

$$\sum_{1 \leq k < l \leq K} \mathbb{E}\, \|\boldsymbol{Y}_{k,l}\|_2^3 \leq \frac{q^{3/2}\, L_B^3\, A_{\max}^{12}}{\xi_{B,\epsilon^\star}^{3/2} \binom{K}{2}^{1/2}} \leq \frac{\sqrt{2}\, q^{3/2}\, L_B^3\, A_{\max}^{12}}{\xi_{B,\epsilon^\star}^{3/2}\, K},$$

noting that there are $\binom{K}{2}$ between-block subgraphs $\boldsymbol{X}_{k,l}$ ($1 \leq k < l \leq K$), as opposed to K within-block subgraphs $\boldsymbol{X}_{k,k}$ ($k = 1, \ldots, K$). Collecting terms and using the inequality $K \geq N \,/\, A_{\max}$,

$$\begin{aligned}\sum_{1 \le k \le l \le K} \mathbb{E}\, \|\boldsymbol{Y}_{k,l}\|_2^3 &\leq A_{\max}^{12} \left[\frac{p^{3/2}\, L_W^3}{\xi_{W,\epsilon^\star}^{3/2}\, K^{1/2}} + \frac{\sqrt{2}\, q^{3/2}\, L_B^3}{\xi_{B,\epsilon^\star}^{3/2}\, K}\right] \\ &\leq \sqrt{2}\, A_{\max}^{12}\, (p+q)^{3/2} \left[\frac{L_W^3}{\xi_{W,\epsilon^\star}^{3/2}\, K^{1/2}} + \frac{L_B^3}{\xi_{B,\epsilon^\star}^{3/2}\, K}\right] \\ &\leq \sqrt{2}\, A_{\max}^{13}\, (p+q)^{3/2} \left[\frac{L_W^3}{\xi_{W,\epsilon^\star}^{3/2}\, N^{1/2}} + \frac{L_B^3}{\xi_{B,\epsilon^\star}^{3/2}\, N}\right].\end{aligned}$$

Thus, there exists a constant $C \coloneqq 58\sqrt{2}\, \max\{L_W^3, L_B^3\} / \min\{\xi_{W,\epsilon^\star}^{3/2}, \xi_{B,\epsilon^\star}^{3/2}\} > 0$, independent of N, p, and q, such that, for all measurable convex sets $\mathcal{A} \subset \mathbb{R}^{p+q}$, the error of the multivariate normal approximation is bounded above by

$$\begin{aligned}|\mathbb{Q}(\mathcal{A}) - \Phi(\mathcal{A})| &\leq 58\, (p+q)^{1/4} \sum_{1 \le k \le l \le K} \mathbb{E}\, \|\boldsymbol{Y}_{k,l}\|_2^3 \\ &\leq C \sqrt{\frac{(p+q)^{7/2}\, A_{\max}^{26}}{N}},\end{aligned}$$

where $\mathbb{Q}$ is the probability distribution of $I(\boldsymbol{\theta}^\star)^{1/2}\, (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star) - \widetilde{\boldsymbol{R}}$.

II. **Demonstrating that $\|I(\boldsymbol{\theta}^\star)^{-1/2}\, \boldsymbol{R}\|_2$ is small with high probability.**

Recall that

$$I(\boldsymbol{\theta}^\star) = \begin{pmatrix} I_W(\boldsymbol{\theta}_W^\star) & \mathbf{0}_{p,q} \\ \mathbf{0}_{q,p} & I_B(\boldsymbol{\theta}_B^\star) \end{pmatrix},$$

where

$$\begin{aligned}I_W(\boldsymbol{\theta}_W^\star) &\coloneqq \sum_{k=1}^K -\mathbb{E}\, \nabla_{\boldsymbol{\theta}_W}^2\, \ell_{k,k}(\boldsymbol{\theta}_W^\star, \boldsymbol{X}_{k,k}) \\ I_B(\boldsymbol{\theta}_B^\star) &\coloneqq \sum_{1 \le k < l \le K} -\mathbb{E}\, \nabla_{\boldsymbol{\theta}_B}^2\, \ell_{k,l}(\boldsymbol{\theta}_B^\star, \boldsymbol{X}_{k,l}),\end{aligned}$$

which implies that

$$I(\boldsymbol{\theta}^\star)^{-1} = \begin{pmatrix} I_W(\boldsymbol{\theta}_W^\star)^{-1} & \mathbf{0}_{p,q} \\ \mathbf{0}_{q,p} & I_B(\boldsymbol{\theta}_B^\star)^{-1} \end{pmatrix}.$$

Applying Weyl's inequality together with Assumption (A.1), we obtain the bounds

$$\lambda_{\min}(I_W(\boldsymbol{\theta}^\star)) \geq \sum_{k=1}^K \lambda_{\min}(-\mathbb{E}\, \nabla_{\boldsymbol{\theta}_W}^2\, \ell_{k,k}(\boldsymbol{\theta}_W^\star, \boldsymbol{X}_{k,k})) \geq \xi_{W,\epsilon^\star}\, K \tag{3.14}$$

and

$$\lambda_{\min}(I_B(\boldsymbol{\theta}_B^\star)) \;\geq\; \sum_{1\leq k<l\leq K} \lambda_{\min}(-\mathbb{E}\,\nabla^2_{\boldsymbol{\theta}_B}\,\ell_{k,l}(\boldsymbol{\theta}_B^\star, \boldsymbol{X}_{k,l})) \;\geq\; \xi_{B,\epsilon^\star}\binom{K}{2}. \tag{3.15}$$

Using (3.14) and (3.15), we can bound $\|I(\boldsymbol{\theta}^\star)^{-1/2}\,\boldsymbol{R}\|_2$ by

$$\begin{aligned}
\|I(\boldsymbol{\theta}^\star)^{-1/2}\,\boldsymbol{R}\|_2^2 \;&=\; \|I_W(\boldsymbol{\theta}^\star)^{-1/2}\,\boldsymbol{R}_W\|_2^2 + \|I_B(\boldsymbol{\theta}^\star)^{-1/2}\,\boldsymbol{R}_B\|_2^2 \\
&\leq\; \|\!|I_W(\boldsymbol{\theta}^\star)^{-1/2}|\!\|_2^2\,\|\boldsymbol{R}_W\|_2^2 + \|\!|I_B(\boldsymbol{\theta}^\star)^{-1/2}|\!\|^2\,\|\boldsymbol{R}_B\|_2^2 \\
&\leq\; \frac{\|\boldsymbol{R}_W\|_2^2}{\xi_{W,\epsilon^\star}\,K} + \frac{\|\boldsymbol{R}_B\|_2^2}{\xi_{B,\epsilon^\star}\,\binom{K}{2}},
\end{aligned}$$

where $\boldsymbol{R}_W \coloneqq (R_1, \ldots, R_p)$ and $\boldsymbol{R}_B \coloneqq (R_{p+1}, \ldots, R_{p+q})$. As a result,

$$\|I(\boldsymbol{\theta}^\star)^{-1/2}\,\boldsymbol{R}\|_2 \;\leq\; \sqrt{\frac{\|\boldsymbol{R}_W\|_2^2}{\xi_{W,\epsilon^\star}\,K} + \frac{\|\boldsymbol{R}_B\|_2^2}{\xi_{B,\epsilon^\star}\,\binom{K}{2}}}.$$

Applying Lemma 4 in the supplementary materials,

$$\begin{aligned}
\|\boldsymbol{R}_W\|_2^2 \;&\leq\; 4\,p\,C_W^2\,A_{\max}^{20}\,(L_W+2)^4\,K^2\,\|\widehat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_W^\star\|_1^4 \\
\|\boldsymbol{R}_B\|_2^2 \;&\leq\; 4\,q\,C_B^2\,A_{\max}^{20}\,(L_B+2)^4\binom{K}{2}^2\|\widehat{\boldsymbol{\theta}}_B - \boldsymbol{\theta}_B^\star\|_1^4.
\end{aligned}$$

Using the inequality $N \geq K$, $\|I(\boldsymbol{\theta}^\star)^{-1/2}\,\boldsymbol{R}\|_2$ is bounded above by

$$\begin{aligned}
&2\,A_{\max}^{10}\sqrt{\frac{p\,C_W^2\,(L_W+2)^4\,K^2\,\|\widehat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_W^\star\|_1^4}{\xi_{W,\epsilon^\star}\,K} + \frac{q\,C_B^2\,(L_B+2)^4\binom{K}{2}^2\|\widehat{\boldsymbol{\theta}}_B - \boldsymbol{\theta}_B^\star\|_1^4}{\xi_{B,\epsilon^\star}\,\binom{K}{2}}} \\
\leq\;& C_3\,A_{\max}^{10}\sqrt{p\,N\,\|\widehat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_W^\star\|_1^4 + q\,N^2\,\|\widehat{\boldsymbol{\theta}}_B - \boldsymbol{\theta}_B^\star\|_1^4},
\end{aligned}$$

where $C_3 \coloneqq 2\,\max\{C_W^2\,(L_W+2)^4,\, C_B\,(L_B+2)^4\}\,/\,\min\{\sqrt{\xi_{W,\epsilon^\star}},\, \sqrt{\xi_{B,\epsilon^\star}}\} > 0$ is a constant independent of N, p, and q. By Theorem 1, there exist constants $C_1 > 0$, $C_2 > 0$, and $N_0 \geq 3$, independent of N, p, and q, such that, for all $N \geq N_0$, the following hold with probability at least $1 - 2\,N^{-2}$: $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}_W, \widehat{\boldsymbol{\theta}}_B)$ exists and satisfies

$$\|\widehat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_W^\star\|_2 \;\leq\; C_1\,A_{\max}^6\sqrt{\frac{p\,\log N}{N}}$$

and

$$\|\widehat{\boldsymbol{\theta}}_B - \boldsymbol{\theta}_B^\star\|^2 \;\leq\; C_2\,A_{\max}^6\sqrt{\frac{q\,\log N}{N^2}},$$

provided each upper bound tends to 0 as $N \to \infty$. As a result,

$$\begin{aligned}\|\widehat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_W^\star\|_1 &\le \sqrt{p}\, \|\widehat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_W^\star\|_2 \le C_1\, p\, A_{\max}^6 \sqrt{\frac{\log N}{N}} \\ \|\widehat{\boldsymbol{\theta}}_B - \boldsymbol{\theta}_B^\star\|_1 &\le \sqrt{q}\, \|\widehat{\boldsymbol{\theta}}_B - \boldsymbol{\theta}_B^\star\|_2 \le C_2\, q\, A_{\max}^6 \sqrt{\frac{\log N}{N^2}},\end{aligned}$$

which leads to the bound

$$\begin{aligned}\|I(\boldsymbol{\theta}^\star)^{-1/2}\, \boldsymbol{R}\|_2 &\le C_3\, A_{\max}^{10} \sqrt{p\, N\, \|\widehat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_W^\star\|_1^4 + q\, N^2\, \|\widehat{\boldsymbol{\theta}}_B - \boldsymbol{\theta}_B^\star\|_1^4} \\ &\le C_3\, A_{\max}^{10} \sqrt{\frac{C_1^4\, p^5\, A_{\max}^{24}\, (\log N)^2}{N} + \frac{C_2^4\, q^5\, A_{\max}^{24}\, (\log N)^2}{N^2}} \\ &\le C_3\, \max\{C_1^2,\, C_2^2\}\, A_{\max}^{22}\, \log N \sqrt{\frac{p^5}{N} + \frac{q^5}{N^2}}.\end{aligned}$$

Thus, there exists a constant $C \coloneqq C_3\, \max\{C_1^2,\, C_2^2\} > 0$, independent of N, p, and q, and a random vector $\widetilde{\boldsymbol{R}} \coloneqq I(\boldsymbol{\theta}^\star)^{-1/2}\, \boldsymbol{R}$, such that

$$\mathbb{P}\left(\|\widetilde{\boldsymbol{R}}\|_2 \le C\, A_{\max}^{22}\, \log N \sqrt{\frac{p^5}{N} + \frac{q^5}{N^2}}\right) \ge 1 - 2\, N^{-2}.$$

Conclusion of proof. We have thus shown—recycling notation of constants—that there exist $C_1 > 0$, $C_2 > 0$, and $N_0 \ge 3$, independent of N, p, and q, and a random vector $\widetilde{\boldsymbol{R}} \in \mathbb{R}^{p+q}$ such that, for all $N > N_0$ and all measurable convex sets $\mathcal{A} \subset \mathbb{R}^{p+q}$,

$$|\mathbb{Q}(\mathcal{A}) - \Phi(\mathcal{A})| \le C_1 \sqrt{\frac{(p+q)^{7/2}\, A_{\max}^{26}}{N}},$$

where $\mathbb{Q}$ is the probability distribution of $I(\boldsymbol{\theta}^\star)^{1/2}\, (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star) - \widetilde{\boldsymbol{R}}$ and

$$\mathbb{P}\left(\|\widetilde{\boldsymbol{R}}\|_2 \le C_2\, A_{\max}^{22}\, \log N \sqrt{\frac{p^5}{N} + \frac{q^5}{N^2}}\right) \ge 1 - 2\, N^{-2}.$$

□

Acknowledgements

This research was supported by the Test Resource Management Center (TRMC) within the Office of the Secretary of Defense (OSD), contract #FA807518D0002. The author thanks Dr. Michael Schweinberger for helpful conversations and comments.

References

[1] Agneessens, F., Trincado-Munoz, F. J. and Koskinen, J. (2022). Network formation in organizational settings: Exploring the importance of local social processes and team-level contextual variables in small groups using bayesian hierarchical ERGMs. *Social Networks*.

[2] Babkin, S., Stewart, J. R., Long, X. and Schweinberger, M. (2020). Large-scale estimation of random graph models with local dependence. *Computational Statistics and Data Analysis* **152** 107029.

[3] Brown, L. D. (1986). *Fundamentals of statistical exponential families: with applications in statistical decision theory.* Institute of Mathematical Statistics.

[4] Chatterjee, S., Diaconis, P. and Sly, A. (2011). Random graphs with a given degree sequence. *The Annals of Applied Probability* **21** 1400-1435.

[5] Chatterjee, S. and Diaconis, P. (2013). Estimating and understanding exponential random graph models. *The Annals of Statistics* **41** 2428-2461.

[6] Chazottes, J. R., Collet, P., Külske, C. and Redig, F. (2007). Concentration inequalities for random fields via coupling. *Probability Theory and Related Fields* **137** 201-225.

[7] Chen, M., Kato, K. and Leng, C. (2021). Analysis of networks via the sparse β-model. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **83** 887-910.

[8] 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*.

[9] 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 Papers and Proceedings* **113** 156-160.

[10] Frank, O. and Strauss, D. (1986). Markov Graphs. *Journal of the American Statistical Association* **81** 832-842.

[11] Holland, P. W., Laskey, K. B. and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks* **5** 109-137.

[12] Hunter, D. R. and Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics* **15** 565-583.

[13] Häggström, O. and Jonasson, J. (1999). Phase Transition in the Random Triangle Model. *Journal of Applied Probability* **36** 1101-1115.

[14] Janková, J. and van de Geer, S. (2018). *Inference in high-dimensional graphical models* 325-346. CRC Press.

[15] Jonasson, J. (1999). The Random Triangle Model. *Journal of Applied Probability* **36** 852-867.

[16] Kontorovich, L. and Ramanan, K. (2008). Concentration inequalities for dependent random variables via the martingale method. *The Annals of Probability* **36** 2126-2158.

[17] Krivitsky, P. N., Handcock, M. S. and Morris, M. (2011). Adjusting for network size and composition effects in exponential-family random graph models.

Statistical Methodology **8** 319-339.

[18] Krivitsky, P. N., Hunter, D. R., Morris, M. and Klumb, C. (2023). ergm 4: New Features for Analyzing Exponential-Family Random Graph Models. *Journal of Statistical Software* **105** 1-44.

[19] Lauritzen, S. L. (1996). *Graphical Models.* Oxford University Press.

[20] Lauritzen, S., Rinaldo, A. and Sadeghi, K. (2018). Random networks, graphical models and exchangeability. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **80** 481-508.

[21] Leskovec, J., Lang, K. J., Dasgupta, A. and Mahoney, M. W. (2008). Statistical properties of community structure in large social and information networks. *Proceedings of the 17th international conference on World Wide Web* 695-704.

[22] Lindvall, T. (2002). *Lectures on the coupling method.* Courier Corporation.

[23] Lusher, D., Koskinen, J. and Robins, G. (2012). *Exponential Random Graph Models for Social Networks.* Cambridge University Press.

[24] Mele, A. (2022). A structural model of homophily and clustering in social networks. *Journal of Business & Economic Statistics* **40** 1377-1389.

[25] Newman, M. E. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E* **69** 026113.

[26] Ortega, J. M. and Rheinboldt, W. C. (2000). *Iterative solution of nonlinear equations in several variables.* Society for Industrial and Applied Mathematics.

[27] Portnoy, S. (1988). Asymptotic Behavior of Likelihood Methods for Exponential Families when the Number of Parameters Tends to Infinity. *The Annals of Statistics* **16** 356-366.

[28] Raič, M. (2019). A multivariate Berry-Esseen theorem with explicit constants. *Bernoulli* **25** 2824-2853.

[29] Ravikumar, P., Wainwright, M. J. and Lafferty, J. D. (2010). High-dimensional ising model selection using ℓ_1-regularized logistic regression. *The Annals of Statistics* **38** 1287-1319.

[30] Rohe, K., Chatterjee, S. and Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics* **39** 1878-1915.

[31] Schubert, D. and Brand, A. (2022). "Whom Should I Talk To?": Role Prescription and Hierarchy Building in Supervised Living Groups. *Social Inclusion* **10** 295-306.

[32] Schweinberger, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *Journal of the American Statistical Association* **106** 1361-1370.

[33] Schweinberger, M. and Handcock, M. S. (2015). Local dependence in random graph models: Characterization, properties and statistical inference. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **77** 647-676.

[34] Schweinberger, M. and Stewart, J. (2020). Concentration and consistency results for canonical and curved exponential-family models of random graphs. *The Annals of Statistics* **48**.

[35] 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. *Statistical Science* **35** 627-662.

[36] 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*.

[37] 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*.

[38] Stewart, J., Schweinberger, M., Bojanowski, M. and Morris, M. (2019). Multilevel network data facilitate statistical inference for curved ERGMs with geometrically weighted terms. *Social Networks* **59** 98-119.

[39] Sundberg, R. (2019). *Statistical Modelling by Exponential Families*. Cambridge University Press.

[40] Tolochko, P. and Boomgaarden, H. G. (2024). Same but different: A comparison of estimation approaches for exponential random graph models for multiple networks. *Social Networks* **76** 1-11.

[41] van den Berg, J. and Maes, C. (1994). Disagreement percolation in the study of Markov fields. *The Annals of Probability* **22** 749-763.

[42] Whetsell, T. A., Kroll, A. and Dehart-Davis, L. (2021). Formal Hierarchies and Informal Networks: How Organizational Structure Shapes Information Search in Local Government. *Journal of Public Administration Research and Theory* **31** 653-669.

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

By Jonathan R. Stewart

Florida State University

We establish a number of auxiliary results in the supplementary materials which are used in the proofs of Theorems 1 and 2, as well as various basic properties of exponential families which are used throughout the main manuscript and this supplement.

Appendix A: Auxiliary results for Theorem 1

Lemma 1. *Consider an exponential-family local dependence random graph. Then,*

$$\mathbb{P}\left(\sup_{\boldsymbol{\theta}_W \in \partial \mathcal{B}_2(\boldsymbol{\theta}_W^\star, \epsilon_W)} \|\nabla_{\boldsymbol{\theta}_W} \ell(\boldsymbol{\theta}, \boldsymbol{X}) - \mathbb{E} \nabla_{\boldsymbol{\theta}_W} \ell(\boldsymbol{\theta}, \boldsymbol{X})\|_\infty \geq \delta\right) \leq 2 \exp\left(-\frac{2\,\delta^2}{K\, A_{\max}^{10}\, L_W^2} + \log p\right)$$

$$\mathbb{P}\left(\sup_{\boldsymbol{\theta}_B \in \partial \mathcal{B}_2(\boldsymbol{\theta}_B^\star, \epsilon_B)} \|\nabla_{\boldsymbol{\theta}_B} \ell(\boldsymbol{\theta}, \boldsymbol{X}) - \mathbb{E} \nabla_{\boldsymbol{\theta}_B} \ell(\boldsymbol{\theta}, \boldsymbol{X})\|_\infty \geq \delta\right) \leq 2 \exp\left(-\frac{2\,\delta^2}{\binom{K}{2}\, A_{\max}^{10}\, L_B^2} + \log q\right)$$

for all $\delta > 0$ and all $(\epsilon_W, \epsilon_B) \in (0, \infty) \times (0, \infty)$.

Proof of Lemma 1. We utilize a general concentration result for dependent random variables due to Chazottes et al. [6] given in Theorem 1 of the cited work. We restate their result in notation and form more useful to our purposes. Let

$$\varphi \;:\; \{1, \ldots, \tbinom{N}{2}\} \;\mapsto\; \{\{i,j\} : i \in \mathcal{N},\; j \in \mathcal{N} \setminus \{i\}\}$$

be a bijective map between the set of dyad indices $\{\{i,j\} : i \in \mathcal{N},\; j \in \mathcal{N} \setminus \{i\}\}$ and the set $\{1, \ldots, \binom{N}{2}\}$. Define, for all $v \in \{1, \ldots, \binom{N}{2}\}$, $\boldsymbol{a} \in \{0,1\}^{\binom{N}{2} - v - 1}$, and $y \in \{0,1\}$,

$$\mathbb{P}_{v,y}^{\boldsymbol{x}}(\boldsymbol{a}) \;\;:=\;\; \mathbb{P}(\boldsymbol{X}_{\varphi(\{v+1, \ldots, \binom{N}{2}\})} = \boldsymbol{a} \,|\, \boldsymbol{X}_{\varphi(\{1, \ldots, v-1\})} = \boldsymbol{x}_{\varphi(\{1, \ldots, v-1\})},\, X_{\varphi(v)} = y),$$

and the upper-triangular matrix $\mathcal{D} \in \mathbb{R}^{\binom{N}{2} \times \binom{N}{2}}$ by defining elements $\mathcal{D}_{v,w}$ of $\mathcal{D}$ by

$$\mathcal{D}_{v,w} \;\; \coloneqq \;\; \begin{cases} 0 & v < w \\ 1 & v = w \\ \sup\limits_{\boldsymbol{x} \in \mathbb{X}} \mathbb{Q}_v^{\boldsymbol{x}}(X^{\star}_{\varphi(w)} \neq X^{\star\star}_{\varphi(w)}) & v > w \end{cases},$$

where $\mathbb{Q}_v^{\boldsymbol{x}}$ is the probability distribution of a coupling $(\boldsymbol{X}^{\star}, \boldsymbol{X}^{\star\star}) \in \{0,1\}^{\binom{N}{2}} \times \{0,1\}^{\binom{N}{2}}$ of the conditional probability distributions $\mathbb{P}_{v,0}^{\boldsymbol{x}}$ and $\mathbb{P}_{v,1}^{\boldsymbol{x}}$. Some background on coupling can be found in Lindvall [22]. Observe that coupling the conditional probability distributions $\mathbb{P}_{v,0}^{\boldsymbol{x}}$ and $\mathbb{P}_{v,1}^{\boldsymbol{x}}$ implies that the following events occur with probability 1:

- $X_v^{\star} = 0$,
- $X_v^{\star\star} = 1$, and
- $\boldsymbol{X}^{\star}_{\varphi(\{1,\ldots,v-1\})} = \boldsymbol{X}^{\star\star}_{\varphi(\{1,\ldots,v-1\})} = \boldsymbol{x}_{\varphi(\{1,\ldots,v-1\})}$.

At this stage of the proof, any coupling will do. When we bound $|||\mathcal{D}|||_2$ in Lemma 2, we will make use of a specific coupling that will provide the desired bound.

Consider a function $g : \mathbb{X} \mapsto \mathbb{R}$ and define $\boldsymbol{\Delta}_g \coloneqq (\Delta_{g,1}, \ldots, \Delta_{g,\binom{N}{2}})$, where

$$\Delta_{g,r} \;\; \coloneqq \;\; \sup_{(\boldsymbol{x},\boldsymbol{x}') \in \mathbb{X} \times \mathbb{X} \,:\, x_{i,j} = x'_{i,j},\, \{i,j\} \neq \varphi(r)} |g(\boldsymbol{x}) - g(\boldsymbol{x}')|, \qquad r = 1, \ldots, \tbinom{N}{2}.$$

With these definitions, we state Theorem 1 of Chazottes et al. [6]: For all $\delta > 0$,

$$\mathbb{P}(|g(\boldsymbol{X}) - \mathbb{E}\, g(\boldsymbol{X})| \;\geq\; \delta) \;\;\leq\;\; 2\, \exp\left(-\frac{2\,\delta^2}{|||\mathcal{D}|||_2^2\, \|\boldsymbol{\Delta}_g\|_2^2}\right),$$

where

$$|||\mathcal{D}|||_2 \;\; \coloneqq \;\; \sup_{\boldsymbol{u} \in \mathbb{R}^{\binom{N}{2}} : \|\boldsymbol{u}\|_2 = 1} \|\mathcal{D}\, \boldsymbol{u}\|_2$$

is the spectral norm of $\mathcal{D}$. By Lemma 2, $|||\mathcal{D}|||_2 \leq A_{\max}^2$, which implies

$$\mathbb{P}(|g(\boldsymbol{X}) - \mathbb{E}\, g(\boldsymbol{X})| \;\geq\; \delta) \;\;\leq\;\; 2\, \exp\left(-\frac{2\,\delta^2}{A_{\max}^4\, \|\boldsymbol{\Delta}_g\|_2^2}\right), \tag{A.1}$$

recalling $A_{\max} \coloneqq \max\{|\mathcal{A}_1|, \ldots, |\mathcal{A}_K|\}$. To prove the desired result, we manipulate the events of interest into forms amenable to (A.1). We are interested in events of the form

$$\|\boldsymbol{g}(\boldsymbol{X}) - \mathbb{E}\, \boldsymbol{g}(\boldsymbol{X})\|_\infty \;\geq\; \delta,$$

for multivariable functions $\boldsymbol{g} : \mathbb{X} \mapsto \mathbb{R}^m$. We use a union bound (over the m components of $\boldsymbol{g}$) and apply the inequality in (A.1):

$$\begin{aligned}\mathbb{P}(\|\boldsymbol{g}(\boldsymbol{X}) - \mathbb{E}\,\boldsymbol{g}(\boldsymbol{X})\|_\infty \geq \delta) &\leq \sum_{l=1}^{m} \mathbb{P}(|g_l(\boldsymbol{X}) - \mathbb{E}\,g_l(\boldsymbol{X})| \geq \delta) \\ &\leq \sum_{l=1}^{m} 2\,\exp\left(-\frac{2\,\delta^2}{A_{\max}^4\,\|\boldsymbol{\Delta}_{g_l}\|_2^2}\right) \\ &\leq 2\,\exp\left(-\frac{2\,\delta^2}{A_{\max}^4\,\max_{1\leq l\leq m}\,\|\boldsymbol{\Delta}_{g_l}\|_2^2} + \log m\right).\end{aligned}$$

The last remaining task is to bound each $\|\boldsymbol{\Delta}_{g_l}\|_2^2$ for each of the cases of interest.

Concentrating gradients of the log-likelihood. First, we recall the following formulas from Section 2.1:

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

Fix $\boldsymbol{\theta} \in \mathbb{R}^{p+q}$ and take

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

For all $t \in \{1, \ldots, p\}$, we have $\Delta_{g_t,r} = 0$ if $\varphi(r) \not\subset \mathcal{A}_k$ for all $k \in \{1, \ldots, K\}$, in which case $X_{\varphi(r)}$ is a between-block edge existing in one of the subgraphs $\boldsymbol{X}_{k,l}$ $(1 \leq k < l \leq K)$. Hence, for all $t \in \{1, \ldots, p\}$,

$$\sum_{r=1}^{\binom{N}{2}} \mathbb{1}\left(\Delta_{g_t,r} \neq 0\right) \leq \sum_{k=1}^{K} \binom{|\mathcal{A}_k|}{2} \leq K\binom{A_{\max}}{2}.$$

Assumption (A.2) assumes that there exists $L_W > 0$, independent of N, p, and q, such that, for all $k \in \{1, \ldots, K\}$,

$$\sup_{(\boldsymbol{v},\boldsymbol{v}')\in\mathbb{X}_{k,k}\times\mathbb{X}_{k,k}\,:\,d_H(\boldsymbol{v},\boldsymbol{v}')=1} \|s_{k,k}(\boldsymbol{v}) - s_{k,k}(\boldsymbol{v}')\|_\infty \leq L_W\binom{|\mathcal{A}_k|}{2} \leq L_W\binom{A_{\max}}{2},$$

which implies $\Delta_{g_t,r} \leq L_W\binom{A_{\max}}{2}$ (for all $t = 1, \ldots, p$; $r = 1, \ldots, \binom{N}{2}$). As a result,

$$\max_{1\leq t\leq p}\,\|\boldsymbol{\Delta}_{g_t}\|_2^2 \leq K\binom{A_{\max}}{2}\left[L_W\binom{A_{\max}}{2}\right]^2 \leq K\,A_{\max}^6\,L_W^2.$$

Thus, we obtain, for all $\delta > 0$ and $\boldsymbol{\theta} \in \mathbb{R}^{p+q}$,

$$\mathbb{P}(\|\nabla_{\boldsymbol{\theta}_W} \ell(\boldsymbol{\theta}, \boldsymbol{X}) - \mathbb{E}\, \nabla_{\boldsymbol{\theta}_W} \ell(\boldsymbol{\theta}, \boldsymbol{X})\|_\infty \;\geq\; \delta) \quad \leq \quad 2 \exp\left(-\frac{2\,\delta^2}{K L_W^2 A_{\max}^{10}} + \log\, p\right).$$

By Lemma 9,

$$\sup_{\boldsymbol{\theta}_W \in \mathbb{R}^p} \|\nabla_{\boldsymbol{\theta}_W} \ell(\boldsymbol{\theta}, \boldsymbol{X}) - \mathbb{E}\, \nabla_{\boldsymbol{\theta}_W} \ell(\boldsymbol{\theta}, \boldsymbol{X})\|_\infty \quad = \quad \left\| \sum_{k=1}^{K} \left[s_{k,k}(\boldsymbol{x}_{k,k}) - \mathbb{E}_{k,k,\boldsymbol{\theta}_W^\star}\, s_{k,k}(\boldsymbol{X}_{k,k})\right] \right\|_\infty,$$

where notably the right-hand side is constant in $\boldsymbol{\theta}_W \in \mathbb{R}^p$. Thus, for all $\epsilon_W \in (0, \infty)$,

$$\mathbb{P}\left(\sup_{\boldsymbol{\theta}_W \in \partial \mathcal{B}_2(\boldsymbol{\theta}_W^\star, \epsilon_W)} \|\nabla_{\boldsymbol{\theta}_W} \ell(\boldsymbol{\theta}, \boldsymbol{X}) - \mathbb{E} \nabla_{\boldsymbol{\theta}_W} \ell(\boldsymbol{\theta}, \boldsymbol{X})\|_\infty \geq \delta\right) \;\leq\; 2 \exp\left(-\frac{2\,\delta^2}{K\, A_{\max}^{10}\, L_W^2} + \log\, p\right).$$

By a similar argument, we have, for all $\delta > 0$ and $\boldsymbol{\theta} \in \mathbb{R}^{p+q}$,

$$\mathbb{P}(\|\nabla_{\boldsymbol{\theta}_B} \ell(\boldsymbol{\theta}, \boldsymbol{X}) - \mathbb{E}\, \nabla_{\boldsymbol{\theta}_B} \ell(\boldsymbol{\theta}, \boldsymbol{X})\|_\infty \;\geq\; \delta) \quad \leq \quad 2 \exp\left(-\frac{2\,\delta^2}{\binom{K}{2} L_B^2 A_{\max}^{10}} + \log\, q\right),$$

where in the application of Assumption (A.2) above, we replace $L_w > 0$ by $L_B > 0$, recalling, for all $\{k, l\} \subseteq \{1, \ldots, K\}$, that

$$\sup_{(\boldsymbol{v}, \boldsymbol{v}') \in \mathbb{X}_{k,l} \times \mathbb{X}_{k,l} \,:\, d_H(\boldsymbol{v}, \boldsymbol{v}') = 1} \|s_{k,l}(\boldsymbol{v}) - s_{k,l}(\boldsymbol{v}')\|_\infty \quad \leq \quad L_B\, |\mathcal{A}_k|\, |\mathcal{A}_l| \quad \leq \quad L_B\, A_{\max}^2.$$

A key difference in proving this case lies in the fact that in considering

$$g_t(\boldsymbol{x}) \quad = \quad \sum_{1 \leq k < l \leq K} \left[s_{k,l,t}(\boldsymbol{x}_{k,l}) - \mathbb{E}_{k,l,\boldsymbol{\theta}_B}\, s_{k,l,t}(\boldsymbol{X}_{k,l})\right], \qquad t = p+1, \ldots, p+q,$$

the quantity $\Delta_{g_t, r} = 0$ for all $t \in \{p{+}1, \ldots, p{+}q\}$ if $\varphi(r) = \{c, d\}$ satisfies $(c, d) \notin \mathcal{A}_k \times \mathcal{A}_l$ for all possible pairs $\{k, l\} \subseteq \{1, \ldots, K\}$, in which case $X_{\varphi(r)}$ is a within-block edge existing in the one of the subgraphs $\boldsymbol{X}_{k,k}$ ($k = 1, \ldots, K$). Hence,

$$\sum_{r=1}^{\binom{N}{2}} \mathbb{1}\left(\Delta_{g_t, r} \neq 0\right) \quad \leq \quad \sum_{1 \leq k < l \leq K} |\mathcal{A}_k|\, |\mathcal{A}_l| \quad \leq \quad \binom{K}{2}\, A_{\max}^2,$$

for all $t \in \{p+1, \ldots, p+q\}$. This results in the bound

$$\max_{p+1 \leq t \leq p+q} \|\boldsymbol{\Delta}_{g_t}\|_2^2 \quad \leq \quad \binom{K}{2} A_{\max}^2 \left[L_B\, A_{\max}^2\right]^2 \quad \leq \quad \binom{K}{2} A_{\max}^6\, L_B^2.$$

By Lemma 9,

$$\sup_{\boldsymbol{\theta}_B \in \mathbb{R}^p} \|\nabla_{\boldsymbol{\theta}_B} \ell(\boldsymbol{\theta}, \boldsymbol{X}) - \mathbb{E}\, \nabla_{\boldsymbol{\theta}_B} \ell(\boldsymbol{\theta}, \boldsymbol{X})\|_\infty \quad = \quad \left\| \sum_{1 \leq k < l \leq K} \left[s_{k,l}(\boldsymbol{x}_{k,l}) - \mathbb{E}_{k,l,\boldsymbol{\theta}_B^\star}\, s_{k,l}(\boldsymbol{X}_{k,l})\right] \right\|_\infty,$$

where notably the right-hand side is constant in $\boldsymbol{\theta}_B \in \mathbb{R}^p$. Thus,

$$\mathbb{P}\left(\sup_{\boldsymbol{\theta}_B \in \partial\mathcal{B}_2(\boldsymbol{\theta}_B^\star, \epsilon_B)} \|\nabla_{\boldsymbol{\theta}_B} \ell(\boldsymbol{\theta}, \boldsymbol{X}) - \mathbb{E}\nabla_{\boldsymbol{\theta}_B} \ell(\boldsymbol{\theta}, \boldsymbol{X})\|_\infty \geq \delta\right) \leq 2 \exp\left(-\frac{2\,\delta^2}{\binom{K}{2} A_{\max}^{10} L_B^2} + \log q\right).$$

for all $\epsilon_B \in (0, \infty)$. □

Lemma 2. *The coupling matrix $\mathcal{D}$ defined in Lemma 1 satisfies $\|\mathcal{D}\|_2 \leq A_{\max}^2$.*

Proof of Lemma 2. We construct a coupling $(\boldsymbol{X}^\star, \boldsymbol{X}^{\star\star}) \in \{0,1\}^{\binom{N}{2}} \times \{0,1\}^{\binom{N}{2}}$ of the conditional probability distributions $\mathbb{P}_{v,0}^{\boldsymbol{x}}$ and $\mathbb{P}_{v,1}^{\boldsymbol{x}}$ defined in Lemma 1 for each $v \in \{1, \dots, \binom{N}{2}\}$. The proof is divided into two parts:

I. Constructing a coupling $(\boldsymbol{X}^\star, \boldsymbol{X}^{\star\star})$ of $\mathbb{P}_{v,0}^{\boldsymbol{x}}$ and $\mathbb{P}_{v,1}^{\boldsymbol{x}}$.

II. Bounding $\|\mathcal{D}\|_2$.

I. Constructing a coupling $(\boldsymbol{X}^\star, \boldsymbol{X}^{\star\star})$ of $\mathbb{P}_{v,0}^{\boldsymbol{x}}$ and $\mathbb{P}_{v,1}^{\boldsymbol{x}}$. Define $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ to be conditional independence graph of the random graph $\boldsymbol{X}$, i.e., the graph with set of vertices $\mathcal{V} = \{1, \dots, \binom{N}{2}\}$, where each vertex $v \in \mathcal{V}$ corresponds to an edge variable $X_{\varphi(v)}$, and the set of edges $\mathcal{E}$, where there exists an edge $\{v, w\} \in \mathcal{E}$ if and only if $X_{\varphi(v)}$ and $X_{\varphi(w)}$ do not satisfy any conditional independence relationships with respect to the edge variables implied by $\mathcal{V} \setminus \{v, w\}$. Further background on undirected graphical models can be found in Lauritzen [19], which have been applied to random graphs in a number of works [e.g., 10, 20, 37].

Let $v \in \{1, \dots, \binom{N}{2}\}$. We construct a coupling $(\boldsymbol{X}^\star, \boldsymbol{X}^{\star\star}) \in \{0,1\}^{\binom{N}{2}} \times \{0,1\}^{\binom{N}{2}}$ due to van den Berg and Maes [41][1] of the conditional probability mass functions $\mathbb{P}_{v,0}^{\boldsymbol{x}}$ and $\mathbb{P}_{v,1}^{\boldsymbol{x}}$ for a given $\boldsymbol{x} \in \mathbb{X}$, which were defined, for all $\boldsymbol{a} \in \{0,1\}^{\binom{N}{2} - v - 1}$, to be

$$\mathbb{P}_{v,y}^{\boldsymbol{x}}(\boldsymbol{a}) \quad := \quad \mathbb{P}(\boldsymbol{X}_{\varphi(\{v+1, \dots, \binom{N}{2}\})} = \boldsymbol{a} \,|\, \boldsymbol{X}_{\varphi(\{1, \dots, v-1\})} = \boldsymbol{x}_{\varphi(\{1, \dots, v-1\})},\, X_{\varphi(v)} = y),$$

through the following algorithm:

1. Initialize $\mathcal{S} = \{1, \dots, v\}$ as a subset of the vertices in $\mathcal{G}$ and set $X_v^\star = 0$, $X_v^{\star\star} = 1$, and $X_w^\star = X_w^{\star\star} = x_w$ for all $w \in \{1, \dots, v-1\}$, with probability 1.
2. Let $w \in \{1, \dots, \binom{N}{2}\} \setminus \mathcal{S}$ be the smallest value for which there exists some $z \in \mathcal{S}$ satisfying $\{w, z\} \in \mathcal{E}$ and $X_{\varphi(z)}^\star \neq X_{\varphi(z)}^{\star\star}$; if no such $w \in \{1, \dots, \binom{N}{2}\} \setminus \mathcal{S}$ exists, let w be the smallest value in $\{1, \dots, \binom{N}{2}\} \setminus \mathcal{S}$.
3. Let $(X_w^\star, X_w^{\star\star})$ be distributed according to an optimal coupling of the following conditional probability distributions of $X_{\varphi(w)}$:

$$\mathbb{P}(X_{\varphi(w)} = \cdot \mid \boldsymbol{X}_{\varphi(\mathcal{S})} = \boldsymbol{X}_{\varphi(\mathcal{S})}^\star) \qquad \text{and} \qquad \mathbb{P}(X_{\varphi(w)} = \cdot \mid \boldsymbol{X}_{\varphi(\mathcal{S})} = \boldsymbol{X}_{\varphi(\mathcal{S})}^{\star\star}).$$

[1] This approach has also been applied elsewhere to random graphs [37].

4. If $\{1,\ldots,\binom{N}{2}\}\setminus(\mathcal{S}\cup\{w\})$ is not empty, replace $\mathcal{S}$ by $\mathcal{S}\cup\{w\}$ and return to step 2.

Denote the joint distribution of $(\boldsymbol{X}^\star, \boldsymbol{X}^{\star\star})$ by $\mathbb{Q}_{v,\boldsymbol{x}}$. It is straightforward to verify that in the case of a random graph, the above algorithm constructs a valid coupling of $\mathbb{P}_{v,0}^{\boldsymbol{x}}$ and $\mathbb{P}_{v,1}^{\boldsymbol{x}}$ [37, Lemma 14 in the supplementary materials].

The above coupling has an important property for our purposes here. If there exists no path between $w \in \mathcal{V}$ and $v \in \mathcal{V}\setminus\{w\}$ in $\mathcal{G}$, then $X_{\varphi(w)}$ is independent of $X_{\varphi(v)}$, by the graph separation property of undirected graphical models [19]. As a result, an optimal coupling under this condition ensures that the total variation distance d_{TV} satisfies:

$$\begin{aligned} & d_{\text{TV}}(\mathbb{P}(X_{\varphi(w)} = \cdot \mid \boldsymbol{X}_{\varphi(\mathcal{S})} = \boldsymbol{X}^\star_{\varphi(\mathcal{S})}), \mathbb{P}(X_{\varphi(w)} = \cdot \mid \boldsymbol{X}_{\varphi(\mathcal{S})} = \boldsymbol{X}^{\star\star}_{\varphi(\mathcal{S})})) \\ = \; & \mathbb{Q}_{v,\boldsymbol{x}}(X^\star_w \neq X^{\star\star}_w) \\ = \; & 0, \end{aligned}$$

for all $\boldsymbol{x} \in \mathbb{X}$, owing to the fact that $X_{\varphi(w)}$ is independent of $X_{\varphi(v)}$ and that, with probability 1, the coupling constructed above ensures that $X^\star_{\varphi(t)} = X^{\star\star}_{\varphi(t)} = x_t$ for all $t \in \{1,\ldots,v-1\}$, $X^\star_v = 0$, and $X^{\star\star}_v = 1$, which implies

$$\mathbb{P}(X_{\varphi(w)} = a \mid \boldsymbol{X}_{\varphi(\mathcal{S})} = \boldsymbol{X}^\star_{\varphi(\mathcal{S})}) \;=\; \mathbb{P}(X_{\varphi(w)} = a \mid \boldsymbol{X}_{\varphi(\mathcal{S})} = \boldsymbol{X}^{\star\star}_{\varphi(\mathcal{S})}), \qquad a \in \{0,1\},$$

resulting in $d_{\text{TV}}(\mathbb{P}(X_{\varphi(w)} = \cdot \mid \boldsymbol{X}_{\varphi(\mathcal{S})} = \boldsymbol{X}^\star_{\varphi(\mathcal{S})}), \mathbb{P}(X_{\varphi(w)} = \cdot \mid \boldsymbol{X}_{\varphi(\mathcal{S})} = \boldsymbol{X}^{\star\star}_{\varphi(\mathcal{S})})) = 0$. Hence, for all pairs $w \in \mathcal{V}\setminus\{v\}$ corresponding to an edge variable $X_{\varphi(w)}$ which is not in the same block-based subgraph $\boldsymbol{X}_{k,l}$ ($1 \leq k \leq l \leq K$) as edge variable $X_{\varphi(v)}$,

$$\mathcal{D}_{v,w} \;=\; \sup_{\boldsymbol{x}\in\mathbb{X}} \mathbb{Q}_{v,\boldsymbol{x}}(X^\star_w \neq X^{\star\star}_w) \;=\; 0.$$

As a result, the number of non-zero elements in a given row of $\mathcal{D}$ is bounded above by the number of edge variables in the block-based subgraph of the edge variable corresponding to that row. We will make use of this fact in the next part of the proof.

II. Bounding $\|\mathcal{D}\|_2$. In order to bound $\|\mathcal{D}\|_2$, we first symmetrize $\mathcal{D}$ by defining:

$$\mathcal{T} \;:=\; \mathcal{D} + \mathcal{D}^\top - \boldsymbol{I}_{\binom{N}{2}},$$

where $\mathcal{D}^\top$ is the matrix transpose of $\mathcal{D}$ and $\boldsymbol{I}_{\binom{N}{2}}$ is the $\binom{N}{2}$-dimensional identity matrix. By Hölder's inequality,

$$\|\mathcal{D}\|_2 \;\leq\; \sqrt{\|\mathcal{D}\|_1 \, \|\mathcal{D}\|_\infty}.$$

Recalling that $\mathcal{D}$ is an upper-triangular matrix, we have $\mathcal{T}_{v,w} \geq \max\{\mathcal{D}_{v,w}, \mathcal{D}_{w,v}\}$ for all $\{v,w\} \subset \mathcal{V}$, which implies that

$$\|\mathcal{D}\|_1 \;\leq\; \|\mathcal{T}\|_1 \qquad \text{and} \qquad \|\mathcal{D}\|_\infty \;\leq\; \|\mathcal{T}\|_\infty.$$

The symmetry of $\mathcal{T}$ implies that $\|\mathcal{T}\|_1 = \|\mathcal{T}^\top\|_1 = \|\mathcal{T}\|_\infty$, which in turn implies

$$\|\mathcal{D}\|_2 \;\leq\; \sqrt{\|\mathcal{T}\|_1 \|\mathcal{T}\|_\infty} \;\leq\; \|\mathcal{T}\|_\infty.$$

Hence, it suffices to bound

$$|||\mathcal{T}|||_\infty \;=\; 1 + \max_{v \in \{1,\ldots,\binom{N}{2}\}} \sum_{w \in \{1,\ldots,\binom{N}{2}\} \setminus \{v\}} \mathcal{T}_{v,w}.$$

As

$$\mathcal{T}_{v,w} \;\coloneqq\; \max\{\mathcal{D}_{v,w},\, \mathcal{D}_{w,v}\} \;=\; \max\left\{\sup_{\boldsymbol{x} \in \mathbb{X}} \mathbb{Q}_{v,\boldsymbol{x}}(X_w^\star \neq X_w^{\star\star}),\; \sup_{\boldsymbol{x} \in \mathbb{X}} \mathbb{Q}_{w,\boldsymbol{x}}(X_v^\star \neq X_v^{\star\star})\right\},$$

we have $\mathcal{T}_{v,w} \coloneqq \max\{\mathcal{D}_{v,w},\, \mathcal{D}_{w,v}\} = 0$ for all pairs $\{v, w\} \subset \mathcal{V}$ which are not connected by a path in $\mathcal{G}$, per the discussion at the end of previous subsection of this proof. Under a local dependence random graph model, this occurs when v and w correspond to pairs of edge variables $X_{\varphi(v)}$ and $X_{\varphi(w)}$ belonging to distinct block-based subgraphs, as the collection of subgraphs $\boldsymbol{X}_{k,l}$ $(1 \leq k \leq l \leq K)$ are mutually independent. Thus,

$$1 + \max_{v \in \{1,\ldots,\binom{N}{2}\}} \sum_{w \in \{1,\ldots,\binom{N}{2}\} \setminus \{v\}} \mathcal{T}_{v,w} \;\leq\; A_{\max}^2,$$

noting that the maximum number of edge variables in any subgraph $\boldsymbol{X}_{k,l}$ $(1 \leq k \leq l \leq K)$ is bounded above by $A_{\max}^2$. We have thus shown that $|||\mathcal{D}|||_2 \leq A_{\max}^2$. □

Appendix B: Auxiliary results for Theorem 2

We first recall a theorem due to Raič [28], restated in Lemma 3.

Lemma 3 (Theorem 1.1, Raič [28])**.** *Consider a sequence of independent random vectors $\boldsymbol{W}_1, \boldsymbol{W}_2, \ldots \in \mathbb{R}^p$ with $\mathbb{E}\, \boldsymbol{W}_i = 0$ for all $i \in \{1, 2, \ldots\}$. Define*

$$\boldsymbol{S}_n \;\coloneqq\; \sum_{i=1}^n \boldsymbol{W}_i, \qquad n \in \{1, 2, \ldots\},$$

and assume that $\mathbb{V}\, \boldsymbol{S}_n = \boldsymbol{I}_p$. Then, for all measurable convex sets $\mathcal{A} \subset \mathbb{R}^p$,

$$|\mathbb{P}(\boldsymbol{S}_n \in \mathcal{A}) - \Phi(\boldsymbol{Z} \in \mathcal{A})| \;\leq\; (42\, p^{1/4} + 16) \sum_{i=1}^n \mathbb{E}\, \|\boldsymbol{W}_i\|_2^3,$$

where $\boldsymbol{Z}$ is a multivariate normal random vector with mean vector $\boldsymbol{0}_p$ and covariance matrix $\boldsymbol{I}_p$ and Φ is the corresponding probability distribution.

Proof of Lemma 3. The lemma is proved as Theorem 1.1 of Raič [28]. □

B.1. Auxiliary results for Part II in the proof of Theorem 2

Lemma 4. *Under the definitions and assumptions of Theorem 2, the remainder terms $\|\boldsymbol{R}_W\|_2^2$ and $\|\boldsymbol{R}_B\|_2^2$ in the proof of Theorem 2 satisfy*

$$\|\boldsymbol{R}_W\|_2^2 \quad \leq \quad \sum_{i=1}^{p} R_i^2 \quad \leq \quad 4\,p\,C_W^2\,A_{\max}^{20}\,(L_W+2)^4\,K^2\,\|\widehat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_W^\star\|_1^4$$

$$\|\boldsymbol{R}_B\|_2^2 \quad \leq \quad \sum_{i=p+1}^{p+q} R_i^2 \quad \leq \quad 4\,q\,C_B^2\,A_{\max}^{20}\,(L_B+2)^4\binom{K}{2}^2\|\widehat{\boldsymbol{\theta}}_B - \boldsymbol{\theta}_B^\star\|_1^4.$$

Proof of Lemma 4. We bound the remainder terms that arose out of the multivariate Taylor approximation in the proof of Theorem 2 using derivatives. Recall that the remainder terms R_i $(i = 1, \ldots, p+q)$ are given by

$$\begin{aligned} R_i \quad = \quad & \sum_{j=1}^{p+q} \frac{1}{2}\left[\frac{\partial^2}{\partial\,\theta_j^2}\left[\nabla_{\boldsymbol{\theta}}\, m(\dot{\boldsymbol{\theta}}^{(i)}, \boldsymbol{X})\right]_i\right](\theta_j - \theta_j^\star)^2 \\ & + \sum_{1\leq j<r\leq p+q} \frac{1}{2}\left[\frac{\partial^2}{\partial\,\theta_j\,\partial\,\theta_r}\left[\nabla_{\boldsymbol{\theta}}\, m(\dot{\boldsymbol{\theta}}^{(i)}, \boldsymbol{X})\right]_i\right](\theta_j - \theta_j^\star)\,(\theta_r - \theta_r^\star), \end{aligned} \tag{B.1}$$

where $\dot{\boldsymbol{\theta}}^{(i)} = t_i\,\boldsymbol{\theta} + (1-t_i)\,\boldsymbol{\theta}^\star$ (for some $t_i \in (0,1)$, $i = 1, \ldots, p+q$). If

$$\sup_{\boldsymbol{\theta}\in\mathbb{R}^{p+q}\,:\,\|\boldsymbol{\theta}-\boldsymbol{\theta}^\star\|_1\leq\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^\star\|_1} \left|\frac{\partial^2}{\partial\,\theta_j\,\partial\,\theta_r}\left[\nabla_{\boldsymbol{\theta}}\, m(\boldsymbol{\theta}, \boldsymbol{X})\right]_i\right| \quad \leq \quad M_i, \qquad 1\leq j\leq r\leq p,$$

for all $i \in \{1, \ldots, p\}$ and

$$\sup_{\boldsymbol{\theta}\in\mathbb{R}^{p+q}\,:\,\|\boldsymbol{\theta}-\boldsymbol{\theta}^\star\|_1\leq\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^\star\|_1} \left|\frac{\partial^2}{\partial\,\theta_j\,\partial\,\theta_r}\left[\nabla_{\boldsymbol{\theta}}\, m(\boldsymbol{\theta}, \boldsymbol{X})\right]_i\right| \quad \leq \quad M_i, \qquad 1+p\leq j\leq r\leq p+q,$$

for all $i \in \{1+p, \ldots, p+q\}$, then the Lagrange remainder is bounded above by

$$|R_i| \quad \leq \quad \begin{cases} \dfrac{M_i}{2}\,\|\widehat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_W^\star\|_1^2, & \text{if } i \in \{1, \ldots, p\} \\[1ex] \dfrac{M_i}{2}\,\|\widehat{\boldsymbol{\theta}}_B - \boldsymbol{\theta}_B^\star\|_1^2 & \text{if } i \in \{p+1, \ldots, p+q\} \end{cases}$$

on the set

$$\{\boldsymbol{\theta}_W \in \mathbb{R}^p \,:\, \|\boldsymbol{\theta}_W - \boldsymbol{\theta}_W^\star\|_1 \leq \|\widehat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_W^\star\|_1\} \times \{\boldsymbol{\theta}_B \in \mathbb{R}^p \,:\, \|\boldsymbol{\theta}_B - \boldsymbol{\theta}_B^\star\|_1 \leq \|\widehat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_B^\star\|_1\}.$$

For the remainder of the proof, assume that $\boldsymbol{\theta}$ belongs to the above set. Under the assumption that there exists $C_W > 0$ and $C_B > 0$, independent of N, p, and q, such that

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

and

$$\sup_{\boldsymbol{x}_{k,l} \in \mathbb{X}_{k,l}} \|s_{k,l}(\boldsymbol{x}_{k,l})\|_\infty \quad \leq \quad C_W\, |\mathcal{A}_k|\, |\mathcal{A}_l|, \qquad \text{for all } 1 \leq k < l \leq K,$$

Lemmas 5 and 6 establish that

$$\left|\frac{\partial^2}{\partial\,\theta_j\;\partial\,\theta_h}\; [\nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}, \boldsymbol{X})]_i\right| \quad \leq \quad \begin{cases} 2\, C_W\, A_{\max}^{10}\, (L_W + 2)^2\, K, & (i,j,h) \in \{1, \dots, p\}^3 \\ 2\, C_B\, A_{\max}^{10}\, (L_B + 2)^2 \dbinom{K}{2}, & (i,j,h) \in \{p+1, \dots, p+q\}^3 \\ 0 & \text{otherwise} \end{cases}.$$

As a result, when $m(\boldsymbol{\theta}, \boldsymbol{X}) = \ell(\boldsymbol{\theta}, \boldsymbol{X})$ in the proof of Theorem 2,

$$|R_i| \quad \leq \quad \begin{cases} 2\, C_W\, A_{\max}^{10}\, (L_W + 2)^2\, K\, \|\widehat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_W^\star\|_1^2, & 1 \leq i \leq p \\ 2\, C_B\, A_{\max}^{10}\, (L_B + 2)^2\, \binom{K}{2}\, \|\widehat{\boldsymbol{\theta}}_B - \boldsymbol{\theta}_B^\star\|_1^2, & p+1 \leq i \leq p+q \end{cases},$$

which implies the bounds

$$\begin{aligned} \|\boldsymbol{R}_W\|_2^2 \quad &\leq \quad \sum_{i=1}^{p} R_i^2 \qquad \leq \quad 4\, p\, C_W^2\, A_{\max}^{20}\, (L_W + 2)^4\, K^2\, \|\widehat{\boldsymbol{\theta}}_W - \boldsymbol{\theta}_W^\star\|_1^4 \\ \|\boldsymbol{R}_B\|_2^2 \quad &\leq \quad \sum_{i=p+1}^{p+q} R_i^2 \quad \leq \quad 4\, q\, C_B^2\, A_{\max}^{20}\, (L_B + 2)^4 \binom{K}{2}^2 \|\widehat{\boldsymbol{\theta}}_B - \boldsymbol{\theta}_B^\star\|_1^4. \end{aligned}$$

□

Lemma 5. *Consider an exponential-family local dependence random graph model. Assume that there exists a constant $C_W > 0$ such that, for all $k \in \{1, \dots, K\}$,*

$$\sup_{\boldsymbol{x}_{k,k} \in \mathbb{X}_{k,k}} \|s_{k,k}(\boldsymbol{x}_{k,k})\|_\infty \quad \leq \quad C_W \binom{|\mathcal{A}_k|}{2}.$$

Then, for all $(i,j,h) \in \{1, \dots, p\}^3$,

$$\left|\frac{\partial^2\, [\nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}, \boldsymbol{x})]_i}{\partial\theta_h\; \partial\theta_j}\right| \quad \leq \quad 2\, C_W\, A_{\max}^{10}\, (L_W + 2)^2\, K.$$

Proof of Lemma 5. By Lemma 8, the second derivatives of the log-likelihood taken with respect to the natural parameters are equal to the variances and covariances of the sufficient statistics of the exponential family, implying, for all $(i,j) \in \{1, \dots, p\}^2$, that

$$\frac{\partial^2\, \ell(\boldsymbol{\theta}, \boldsymbol{x})}{\partial\theta_j\; \partial\theta_i} \quad = \quad \frac{\partial\, [\nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}, \boldsymbol{x})]_i}{\partial\theta_j} \quad = \quad \mathbb{C}_{\boldsymbol{\theta}} \left(\sum_{k=1}^{K} s_{k,k,i}(\boldsymbol{X}_{k,k}),\; \sum_{k=1}^{K} s_{k,k,j}(\boldsymbol{X}_{k,k})\right),$$

where $\mathbb{C}_{\boldsymbol{\theta}}$ denotes the covariance operator corresponding to the probability distribution $\mathbb{P}_{\boldsymbol{\theta}}$. By the independence of the block-based subgraphs $\boldsymbol{X}_{k,k}$ $(k = 1, \ldots, K)$,

$$\mathbb{C}_{\boldsymbol{\theta}}\left(\sum_{k=1}^{K} s_{k,k,i}(\boldsymbol{X}_{k,k}),\ \sum_{k=1}^{K} s_{k,k,j}(\boldsymbol{X}_{k,k})\right) \;=\; \sum_{k=1}^{K} \mathbb{C}_{\boldsymbol{\theta}}(s_{k,k,i}(\boldsymbol{X}_{k,k}),\ s_{k,k,j}(\boldsymbol{X}_{k,k})).$$

Taking $h \in \{1, \ldots, p\}$ and using the triangle inequality, we obtain the bound

$$\left|\frac{\partial^2\, [\nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}, \boldsymbol{x})]_i}{\partial\theta_h\, \partial\theta_j}\right| \;\leq\; \sum_{k=1}^{K} \left|\frac{\partial}{\partial\, \theta_h}\, \mathbb{C}_{\boldsymbol{\theta}}(s_{k,k,i}(\boldsymbol{X}_{k,k}),\ s_{k,k,j}(\boldsymbol{X}_{k,k}))\right|. \tag{B.2}$$

To proceed from here, we apply Lemma 10. In order to do so, we verify that the assumptions of Lemma 10 are met. By Assumption (A.2), $L_W > 0$ satisfies

$$\sup_{(\boldsymbol{v},\boldsymbol{v}') \in \mathbb{X}_{k,k} \times \mathbb{X}_{k,k}\, :\, d_H(\boldsymbol{v},\boldsymbol{v}')=1} \|s_{k,k}(\boldsymbol{v}) - s_{k,k}(\boldsymbol{v}')\|_\infty \;\leq\; L_W \binom{|\mathcal{A}_k|}{2},$$

which implies, for all $k \in \{1, \ldots, K\}$, that

$$\begin{aligned}\sup_{(\boldsymbol{v},\boldsymbol{v}') \in \mathbb{X}_{k,k} \times \mathbb{X}_{k,k}} \|s_{k,k}(\boldsymbol{v}) - s_{k,k}(\boldsymbol{v}')\|_\infty \;&\leq\; \sup_{(\boldsymbol{v},\boldsymbol{v}') \in \mathbb{X}_{k,k} \times \mathbb{X}_{k,k}} L_W \binom{|\mathcal{A}_k|}{2}\, d_H(\boldsymbol{v}, \boldsymbol{v}') \\ &\leq\; L_W \binom{|\mathcal{A}_k}{2}^2 \;\leq\; L_W\, A_{\max}^4.\end{aligned}$$

As a result,

$$\sup_{1 \leq k \leq K}\, \sup_{\boldsymbol{v} \in \mathbb{X}_{k,k}} \|s_{k,k}(\boldsymbol{v}) - \mathbb{E}_{\boldsymbol{\theta}}\, s_{k,k}(\boldsymbol{X}_{k,k})\|_\infty \;\leq\; L_W\, A_{\max}^4.$$

By assumption, the constant $C_W > 0$ is such that, for all $k \in \{1, \ldots, K\}$,

$$\sup_{\boldsymbol{x}_{k,k} \in \mathbb{X}_{k,k}} \|s_{k,k}(\boldsymbol{x}_{k,k})\|_\infty \;\leq\; C_W \binom{|\mathcal{A}_k|}{2} \;\leq\; \frac{C_W\, A_{\max}^2}{2} \;\leq\; C_W\, A_{\max}^2.$$

Taking $U_1 = C_W\, A_{\max}^2 > 0$ and $U_2 = L_W\, A_{\max}^4 > 0$ verifies the assumptions of Lemma 10, and applying Lemma 10 provides the bound

$$\begin{aligned}\left|\frac{\partial}{\partial\, \theta_h}\, \mathbb{C}_{\boldsymbol{\theta}}(s_{k,k,i}(\boldsymbol{X}_{k,k}),\ s_{k,k,j}(\boldsymbol{X}_{k,k}))\right| \;&\leq\; 2\, \left(C_W\, A_{\max}^2\right)\, \left(L_W\, A_{\max}^4\right)\, \left(L_W\, A_{\max}^4 + 2\right) \\ &\leq\; 2\, A_{\max}^{10}\, C_W\, (L_W + 2)^2.\end{aligned}$$

Hence, for all $\{i, j, h\} \subseteq \{1, \ldots, p\}$,

$$\left|\frac{\partial^2\, [\nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}, \boldsymbol{x})]_i}{\partial\theta_h\, \partial\theta_j}\right| \;\leq\; \sum_{k=1}^{K} 2\, C_W\, A_{\max}^{10}\, (L_W + 2)^2 \;\leq\; 2\, C_W\, A_{\max}^{10}\, (L_W + 2)^2\, K.$$

□

Lemma 6. *Consider an exponential-family local dependence random graph model. Assume that there exists a constant $C_B > 0$ such that, for all $1 \leq k < l \leq K$,*

$$\sup_{\boldsymbol{x}_{k,l} \in \mathbb{X}_{k,l}} \|s_{k,l}(\boldsymbol{x}_{k,l})\|_\infty \quad \leq \quad C_B \, |\mathcal{A}_k| \, |\mathcal{A}_l|.$$

Then, for all $(i, j, h) \in \{p+1, \ldots, p+q\}^3$,

$$\left| \frac{\partial^2 \left(\nabla_{\boldsymbol{\theta}} \, \ell(\boldsymbol{\theta}; \boldsymbol{x})\right)_i}{\partial \theta_h \, \partial \theta_j} \right| \quad \leq \quad 2 \, C_B \, A_{\max}^{10} \, (L_B + 2)^2 \binom{K}{2}.$$

Proof of Lemma 6. The lemma is proved similarly to Lemma 5, with the notable exception that the sum in (B.2) is over the index set $1 \leq k < l \leq K$, for the between-block subgraphs. As a result, the factor of K in the bound in Lemma 5 is replaced with $\binom{K}{2}$. Otherwise, using the bound $|\mathcal{A}_k| \, |\mathcal{A}_l| \leq A_{\max}^2$ in place of $\binom{|\mathcal{A}_k|}{2} \leq A_{\max}^2$, the rest of the proof remains valid for the between-block case, making the appropriate adjustments to indexing (i.e., using $C_B > 0$ and $L_B > 0$ in place of their within-block equivalents). □

Lemma 7. *Let $a_1, a_2, b_1, b_2 \in \mathbb{R}$. Then*

$$|a_1 \, b_1 - a_2 \, b_2| \quad \leq \quad |a_1| \, |b_1 - b_2| + |b_2| \, |a_1 - a_2|.$$

Proof of Lemma 7. Write

$$\begin{aligned} |a_1 \, b_1 - a_2 \, b_2| \quad &= \quad |a_1 \, b_1 - (a_2 - a_1 + a_1) \, b_2| \quad &&= \quad |a_1 \, b_1 - b_2 \, (a_2 - a_1) - a_1 \, b_2| \\ &= \quad |a_1 \, (b_1 - b_2) \, - b_2 \, (a_2 - a_1)| \quad &&\leq \quad |a_1| \, |b_1 - b_2| + |b_2| \, |a_1 - a_2|. \end{aligned}$$

□

Appendix C: Auxiliary results for exponential families

We prove various auxiliary results which establish some properties of exponential families.

Lemma 8. *Consider a random vector $\boldsymbol{Y}$ with finite support $\mathbb{Y}$ (i.e., $|\mathbb{Y}| < \infty$) and assume that the probability mass function $f_{\boldsymbol{\theta}} : \mathbb{Y} \mapsto (0, 1)$ belongs to an m-dimensional exponential family, i.e., $f_{\boldsymbol{\theta}}(\boldsymbol{y}) = h(\boldsymbol{y}) \, \exp\left(\langle \boldsymbol{\theta}, \, s(\boldsymbol{y}) \rangle - \psi(\boldsymbol{\theta})\right)$ ($\boldsymbol{\theta} \in \mathbb{R}^m$). Then*

$$\begin{aligned} \nabla_{\boldsymbol{\theta}} \, \psi(\boldsymbol{\theta}) \quad &= \quad \mathbb{E}_{\boldsymbol{\theta}} \, s(\boldsymbol{Y}) \\ \nabla_{\boldsymbol{\theta}} \, \ell(\boldsymbol{\theta}, \boldsymbol{y}) \quad &= \quad s(\boldsymbol{y}) - \mathbb{E}_{\boldsymbol{\theta}} \, s(\boldsymbol{Y}) \\ \nabla_{\boldsymbol{\theta}}^2 \, \psi(\boldsymbol{\theta}) \quad &= \quad \nabla_{\boldsymbol{\theta}}^2 \, \ell(\boldsymbol{\theta}, \boldsymbol{y}) \quad = \quad \mathbb{V}_{\boldsymbol{\theta}} \, s(\boldsymbol{Y}). \end{aligned}$$

Proof of Lemma 8. All results follow from Propositions 3.8 and 3.10 of [39]. □

Lemma 9. *Consider a random vector $\boldsymbol{Y}$ with finite support $\mathbb{Y}$ (i.e., $|\mathbb{Y}| < \infty$) and assume that the probability mass function $f_{\boldsymbol{\theta}} : \mathbb{Y} \mapsto (0,1)$ belongs to an m-dimensional exponential family, i.e.,*

$$f_{\boldsymbol{\theta}}(\boldsymbol{y}) \;=\; h(\boldsymbol{y}) \exp\left(\langle \boldsymbol{\theta},\, s(\boldsymbol{y})\rangle - \psi(\boldsymbol{\theta})\right), \qquad \boldsymbol{\theta} \in \mathbb{R}^m.$$

Then

$$\nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}, \boldsymbol{Y}) - \mathbb{E}\, \nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}, \boldsymbol{Y}) \;=\; s(\boldsymbol{Y}) - \mathbb{E}\, s(\boldsymbol{Y}), \qquad \boldsymbol{\theta} \in \mathbb{R}^m,$$

and

$$\sup_{\boldsymbol{\theta} \in \mathbb{R}^m} \|\nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}, \boldsymbol{Y}) - \mathbb{E}\, \nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}, \boldsymbol{Y})\|_\infty \;=\; \|s(\boldsymbol{Y}) - \mathbb{E}\, s(\boldsymbol{Y})\|_\infty.$$

Proof of Lemma 9. Applying Lemma 8,

$$\nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}, \boldsymbol{y}) \;=\; s(\boldsymbol{y}) - \mathbb{E}_{\boldsymbol{\theta}}\, s(\boldsymbol{Y}).$$

As a result, for all $\boldsymbol{\theta} \in \mathbb{R}^m$,

$$\begin{aligned}\nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}, \boldsymbol{Y}) - \mathbb{E}\, \nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}, \boldsymbol{Y}) &= s(\boldsymbol{Y}) - \mathbb{E}_{\boldsymbol{\theta}}\, s(\boldsymbol{Y}) - \mathbb{E}\, s(\boldsymbol{Y}) + \mathbb{E}\, \mathbb{E}_{\boldsymbol{\theta}}\, s(\boldsymbol{Y}) \\ &= s(\boldsymbol{Y}) - \mathbb{E}\, s(\boldsymbol{Y}).\end{aligned}$$

Hence,

$$\begin{aligned}\sup_{\boldsymbol{\theta} \in \mathbb{R}^m} \|\nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}, \boldsymbol{Y}) - \mathbb{E}\, \nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}, \boldsymbol{Y})\|_\infty &= \sup_{\boldsymbol{\theta} \in \mathbb{R}^m} \|s(\boldsymbol{Y}) - \mathbb{E}\, s(\boldsymbol{Y})\|_\infty \\ &= \|s(\boldsymbol{Y}) - \mathbb{E}\, s(\boldsymbol{Y})\|_\infty.\end{aligned}$$

□

Lemma 10. *Let $\boldsymbol{Y}$ be an m-dimensional random vector with finite support $\mathbb{Y}$. Assume that the distribution of $\boldsymbol{Y}$ belongs to an exponential family with with probability mass functions of the form*

$$f_{\boldsymbol{\theta}}(\boldsymbol{y}) \;=\; h(\boldsymbol{y}) \exp(\langle \boldsymbol{\theta},\, \boldsymbol{y}\rangle - \psi(\boldsymbol{\theta})), \qquad \boldsymbol{\theta} \in \mathbb{R}^m.$$

Assume that there exist constants $U_1 > 0$ and $U_2 > 0$ such that, for all $t \in \{1, \dots, m\}$,

$$|Y_t| \;\leq\; U_1 \qquad \text{and} \qquad |Y_t - \mathbb{E}_{\boldsymbol{\theta}}\, Y_t| \;\leq\; U_2$$

hold with probability 1. Then, for all $(i, j, h) \in \{1, \dots, m\}^3$,

$$\left|\frac{\partial}{\partial\, \theta_h}\, \mathbb{C}_{\boldsymbol{\theta}}(Y_i,\, Y_j)\right| \;\leq\; 2\, U_1\, U_2\, (U_2 + 2).$$

Proof of Lemma 10. Let $(i, j, h) \in \{1, \ldots, m\}^3$ and define $\mu_t(\boldsymbol{\theta}) \coloneqq \mathbb{E}_{\boldsymbol{\theta}}\, Y_t$. Then

$$\begin{aligned}
\frac{\partial}{\partial\,\theta_h}\, \mathbb{C}_{\boldsymbol{\theta}}(Y_i,\, Y_j) &= \frac{\partial}{\partial\,\theta_h}\, \mathbb{E}_{\boldsymbol{\theta}}[Y_i\, Y_j - \mu_i(\boldsymbol{\theta})\, \mu_j(\boldsymbol{\theta})] \\
&= \frac{\partial}{\partial\,\theta_h} \sum_{\boldsymbol{y} \in \mathbb{Y}} [y_i\, y_j - \mu_i(\boldsymbol{\theta})\, \mu_j(\boldsymbol{\theta})]\; f_{\boldsymbol{\theta}}(\boldsymbol{y}) \\
&= \sum_{\boldsymbol{y} \in \mathbb{Y}} \left[f_{\boldsymbol{\theta}}(\boldsymbol{y}) \frac{\partial}{\partial\,\theta_h} [y_i\, y_j - \mu_i(\boldsymbol{\theta})\, \mu_j(\boldsymbol{\theta})] + [y_i\, y_j - \mu_i(\boldsymbol{\theta})\, \mu_j(\boldsymbol{\theta})]\; \frac{\partial}{\partial\,\theta_h}\, f_{\boldsymbol{\theta}}(\boldsymbol{y}) \right].
\end{aligned}$$

By Lemma 8,

$$\frac{\partial}{\partial\,\theta_h}\, f_{\boldsymbol{\theta}}(\boldsymbol{y}) \;=\; \frac{\partial}{\partial\,\theta_h}\, h(\boldsymbol{y})\, \exp(\langle \boldsymbol{\theta},\, \boldsymbol{y} \rangle - \psi(\boldsymbol{\theta})) \;=\; [y_h - \mu_h(\boldsymbol{\theta})]\; f_{\boldsymbol{\theta}}(\boldsymbol{y}).$$

Hence,

$$\frac{\partial}{\partial\,\theta_h}\, \mathbb{C}_{\boldsymbol{\theta}}(Y_i,\, Y_j) \;=\; \mathbb{E}_{\boldsymbol{\theta}}[(Y_i\, Y_j - \mu_i(\boldsymbol{\theta})\, \mu_j(\boldsymbol{\theta}))\, (Y_h - \mu_h(\boldsymbol{\theta}))] - \frac{\partial}{\partial\,\theta_h}\, [\mu_i(\boldsymbol{\theta})\, \mu_j(\boldsymbol{\theta})]\,.$$

We next compute

$$\frac{\partial}{\partial\,\theta_h}\, [\mu_i(\boldsymbol{\theta})\, \mu_j(\boldsymbol{\theta})] \;=\; \mu_i(\boldsymbol{\theta})\, \mathbb{E}_{\boldsymbol{\theta}}[Y_j\, Y_h - \mu_j(\boldsymbol{\theta})\, \mu_h(\boldsymbol{\theta})] + \mu_j(\boldsymbol{\theta})\, \mathbb{E}_{\boldsymbol{\theta}}[Y_i\, Y_h - \mu_i(\boldsymbol{\theta})\, \mu_h(\boldsymbol{\theta})].$$

By the triangle inequality

$$\begin{aligned}
\left| \frac{\partial}{\partial\,\theta_h}\, \mathbb{C}_{\boldsymbol{\theta}}(Y_i,\, Y_j) \right| &\leq \mathbb{E}_{\boldsymbol{\theta}}\, [|(Y_i\, Y_j - \mu_i(\boldsymbol{\theta})\, \mu_j(\boldsymbol{\theta}))|\; |(Y_h - \mu_h(\boldsymbol{\theta}))|] \\
&\quad + \mu_i(\boldsymbol{\theta})\, \mathbb{E}_{\boldsymbol{\theta}}\, |Y_j\, Y_h - \mu_j(\boldsymbol{\theta})\, \mu_h(\boldsymbol{\theta})| \\
&\quad + \mu_j(\boldsymbol{\theta})\, \mathbb{E}_{\boldsymbol{\theta}}\, |Y_i\, Y_h - \mu_i(\boldsymbol{\theta})\, \mu_h(\boldsymbol{\theta})|\,.
\end{aligned}$$

The assumption that there exist constants $U_1 > 0$ and $U_2 > 0$ such that $|Y_t| \leq U_1$ for all $t \in \{1, \ldots, m\}$ and $|Y_t - \mu_t(\boldsymbol{\theta})| \leq U_2$ ($t \in \{1, \ldots, m\}$) hold with probability 1 implies that $|\mu_t(\boldsymbol{\theta})| \leq U_1$ for all $t \in \{1, \ldots, m\}$ and, through an application of Lemma 7, that

$$|Y_j\, Y_h - \mu_j(\boldsymbol{\theta})\, \mu_h(\boldsymbol{\theta})| \;\leq\; |Y_j|\, |Y_h - \mu_h(\boldsymbol{\theta})| + |\mu_h(\boldsymbol{\theta})|\, |Y_j - \mu_j(\boldsymbol{\theta})| \;\leq\; 2\, U_1\, U_2.$$

Hence,

$$\left| \frac{\partial}{\partial\,\theta_h}\, \mathbb{C}_{\boldsymbol{\theta}}(Y_i,\, Y_j) \right| \;\leq\; 2\, U_1\, U_2^2 + 4\, U_1\, U_2 \;=\; 2\, U_1\, U_2\, (U_2 + 2).$$

□

Lemma 11. *Consider a random vector* $\boldsymbol{Y}$ *with finite support* $\mathbb{Y}$ *(i.e.,* $|\mathbb{Y}| < \infty$*). Assume that* $\boldsymbol{Y}$ *belongs to an exponential family:*

$$\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{Y} = \boldsymbol{y}) \quad = \quad h(\boldsymbol{y}) \exp\left(\langle \boldsymbol{\theta}, s(\boldsymbol{y})\rangle - \psi(\boldsymbol{\theta})\right), \qquad \boldsymbol{y} \in \mathbb{Y},$$

with $\dim(\boldsymbol{\theta}) = m$*. Then for all functions* $f : \mathbb{Y} \mapsto \mathbb{R}$*,*

$$\frac{\partial}{\partial\, \theta_i}\, \mathbb{E}_{\boldsymbol{\theta}}\, f(\boldsymbol{Y}) \quad = \quad \mathbb{E}_{\boldsymbol{\theta}}[f(\boldsymbol{Y})\,(s_i(\boldsymbol{Y}) - \mathbb{E}_{\boldsymbol{\theta}}\, s_i(\boldsymbol{Y}))],$$

for all $i \in \{1, \ldots, m\}$*.*

Proof of Lemma 11. Write

$$\begin{aligned}
\frac{\partial}{\partial\, \theta_i}\, \mathbb{E}_{\boldsymbol{\theta}}\, f(\boldsymbol{Y}) \quad &= \quad \frac{\partial}{\partial\, \theta_i} \sum_{\boldsymbol{y} \in \mathbb{Y}} f(\boldsymbol{y})\, h(\boldsymbol{y}) \exp\left(\langle \boldsymbol{\theta}, s(\boldsymbol{y})\rangle - \psi(\boldsymbol{\theta})\right) \\
&= \quad \sum_{\boldsymbol{y} \in \mathbb{Y}} f(\boldsymbol{y})\, h(\boldsymbol{y}) \left[\frac{\partial}{\partial\, \theta_i} \exp\left(\langle \boldsymbol{\theta}, s(\boldsymbol{y})\rangle - \psi(\boldsymbol{\theta})\right)\right] \\
&= \quad \sum_{\boldsymbol{y} \in \mathbb{Y}} f(\boldsymbol{y})\, h(\boldsymbol{y}) \exp\left(\langle \boldsymbol{\theta}, s(\boldsymbol{y})\rangle - \psi(\boldsymbol{\theta})\right) \left(s_i(\boldsymbol{y}) - \mathbb{E}_{\boldsymbol{\theta}}\, s_i(\boldsymbol{Y})\right) \\
&= \quad \mathbb{E}_{\boldsymbol{\theta}}[f(\boldsymbol{Y})\,(s_i(\boldsymbol{Y}) - \mathbb{E}_{\boldsymbol{\theta}}\, s_i(\boldsymbol{Y}))],
\end{aligned}$$

as applying Lemma 8 shows that

$$\frac{\partial}{\partial\, \theta_i}\, \psi(\boldsymbol{\theta}) \quad = \quad \mathbb{E}_{\boldsymbol{\theta}}\, s_i(\boldsymbol{X}), \qquad i = 1, \ldots, m.$$

□

References

[1] Agneessens, F., Trincado-Munoz, F. J. and Koskinen, J. (2022). Network formation in organizational settings: Exploring the importance of local social processes and team-level contextual variables in small groups using bayesian hierarchical ERGMs. *Social Networks.*

[2] Babkin, S., Stewart, J. R., Long, X. and Schweinberger, M. (2020). Large-scale estimation of random graph models with local dependence. *Computational Statistics and Data Analysis* **152** 107029.

[3] Brown, L. D. (1986). *Fundamentals of statistical exponential families: with applications in statistical decision theory.* Institute of Mathematical Statistics.

[4] Chatterjee, S., Diaconis, P. and Sly, A. (2011). Random graphs with a given degree sequence. *The Annals of Applied Probability* **21** 1400-1435.

[5] Chatterjee, S. and Diaconis, P. (2013). Estimating and understanding exponential random graph models. *The Annals of Statistics* **41** 2428-2461.
[6] Chazottes, J. R., Collet, P., Külske, C. and Redig, F. (2007). Concentration inequalities for random fields via coupling. *Probability Theory and Related Fields* **137** 201-225.
[7] Chen, M., Kato, K. and Leng, C. (2021). Analysis of networks via the sparse β-model. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **83** 887-910.
[8] 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*.
[9] 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 Papers and Proceedings* **113** 156-160.
[10] Frank, O. and Strauss, D. (1986). Markov Graphs. *Journal of the American Statistical Association* **81** 832-842.
[11] Holland, P. W., Laskey, K. B. and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks* **5** 109-137.
[12] Hunter, D. R. and Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics* **15** 565-583.
[13] Häggström, O. and Jonasson, J. (1999). Phase Transition in the Random Triangle Model. *Journal of Applied Probability* **36** 1101-1115.
[14] Janková, J. and van de Geer, S. (2018). *Inference in high-dimensional graphical models* 325-346. CRC Press.
[15] Jonasson, J. (1999). The Random Triangle Model. *Journal of Applied Probability* **36** 852-867.
[16] Kontorovich, L. and Ramanan, K. (2008). Concentration inequalities for dependent random variables via the martingale method. *The Annals of Probability* **36** 2126-2158.
[17] Krivitsky, P. N., Handcock, M. S. and Morris, M. (2011). Adjusting for network size and composition effects in exponential-family random graph models. *Statistical Methodology* **8** 319-339.
[18] Krivitsky, P. N., Hunter, D. R., Morris, M. and Klumb, C. (2023). ergm 4: New Features for Analyzing Exponential-Family Random Graph Models. *Journal of Statistical Software* **105** 1-44.
[19] Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press.
[20] Lauritzen, S., Rinaldo, A. and Sadeghi, K. (2018). Random networks, graphical models and exchangeability. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **80** 481-508.
[21] Leskovec, J., Lang, K. J., Dasgupta, A. and Mahoney, M. W. (2008). Statistical properties of community structure in large social and information networks. *Proceedings of the 17th international conference on World Wide Web* 695-704.
[22] Lindvall, T. (2002). *Lectures on the coupling method*. Courier Corporation.
[23] Lusher, D., Koskinen, J. and Robins, G. (2012). *Exponential Random Graph*

Models for Social Networks. Cambridge University Press.

[24] Mele, A. (2022). A structural model of homophily and clustering in social networks. *Journal of Business & Economic Statistics* **40** 1377-1389.

[25] Newman, M. E. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E* **69** 026113.

[26] Ortega, J. M. and Rheinboldt, W. C. (2000). *Iterative solution of nonlinear equations in several variables*. Society for Industrial and Applied Mathematics.

[27] Portnoy, S. (1988). Asymptotic Behavior of Likelihood Methods for Exponential Families when the Number of Parameters Tends to Infinity. *The Annals of Statistics* **16** 356-366.

[28] Raič, M. (2019). A multivariate Berry-Esseen theorem with explicit constants. *Bernoulli* **25** 2824-2853.

[29] Ravikumar, P., Wainwright, M. J. and Lafferty, J. D. (2010). High-dimensional ising model selection using ℓ_1-regularized logistic regression. *The Annals of Statistics* **38** 1287-1319.

[30] Rohe, K., Chatterjee, S. and Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics* **39** 1878-1915.

[31] Schubert, D. and Brand, A. (2022). "Whom Should I Talk To?": Role Prescription and Hierarchy Building in Supervised Living Groups. *Social Inclusion* **10** 295-306.

[32] Schweinberger, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *Journal of the American Statistical Association* **106** 1361-1370.

[33] Schweinberger, M. and Handcock, M. S. (2015). Local dependence in random graph models: Characterization, properties and statistical inference. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **77** 647-676.

[34] Schweinberger, M. and Stewart, J. (2020). Concentration and consistency results for canonical and curved exponential-family models of random graphs. *The Annals of Statistics* **48**.

[35] 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. *Statistical Science* **35** 627-662.

[36] 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*.

[37] 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*.

[38] Stewart, J., Schweinberger, M., Bojanowski, M. and Morris, M. (2019). Multilevel network data facilitate statistical inference for curved ERGMs with geometrically weighted terms. *Social Networks* **59** 98-119.

[39] Sundberg, R. (2019). *Statistical Modelling by Exponential Families*. Cambridge University Press.

[40] Tolochko, P. and Boomgaarden, H. G. (2024). Same but different: A comparison of estimation approaches for exponential random graph models for multiple

networks. *Social Networks* **76** 1-11.

[41] van den Berg, J. and Maes, C. (1994). Disagreement percolation in the study of Markov fields. *The Annals of Probability* **22** 749-763.

[42] Whetsell, T. A., Kroll, A. and DeHart-Davis, L. (2021). Formal Hierarchies and Informal Networks: How Organizational Structure Shapes Information Search in Local Government. *Journal of Public Administration Research and Theory* **31** 653-669.