PSEUDO-LIKELIHOOD-BASED M-ESTIMATION OF RANDOM GRAPHS WITH DEPENDENT EDGES AND PARAMETER VECTORS OF INCREASING DIMENSION

By Jonathan R. Stewart and Michael Schweinberger

Florida State University and Penn State University

An important question in statistical network analysis is how to estimate models of discrete and dependent network data with intractable likelihood functions, without sacrificing computational scalability and statistical guarantees. We demonstrate that scalable estimation of random graph models with dependent edges is possible, by establishing convergence rates of pseudo-likelihood-based M-estimators for discrete undirected graphical models with exponential parameterizations and parameter vectors of increasing dimension in single-observation scenarios. We highlight the impact of two complex phenomena on the convergence rate: phase transitions and model near-degeneracy. The main results have possible applications to discrete and dependent network, spatial, and temporal data. To showcase convergence rates, we introduce a novel class of generalized β-models with dependent edges and parameter vectors of increasing dimension, which leverage additional structure in the form of overlapping subpopulations to control dependence. We establish convergence rates of pseudo-likelihood-based M-estimators for generalized β-models in dense- and sparse-graph settings.

1. Introduction.

Network data have garnered considerable attention in recent years, driven by the growth of the internet and online social networks that can serve as echo chambers and facilitate polarization, and applications in science, technology, and public health (e.g., pandemics).

During the past two decades, substantial progress has been made on models of network data, including β- and p_1-models [e.g., 23, 9, 42, 33, 24, 28, 11]; exchangeable random graph models [e.g., 6, 13]; stochastic block models [e.g., 3, 34, 1, 17]; latent space models [e.g., 22]; and exponential-family models of random graphs [e.g., 19, 35, 8, 29, 37]. Other models are small-world networks [40] and scale-free networks with power law degree distributions [4]. That said, despite strides in modeling and inference, fundamental questions arising from the statistical analysis of non-standard and dependent network data have remained unanswered.

[*]MSC2010 subject classifications. Primary 05C80; secondary 62B05, 62F10, 91D30.

Keywords and phrases: Markov random fields, graphical models, conditional independence, statistical exponential families, phase transitions, model near-degeneracy

1.1. *Three questions.* Since the dawn of statistical network analysis in the 1980s [23, 15], three questions have loomed large:

I. How can one construct models that allow the propensities of nodes to form edges and other subgraphs to vary across nodes?
II. How can one construct models that do justice to the fact that network data are dependent data?
III. How can one learn models from a single observation of a random graph with dependent edges and parameter vectors of increasing dimension, regardless of whether the likelihood function is tractable?

We take steps to answer these questions by building on the statistical exponential-family platform [5], which has long served as a convenient mathematical platform for obtaining first answers to statistical questions involving discrete and dependent data and hosts Bernoulli random graphs, β- and p_1-models [23, 9], generalized linear models of random graphs, and undirected graphical models of random graphs [15, 25]. An alternative route, not considered here, is provided by the Hoover-Aldous representation theorem [3] via exchangeable random graphs [6, 13], which can likewise induce dependence (as demonstrated by stochastic block and latent space models).

On the statistical exponential-family platform, research has focused on β- and p_1-models, which provide answers to the first question but assume that edges are independent; and exponential-family random graph models, which allow edges to be dependent and can capture observed heterogeneity via covariates, but are less suited to capturing unobserved heterogeneity and often give rise to intractable likelihood functions. An additional issue is that theoretical properties of statistical procedures – well-established in the literature on β- and p_1-models [e.g., 9, 42, 33, 24, 41, 28, 11] – are scarce in the literature on exponential-family random graph models, with two recent exceptions. Mukherjee [29] considered models with functions of degrees as sufficient statistics, which allow edges to be dependent, but have two parameters and do not capture network features other than degrees. Schweinberger and Stewart [37] considered models with dependent edges, but constrained dependence to non-overlapping subpopulations of nodes. While both works provide statistical guarantees, these works focus on the second question rather than the first question.

We aim to provide tentative answers to all three questions, leveraging the statistical exponential-family platform.

1.2. *Probabilistic framework.* On the modeling side, we consider a flexible approach to specifying random graph models with complex dependence from simple building blocks. We demonstrate the probabilistic framework by extending the β-model of Chatterjee et al. [9] – studied by Rinaldo et al. [33], Yan and Xu [42], Karwa and Slavković [24], Mukherjee et al. [28], Chen et al. [11], and others – to generalized β-models capturing dependence among edges along with heterogeneity in the propensities of nodes to form edges. To control the dependence among edges, generalized β-models leverage additional structure in the form of overlapping subpopulations. The β-model and generalized β-models have in common that the number of parameters increases with the number of nodes. Having said that, the closest relative of generalized β-models with dependent edges is not the β-model with independent edges, but are statistical exponential-family models for discrete and dependent random variables: e.g., Ising models, Markov random fields, and undirected graphical models for discrete and dependent network, spatial, and temporal data [e.g., 18].

1.3. *Computational scalability and statistical guarantees.* On the statistical side, we demonstrate that computational scalability and statistical guarantees need not be sacrificed in order to estimate random graph models with dependent edges and parameter vectors of increasing dimension.

We do so by focusing on pseudo-likelihood-based M-estimators, which possess convenient factorization properties and are more scalable than estimators based on intractable likelihood functions. Despite computational advantages, the properties of pseudo-likelihood-based M-estimators for random graphs with dependent edges and parameter vectors of increasing dimension are unknown. In the related literature on Ising models and discrete Markov random fields in single-observation scenarios, consistency of maximum pseudo-likelihood estimators has been established [12, 7, 2, 18], but those results are limited to a fixed number of parameters.

We demonstrate that scalable estimation of random graph models with dependent edges is possible, by establishing convergence rates of pseudo-likelihood-based M-estimators for discrete undirected graphical models with exponential parameterizations and parameter vectors of increasing dimension in single-observation scenarios. In contrast to Ravikumar et al. [31] and other works on high-dimensional Ising models and discrete Markov random fields, we do not assume that independent replications are available. The main results have possible applications to discrete and dependent network, spatial, and temporal data. We highlight the impact of two complex phenomena on the convergence rate: phase transitions and model near-degeneracy. To showcase convergence rates, we establish convergence rates for generalized β-models with dependent edges and parameter vectors of increasing dimension in dense- and sparse-graph settings.

1.4. *Structure.* Section 2 introduces the probabilistic framework. Section 3 establishes convergence rates for pseudo-likelihood-based M-estimators.

1.5. *Notation.* Let $\mathcal{N} \coloneqq \{1, \dots, N\}$ $(N \geq 3)$ be a finite set of nodes and $\boldsymbol{X}$ be a random graph defined on $\mathcal{N}$ with sample space $\mathbb{X} \coloneqq \{0, 1\}^{\binom{N}{2}}$, where $X_{i,j} = 1$ if nodes $i \in \mathcal{N}$ and $j \in \mathcal{N}$ are connected by an edge and $X_{i,j} = 0$ otherwise. We focus on random graphs with undirected edges and without self-edges, although our results can be extended to directed random graphs. The set $\mathbb{R}^+ \coloneqq (0,\, \infty)$ denotes the set of positive real numbers, and the vector $\mathbf{0} \in \mathbb{R}^d$ denotes the d-dimensional null vector in $\mathbb{R}^d$ $(d \geq 1)$. We denote the ℓ_1-, ℓ_2-, and ℓ_∞-norm of vectors in $\mathbb{R}^d$ by $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$, respectively. For any matrix $\boldsymbol{A} \in \mathbb{R}^{d \times d}$, let $|||\boldsymbol{A}|||_1 \coloneqq \max_{1 \leq j \leq d} \sum_{i=1}^{d} |A_{i,j}|$, $|||\boldsymbol{A}|||_\infty \coloneqq \max_{1 \leq i \leq d} \sum_{j=1}^{d} |A_{i,j}|$, and $|||\boldsymbol{A}|||_2 \coloneqq \sup_{\boldsymbol{u} \in \mathbb{R}^d :\, \|\boldsymbol{u}\|_2 = 1} \|\boldsymbol{A}\, \boldsymbol{u}\|_2$. The open hypercube in $\mathbb{R}^d$ centered at $\boldsymbol{c} \in \mathbb{R}^d$ with radius $\rho > 0$ is denoted by $\mathcal{B}_\infty(\boldsymbol{c},\, \rho) \coloneqq \{\boldsymbol{a} \in \mathbb{R}^d : \|\boldsymbol{a} - \boldsymbol{c}\|_\infty < \rho\}$. For any subset $\mathcal{S} \subset \mathbb{R}^d$, $\operatorname{int}(\mathcal{S})$ denotes the interior of $\mathcal{S}$. The total variation distance between two probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ defined on a common measurable space is denoted by $\|\mathbb{P}_1 - \mathbb{P}_2\|_{\text{TV}}$. Expectations, variances, and covariances are denoted by $\mathbb{E}$, $\mathbb{V}$, and $\mathbb{C}$, respectively. For any finite set $\mathcal{S}$, the number of elements of $\mathcal{S}$ is denoted by $|\mathcal{S}|$. The function $\mathbb{1}(\,\cdot\,)$ is an indicator function, which is 1 if its argument is true and is 0 otherwise. Uppercase letters $A, B, C, \dots$ denote finite constants. We write $a(n) = O(b(n))$ if there exists a finite constant $C > 0$ such that $|a(n) \,/\, b(n)| \leq C$ for all large enough n, and write $a(n) = o(b(n))$ if, for all $\epsilon > 0$, $|a(n) \,/\, b(n)| \;< \epsilon$ for all large enough n.

2. Probabilistic framework.

We consider a simple and flexible approach to specifying random graph models with complex dependence from simple building blocks. Let $\{\mathbb{P}_{\boldsymbol{\theta}},\, \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ be a family of probability measures dominated by a σ-finite measure ν, with densities of the form

$$f_{\boldsymbol{\theta}}\,(\boldsymbol{x}) \;\; \propto \;\; \prod_{i<j}^{N} \varphi_{i,j}(x_{i,j},\, \boldsymbol{x}_{\mathcal{S}_{i,j}};\, \boldsymbol{\theta}), \qquad \boldsymbol{x} \in \mathbb{X}, \tag{2.1}$$

where $\varphi_{i,j} : \{0,1\}^{|\mathcal{S}_{i,j}|+1} \times \boldsymbol{\Theta} \mapsto [0,\, \infty)$ is a function that specifies how edge variable $X_{i,j}$ depends on a subset of edge variables $\boldsymbol{X}_{\mathcal{S}_{i,j}}$. Here, $\mathcal{S}_{i,j}$ denotes a subset of unordered pairs of nodes $\{a, b\} \subset \mathcal{N}$, and $\boldsymbol{X}_{\mathcal{S}_{i,j}}$ denotes a set of indicators of edges between the unordered pairs of nodes in $\mathcal{S}_{i,j}$. We allow the dimension $p \geq 1$ of parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$ to increase as a function of the number of nodes N, i.e., $p \to \infty$ as $N \to \infty$. A natural choice of reference measure ν is the counting measure.

It is worth noting that the factorization of (2.1) does not imply that edges are independent, because each $\varphi_{i,j}$ can be a function of multiple edges and can hence induce dependence among edges. That said, the factorization of (2.1) implies conditional independence properties [14], and the resulting models can be viewed as undirected graphical models of random graphs [15, 25]. In contrast to the undirected graphical models of random graphs by Frank and Strauss [15], which allow edges to depend on many other edges and can give rise to undesirable behavior [e.g., model near-degeneracy, 19, 32, 35, 8], we leverage additional structure to control dependence among edges. The additional structure consists of a population with overlapping subpopulations and comes with two benefits. First, it facilitates the construction of novel models with non-trivial dependence. Second, it helps control the dependence among edges. To demonstrate, we introduce a novel class of generalized β-models with dependent edges in Sections 2.2–2.4.

2.1. *Parameterizations.* It is convenient to parameterize the functions of edges $\varphi_{i,j}$ by using exponential parameterizations. Exponential parameterizations are widely used in the literature on undirected graphical models: see, e.g., Lauritzen et al. [25]. We therefore assume that

$$\varphi_{i,j}(x_{i,j},\, \boldsymbol{x}_{\mathcal{S}_{i,j}};\, \boldsymbol{\theta}) \;\; := \;\; a_{i,j}(x_{i,j},\, \boldsymbol{x}_{\mathcal{S}_{i,j}})\, \exp(\langle \boldsymbol{\theta},\, s_{i,j}(x_{i,j},\, \boldsymbol{x}_{\mathcal{S}_{i,j}}) \rangle), \tag{2.2}$$

where $a_{i,j} : \{0,1\}^{|\mathcal{S}_{i,j}|+1} \mapsto [0,\, \infty)$ is a function of $x_{i,j}$ and $\boldsymbol{x}_{\mathcal{S}_{i,j}}$, which can be used to induce sparsity by penalizing edges, and $\langle \boldsymbol{\theta},\, s_{i,j}(x_{i,j},\, \boldsymbol{x}_{\mathcal{S}_{i,j}})\rangle$ is the inner product of a vector of parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$ and a vector of statistics $s_{i,j} : \{0,1\}^{|\mathcal{S}_{i,j}|+1} \mapsto \mathbb{R}^p$ ($\{i,j\} \subset \mathcal{N}$). The probability density function (2.1) with parameterization (2.2) can be written in exponential-family form:

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) \;\; = \;\; a(\boldsymbol{x})\, \exp\left(\langle \boldsymbol{\theta},\, s(\boldsymbol{x})\rangle - \psi(\boldsymbol{\theta})\right), \qquad \boldsymbol{x} \in \mathbb{X}, \tag{2.3}$$

where $a : \mathbb{X} \mapsto [0,\, \infty)$ is given by $a(\boldsymbol{x}) := \prod_{i<j}^N a_{i,j}(x_{i,j},\, \boldsymbol{x}_{\mathcal{S}_{i,j}})$ and the vector of sufficient statistics $s : \mathbb{X} \mapsto \mathbb{R}^p$ is given by

$$s(\boldsymbol{x}) \;\; := \;\; \sum_{i<j}^{N} s_{i,j}(x_{i,j},\, \boldsymbol{x}_{\mathcal{S}_{i,j}}). \tag{2.4}$$

The function $\psi : \boldsymbol{\Theta} \mapsto (0, \infty)$ ensures that $\int_{\mathbb{X}} f_{\boldsymbol{\theta}}(\boldsymbol{x})\, \mathrm{d}\,\nu(\boldsymbol{x}) = 1$:

$$\psi(\boldsymbol{\theta}) \;\; := \;\; \log \int\limits_{\mathbb{X}} a(\boldsymbol{x})\, \exp\left(\langle \boldsymbol{\theta},\, s(\boldsymbol{x})\rangle\right) \mathrm{d}\,\nu(\boldsymbol{x}), \qquad \boldsymbol{\theta} \;\in\; \boldsymbol{\Theta}.$$

The parameter space is $\boldsymbol{\Theta} := \{\boldsymbol{\theta} \in \mathbb{R}^p \,:\, \psi(\boldsymbol{\theta}) < \infty\} = \mathbb{R}^p$, because the family of densities is an exponential family of densities with respect to a σ-finite measure with a finite support [5]. To ensure that $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is identifiable, we assume that the exponential family is minimal in the sense of Brown [5, p. 2]. The assumption of a minimal exponential family involves no loss of generality, because all non-minimal exponential families can be reduced to minimal exponential families [5, Theorem 1.9, p. 13].

We demonstrate the probabilistic framework by developing a novel class of generalized β-models with dependent edges and $p \geq N \to \infty$ parameters.

2.2. *Model 1: β-model with independent edges.* To introduce generalized β-models with dependent edges, we first review the β-model with independent edges [9]. The β-model assumes that edges between nodes $i \in \mathcal{N}$ and $j \in \mathcal{N}$ are independent Bernoulli$(\mu_{i,j})$ $(\mu_{i,j} \in (0,1))$ random variables, where

$$\log \frac{\mu_{i,j}}{1 - \mu_{i,j}} \;\;=\;\; \theta_i + \theta_j, \qquad \theta_i \in \mathbb{R}, \qquad \theta_j \in \mathbb{R}.$$

The parameters θ_i and θ_j can be interpreted as the propensities of nodes i and j to form edges. The β-model is a special case of the probabilistic framework introduced above, corresponding to

$$\varphi_{i,j}(x_{i,j};\, \boldsymbol{\theta}) \;\;=\;\; a_{i,j}(x_{i,j})\, \exp((\theta_i + \theta_j)\, x_{i,j}), \quad \boldsymbol{\theta} = (\theta_1, \ldots, \theta_N) \in \mathbb{R}^N,$$

where $a_{i,j}(x_{i,j})$ is 1 if $x_{i,j} \in \{0, 1\}$ and is 0 otherwise. The β-model captures heterogeneity in the propensities of nodes to form edges, but assumes that edges are independent.

2.3. *Model 2: generalized β-model with dependent edges.* We introduce a generalization of the β-model, which captures dependence among edges induced by brokerage in networks, in addition to heterogeneity in the propensities of nodes to form edges. Brokerage can influence economic and political outcomes of interest and has therefore been studied by economists, political scientists, and other network scientists since at least the 1980s. An example of brokerage is given by faculty members of universities with appointments in both computer science and statistics, who can facilitate collaborations between faculty members in computer science and faculty members in statistics and can hence facilitate interdisciplinary research.

To capture dependence among edges induced by brokerage in networks, consider a finite population of nodes $\mathcal{N}$ consisting of $K \geq 2$ known subpopulations $\mathcal{A}_1, \ldots, \mathcal{A}_K$, which may overlap in the sense that the intersections of subpopulations are non-empty. As a consequence, nodes may belong to

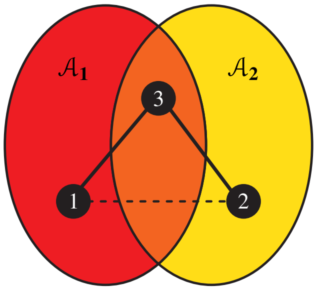


FIG 1. *A graphical representation of the dependencies among edges induced by brokerage. Consider two overlapping subpopulations $\mathcal{A}_1$ and $\mathcal{A}_2$. The nodes $1 \in \mathcal{A}_1 \setminus \mathcal{A}_2$ and $2 \in \mathcal{A}_2 \setminus \mathcal{A}_1$ do not belong to the same subpopulation, but the shared partner $3 \in \mathcal{A}_1 \cap \mathcal{A}_2$ in the intersection of subpopulations $\mathcal{A}_1$ and $\mathcal{A}_2$ can facilitate an edge between nodes 1 and 2, indicated by the dashed line between nodes 1 and 2.*

multiple subpopulations: e.g., faculty members of universities may have appointments in multiple departments, which implies that the faculties of departments overlap. Subpopulation structure is inherent to many real-world networks, in part because people tend to build communities, and in part because organizations tend to divide large bodies of people into small bodies of people (e.g., divisions, subdivisions). It is worth noting that we focus on known subpopulations that can overlap, in contrast to the literature on stochastic block models [3]. In applications, it is often possible to observe subpopulation structure: e.g., the appointments of faculty members can be determined by scraping the websites of universities.

Define, for each node $i \in \mathcal{N}$, its neighborhood $\mathcal{N}_i$ as the subset of all other nodes $j \in \mathcal{N} \setminus \{i\}$ that share at least one subpopulation with node $i \in \mathcal{N}$:

$$\mathcal{N}_i \;\coloneqq\; \{j \in \mathcal{N} \setminus \{i\} : \exists\, k \in \{1, \dots, K\} \text{ such that } i \in \mathcal{A}_k \text{ and } j \in \mathcal{A}_k\}\,.$$

To capture dependence among edges induced by shared partners in the intersections of neighborhoods, we consider functions of edges $\varphi_{i,j}$ of the form

$$\varphi_{i,j}(x_{i,j},\, \boldsymbol{x}_{\mathcal{S}_{i,j}};\, \boldsymbol{\theta}) \;\coloneqq\; a_{i,j}(x_{i,j})\, \exp\left((\theta_i + \theta_j)\, x_{i,j} + \theta_{N+1}\, b_{i,j}(x_{i,j},\, \boldsymbol{x}_{\mathcal{S}_{i,j}})\right),$$

where $\mathcal{S}_{i,j} \subset \mathcal{N}$ is the set of unordered pairs of nodes such that one node is an element of $\{i,\, j\}$ and the other node is an element of $\mathcal{N}_i \,\cap\, \mathcal{N}_j$, $a_{i,j}(x_{i,j})$ is 1 if $x_{i,j} \in \{0, 1\}$ and is 0 otherwise, and

$$b_{i,j}(x_{i,j}, \boldsymbol{x}_{\mathcal{S}_{i,j}}) \coloneqq \begin{cases} 0 & \text{if } \mathcal{N}_i \cap \mathcal{N}_j = \emptyset \\ x_{i,j}\, \mathbb{1}\left(\displaystyle\sum_{h \in \mathcal{N}_i \cap \mathcal{N}_j} x_{i,h}\, x_{j,h} \geq 1\right) & \text{if } \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset. \end{cases} \tag{2.5}$$

Here, $\mathbb{1}(\sum_{h \in \mathcal{N}_i \cap \mathcal{N}_j} x_{i,h}\, x_{j,h} \geq 1)$ is an indicator function, which is 1 if nodes i and j have at least one shared partner in the intersection of neighborhoods $\mathcal{N}_i$ and $\mathcal{N}_j$ and is 0 otherwise, and $\boldsymbol{\theta} \coloneqq (\theta_1, \ldots, \theta_{N+1}) \in \mathbb{R}^{N+1}$.

Remark. Generalized β-model captures brokerage in networks. The generalized β-model captures brokerage in networks, along with heterogeneity in the propensities of nodes to form edges. To demonstrate, consider the two overlapping subpopulations $\mathcal{A}_1$ and $\mathcal{A}_2$ shown in Figure 1. The nodes $1 \in \mathcal{A}_1 \setminus \mathcal{A}_2$ and $2 \in \mathcal{A}_2 \setminus \mathcal{A}_1$ do not belong to the same subpopulation, but the shared partner $3 \in \mathcal{A}_1 \cap \mathcal{A}_2$ in the intersection of subpopulations $\mathcal{A}_1$ and $\mathcal{A}_2$ can facilitate an edge between nodes 1 and 2, provided $\theta_{N+1} > 0$. In the language of network science, nodes in the intersection $\mathcal{A}_1 \cap \mathcal{A}_2$ of subpopulations $\mathcal{A}_1$ and $\mathcal{A}_2$ can act as brokers, facilitating edges between nodes in $\mathcal{A}_1 \setminus \mathcal{A}_2$ and nodes in $\mathcal{A}_2 \setminus \mathcal{A}_1$. In fact, the generalized β-model can capture an excess in the expected number of brokered edges relative to the β-model, in the sense that

$$\underbrace{\mathbb{E}_{\theta_1,\ldots,\theta_N,\theta_{N+1}>0}\; b(\boldsymbol{X})}_{\textit{generalized } \beta\textit{-model}} \quad > \quad \underbrace{\mathbb{E}_{\theta_1,\ldots,\theta_N,\theta_{N+1}=0}\; b(\boldsymbol{X})}_{\beta\textit{-model}}, \tag{2.6}$$

where $b(\boldsymbol{X}) = \sum_{i<j}^{N} b_{i,j}(X_{i,j},\, \boldsymbol{X}_{\mathcal{S}_{i,j}})$ and $\mathbb{E}_{\theta_1,\ldots,\theta_N,\theta_{N+1}}\, b(\boldsymbol{X})$ is the expectation of $b(\boldsymbol{X})$ under $(\theta_1, \ldots, \theta_N, \theta_{N+1}) \in \mathbb{R}^{N+1}$. In other words, the generalized β-model with $\theta_{N+1} > 0$ generates graphs that have, on average, more brokered edges than the β-model, assuming that the propensities $\theta_1, \ldots, \theta_N$ of nodes $1, \ldots, N$ to form edges are the same under both models. The inequality in (2.6) follows from the fact that the generalized β-model is an exponential-family model along with Corollary 2.5 of Brown [5, p. 37].

2.4. *Model 3: sparse generalized β-models with dependent edges.* Sparse random graphs have been studied since the pioneering work of Erdős and Rényi [e.g., 33, 28, 29, 11]. To develop sparse versions of generalized β-models, it makes sense to penalize edges between nodes $i \in \mathcal{N}$ and $j \in \mathcal{N}$ that are distant in the sense that $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset$, without penalizing edges between nodes that are close in the sense that $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$. We therefore induce sparsity by considering Model 2 with

$$a_{i,j}(x_{i,j}) \quad \coloneqq \quad \begin{cases} N^{-\alpha\, x_{i,j}\, \mathbb{1}(\mathcal{N}_i \cap \mathcal{N}_j = \emptyset)} & \text{if } x_{i,j} \in \{0, 1\} \\ 0 & \text{otherwise,} \end{cases}$$

where $\alpha \in (0,\, 1]$ is called the level of sparsity of the random graph.

To demonstrate that Model 3 encourages random graphs to be sparse, we bound the expected degrees of nodes.

Proposition 1. *Consider Model 3 with $\boldsymbol{\theta} \in \mathbb{R}^{N+1}$ and $\alpha \in (0, 1]$. Then*

$$\max_{1 \leq i \leq N} \, \mathbb{E}_{\boldsymbol{\theta}} \left(\sum_{j \neq i}^{N} X_{i,j} \right) \; \leq \; 2 \, \exp(3 \, \|\boldsymbol{\theta}\|_\infty) \, \left(\left(\max_{1 \leq h \leq N} |\mathcal{N}_h| \right)^2 + N^{1-\alpha} \right).$$

Proposition 1 reveals that when the neighborhoods $\mathcal{N}_h$ of nodes $h \in \mathcal{N}$ are not too large, the random graph is sparse in the sense that the expected degrees of all nodes are $o(N)$. For example, if $\max_{1 \leq h \leq N} |\mathcal{N}_h|$ and $\|\boldsymbol{\theta}\|_\infty$ are bounded above, the expected degrees of nodes are $O(N^{1-\alpha})$.

3. Statistical guarantees.

We establish consistency results and convergence rates of maximum likelihood and pseudo-likelihood-based M-estimators in Sections 3.2 and 3.3, respectively. We then present applications to β- and generalized β-models with dependent edges in Section 3.4. To prepare the ground, we first discuss how the dependence among edges and the smoothness of sufficient statistics can be quantified. To ease the presentation, we replace the double subscripts of edge variables by single subscripts and write $(X_m)_{1 \leq m \leq M}$ instead of $(X_{i,j})_{i<j:\, i \in \mathcal{N},\, j \in \mathcal{N}}$, where $M := \binom{N}{2}$. The data-generating parameter vector is denoted by $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta} = \mathbb{R}^p$.

3.1. *Controlling dependence and smoothness.*

To obtain consistency results and convergence rates based on a single observation of a random graph with dependent edges, we need to control the dependence among edges along with the smoothness of the sufficient statistics of the model.

The dependence among edges can be controlled by bounding the total variation distance between conditional probability mass functions of edge variables, quantifying how much the conditional probability mass functions of edge variables are affected by changes of other edge variables. Define $\boldsymbol{X}_{a:b} := (X_a, \ldots, X_b) \in \{0,\, 1\}^{b-a+1}$, where $a \leq b$ and $a,\, b \in \{1, \ldots, M\}$. For each $i \in \{1, \ldots, M\}$, we denote the conditional probability mass function of subgraph $\boldsymbol{X}_{i+1:M}$ given subgraph $(\boldsymbol{X}_{1:i-1},\, X_i) = (\boldsymbol{x}_{1:i-1},\, x_i)$ by $\mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{x}_{1:i-1}, x_i}$:

$$\mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{x}_{1:i-1}, x_i}(\boldsymbol{X}_{i+1:M} = \boldsymbol{a}) \; := \; \mathbb{P}_{\boldsymbol{\theta}^\star}(\boldsymbol{X}_{i+1:M} = \boldsymbol{a} \mid (\boldsymbol{X}_{1:i-1}, X_i) = (\boldsymbol{x}_{1:i-1}, x_i)),$$

where $\boldsymbol{a} \in \{0, 1\}^{M-i}$. We quantify the dependence among edges by bounding the total variation distance between the conditional probability mass functions $\mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{x}_{1:i-1}, 0}$ and $\mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{x}_{1:i-1}, 1}$ by using coupling methods [27]:

$$\|\mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{x}_{1:i-1}, 0} - \mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{x}_{1:i-1}, 1}\|_{\text{TV}} \; \leq \; \mathbb{Q}_{\boldsymbol{\theta}^\star, i, \boldsymbol{x}_{1:i-1}}(\boldsymbol{X}^\star_{i+1:M} \neq \boldsymbol{X}^{\star\star}_{i+1:M}),$$

where the pair of random vectors $(\boldsymbol{X}^{\star}_{i+1:M}, \boldsymbol{X}^{\star\star}_{i+1:M}) \in \{0,1\}^{M-i} \times \{0,1\}^{M-i}$ with joint probability mass function $\mathbb{Q}_{\boldsymbol{\theta}^\star, i, \boldsymbol{x}_{1:i-1}}$ is a coupling of $\mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{x}_{1:i-1}, 0}$ and $\mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{x}_{1:i-1}, 1}$ [27]. The coupling $\mathbb{Q}_{\boldsymbol{\theta}^\star, i, \boldsymbol{x}_{1:i-1}}$ is constructed in Lemma 12 in the supplement [38]. Based on the coupling $\mathbb{Q}_{\boldsymbol{\theta}^\star, i, \boldsymbol{x}_{1:i-1}}$, we quantify the dependence among edges by the spectral norm $|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2$ of the upper triangular $M \times M$ coupling matrix $\mathcal{D}_N(\boldsymbol{\theta}^\star)$ with elements

$$\mathcal{D}_{i,j}(\boldsymbol{\theta}^\star) \;\; := \;\; \begin{cases} 0 & \text{if } j < i \\ 1 & \text{if } j = i \\ \displaystyle\max_{\boldsymbol{x}_{1:i-1} \in \{0,1\}^{i-1}} \mathbb{Q}_{\boldsymbol{\theta}^\star, i, \boldsymbol{x}_{1:i-1}}(X^\star_j \neq X^{\star\star}_j) & \text{if } j > i. \end{cases}$$

While the definition of $\mathcal{D}_N(\boldsymbol{\theta}^\star)$ depends on the ordering of edge variables, it is possible to obtain bounds on the spectral norm $|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2$ of $\mathcal{D}_N(\boldsymbol{\theta}^\star)$ that hold for all orderings. We describe in Section 3.3.2 how $|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2$ can be bounded by using coupling methods from percolation theory [39].

To control the smoothness of the sufficient statistics of the model, define

$$\Xi_{i,j} \;\; := \max_{(\boldsymbol{x}, \boldsymbol{x}') \in \mathbb{X} \times \mathbb{X}:\; x_k = x_k' \text{ for all } k \neq j} |s_i(\boldsymbol{x}) - s_i(\boldsymbol{x}')|, \quad j = 1, \dots, M,$$

where $s_1(\boldsymbol{x}), \dots, s_p(\boldsymbol{x})$ are the coordinates of the sufficient statistic vector $s(\boldsymbol{x}) \in \mathbb{R}^p$ defined in (2.4). Let $\boldsymbol{\Xi}_i = (\Xi_{i,1}, \dots, \Xi_{i,M})$ and define

$$\Psi_N \;\; := \;\; \max_{1 \leq i \leq p} \|\boldsymbol{\Xi}_i\|_2.$$

To exclude the trivial case where $\Psi_N = 0$, we assume that there exists an integer $N_0 \geq 3$ such that $\Psi_N > 0$ for all $N > N_0$.

3.2. *Maximum likelihood estimators.* Consider a single observation $\boldsymbol{x}$ of a random graph $\boldsymbol{X}$ with dependent edges. Let $\ell(\boldsymbol{\theta};\, \boldsymbol{x}) := \log f_{\boldsymbol{\theta}}(\boldsymbol{x})$ and

$$\widehat{\boldsymbol{\Theta}} \;\; := \;\; \{\boldsymbol{\theta} \in \mathbb{R}^p : \; \|\nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta};\, \boldsymbol{x})\|_\infty \; = \; 0\,\}\,.$$

We develop a novel approach to establishing consistency results and convergence rates of maximum likelihood estimators for discrete undirected graphical models with exponential parameterizations and parameter vectors of increasing dimension in single-observation scenarios. These results serve as a stepping stone for establishing consistency results and convergence rates of pseudo-likelihood-based M-estimators in Section 3.3.

Let $\mathcal{I}(\boldsymbol{\theta}) := \nabla^2_{\boldsymbol{\theta}}\, \psi(\boldsymbol{\theta}) = \mathbb{C}_{\boldsymbol{\theta}}\, s(\boldsymbol{X}) = -\mathbb{E}_{\boldsymbol{\theta}}\, \nabla^2_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta};\, \boldsymbol{X})$ [26, Theorem 2.7.1, p. 49]. Assume that there exists a constant $\epsilon^\star \in (0, \infty)$, independent of N

and p, such that $\mathcal{I}(\boldsymbol{\theta})$ is invertible for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$. Define

$$\begin{aligned} \Lambda_N(\boldsymbol{\theta}^\star) &\coloneqq \sup_{\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)} \lVert\!| \mathcal{I}(\boldsymbol{\theta})^{-1} |\!\rVert_\infty \\ \Phi_N(\boldsymbol{\theta}^\star) &\coloneqq \Lambda_N(\boldsymbol{\theta}^\star)\, \lVert\!| \mathcal{D}_N(\boldsymbol{\theta}^\star) |\!\rVert_2\, \Psi_N \sqrt{\log \max\{N, p\}}. \end{aligned} \tag{3.1}$$

Theorem 1. *Consider a single observation of a random graph with N nodes and dependent edges. Assume that $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta} = \mathbb{R}^p$, where $p \to \infty$ as $N \to \infty$ is allowed. If $\Phi_N(\boldsymbol{\theta}^\star) \to 0$ as $N \to \infty$, there exists an integer $N_0 \geq 3$ such that, for all $N > N_0$, the random set $\widehat{\boldsymbol{\Theta}}$ is non-empty and its unique element $\widehat{\boldsymbol{\theta}}$ satisfies*

$$\lVert \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star \rVert_\infty \;\leq\; \sqrt{3/2}\; \Phi_N(\boldsymbol{\theta}^\star)$$

with probability at least $1 - 2 \,/\, \max\{N,\, p\}^2$.

While Theorem 1 is stated in terms of random graphs, Theorem 1 covers discrete undirected graphical models with exponential parameterizations and parameter vectors of increasing dimension in single-observation scenarios. Theorem 1 suggests that the convergence rate of maximum likelihood estimators depends on the dimension p of the parameter space $\boldsymbol{\Theta} = \mathbb{R}^p$ and

- the inverse Fisher information matrix in a neighborhood of the data-generating parameter vector $\boldsymbol{\theta}^\star \in \mathbb{R}^p$, quantified by $\Lambda_N(\boldsymbol{\theta}^\star)$;
- the dependence induced by the model, quantified by $\lVert\!| \mathcal{D}_N(\boldsymbol{\theta}^\star) |\!\rVert_2$;
- the sensitivity of sufficient statistics, quantified by Ψ_N.

We highlight the impact of two complex phenomena on the convergence rate: phase transitions and model near-degeneracy [19, 32, 35, 8, 36]. It is known that some random graph models with dependent edges [e.g., the ill-posed edge-and-triangle model, 19, 32, 35, 8, 36] exhibit phase transitions and model near-degeneracy. To examine the impact of phase transitions and model near-degeneracy on the convergence rate, consider a model with a parameter space $\boldsymbol{\Theta} = \mathbb{R}^p$ divided into two or more subsets (regimes) inducing very different distributions, some of which may place almost all mass on a small subset of graphs (e.g., near-empty or near-complete graphs).

Phase transitions. On subsets of $\boldsymbol{\Theta}$ where transitions between such regimes occur, small changes of natural parameters $\boldsymbol{\theta}$ can lead to large changes of mean-value parameters $\boldsymbol{\mu}(\boldsymbol{\theta}) \coloneqq \nabla_{\boldsymbol{\theta}}\, \psi(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}\, s(\boldsymbol{X})$. In such cases, $\mathcal{I}(\boldsymbol{\theta}) \coloneqq \nabla_{\boldsymbol{\theta}}^2\, \psi(\boldsymbol{\theta})$ can become ill-posed and non-invertible, in which case Theorem 1 does not establish consistency.

Model near-degeneracy. On subsets of $\mathbf{\Theta}$ inducing near-degenerate distributions, the variances of sufficient statistics (e.g., the number of edges) can be small, so that the elements on the main diagonal of $\mathcal{I}(\boldsymbol{\theta}) = \mathbb{C}_{\boldsymbol{\theta}}\, s(\boldsymbol{X})$ can be small for some or all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$. In such cases, the convergence rate is reduced via $\Lambda_N(\boldsymbol{\theta}^\star)$. In addition, model near-degeneracy is sometimes associated with strong dependence and high sensitivity of sufficient statistics [35], depressing the convergence rate via $|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2$ and Ψ_N. An example is the ill-posed edge-and-triangle model [19, 32, 35, 8, 36]. We are interested in well-posed models that are amenable to scalable estimation with statistical guarantees. Therefore, the applications in Section 3.4 focus on models that leverage additional structure to control all relevant quantities.

Proof of Theorem 1. Let $\boldsymbol{\mu}(\boldsymbol{\theta}) \coloneqq \mathbb{E}_{\boldsymbol{\theta}}\, s(\boldsymbol{X})$ be the mean-value parameter vector of the exponential family (2.3) parameterized by (2.1) and (2.2). The map $\boldsymbol{\mu} : \mathbf{\Theta} \mapsto \mathbb{M}$ from the natural parameter space $\mathbf{\Theta} \coloneqq \{\boldsymbol{\theta} \in \mathbb{R}^p :\, \psi(\boldsymbol{\theta}) < \infty\} = \mathbb{R}^p$ to the mean-value parameter space $\mathbb{M} \coloneqq \boldsymbol{\mu}(\mathbf{\Theta})$ is a homeomorphism [5, Theorem 3.6, p. 74]. As a result, the inverse map $\boldsymbol{\mu}^{-1} : \mathbb{M} \mapsto \mathbf{\Theta}$ exists and $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^{-1}$ are continuous, one-to-one, and onto.

By assumption, there exists a constant $\epsilon^\star \in (0,\, \infty)$, independent of N and p, such that $\mathcal{I}(\boldsymbol{\theta})$ is invertible for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$. The proof of Theorem 1 will focus on the subset $\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$ of $\mathbf{\Theta}$. We will show in (3.12) that it is legitimate to focus on $\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$, because the event $\widehat{\boldsymbol{\theta}} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$ occurs with probability at least $1 - 2 \,/\, \max\{N,\, p\}^2$ for all large enough N.

Consider any $\epsilon \in (0,\, \epsilon^\star)$ and define

$$\delta(\epsilon) \;\coloneqq\; \sup\left\{\rho \,\in\, (0,\, \infty) :\, \mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, \rho) \;\subseteq\; \boldsymbol{\mu}(\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon))\right\}.$$

By definition of $\delta(\epsilon)$, the hypercube $\mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, \delta(\epsilon))$ is the largest hypercube centered at $\boldsymbol{\mu}(\boldsymbol{\theta}^\star)$ and contained in the image $\boldsymbol{\mu}(\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon))$ of $\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon)$. As a result, $\delta(\epsilon)$ is the greatest real number $\rho \in (0,\, \infty)$ with the property that every element $\boldsymbol{\mu}(\boldsymbol{\theta}) \in \mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, \rho)$ pulls back to an element $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon)$:

$$\delta(\epsilon) \;=\; \sup\left\{\rho \,\in\, (0,\, \infty) :\, \boldsymbol{\mu}^{-1}\left(\mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, \rho)\right) \subseteq \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon)\right\}. \tag{3.2}$$

A direct consequence of (3.2) is the following observation:

$$\begin{gathered}\text{If } \rho \,\in\, (0,\, \infty) \text{ satisfies } \boldsymbol{\mu}^{-1}\left(\mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, \rho)\right) \;\subseteq\; \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon),\\ \text{then } \rho \;\leq\; \delta(\epsilon).\end{gathered} \tag{3.3}$$

The observation (3.3) paves the way for establishing convergence rates for $\widehat{\boldsymbol{\theta}}$ based on $\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}})$, because we will prove that

$$\boldsymbol{\mu}^{-1}\left(\mathcal{B}_\infty\left(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, \frac{\epsilon}{\Lambda_N(\boldsymbol{\theta}^\star)}\right)\right) \;\subseteq\; \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon), \tag{3.4}$$

which will establish the following fundamental relation between ϵ and $\delta(\epsilon)$:

$$\frac{\epsilon}{\Lambda_N(\boldsymbol{\theta}^\star)} \quad\leq\quad \delta(\epsilon). \tag{3.5}$$

To prove (3.4) and (3.5) and establish convergence rates for $\widehat{\boldsymbol{\theta}}$ based on $\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}})$, note that $\mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon) \subset \boldsymbol{\Theta} = \mathbb{R}^p$ implies $\boldsymbol{\mu}(\mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon)) \subset \mathbb{M}$ and that the random set $\widehat{\boldsymbol{\Theta}}$ is non-empty in the event $s(\boldsymbol{X}) \in \boldsymbol{\mu}(\mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon)) \subset \mathbb{M}$, with its unique element $\widehat{\boldsymbol{\theta}} \in \widehat{\boldsymbol{\Theta}}$ solving $\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}}) = s(\boldsymbol{X})$ [5, Theorem 5.5, p. 148]. We can thus bound the probability of event $\widehat{\boldsymbol{\theta}} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon)$ as follows:

$$\begin{aligned} \mathbb{P}(\widehat{\boldsymbol{\theta}} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon)) &\;=\; \mathbb{P}(\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}}) \in \boldsymbol{\mu}(\mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon)))\\ &\;\geq\; \mathbb{P}(\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}}) \in \mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star), \delta(\epsilon)))\\ &\;\geq\; 1 - 2\,\exp\left(-\frac{2\,\delta(\epsilon)^2}{|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2^2\;\Psi_N^2} + \log p\right), \end{aligned} \tag{3.6}$$

where the first line follows from the fact that $\boldsymbol{\mu} : \boldsymbol{\Theta} \mapsto \mathbb{M}$ is a homeomorphism and $\widehat{\boldsymbol{\theta}}$ exists, is unique, and solves $\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}}) = s(\boldsymbol{X})$ in the event $s(\boldsymbol{X}) \in \boldsymbol{\mu}(\mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon)) \subset \mathbb{M}$ [5, Theorem 5.5, p. 148], the second line follows from $\mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star), \delta(\epsilon)) \subseteq \boldsymbol{\mu}(\mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon))$, and the third line follows from $\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}}) = s(\boldsymbol{X})$ along with Lemma 1 in the supplement [38].

We bound the right-hand side of (3.6) by bounding $\delta(\epsilon)$, taking advantage of observation (3.3). Consider any $\rho \leq \delta(\epsilon)$ and any $\boldsymbol{\mu}' \in \mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star), \rho)$. The assumption $\rho \leq \delta(\epsilon)$, together with the definition of $\delta(\epsilon)$, implies that

$$\boldsymbol{\mu}' \;\in\; \mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star), \rho) \;\subseteq\; \mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star), \delta(\epsilon)) \;\subseteq\; \boldsymbol{\mu}(\mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon)).$$

Therefore, the natural parameter vector $\boldsymbol{\theta}' := \boldsymbol{\mu}^{-1}(\boldsymbol{\mu}')$ corresponding to the mean-value parameter vector $\boldsymbol{\mu}'$ falls into $\mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon)$:

$$\boldsymbol{\theta}' \;\in\; \boldsymbol{\mu}^{-1}(\mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star), \rho)) \;\subseteq\; \boldsymbol{\mu}^{-1}(\mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star), \delta(\epsilon))) \;\subseteq\; \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon).$$

In addition, $\boldsymbol{\theta}'$ falls into $\mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$, because $\epsilon \in (0, \epsilon^\star)$:

$$\boldsymbol{\theta}' \;\in\; \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon) \;\subset\; \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star).$$

By the multivariate mean-value theorem [16, Theorem 5], there exists

$$\boldsymbol{\theta}'' \;:=\; \lambda\,\boldsymbol{\theta}^\star + (1-\lambda)\,\boldsymbol{\theta}' \;\in\; \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star), \quad \lambda \in (0,1)$$

such that

$$\boldsymbol{\mu}(\boldsymbol{\theta}') - \boldsymbol{\mu}(\boldsymbol{\theta}^\star) \;=\; \mathcal{I}(\boldsymbol{\theta}'')\,(\boldsymbol{\theta}' - \boldsymbol{\theta}^\star).$$

By assumption, $\mathcal{I}(\boldsymbol{\theta})$ is invertible for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$, which includes $\boldsymbol{\theta}'' \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$. As a result, $\boldsymbol{\theta}' - \boldsymbol{\theta}^\star = \mathcal{I}(\boldsymbol{\theta}'')^{-1}\,(\boldsymbol{\mu}(\boldsymbol{\theta}') - \boldsymbol{\mu}(\boldsymbol{\theta}^\star))$ and

$$\|\boldsymbol{\theta}' - \boldsymbol{\theta}^\star\|_\infty \;\leq\; |||\mathcal{I}(\boldsymbol{\theta}'')^{-1}|||_\infty\, \|\boldsymbol{\mu}(\boldsymbol{\theta}') - \boldsymbol{\mu}(\boldsymbol{\theta}^\star)\|_\infty \;<\; \rho\; |||\mathcal{I}(\boldsymbol{\theta}'')^{-1}|||_\infty, \tag{3.7}$$

recalling that $\|\boldsymbol{\mu}(\boldsymbol{\theta}') - \boldsymbol{\mu}(\boldsymbol{\theta}^\star)\|_\infty < \rho$ for all $\boldsymbol{\mu}(\boldsymbol{\theta}') \in \mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, \rho)$; note that $\boldsymbol{\mu}(\boldsymbol{\theta}')$ is identical to $\boldsymbol{\mu}' \in \mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, \rho)$. As a consequence, all $\rho \in (0, \infty)$ satisfying $\rho \leq \epsilon \,/\, |||\mathcal{I}(\boldsymbol{\theta}'')^{-1}|||_\infty$ guarantee that $\|\boldsymbol{\theta}' - \boldsymbol{\theta}^\star\|_\infty < \epsilon$. Choosing $\rho \coloneqq \epsilon \,/\, \Lambda_N(\boldsymbol{\theta}^\star)$ with $\Lambda_N(\boldsymbol{\theta}^\star) \coloneqq \sup_{\boldsymbol{\theta} \,\in\, \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)} |||\mathcal{I}(\boldsymbol{\theta})^{-1}|||_\infty$ and applying (3.7) with the chosen ρ establishes the fundamental relation

$$\boldsymbol{\mu}^{-1}\left(\mathcal{B}_\infty\left(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, \frac{\epsilon}{\Lambda_N(\boldsymbol{\theta}^\star)}\right)\right) \;\subseteq\; \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon). \tag{3.8}$$

The fundamental relation (3.8), combined with the observation (3.3), implies that ϵ is related to $\delta(\epsilon)$ as follows:

$$\frac{\epsilon}{\Lambda_N(\boldsymbol{\theta}^\star)} \;\leq\; \delta(\epsilon). \tag{3.9}$$

Armed with inequality (3.9), we revisit (3.6) to obtain

$$\begin{aligned}\mathbb{P}(\widehat{\boldsymbol{\theta}} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon)) \;&\geq\; 1 - 2\, \exp\left(-\frac{2\,\delta(\epsilon)^2}{|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2^2\, \Psi_N^2} + \log\, p\right)\\ &\geq\; 1 - 2\, \exp\left(-\frac{2\,\epsilon^2}{\Lambda_N(\boldsymbol{\theta}^\star)^2\, |||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2^2\, \Psi_N^2} + \log\, p\right).\end{aligned}$$

Choosing $\epsilon \coloneqq \sqrt{3/2}\;\, \Phi_N(\boldsymbol{\theta}^\star)$ establishes

$$\mathbb{P}(\widehat{\boldsymbol{\theta}} \,\in\, \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \sqrt{3/2}\; \Phi_N(\boldsymbol{\theta}^\star))) \;\geq\; 1 - \frac{2}{\max\{N,\, p\}^2}. \tag{3.10}$$

To complete the proof, we show that the focus on the subset $\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$ of $\boldsymbol{\Theta}$ is legitimate, by noting that the assumption $\Phi_N(\boldsymbol{\theta}^\star) \to 0$ as $N \to \infty$ implies that there exists an integer $N_0 \geq 3$ such that $\epsilon \coloneqq \sqrt{3/2}\;\, \Phi_N(\boldsymbol{\theta}^\star) < \epsilon^\star$ for all $N > N_0$. The fact that $\sqrt{3/2}\; \Phi_N(\boldsymbol{\theta}^\star) < \epsilon^\star$ for all $N > N_0$ implies

$$\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \sqrt{3/2}\; \Phi_N(\boldsymbol{\theta}^\star)) \;\subset\; \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star) \quad \text{for all} \quad N > N_0. \tag{3.11}$$

The relation (3.11), along with the lower bound on the probability of event $\widehat{\boldsymbol{\theta}} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \sqrt{3/2}\; \Phi_N(\boldsymbol{\theta}^\star))$ in (3.10), shows that

$$\begin{aligned}\mathbb{P}(\widehat{\boldsymbol{\theta}} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)) \;&\geq\; \mathbb{P}(\widehat{\boldsymbol{\theta}} \,\in\, \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \sqrt{3/2}\; \Phi_N(\boldsymbol{\theta}^\star)))\\ &\geq\; 1 - \frac{2}{\max\{N,\, p\}^2} \quad \text{for all} \quad N > N_0.\end{aligned} \tag{3.12}$$

We conclude that, for all $N > N_0$, the random set $\widehat{\mathbf{\Theta}}$ is non-empty and its unique element $\widehat{\boldsymbol{\theta}}$ satisfies

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star\|_\infty \quad \leq \quad \sqrt{3/2}\ \Phi_N(\boldsymbol{\theta}^\star)$$

with probability at least $1 - 2\,/\max\{N,\, p\}^2$. □

3.3. *Pseudo-likelihood-based M-estimators.* Maximum likelihood estimators are unappealing on computational grounds, because evaluating $\ell(\boldsymbol{\theta};\, \boldsymbol{x})$ requires evaluating the normalizing constant of $f_{\boldsymbol{\theta}}(\boldsymbol{x})$. The normalizing constant of $f_{\boldsymbol{\theta}}(\boldsymbol{x})$ is a sum over $\exp(M \log 2)$ possible graphs and cannot be computed unless $M \coloneqq \binom{N}{2}$ is small or the model makes restrictive independence assumptions. As a scalable alternative, consider M-estimators

$$\widetilde{\mathbf{\Theta}}(\gamma_N) \quad \coloneqq \quad \left\{\boldsymbol{\theta} \in \mathbb{R}^p :\ \|\nabla_{\boldsymbol{\theta}}\, \widetilde{\ell}(\boldsymbol{\theta};\, \boldsymbol{x})\|_\infty \ \leq\ \gamma_N\right\}, \qquad \gamma_N \in [0,\, \infty)$$

based on the pseudo-loglikelihood function

$$\widetilde{\ell}(\boldsymbol{\theta};\, \boldsymbol{x}) \quad \coloneqq \quad \log \prod_{i=1}^{M} f_{\boldsymbol{\theta}}(x_i \mid \boldsymbol{x}_{-i}),$$

where $f_{\boldsymbol{\theta}}(x_i \mid \boldsymbol{x}_{-i})$ is the conditional probability of $X_i = x_i$ given all other edge variables $\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i}$ $(i = 1, \dots, M)$.

To bound the statistical error of pseudo-likelihood-based M-estimators in single-observation scenarios with $p \to \infty$ parameters, let $i \in \{1, \dots, M\}$ and $\mathfrak{N}_i \ \subseteq\ \{1, \dots, M\} \setminus \{i\}$ be the smallest subset of $\{1, \dots, M\} \setminus \{i\}$ such that

$$X_i \quad \perp\!\!\!\perp \quad \boldsymbol{X}_{\{1,\dots,M\}\,\setminus\,(\{i\}\,\cup\,\mathfrak{N}_i)} \mid \boldsymbol{X}_{\mathfrak{N}_i}.$$

Let $\epsilon^\star \in (0,\, \infty)$ be a constant, independent of N and p, and assume that $-\mathbb{E}\, \nabla^2_{\boldsymbol{\theta}}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$ is invertible for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$. Define

$$\widetilde{\Lambda}_N(\boldsymbol{\theta}^\star) \quad \coloneqq \quad \sup_{\boldsymbol{\theta}\, \in\, \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)} \ \vert\!\vert\!\vert(-\mathbb{E}\, \nabla^2_{\boldsymbol{\theta}}\ \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}))^{-1}\vert\!\vert\!\vert_\infty$$

$$\widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \quad \coloneqq \quad \widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)\ (1 + D_N)\ \vert\!\vert\!\vert\mathcal{D}_N(\boldsymbol{\theta}^\star)\vert\!\vert\!\vert_2\ \Psi_N\ \sqrt{\log \max\{N,\, p\}},$$

where $D_N \coloneqq \max\{|\mathfrak{N}_1|, \dots, |\mathfrak{N}_M|\} \in \{0, \dots, M-1\}$.

We can then bound the statistical error of pseudo-likelihood-based M-estimators in single-observation scenarios with $p \to \infty$ parameters as follows.

Theorem 2. *Consider a single observation of a random graph with N nodes and dependent edges. Assume that $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta} = \mathbb{R}^p$, where $p \to \infty$ as $N \to \infty$ is allowed. If $\widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \to 0$ as $N \to \infty$, there exists an integer $N_0 \geq 3$ such that, for all $N > N_0$, the random set $\widetilde{\boldsymbol{\Theta}}(\gamma_N)$ is non-empty and any element $\widetilde{\boldsymbol{\theta}}$ of $\widetilde{\boldsymbol{\Theta}}(\gamma_N)$ satisfies*

$$\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star\|_\infty \;\; \leq \;\; \sqrt{96}\; \widetilde{\Phi}_N(\boldsymbol{\theta}^\star)$$

with probability at least $1 - 2\,/\max\{N,\, p\}^2$, provided

$$\gamma_N \;\; = \;\; \sqrt{24}\;\, (1 + D_N)\; \|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2\; \Psi_N\; \sqrt{\log \max\{N,\, p\}}.$$

A proof of Theorem 2 is provided in the supplement [38]. While stated in terms of random graphs, Theorems 1 and 2 cover discrete undirected graphical models with exponential parameterizations and parameter vectors of increasing dimension in single-observation scenarios. As a result, Theorems 1 and 2 have possible applications to discrete and dependent network, spatial, and temporal data.

We first provide a simple application of Theorems 1 and 2 in Section 3.3.1 and explore how fast the dimension p of the parameter space $\boldsymbol{\Theta} = \mathbb{R}^p$ can grow as a function of the number of nodes N. We then explain in Sections 3.3.2, 3.3.3, and 3.3.4 how $\widetilde{\Phi}_N(\boldsymbol{\theta}^\star)$ can be bounded. Applications to generalized β-models with dependent edges and $p \geq N \to \infty$ parameters are presented in Section 3.4. These applications demonstrate that $\widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \to 0$ as $N \to \infty$ provided D_N does not grow too fast. We conclude Section 3.4 with a comparison with related statistical exponential-family models for discrete and dependent random variables in single-observation scenarios.

3.3.1. *Example: growth of p as a function of N.* To showcase Theorems 1 and 2 in one of the simplest possible scenarios and explore how fast the dimension p of the parameter space $\boldsymbol{\Theta} = \mathbb{R}^p$ can grow as a function of N, we consider inhomogeneous Bernoulli random graphs in the dense-graph regime. Inhomogeneous Bernoulli random graphs assume that edge variables X_i are independent Bernoulli(μ_i) random variables, with edge probabilities $\mu_i := \mathbb{E}\, X_i$ satisfying $0 < C_1 < \mu_i < C_2 < 1$ for finite constants C_1 and C_2, independent of N. Suppose that each edge variable X_i belongs to one of $p \leq M$ distinct categories $k \in \{1, \ldots, p\}$ with edge probabilities $\pi_k \in (0,\, 1)$, and that $\mu_i = \pi_k$ if edge variable X_i is assigned to category k. Inhomogeneous Bernoulli random graphs are statistical exponential families with natural parameters $\theta_k := \text{logit}(\pi_k)$ and sufficient statistics $s_k(\boldsymbol{x}) := \sum_{i=1}^M \mathbb{1}_k(i)\, x_i$ $(k = 1, \ldots, p)$, where $\boldsymbol{\theta} := (\theta_1, \ldots, \theta_p) \in \mathbb{R}^p$, $s(\boldsymbol{x}) := (s_1(\boldsymbol{x}), \ldots, s_p(\boldsymbol{x})) \in \mathbb{R}^p$, and $\mathbb{1}_k(i)$ is 1 if edge variable X_i is assigned to category k and is 0 otherwise. Since edges are independent, the pseudo-loglikelihood function reduces to the loglikelihood and its negative expected Hessian is $\mathcal{I}(\boldsymbol{\theta}) = \mathbb{C}_{\boldsymbol{\theta}}\, s(\boldsymbol{X})$. By the independence of edges, $\mathbb{C}_{\boldsymbol{\theta}}\, s(\boldsymbol{X})$ is a diagonal matrix, so the variances $\mathbb{V}_{\boldsymbol{\theta}}\, s_1(\boldsymbol{X}), \ldots, \mathbb{V}_{\boldsymbol{\theta}}\, s_p(\boldsymbol{X})$ are the eigenvalues of $\mathbb{C}_{\boldsymbol{\theta}}\, s(\boldsymbol{X})$. To bound them, assume that there exist finite constants $0 < C_3 < C_4$ such that

$$\frac{C_3\, N^2}{p} \;\leq\; \sum_{i=1}^{M} \mathbb{1}_k(i) \;\leq\; \frac{C_4\, N^2}{p}, \qquad k = 1, \ldots, p,$$

that is, the p categories are balanced, in the sense that the sizes of the p categories are of the same order of magnitude. Then there exists a finite constant $C_5 > 0$, independent of N and p, such that

$$\begin{aligned} \Lambda_N(\boldsymbol{\theta}^\star) \;&=\; \widetilde{\Lambda}_N(\boldsymbol{\theta}^\star) \;=\; \sup_{\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)} \; |||\mathcal{I}(\boldsymbol{\theta})^{-1}|||_\infty \\ &=\; \sup_{\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)} \; \max_{1 \leq k \leq p} \frac{1}{\mathbb{V}_{\boldsymbol{\theta}}\, s_k(\boldsymbol{X})} \;\leq\; \frac{C_5\; p}{N^2}. \end{aligned}$$

By the independence of edges, $D_N = 0$ and the coupling matrix $\mathcal{D}_N(\boldsymbol{\theta}^\star)$ is the $M \times M$ identity matrix with spectral norm $|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2 = 1$. The quantity $\Psi_N \coloneqq \max_{1 \leq k \leq p} |||\Xi_k|||_2$ can be bounded as follows. First, adding or deleting an edge in any category k can change the number of edges $s_k(\boldsymbol{x})$ in category k by -1 or $+1$, while changes of edges in other categories leave $s_k(\boldsymbol{x})$ unchanged. Second, each category k contains at most $C_4\, N^2/p$ edges, so $|||\Xi_k|||_2 \leq \sqrt{C_4\, N^2/p}$ for all k and hence $\Psi_N \leq \sqrt{C_4\, N^2/p}$. Thus, there exists a finite constant $C > 0$, independent of N and p, such that

$$\Phi_N(\boldsymbol{\theta}^\star) \;=\; \widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \;\leq\; \frac{\sqrt{p\, \log \max\{N,\, p\}}}{C\, N}.$$

If $p = o(N^2/\log N)$, then $\Phi_N(\boldsymbol{\theta}^\star) = \widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \to 0$ and the maximum likelihood and pseudo-likelihood estimators $\widehat{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\theta}}$ are consistent estimators of $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta} = \mathbb{R}^p$ by Theorem 1; note that $\widehat{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\theta}}$ are equal with probability 1 when edges are independent. Thus, Theorems 1 and 2 confirm the intuition that the number of parameters p we can estimate (without assuming $\boldsymbol{\theta}^\star$ to be sparse) is less than N^2 (ignoring logarithmic terms). These results dovetail with the results of Portnoy [30, Theorem 2.1] based on $n \to \infty$ independent observations from a statistical exponential family with $p \to \infty$ parameters, which suggest that consistency results can be obtained as long as $p = o(n)$; note that the number of independent observations under inhomogeneous Bernoulli random graphs is $n = \binom{N}{2}$. While the example is limited to inhomogeneous Bernoulli random graphs, we conjecture that p can grow at most as fast when edges are dependent and the random graph is sparse, because dependence increases $|\!|\!|\mathcal{D}_N(\boldsymbol{\theta}^\star)|\!|\!|_2$ while sparsity decreases information $\mathcal{I}(\boldsymbol{\theta})$ and hence increases $\Lambda_N(\boldsymbol{\theta}^\star)$.

3.3.2. *Bounding the spectral norm of the coupling matrix.* If edges are independent, the spectral norm $|\!|\!|\mathcal{D}_N(\boldsymbol{\theta}^\star)|\!|\!|_2$ of the coupling matrix $\mathcal{D}_N(\boldsymbol{\theta}^\star)$ is 1, otherwise $|\!|\!|\mathcal{D}_N(\boldsymbol{\theta}^\star)|\!|\!|_2$ needs to be bounded from above. We transform the hard problem of bounding $|\!|\!|\mathcal{D}_N(\boldsymbol{\theta}^\star)|\!|\!|_2$ into the more convenient problem of studying paths in a conditional independence graph $\mathcal{G}$ that represents the conditional independence structure of a random graph [15, 25]. A conditional independence graph $\mathcal{G}$ consists of a set of vertices $\mathcal{V} \coloneqq \{X_1, \dots, X_M\}$ and a set of undirected edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ indicating the absence of conditional independencies among edge variables $X_1, \dots, X_M$ [see, e.g., 15, 25].

We begin with the observation that the concentration results of Chazottes et al. [10] leveraged in Theorems 1 and 2 hold for all possible couplings $\mathbb{Q}_{\boldsymbol{\theta}^\star,i,\boldsymbol{x}_{1:i-1}}$ of $\mathbb{P}_{\boldsymbol{\theta}^\star,\boldsymbol{x}_{1:i-1},0}$ and $\mathbb{P}_{\boldsymbol{\theta}^\star,\boldsymbol{x}_{1:i-1},1}$, and all possible couplings bound the total variation distance between $\mathbb{P}_{\boldsymbol{\theta}^\star,\boldsymbol{x}_{1:i-1},0}$ and $\mathbb{P}_{\boldsymbol{\theta}^\star,\boldsymbol{x}_{1:i-1},1}$:

$$\begin{aligned}\|\mathbb{P}_{\boldsymbol{\theta}^\star,\boldsymbol{x}_{1:i-1},0} - \mathbb{P}_{\boldsymbol{\theta}^\star,\boldsymbol{x}_{1:i-1},1}\|_{\text{TV}} \;&\leq\; \mathbb{Q}_{\boldsymbol{\theta}^\star,i,\boldsymbol{x}_{1:i-1}}(\boldsymbol{X}^\star_{i+1:M} \neq \boldsymbol{X}^{\star\star}_{i+1:M})\\ =\; \mathbb{Q}_{\boldsymbol{\theta}^\star,i,\boldsymbol{x}_{1:i-1}}\left(\bigcup_{j=i+1}^{M} \{X_j^\star \neq X_j^{\star\star}\}\right) \;&\leq\; \sum_{j=i+1}^{M} \mathcal{D}_{i,j}(\boldsymbol{\theta}^\star).\end{aligned}$$

We can therefore replace optimal couplings (which provide the tightest bounds on the total variation distance) by suboptimal but more convenient couplings that facilitate bounds on the spectral norm $|\!|\!|\mathcal{D}_N(\boldsymbol{\theta}^\star)|\!|\!|_2$ of $\mathcal{D}_N(\boldsymbol{\theta}^\star)$. To do so, we adapt the coupling approach of van den Berg and Maes [39, pp. 759–760] from Markov random fields to random graphs. The resulting coupling $\mathbb{Q}_{\boldsymbol{\theta}^\star,i,\boldsymbol{x}_{1:i-1}}$ is described in Lemma 12 in the supplement [38] and may not be optimal, but it helps translate the hard problem of bounding the spectral norm $|\!|\!|\mathcal{D}_N(\boldsymbol{\theta}^\star)|\!|\!|_2$ of $\mathcal{D}_N(\boldsymbol{\theta}^\star)$ into the more convenient problem of studying paths in the conditional independence graph $\mathcal{G}$.

We start with the inequality

$$|\!|\!|\mathcal{D}_N(\boldsymbol{\theta}^\star)|\!|\!|_2 \;\;\leq\;\; \sqrt{|\!|\!|\mathcal{D}_N(\boldsymbol{\theta}^\star)|\!|\!|_1 \; |\!|\!|\mathcal{D}_N(\boldsymbol{\theta}^\star)|\!|\!|_\infty}.$$

We then bound the quantities $|\!|\!|\mathcal{D}_N(\boldsymbol{\theta}^\star)|\!|\!|_1$ and $|\!|\!|\mathcal{D}_N(\boldsymbol{\theta}^\star)|\!|\!|_\infty$ by bounding the above-diagonal elements $\mathcal{D}_{i,j}(\boldsymbol{\theta}^\star)$ of $\mathcal{D}_N(\boldsymbol{\theta}^\star)$, using paths of disagreement $i \not\longleftrightarrow j$ between vertices X_i and X_j in the conditional independence graph $\mathcal{G}$; note that the below-diagonal and diagonal elements of $\mathcal{D}_N(\boldsymbol{\theta}^\star)$ are 0 and

1. A path of disagreement $i \not\leftrightarrow j$ between vertices X_i and X_j is a sequence of two or more distinct vertices $(X_i, \ldots, X_j)$ in the conditional independence graph $\mathcal{G}$ starting at vertex X_i and ending at vertex X_j, such that

- each subsequent pair of vertices (X_v, X_w) in the sequence is connected by an edge in the conditional independence graph $\mathcal{G}$, which indicates the absence of conditional independence of vertices X_v and X_w;
- the coupling $(\boldsymbol{X}_{i+1:M}^{\star}, \boldsymbol{X}_{i+1:M}^{\star\star}) \in \{0,1\}^{M-i} \times \{0,1\}^{M-i}$ with joint probability mass function $\mathbb{Q}_{\boldsymbol{\theta}^\star, i, \boldsymbol{x}_{1:i-1}}$ disagrees at each vertex X_v in the sequence, in the sense that $X_v^\star \neq X_v^{\star\star}$.

Theorem 1 of van den Berg and Maes [39] implies that the coupling $\mathbb{Q}_{\boldsymbol{\theta}^\star, i, \boldsymbol{x}_{1:i-1}}$ constructed in the supplement [38] satisfies

$$\mathbb{Q}_{\boldsymbol{\theta}^\star, i, \boldsymbol{x}_{1:i-1}}(X_j^\star \neq X_j^{\star\star}) = \mathbb{Q}_{\boldsymbol{\theta}^\star, i, \boldsymbol{x}_{1:i-1}}(i \not\leftrightarrow j) \leq \mathbb{B}_{\boldsymbol{\pi}(\boldsymbol{\theta}^\star)}(i \not\leftrightarrow j), \tag{3.13}$$

where $\mathbb{B}_{\boldsymbol{\pi}(\boldsymbol{\theta}^\star)}$ is a Bernoulli product measure on $\{0,1\}^M$ with probability vector $\boldsymbol{\pi}(\boldsymbol{\theta}^\star) \in [0,1]^M$. The coordinates $\pi_v(\boldsymbol{\theta}^\star)$ of $\boldsymbol{\pi}(\boldsymbol{\theta}^\star)$ are given by

$$\pi_v(\boldsymbol{\theta}^\star) := \begin{cases} 0 & \text{if } v \leq i-1 \\ 1 & \text{if } v = i \\ \max\limits_{(\boldsymbol{x}_{-v}, \boldsymbol{x}'_{-v}) \in \{0,1\}^{M-1} \times \{0,1\}^{M-1}} \pi_{v, \boldsymbol{x}_{-v}, \boldsymbol{x}'_{-v}}(\boldsymbol{\theta}^\star) & \text{if } v \geq i+1, \end{cases}$$

where

$$\pi_{v, \boldsymbol{x}_{-v}, \boldsymbol{x}'_{-v}}(\boldsymbol{\theta}^\star) \quad := \quad \|\mathbb{P}_{\boldsymbol{\theta}^\star}(\,\cdot \mid \boldsymbol{X}_{-v} = \boldsymbol{x}_{-v}) - \mathbb{P}_{\boldsymbol{\theta}^\star}(\,\cdot \mid \boldsymbol{X}_{-v} = \boldsymbol{x}'_{-v})\|_{\mathrm{TV}}$$

is the total variation distance between the conditional probability mass functions of vertex X_v given $\boldsymbol{X}_{-v} = \boldsymbol{x}_{-v}$ and $\boldsymbol{X}_{-v} = \boldsymbol{x}'_{-v}$. Leveraging (3.13), we can bound the above-diagonal elements $\mathcal{D}_{i,j}(\boldsymbol{\theta}^\star)$ of $\mathcal{D}_N(\boldsymbol{\theta}^\star)$ as follows:

$$\mathcal{D}_{i,j}(\boldsymbol{\theta}^\star) \quad := \quad \max_{\boldsymbol{x}_{1:i-1} \in \{0,1\}^{i-1}} \mathbb{Q}_{\boldsymbol{\theta}^\star, i, \boldsymbol{x}_{1:i-1}}(X_j^\star \neq X_j^{\star\star}) \quad \leq \quad \mathbb{B}_{\boldsymbol{\pi}(\boldsymbol{\theta}^\star)}(i \not\leftrightarrow j).$$

In other words, the spectral norm $\|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2$ of $\mathcal{D}_N(\boldsymbol{\theta}^\star)$ can be bounded by using paths of disagreement $i \not\leftrightarrow j$ in the conditional independence graph $\mathcal{G}$, and by bounding the probabilities of those paths by Bernoulli product measures. Specific bounds depend on the data-generating model with parameter vector $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta} = \mathbb{R}^p$. Applications to generalized β-models with dependent edges can be found in the supplement [38].

3.3.3. *Bounding the ℓ_∞-norm of inverse negative expected Hessians.* To establish convergence rates, $\Lambda_N(\boldsymbol{\theta}^\star)$ and $\widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)$ need to be bounded, which amounts to bounding the suprema of $|||(-\mathbb{E}\,\nabla_{\boldsymbol{\theta}}^2\,\ell(\boldsymbol{\theta};\boldsymbol{X}))^{-1}|||_\infty$ and $|||(-\mathbb{E}\,\nabla_{\boldsymbol{\theta}}^2\,\widetilde{\ell}(\boldsymbol{\theta};\boldsymbol{X}))^{-1}|||_\infty$ on $\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\,\epsilon^\star)\subset\boldsymbol{\Theta}=\mathbb{R}^p$.

In general, bounds on the ℓ_∞-induced matrix norm of inverse matrices are non-trivial. Standard matrix norm inequalities reveal that

$$\begin{aligned}|||(-\mathbb{E}\,\nabla_{\boldsymbol{\theta}}^2\,\widetilde{\ell}(\boldsymbol{\theta};\boldsymbol{X}))^{-1}|||_\infty &\leq \sqrt{p}\;|||(-\mathbb{E}\,\nabla_{\boldsymbol{\theta}}^2\,\widetilde{\ell}(\boldsymbol{\theta};\boldsymbol{X}))^{-1}|||_2\\ &= \frac{\sqrt{p}}{\lambda_{\min}(-\mathbb{E}\,\nabla_{\boldsymbol{\theta}}^2\,\widetilde{\ell}(\boldsymbol{\theta};\boldsymbol{X}))},\end{aligned}\tag{3.14}$$

where $\lambda_{\min}(-\mathbb{E}\,\nabla_{\boldsymbol{\theta}}^2\,\widetilde{\ell}(\boldsymbol{\theta};\boldsymbol{X}))>0$ is the smallest eigenvalue of $-\mathbb{E}\,\nabla_{\boldsymbol{\theta}}^2\,\widetilde{\ell}(\boldsymbol{\theta};\boldsymbol{X})$. That said, bounds of $\widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)$ based on (3.14) may be loose when $p\to\infty$ as $N\to\infty$, as is the case with generalized β-models with dependent edges.

To establish bounds on $\widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)$ in scenarios with $p\to\infty$ parameters, we leverage the fact that generalized β-models with dependent edges and $p=N+1\to\infty$ parameters include the β-model with independent edges and $p=N\to\infty$ parameters as a special case, along with the fact that the negative expected Hessian of the β-model is diagonally dominant in the sense of Hillar and Wibisono [21]. By leveraging these properties, Lemma 6 in the supplement [38] establishes the bound $\widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)\leq C\,D_N^9\,/\,N^{1-(\alpha+\vartheta)}$, where the constants $C\in(0,\,\infty)$, $\alpha\in[0,\,1/2)$, and $\vartheta\in[0,\,1/2-\alpha)$ are independent of N and p, while D_N satisfies $D_N=O(\log N)$.

3.3.4. *Bounding the smoothness of the sufficient statistics.* The quantity $\Psi_N\coloneqq\max_{1\leq i\leq p}\,\|\boldsymbol{\Xi}_i\|_2$ can be bounded by bounding the coordinates $\Xi_{i,j}$ of $\boldsymbol{\Xi}_i$. Bounding $\Xi_{i,j}$ amounts to bounding changes of sufficient statistics.

3.4. *Applications.* We present applications of pseudo-likelihood-based M-estimators to β- and generalized β-models with dependent edges and $p\geq N\to\infty$ parameters, in dense- and sparse-graph settings. Throughout, we assume that the data-generating parameter vector $\boldsymbol{\theta}^\star\in\boldsymbol{\Theta}=\mathbb{R}^p$ satisfies

$$\|\boldsymbol{\theta}^\star\|_\infty \;\leq\; \frac{L+\vartheta\,\log N}{14\,(3+D_N)}-\epsilon^\star,\tag{3.15}$$

where $L\in[0,\,\infty)$, $\vartheta\in[0,\,\infty)$, and $\epsilon^\star\in(0,\,\infty)$ are constants, independent of N and p. The constant $\epsilon^\star\in(0,\,\infty)$ is identical to the constant $\epsilon^\star$ in the definition of $\widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)$ and Theorem 2. The quantity $D_N\coloneqq\max\{|\mathfrak{N}_1|,\ldots,|\mathfrak{N}_M|\}$ is identical to the quantity D_N in the definition of $\widetilde{\Phi}_N(\boldsymbol{\theta}^\star)$ and satisfies

$D_N = 0$ under Model 1, but can increase as a function of N under Models 2 and 3. To ensure that $\|\boldsymbol{\theta}^\star\|_\infty > 0$, we assume that D_N satisfies

$$1 \;\; \leq \;\; D_N \;\; < \;\; \frac{L + \vartheta \log N}{14\, \epsilon^\star} - 3$$

under Models 2 and 3.

We start with the β-model [9], because its theoretical properties have been studied and it is therefore a convenient benchmark.

Corollary 1. **β-model.** *Consider Model 1 with $\boldsymbol{\theta}^\star \in \mathbb{R}^N$ satisfying (3.15) with $\vartheta \in [0,\, 7/2)$. Then there exist finite constants $C > 0$ and $N_0 \geq 3$, independent of N and p, such that, for all $N > N_0$,*

$$\Phi_N(\boldsymbol{\theta}^\star) \;\; = \;\; \widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \;\; \leq \;\; C \sqrt{\frac{\log N}{N^{1 - 2\,\vartheta/7}}}.$$

Corollary 1 shows that the convergence rate is highest when $\|\boldsymbol{\theta}^\star\|_\infty$ is bounded above ($\vartheta = 0$). Condition (3.15) is the weakest known condition on $\|\boldsymbol{\theta}^\star\|_\infty$: Chatterjee et al. [9, Theorem 1.3] report a non-asymptotic error bound of the form $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star\|_\infty \leq C \sqrt{\log N \,/\, N}$ assuming that $\|\boldsymbol{\theta}^\star\|_\infty$ is bounded above ($\vartheta = 0$), while Yan and Xu [42, Theorem 1] report asymptotic consistency and normality results assuming that $\|\boldsymbol{\theta}^\star\|_\infty = o(\log\log N)$. By contrast, condition (3.15) assumes that $\|\boldsymbol{\theta}^\star\|_\infty \; < \; (1/12)\log N$ ($\boldsymbol{\theta}^\star \in \mathbb{R}^N$, $L = 0$, $\vartheta < 7/2$, $D_N = 0$), which dovetails with the condition $\|\boldsymbol{\theta}^\star\|_\infty < (1/24)\log N$ ($\boldsymbol{\theta}^\star \in \mathbb{R}^{2\,N-1}$) of Yan et al. [41, Theorem 1] based on the p_1-model for directed random graphs; note that the β-model for undirected random graphs can be viewed as a relative of the p_1-model for directed random graphs, because both models are statistical exponential-family models of degree sequences. These results, along with the results on the dimension p of the parameter space $\boldsymbol{\Theta} = \mathbb{R}^p$ in Section 3.3.1, demonstrate that Theorems 1 and 2 recover the sharpest known results for random graphs with independent edges and $p \to \infty$ parameters, suggesting that the generality of Theorems 1 and 2 comes at a low cost. It is worth noting that it is unknown whether the constants mentioned above are sharp. While it would be of interest to investigate whether these constants are sharp, constants do not affect convergence rates and the question of whether these constants are sharp is therefore not pertinent to the main results of the paper.

To demonstrate that Theorem 2 covers random graph models with non-trivial dependence, we turn to generalized β-models with dependent edges. Throughout, we assume that the size $|\mathcal{A}_k|$ of each subpopulation $\mathcal{A}_k$ satisfies $|\mathcal{A}_k| \geq 3$ ($k = 1, \ldots, K$). We start with non-overlapping subpopulations in Corollary 2 and deal with overlapping subpopulations in Corollary 3.

Corollary 2. Generalized β-models with dependent edges. *Consider Models 2 and 3 with non-overlapping subpopulations, level of sparsity $\alpha \in [0,\, 1/2)$, and $\boldsymbol{\theta}^\star \in \mathbb{R}^{N+1}$ satisfying (3.15) with $\vartheta \in [0,\, 1/2 - \alpha)$. Then*

$$4\, D_N \sqrt{N \log N} \quad \leq \quad \gamma_N \quad \leq \quad 28\, D_N^5 \sqrt{N \log N},$$

and there exist finite constants $C > 0$ and $N_0 \geq 3$, independent of N and p, such that, for all $N > N_0$,

$$\widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \quad \leq \quad C\, D_N^{14} \sqrt{\frac{\log N}{N^{1-2\,(\alpha+\vartheta)}}}.$$

Corollary 2 shows that the convergence rate of pseudo-likelihood-based M-estimators under generalized β-models with dependent edges and non-overlapping subpopulations resembles the convergence rate under the β-model with independent edges when the random graph is dense ($\alpha = 0$) and $\|\boldsymbol{\theta}^\star\|_\infty$ is bounded above ($\vartheta = 0$), ignoring logarithmic terms; note that D_N needs to satisfy $D_N = O(\log N)$ to ensure $\|\boldsymbol{\theta}^\star\|_\infty > 0$. In addition, Corollary 2 reveals a trade-off between the sparsity of the random graph controlled by $\alpha \in [0,\, 1/2)$ and the growth of $\|\boldsymbol{\theta}^\star\|_\infty$ controlled by $\vartheta \in [0,\, 1/2 - \alpha)$.

We turn to overlapping subpopulations. To bound $|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2$ in scenarios with overlapping subpopulations, we need to control the amount of overlap of subpopulations, because the dependence among edges can propagate through overlapping subpopulations. To do so, we introduce a subpopulation graph $\mathcal{G}_{\mathcal{A}}$ with a set of vertices $\mathcal{V}_{\mathcal{A}} \coloneqq \{\mathcal{A}_1, \ldots, \mathcal{A}_K\}$, where a pair of distinct subpopulations $\mathcal{A}_k$ and $\mathcal{A}_l$ is connected by an edge if $\mathcal{A}_k \cap \mathcal{A}_l \neq \emptyset$. Denote by $d_{\mathcal{G}_{\mathcal{A}}} : \mathcal{V}_{\mathcal{A}} \times \mathcal{V}_{\mathcal{A}} \mapsto \{0, 1, \ldots\} \cup \{\infty\}$ the length of the shortest path between pairs of subpopulations in $\mathcal{G}_{\mathcal{A}}$, called the graph distance; note that $d_{\mathcal{G}_{\mathcal{A}}}(\mathcal{A}_k,\, \mathcal{A}_k) \coloneqq 0$ and $d_{\mathcal{G}_{\mathcal{A}}}(\mathcal{A}_k,\, \mathcal{A}_l) \coloneqq \infty$ if there is no path of finite length between two distinct subpopulations $\mathcal{A}_k$ and $\mathcal{A}_l$. Let $\mathcal{V}_{\mathcal{A}_k,l}$ be the subset of subpopulations at graph distance l from a given subpopulation $\mathcal{A}_k$:

$$\mathcal{V}_{\mathcal{A}_k,l} \quad \coloneqq \quad \{\mathcal{A}^\star \in \{\mathcal{A}_1, \ldots, \mathcal{A}_K\} \setminus \{\mathcal{A}_k\} :\; d_{\mathcal{G}_{\mathcal{A}}}(\mathcal{A}_k,\, \mathcal{A}^\star) = l\}\,.$$

Assumption A. *Define $U \coloneqq 1\,/\,(1 + \exp(-L))$, where $L \in [0,\, \infty)$ is identical to the constant L in (3.15) and is independent of N and p. Assume that $D_N \in [1,\, \infty)$ and that there exist finite constants $\omega_1 \in [0,\, \infty)$ and*

$$\omega_2 \quad \leq \quad \min\left\{\omega_1,\; \frac{1}{(\omega_1 + 1)\, |\log(1 - U)|}\right\},$$

independent of N and p, such that

$$\max_{1 \leq k \leq K} |\mathcal{V}_{\mathcal{A}_k,l}| \quad \leq \quad \omega_1 + \frac{\omega_2}{2\, D_N^3} \log l, \qquad l \in \{1, \ldots, K-1\}.$$

Assumption A covers tree- and non-tree subpopulation graphs in which, for each subpopulation, the number of subpopulations at graph distance l is either constant or grows slowly as a function of l (depending on D_N).

Corollary 3. Generalized β-models with dependent edges. *Consider Models 2 and 3 with overlapping subpopulations and level of sparsity $\alpha \in [0,\, 1/2)$. Assume that $\boldsymbol{\theta}^\star \in \mathbb{R}^{N+1}$ satisfies (3.15) with $\vartheta = 0$ and that Assumption A is satisfied. Then there exist finite constants $A > 0$, $B > 0$, $C > 0$, and $N_0 \geq 3$, independent of N and p, such that, for all $N > N_0$,*

$$4\, D_N \sqrt{N \log N} \quad \leq \quad \gamma_N \quad \leq \quad B\, \exp(A\, D_N^3) \sqrt{N \log N},$$

and

$$\widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \quad \leq \quad C\, \exp(A\, D_N^3) \sqrt{\frac{\log N}{N^{1-2\,\alpha}}}.$$

A comparison of Corollaries 2 and 3 reveals that, when both D_N and $\|\boldsymbol{\theta}^\star\|_\infty$ are bounded above ($\vartheta = 0$), the convergence rate of pseudo-likelihood-based M-estimators under generalized β-models with dependent edges is the same, regardless of whether subpopulations overlap (as long as Assumption A is satisfied). If D_N increases as a function of N, overlap comes at a cost. First, the convergence rate is lower due to the factor $\exp(A\, D_N^3)$ in the overlapping subpopulation scenario, compared with the factor D_N^{14} in the non-overlapping subpopulation scenario. Second, overlap requires stronger restrictions on D_N. For example, consider the best-case scenario when the random graph is dense ($\alpha = 0$) and $\|\boldsymbol{\theta}^\star\|_\infty$ is bounded above ($\vartheta = 0$). Then, to ensure $\|\boldsymbol{\theta}^\star\|_\infty > 0$ and $\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star\|_\infty \xrightarrow{\text{p}} 0,$ D_N needs to satisfy

- $D_N = O(\log N)$ when the subpopulations do not overlap;
- $D_N = o((\log(N \,/\, \log N))^{1/3})$ when the subpopulations do overlap.

These results dovetail with results on other statistical exponential-family models for discrete and dependent random variables in single-observation scenarios. For example, Chatterjee and Diaconis [8] considered the edge-and-triangle model with $p = 2$ parameters and unbounded D_N of order $O(N)$, but concluded that the edge-and-triangle model possesses undesirable properties and did not report consistency results. Likewise, the recent results of Ghosal and Mukherjee [18] on Ising models with $p = 2$ parameters suggest that consistency results may not be obtainable unless D_N is bounded or other restrictions are imposed. By contrast,

- we allow $D_N \to \infty$ as $N \to \infty$ provided $D_N = O(\log N)$ (non-overlapping subpopulations) or $D_N = o((\log(N \,/\, \log N))^{1/3})$ (overlapping subpopulations), as discussed above;

- we allow $p \to \infty$ as $N \to \infty$ provided $p = o(N^2/\log N)$, as discussed in Section 3.3.1;
- we cover a wide range of model specifications, beyond the pairwise interaction terms of discrete graphical models (e.g., Ising models).

Supplementary materials. We present proofs of all theoretical results along with simulation results in the supplement [38].

Acknowledgements. We acknowledge support from NSF awards DMS-1513644 and DMS-1812119 and DoD award ARO W911NF-21-1-0335.

References.

[1] Amini, A. A., Chen, A., Bickel, P. J., and Levina, E. (2013), "Pseudo-likelihood methods for community detection in large sparse networks," *The Annals of Statistics*, 41, 2097–2122.

[2] Bhattacharya, B. B., and Mukherjee, S. (2018), "Inference in Ising models," *Bernoulli*, 24, 493–525.

[3] Bickel, P. J., and Chen, A. (2009), "A nonparametric view of network models and Newman-Girvan and other modularities," in *Proceedings of the National Academy of Sciences*, Vol. 106, pp. 21068–21073.

[4] Bollobás, B., Riordan, O., Spencer, J., and Tusnády, G. (2001), "The degree sequence of a scale-free random graph process," *Random Structures & Algorithms*, 18, 279–290.

[5] Brown, L. (1986), *Fundamentals of Statistical Exponential Families: With Applications in Statistical Decision Theory*, Hayworth, CA, USA: Institute of Mathematical Statistics.

[6] Caron, F., and Fox, E. B. (2017), "Sparse graphs using exchangeable random measures," *Journal of the Royal Statistical Society, Series B (with discussion)*, 79, 1–44.

[7] Chatterjee, S. (2007), "Estimation in spin glasses: A first step," *The Annals of Statistics*, 35, 1931–1946.

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

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

[10] 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.

[11] 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.

[12] Comets, F. (1992), "On consistency of a class of estimators for exponential families of Markov random fields on the lattice," *The Annals of Statistics*, 20, 455–468.

[13] Crane, H., and Dempsey, W. (2018), "Edge exchangeable models for interaction networks," *Journal of the American Statistical Association*, 113, 1311–1326.

[14] Dawid, A. P. (1979), "Conditional independence in statistical theory," *Journal of the Royal Statistical Society, Series B*, 41, 1–31.

[15] Frank, O., and Strauss, D. (1986), "Markov graphs," *Journal of the American Statistical Association*, 81, 832–842.

[16] Furi, M., and Martelli, M. (1991), "On the mean value theorem, inequality, and inclusion," *The American Mathematical Monthly*, 98, 840–846.

[17] Gao, C., Ma, Z., Zhang, A. Y., and Zhou, H. H. (2018), "Community detection in degree-corrected block models," *The Annals of Statistics*, 46, 2153–2185.

[18] Ghosal, P., and Mukherjee, S. (2020), "Joint estimation of parameters in Ising model," *The Annals of Statistics*, 48, 785–810.

[19] Handcock, M. S. (2003), "Statistical Models for Social Networks: Inference and Degeneracy," in *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, eds. Breiger, R., Carley, K., and Pattison, P., Washington, D.C.: National Academies Press, pp. 1–12.

[20] Harville, D. A. (1997), *Matrix algebra from a statistician's perspective*, New York: Springer.

[21] Hillar, C. J., and Wibisono, A. (2015), "A Hadamard-type lower bound for symmetric diagonally dominant positive matrices," *Linear Algebra and its Applications*, 472, 135–141.

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

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

[24] Karwa, V., and Slavković, A. B. (2016), "Inference using noisy degrees: Differentially private β-model and synthetic graphs," *The Annals of Statistics*, 44, 87–112.

[25] 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.

[26] Lehmann, E. L., and Romano, J. P. (2005), *Testing Statistical Hypotheses*, New York: Springer-Verlag, 3rd ed.

[27] Lindvall, T. (2002), *Lectures On The Coupling Method*, Courier Corporation.

[28] Mukherjee, R., Mukherjee, S., and Sen, S. (2018), "Detection thresholds for the β-model on sparse graphs," *The Annals of Statistics*, 46, 1288–1317.

[29] Mukherjee, S. (2020), "Degeneracy in sparse ERGMs with functions of degrees as sufficient statistics," *Bernoulli*, 26, 1016–1043.

[30] 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.

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

[32] Rinaldo, A., Fienberg, S. E., and Zhou, Y. (2009), "On the geometry of discrete exponential families with application to exponential random graph models," *Electronic Journal of Statistics*, 3, 446–484.

[33] Rinaldo, A., Petrović, S., and Fienberg, S. E. (2013), "Maximum likelihood estimation in the β-model," *The Annals of Statistics*, 41, 1085–1110.

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

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

[36] Schweinberger, M., Krivitsky, P. N., Butts, C. T., and Stewart, J. R. (2020), "Exponential-family models of random graphs: Inference in finite, super, and infi-

nite population scenarios," *Statistical Science*, 35, 627–662.

[37] Schweinberger, M., and Stewart, J. R. (2020), "Concentration and consistency results for canonical and curved exponential-family models of random graphs," *The Annals of Statistics*, 48, 374–396.

[38] Stewart, J. R., and Schweinberger, M. (2023), "Supplement to: Pseudo-likelihood-based M-estimators for random graphs with dependent edges and parameter vectors of increasing dimension," *Department of Statistics, Florida State University.*

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

[40] Watts, D. J. (2003), *Six Degrees. The Science of a Connected Age*, Norton.

[41] Yan, T., Leng, C., and Zhu, J. (2016), "Asymptotics in directed exponential random graph models with an increasing bi-degree sequence," *The Annals of Statistics*, 44, 31–57.

[42] Yan, T., and Xu, J. (2013), "A central limit theorem in the β-model for undirected random graphs with a diverging number of vertices," *Biometrika*, 100, 519–524.

Supplement to: Pseudo-likelihood-based M-estimation of random graphs with dependent edges and parameter vectors of increasing dimension

By Jonathan R. Stewart and Michael Schweinberger

Florida State University and Penn State University

In Appendices A, B, C.2.3 and C.2.4, we adopt the notation used in Section 3 of the manuscript, by denoting the number of edge variables by $M := \binom{N}{2}$ and edge variables by $X_1, \ldots, X_M$. In addition, we denote the data-generating parameter vector by $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta} = \mathbb{R}^p$ and the data-generating probability measure and expectation by $\mathbb{P} \equiv \mathbb{P}_{\boldsymbol{\theta}^\star}$ and $\mathbb{E} \equiv \mathbb{E}_{\boldsymbol{\theta}^\star}$, respectively. Throughout, we assume that $\min_{1 \leq k \leq K} |\mathcal{A}_k| \geq 3$.

APPENDIX A: AUXILIARY RESULTS FOR THEOREM 1

A proof of Theorem 1 can be found in Section 3.2 of the manuscript. Here, we state and prove Lemma 1, which is used in the proof of Theorem 1.

Lemma 1. *Under the assumptions of Theorem 1, for all $t > 0$,*

$$\mathbb{P}\left(s(\boldsymbol{X}) \;\in\; \mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, t)\right) \;\;\geq\;\; 1 - 2\, \exp\left(-\frac{2\, t^2}{|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2^2\; \Psi_N^2} + \log p\right),$$

where $|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2 \geq 1$ and $\Psi_N > 0$ provided N is large enough.

Proof of Lemma 1. By Theorem 1 of Chazottes et al. [10, p. 207],

$$\mathbb{P}\left(|s_i(\boldsymbol{X}) - \mathbb{E}\, s_i(\boldsymbol{X})| \;\geq\; t\right) \;\leq\; 2\, \exp\left(-\frac{2\, t^2}{|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2^2\; \|\boldsymbol{\Xi}_i\|_2^2}\right),\; i = 1, \ldots, p.$$

A union bound over the p coordinates of $s(\boldsymbol{X})$ shows that

$$\mathbb{P}\left(\|s(\boldsymbol{X}) - \mathbb{E}\, s(\boldsymbol{X})\|_\infty \;\geq\; t\right) \;\;\leq\;\; 2\,\exp\left(-\frac{2\,t^2}{\|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2^2\;\;\Psi_N^2} + \log p\right),$$

where $\Psi_N \coloneqq \max_{1 \leq i \leq p}\; \|\boldsymbol{\Xi}_i\|_2$. As a result, we obtain

$$\mathbb{P}\left(s(\boldsymbol{X}) \;\in\; \mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, t)\right) \;\;\geq\;\; 1 - 2\;\exp\left(-\frac{2\,t^2}{\|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2^2\;\;\Psi_N^2} + \log p\right),$$

using $\boldsymbol{\mu}(\boldsymbol{\theta}^\star) \coloneqq \mathbb{E}_{\boldsymbol{\theta}^\star}\, s(\boldsymbol{X})$. □

Remark. Extensions to dependent random variables with countable and uncountable sample spaces. Theorem 1 is not restricted to random graphs with dependent edges. It covers models of dependent random variables with finite sample spaces, and can be extended to countable sample spaces: e.g., the concentration result of Chazottes et al. [10] used in Theorem 1 assumes that the sample spaces are finite—motivated by applications to Ising models—but could be extended to countable sample spaces. Uncountable sample spaces could be accommodated by replacing the concentration result of Chazottes et al. [10] by other suitable concentration results, e.g., Subgaussian concentration results. Likewise, the exponential-family properties used in Theorem 1 are neither restricted to finite nor countable sample spaces [5].

APPENDIX B: PROOF OF THEOREM 2

We prove Theorem 2 stated in Section 3.3 of the manuscript. Auxiliary results are proved in Appendix B.1.

Proof of Theorem 2. To prepare the ground, we first review basic facts that help prove Theorem 2. Define

$$\begin{aligned} \boldsymbol{g}(\boldsymbol{\theta}; \boldsymbol{X}) \;\;&\coloneqq\;\; \nabla_{\boldsymbol{\theta}}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}) \\ \boldsymbol{g}(\boldsymbol{\theta}) \;\;&\coloneqq\;\; \nabla_{\boldsymbol{\theta}}\, \mathbb{E}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}). \end{aligned}$$

By Lemma 3, the function $\mathbb{E}\;\widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$ is a strictly concave function on the convex set $\boldsymbol{\Theta} = \mathbb{R}^p$. In addition, the maximizer of $\mathbb{E}\;\widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$ exists and is unique, and is given by $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta}$. By Lemma 4, its gradient

$$\boldsymbol{g}(\boldsymbol{\theta}) \;\;\coloneqq\;\; \nabla_{\boldsymbol{\theta}}\, \mathbb{E}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}) \;\;=\;\; \mathbb{E}\, \nabla_{\boldsymbol{\theta}}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}) \tag{B.1}$$

exists, is continuous, and is one-to-one, so the inverse $\boldsymbol{g}^{-1}$ of $\boldsymbol{g}$ exists and is continuous. The interchange of differentiation and integration in (B.1) is admissible by Lemma 5.

By assumption, there exists a constant $\epsilon^\star \in (0, \infty)$, independent of N and p, such that $-\mathbb{E}\, \nabla_{\boldsymbol{\theta}}^2\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$ is invertible for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$. The proof of Theorem 2 will focus on the subset $\mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star) \subset \boldsymbol{\Theta} = \mathbb{R}^p$. We will show in (B.10) and (B.11) that it is legitimate to focus on the subset $\mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$ of $\boldsymbol{\Theta}$, because the event $\widetilde{\boldsymbol{\Theta}}(\gamma_N) \subset \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$ occurs with probability at least $1 - 2 / \max\{N, p\}^2$ for all large enough N.

Consider any $\epsilon \in (0, \epsilon^\star)$. By the continuity of $\boldsymbol{g}$ and its inverse $\boldsymbol{g}^{-1}$, there exists a real number $\delta(\epsilon) \in (0, \infty)$ such that

$$\boldsymbol{g}(\boldsymbol{\theta}) \;\in\; \mathcal{B}_\infty(\boldsymbol{g}(\boldsymbol{\theta}^\star), \delta(\epsilon)) \qquad \text{implies} \qquad \boldsymbol{\theta} \;\in\; \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon).$$

Consider any $\gamma_N \in (0, \delta(\epsilon)/2]$ and define

$$\mathbb{G}(\gamma_N) \;\;:=\;\; \left\{\boldsymbol{x} \in \mathbb{X} :\; \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta}; \boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty \;<\; \gamma_N\right\} \;\;\subseteq\;\; \mathbb{X}.$$

We will prove that the event $\boldsymbol{X} \in \mathbb{G}(\gamma_N)$ occurs with high probability for all large enough N and divide the remainder of the proof into four parts:

I. In the event $\boldsymbol{X} \in \mathbb{G}(\gamma_N)$, the set $\widetilde{\boldsymbol{\Theta}}(\gamma_N)$ is non-empty.

II. In the event $\boldsymbol{X} \in \mathbb{G}(\gamma_N)$, the set $\widetilde{\boldsymbol{\Theta}}(\gamma_N)$ is a subset of $\mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon)$.

III. Convergence rate of $\widetilde{\boldsymbol{\Theta}}(\gamma_N)$.

IV. The event $\widetilde{\boldsymbol{\Theta}}(\gamma_N) \subset \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$ occurs with probability at least $1 - 2 / \max\{N, p\}^2$ for all large enough N.

I. In the event $\boldsymbol{X} \in \mathbb{G}(\gamma_N)$, the set $\widetilde{\boldsymbol{\Theta}}(\gamma_N)$ is non-empty. Consider any element $\boldsymbol{x} \in \mathbb{G}(\gamma_N)$. Then

$$\begin{aligned}\|\boldsymbol{g}(\boldsymbol{\theta}^\star; \boldsymbol{x})\|_\infty \;&=\; \|\boldsymbol{g}(\boldsymbol{\theta}^\star; \boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{\theta}^\star)\|_\infty \\ &\leq\; \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta}; \boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty \;<\; \gamma_N,\end{aligned}$$

because $\boldsymbol{g}(\boldsymbol{\theta}^\star) = \boldsymbol{0}$ by Lemma 3 and $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta}; \boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty < \gamma_N$ for all $\boldsymbol{x} \in \mathbb{G}(\gamma_N)$. As a result, the set

$$\widetilde{\boldsymbol{\Theta}}(\gamma_N) \;\;:=\;\; \{\boldsymbol{\theta} \;\in\; \boldsymbol{\Theta} :\; \|\boldsymbol{g}(\boldsymbol{\theta}; \boldsymbol{x})\|_\infty \;\leq\; \gamma_N\}$$

contains the data-generating parameter vector $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta} = \mathbb{R}^p$ and is hence non-empty in the event $\boldsymbol{X} \in \mathbb{G}(\gamma_N)$.

II. In the event $\boldsymbol{X} \in \mathbb{G}(\gamma_N)$, the set $\widetilde{\boldsymbol{\Theta}}(\gamma_N)$ is a subset of $\mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon)$. We have shown that the set $\widetilde{\boldsymbol{\Theta}}(\gamma_N)$ is non-empty in the event $\boldsymbol{X} \in \mathbb{G}(\gamma_N)$.

Consider any $\boldsymbol{x} \in \mathbb{G}(\gamma_N)$. For all $\widetilde{\boldsymbol{\theta}} \in \widetilde{\boldsymbol{\Theta}}(\gamma_N)$, we have

$$\begin{aligned} \|\boldsymbol{g}(\widetilde{\boldsymbol{\theta}}) - \boldsymbol{g}(\boldsymbol{\theta}^\star)\|_\infty &\leq \|\boldsymbol{g}(\widetilde{\boldsymbol{\theta}}) - \boldsymbol{g}(\widetilde{\boldsymbol{\theta}};\, \boldsymbol{x})\|_\infty + \|\boldsymbol{g}(\widetilde{\boldsymbol{\theta}};\, \boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{\theta}^\star)\|_\infty \\ &= \|\boldsymbol{g}(\widetilde{\boldsymbol{\theta}};\, \boldsymbol{x}) - \boldsymbol{g}(\widetilde{\boldsymbol{\theta}})\|_\infty + \|\boldsymbol{g}(\widetilde{\boldsymbol{\theta}};\, \boldsymbol{x})\|_\infty \\ &\leq \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty + \gamma_N, \end{aligned}$$

because $\boldsymbol{g}(\boldsymbol{\theta}^\star) = \boldsymbol{0}$ by Lemma 3 and $\|\boldsymbol{g}(\widetilde{\boldsymbol{\theta}};\, \boldsymbol{x})\|_\infty \leq \gamma_N$ for any element $\widetilde{\boldsymbol{\theta}}$ of $\widetilde{\boldsymbol{\Theta}}(\gamma_N)$ by construction of the set $\widetilde{\boldsymbol{\Theta}}(\gamma_N)$. In addition, by construction of the set $\mathbb{G}(\gamma_N)$, we know that

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty \;<\; \gamma_N,$$

which, combined with the choice $\gamma_N \in (0,\, \delta(\epsilon)\,/\,2]$, implies that

$$\|\boldsymbol{g}(\widetilde{\boldsymbol{\theta}}) - \boldsymbol{g}(\boldsymbol{\theta}^\star)\|_\infty \;\leq\; \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty + \gamma_N \;<\; \delta(\epsilon). \tag{B.2}$$

In other words, in the event $\boldsymbol{X} \in \mathbb{G}(\gamma_N)$, any element $\widetilde{\boldsymbol{\theta}} \in \widetilde{\boldsymbol{\Theta}}(\gamma_N)$ satisfies

$$\boldsymbol{g}(\widetilde{\boldsymbol{\theta}}) \;\in\; \mathcal{B}_\infty(\boldsymbol{g}(\boldsymbol{\theta}^\star),\, \delta(\epsilon)),$$

which implies that

$$\widetilde{\boldsymbol{\theta}} \;\in\; \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon)$$

and hence

$$\widetilde{\boldsymbol{\Theta}}(\gamma_N) \;\subseteq\; \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon)$$

by the continuity of $\boldsymbol{g}$ and its inverse $\boldsymbol{g}^{-1}$.

III. Convergence rate of $\widetilde{\boldsymbol{\Theta}}(\gamma_N)$. To establish convergence rates, we leverage the method of proof used in Theorem 1 in Section 3.2 of the manuscript, which relates ϵ to $\delta(\epsilon)$. Adapting the argument from maximum likelihood to pseudo-likelihood-based M-estimators establishes

$$\frac{\epsilon}{\widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)} \;\leq\; \delta(\epsilon), \tag{B.3}$$

provided $-\mathbb{E}\, \nabla_{\boldsymbol{\theta}}^2\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$ is invertible for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$. We take advantage of (B.3) in the event $\widetilde{\boldsymbol{\Theta}}(\gamma_N) \subset \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$, in which case $-\mathbb{E}\, \nabla_{\boldsymbol{\theta}}^2\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$ is invertible by assumption. To invoke (B.3), we start with the following observation, which follows from (B.2): If $\boldsymbol{x} \in \mathbb{X}$ satisfies

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty + \gamma_N \;<\; \delta(\epsilon), \tag{B.4}$$

then $\boldsymbol{x} \in \mathbb{G}(\delta(\epsilon) - \gamma_N) \supseteq \mathbb{G}(\gamma_N)$ because $\gamma_N \in (0, \delta(\epsilon)/2]$, which implies that the set $\widetilde{\boldsymbol{\Theta}}(\gamma_N)$ is non-empty and satisfies

$$\widetilde{\boldsymbol{\Theta}}(\gamma_N) \subseteq \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon). \tag{B.5}$$

Since (B.4) implies (B.5), we obtain

$$\begin{aligned}
&\mathbb{P}\left(\widetilde{\boldsymbol{\Theta}}(\gamma_N) \subseteq \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon)\right) \\
\geq\ &\mathbb{P}\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty + \gamma_N < \delta(\epsilon)\right).
\end{aligned}$$

By assumption, for all $\epsilon \in (0, \epsilon^\star)$ and all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon) \subset \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$, $-\mathbb{E}\, \nabla^2_{\boldsymbol{\theta}}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$ is invertible. As a result, the inequality $\delta(\epsilon) \geq \epsilon \,/\, \widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)$ can be invoked in the event $\widetilde{\boldsymbol{\Theta}}(\gamma_N) \subseteq \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon)$, which implies that

$$\begin{aligned}
&\mathbb{P}\left(\widetilde{\boldsymbol{\Theta}}(\gamma_N) \subseteq \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon)\right) \\
\geq\ &\mathbb{P}\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty + \gamma_N < \delta(\epsilon)\right) \\
\geq\ &\mathbb{P}\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty + \gamma_N < \frac{\epsilon}{\widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)}\right).
\end{aligned} \tag{B.6}$$

We bound the probability of the complement of the event on the right-hand side of (B.6) by bounding

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty + \gamma_N \geq \frac{\epsilon}{\widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)}\right) \\
\leq\ &\mathbb{P}\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty \geq \frac{\epsilon}{2\, \widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)}\right) \\
+\ &\mathbb{P}\left(\gamma_N > \frac{\epsilon}{2\, \widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)}\right).
\end{aligned} \tag{B.7}$$

The first term on the right-hand side of (B.7) can be bounded by using Lemma 2, which shows that

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty \geq \frac{\epsilon}{2\, \widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)}\right) \\
\leq\ &2\, \exp\left(-\frac{\epsilon^2}{32\, \widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)^2\, (1 + D_N)^2\, \|\!|\mathcal{D}_N(\boldsymbol{\theta}^\star)|\!\|_2^2\, \Psi_N^2} + \log p\right).
\end{aligned}$$

Choosing

$$\begin{aligned} \epsilon &:= \sqrt{96}\; \widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \\ &= \sqrt{96}\; \widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)\; (1 + D_N)\; |||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2\; \Psi_N\; \sqrt{\log \max\{N,\, p\}} \end{aligned}$$

gives

$$\mathbb{P}\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty \;\geq\; \frac{\epsilon}{2\, \widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)}\right) \;\leq\; \frac{2}{\max\{N,\, p\}^2}. \tag{B.8}$$

The second term on the right-hand side of (B.7) can be bounded above as follows. Choosing

$$\gamma_N \;=\; \sqrt{24}\; (1 + D_N)\; |||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2\; \Psi_N\; \sqrt{\log \max\{N,\, p\}} \;=\; \frac{\epsilon}{2\, \widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)}$$

implies that

$$\mathbb{P}\left(\gamma_N \;>\; \frac{\epsilon}{2\, \widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)}\right) \;=\; 0. \tag{B.9}$$

The choice of $\gamma_N = \epsilon \,/\, (2\, \widetilde{\Lambda}_N(\boldsymbol{\theta}^\star))$ guarantees that $\gamma_N \leq \delta(\epsilon) \,/\, 2$, because $\epsilon \,/\, \widetilde{\Lambda}_N(\boldsymbol{\theta}^\star) \leq \delta(\epsilon)$ according to (B.3).

By combining (B.6), (B.7), (B.8), and (B.9) along with $\epsilon := \sqrt{96}\; \widetilde{\Phi}_N(\boldsymbol{\theta}^\star)$, we conclude that

$$\begin{aligned} &\mathbb{P}\left(\widetilde{\boldsymbol{\Theta}}(\gamma_N) \;\subseteq\; \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon)\right) \\ \geq\;& \mathbb{P}\left(\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty + \gamma_N \;<\; \frac{\epsilon}{\widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)}\right) \\ \geq\;& 1 - \frac{2}{\max\{N,\, p\}^2}. \end{aligned}$$

IV. The event $\widetilde{\boldsymbol{\Theta}}(\gamma_N) \subset \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$ occurs with probability at least $1 - 2 \,/\, \max\{N,\, p\}^2$ for all large enough N. To show that the focus on the subset $\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$ of $\boldsymbol{\Theta}$ is legitimate, note that the assumption that $\widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \to 0$ as $N \to \infty$ implies that there exists an integer $N_0 \geq 3$ such that

$$\epsilon \;:=\; \sqrt{96}\; \widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \;<\; \epsilon^\star \quad \text{for all} \quad N > N_0. \tag{B.10}$$

The fact that $\sqrt{96}\ \widetilde{\Phi}_N(\boldsymbol{\theta}^\star) < \epsilon^\star$ for all $N > N_0$ implies that $\widetilde{\boldsymbol{\Theta}}(\gamma_N) \subset \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$ with probability at least $1 - 2\,/\max\{N,\, p\}^2$ for all $N > N_0$:

$$\begin{aligned}
&\mathbb{P}(\widetilde{\boldsymbol{\Theta}}(\gamma_N) \subset \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star))\\
\geq\ &\mathbb{P}(\widetilde{\boldsymbol{\Theta}}(\gamma_N) \subset \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \sqrt{96}\ \widetilde{\Phi}_N(\boldsymbol{\theta}^\star)))\\
\geq\ &1 - \frac{2}{\max\{N,\, p\}^2}.
\end{aligned} \tag{B.11}$$

Conclusion. Combining the above results implies that, for all $N > N_0$, the random set $\widetilde{\boldsymbol{\Theta}}(\gamma_N)$ is non-empty and any element $\widetilde{\boldsymbol{\theta}}$ of $\widetilde{\boldsymbol{\Theta}}(\gamma_N)$ satisfies

$$\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star\|_\infty \quad\leq\quad \sqrt{96}\ \widetilde{\Phi}_N(\boldsymbol{\theta}^\star)$$

with probability at least $1 - 2\ /\ \max\{N,\, p\}^2$, provided

$$\gamma_N \quad\coloneqq\quad \sqrt{24}\ (1 + D_N)\ \|\!|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|\!|_2\ \Psi_N\ \sqrt{\log\max\{N,\, p\}}.$$

□

B.1. Auxiliary results for Theorem 2.

Lemma 2. *Under the assumptions of Theorem 2, for all $t > 0$,*

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\boldsymbol{\theta}\,\in\,\boldsymbol{\Theta}}\ \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty\ \geq\ t\right)\\
\leq\ \ &2\ \exp\left(-\frac{t^2}{8\,(1 + D_N)^2\ \|\!|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|\!|_2^2\ \Psi_N^2} + \log\, p\right),
\end{aligned}$$

where $\boldsymbol{g}(\boldsymbol{\theta}; \boldsymbol{X})$ and $\boldsymbol{g}(\boldsymbol{\theta})$ are defined by

$$\begin{aligned}
\boldsymbol{g}(\boldsymbol{\theta}; \boldsymbol{X}) &\quad\coloneqq\quad \nabla_{\boldsymbol{\theta}}\ \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})\\
\boldsymbol{g}(\boldsymbol{\theta}) &\quad\coloneqq\quad \nabla_{\boldsymbol{\theta}}\ \mathbb{E}\ \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}),
\end{aligned}$$

while $D_N \geq 0$, $\|\!|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|\!|_2 \geq 1$, and $\Psi_N > 0$ provided N is large enough.

Proof of Lemma 2. We prove Lemma 2 by leveraging concentration results of Chazottes et al. [10] along with conditional independence properties of models with factorization properties of the form (2.1).

Consider any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and any $\boldsymbol{x} \in \mathbb{X}$. By definition,

$$\widetilde{\ell}(\boldsymbol{\theta};\, \boldsymbol{x}) \quad\coloneqq\quad \sum_{i=1}^{M} \log\, \mathbb{P}_{\boldsymbol{\theta}}(X_i = x_i \,|\, \boldsymbol{X}_{-i} = \boldsymbol{x}_{-i}),$$

which implies that

$$\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{x}) \;\; \coloneqq \;\; \nabla_{\boldsymbol{\theta}}\, \widetilde{\ell}(\boldsymbol{\theta};\, \boldsymbol{x}) \;\; = \;\; \sum_{i=1}^{M} \nabla_{\boldsymbol{\theta}} \log \mathbb{P}_{\boldsymbol{\theta}}(X_i = x_i \,|\, \boldsymbol{X}_{-i} = \boldsymbol{x}_{-i}).$$

Observe that

$$\nabla_{\boldsymbol{\theta}} \log \mathbb{P}_{\boldsymbol{\theta}}(X_i = x_i \,|\, \boldsymbol{X}_{-i} = \boldsymbol{x}_{-i}) \;\; = \;\; s(\boldsymbol{x}) - \mathbb{E}_{\boldsymbol{\theta}, \boldsymbol{x}_{-i}}\, s(\boldsymbol{X}), \tag{B.12}$$

where $\mathbb{E}_{\boldsymbol{\theta}, \boldsymbol{x}_{-i}}$ denotes the expectation with respect to the conditional probability distribution of X_i given $\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i}$. The result in (B.12) follows from exponential-family properties [5], because the conditional distribution of X_i given $\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i}$ is an exponential-family distribution with sufficient statistic vector $s(\boldsymbol{x})$ and natural parameter vector $\boldsymbol{\theta}$.

We are interested in events of the form

$$\left\{ \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \; \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_{\infty} \, \geq \, t \right\}, \qquad t > 0. \tag{B.13}$$

To bound the probabilities of events of the form (B.13), we leverage concentration results of Chazottes et al. [10]. Theorem 1 of Chazottes et al. [10] states that, for each $k \in \{1, \ldots, p\}$ and $t > 0$,

$$\mathbb{P}\left(|g_k(\boldsymbol{\theta}, \boldsymbol{X}) - \mathbb{E}\, g_k(\boldsymbol{\theta}, \boldsymbol{X})| \, \geq \, t\right) \;\; \leq \;\; 2\, \exp\left(- \frac{2\, t^2}{\|\boldsymbol{\Delta}_k\|_2^2 \; \|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2^2}\right),$$

where $\mathcal{D}_N(\boldsymbol{\theta}^\star)$ is defined in Section 3.1 and $\boldsymbol{\Delta}_k \in [0,\, \infty)^M$ is defined by

$$\Delta_{k,i} \;\; \coloneqq \;\; \max_{(\boldsymbol{x}, \boldsymbol{x}') \,\in\, \mathbb{X} \times \mathbb{X}:\; x_l = x_l' \text{ for all } l \neq i} \; |g_k(\boldsymbol{\theta}; \boldsymbol{x}) - g_k(\boldsymbol{\theta}; \boldsymbol{x}')|, \quad i \in \{1, \ldots, M\}.$$

We bound the probability of event (B.13) by bounding

$$\|\boldsymbol{\Delta}_k\|_2^2 \;\; = \;\; \sum_{i=1}^{M} \left(\max_{(\boldsymbol{x}, \boldsymbol{x}') \,\in\, \mathbb{X} \times \mathbb{X}:\; x_l = x_l' \text{ for all } l \neq i} \; |g_k(\boldsymbol{\theta}; \boldsymbol{x}) - g_k(\boldsymbol{\theta}; \boldsymbol{x}')| \right)^2.$$

Consider any $i \in \{1, \ldots, M\}$ and any $(\boldsymbol{x}, \boldsymbol{x}') \in \mathbb{X} \times \mathbb{X}$ such that $x_i = 0$ and $x_i' = 1$ while $x_l = x_l'$ for all $l \neq i$. Write

$$\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{x}') \;\; = \;\; \sum_{j=1}^{M} \nabla_{\boldsymbol{\theta}}\, \lambda_j(\boldsymbol{\theta}; \boldsymbol{x}, \boldsymbol{x}'),$$

where

$$\lambda_j(\boldsymbol{\theta}; \boldsymbol{x}, \boldsymbol{x}') \;\; \coloneqq \;\; \log \frac{\mathbb{P}_{\boldsymbol{\theta}}(X_j = x_j \mid \boldsymbol{X}_{-j} = \boldsymbol{x}_{-j})}{\mathbb{P}_{\boldsymbol{\theta}}(X_j = x_j' \mid \boldsymbol{X}_{-j} = \boldsymbol{x}'_{-j})}, \qquad j \in \{1, \dots, M\}.$$

By definition, for any given $j \in \{1, \dots, M\}$, the set $\mathfrak{N}_j \subseteq \{1, \dots, M\} \setminus \{j\}$ is the smallest subset of indices such that

$$X_j \;\; \perp\!\!\!\perp \;\; \boldsymbol{X}_{\{1,\dots,M\} \setminus (\{j\} \,\cup\, \mathfrak{N}_j)} \mid \boldsymbol{X}_{\mathfrak{N}_j}. \tag{B.14}$$

Therefore, for all $j \in \{1, \dots, M\} \setminus (\{i\} \cup \mathfrak{N}_i)$, the conditional probability mass function of X_j is unaffected by X_i, so (B.14) implies that

$$\mathbb{P}_{\boldsymbol{\theta}}(X_j = x_j \mid \boldsymbol{X}_{-j} = \boldsymbol{x}_{-j}) \;\; = \;\; \mathbb{P}_{\boldsymbol{\theta}}(X_j = x_j' \mid \boldsymbol{X}_{-j} = \boldsymbol{x}'_{-j}),$$

which in turn implies that

$$\lambda_j(\boldsymbol{\theta}; \boldsymbol{x}, \boldsymbol{x}') \;\; \coloneqq \;\; \log \frac{\mathbb{P}_{\boldsymbol{\theta}}(X_j = x_j \mid \boldsymbol{X}_{-j} = \boldsymbol{x}_{-j})}{\mathbb{P}_{\boldsymbol{\theta}}(X_j = x_j' \mid \boldsymbol{X}_{-j} = \boldsymbol{x}'_{-j})} \;\; = \;\; 0,$$

noting that $x_j = x_j'$ for all $j \in \{1, \dots, M\} \setminus (\{i\} \cup \mathfrak{N}_i)$. As a result,

$$\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{x}') \;\; = \;\; \sum_{j=1}^{M} \nabla_{\boldsymbol{\theta}}\, \lambda_j(\boldsymbol{\theta}; \boldsymbol{x}, \boldsymbol{x}') \;\; = \;\; \sum_{j \,\in\, \{i\} \,\cup\, \mathfrak{N}_i} \nabla_{\boldsymbol{\theta}}\, \lambda_j(\boldsymbol{\theta}; \boldsymbol{x}, \boldsymbol{x}').$$

The triangle inequality implies that, for each $k \in \{1, \dots, p\}$,

$$\begin{aligned} & |g_k(\boldsymbol{\theta};\, \boldsymbol{x}) - g_k(\boldsymbol{\theta};\, \boldsymbol{x}')| \\ \leq & \sum_{j \,\in\, \{i\} \,\cup\, \mathfrak{N}_i} \left(\left| s_k(\boldsymbol{x}) - s_k(\boldsymbol{x}') \right| + \left| \mathbb{E}_{\boldsymbol{\theta}, \boldsymbol{x}_{-j}}\, s_k(\boldsymbol{X}) - \mathbb{E}_{\boldsymbol{\theta}, \boldsymbol{x}'_{-j}}\, s_k(\boldsymbol{X}) \right| \right). \end{aligned}$$

We bound the terms of the above sum one by one.

Bounding $|s_k(\boldsymbol{x}) - s_k(\boldsymbol{x}')|$. Consider any $i \in \{1, \dots, M\}$ and any $(\boldsymbol{x}, \boldsymbol{x}') \in \mathbb{X} \times \mathbb{X}$ such that $x_i = 0$ and $x_i' = 1$ while $x_l = x_l'$ for all $l \neq i$.

By definition,

$$\Xi_{k,i} \;\; \coloneqq \;\; \max_{(\boldsymbol{x}, \boldsymbol{x}') \,\in\, \mathbb{X} \times \mathbb{X}:\; x_l = x_l' \;\text{ for all } l \neq i} \; |s_k(\boldsymbol{x}) - s_k(\boldsymbol{x}')|,$$

providing the following bound:

$$|s_k(\boldsymbol{x}) - s_k(\boldsymbol{x}')| \;\; \leq \;\; \Xi_{k,i},$$

provided $(\boldsymbol{x}, \boldsymbol{x}') \in \mathbb{X} \times \mathbb{X}$ satisfies $x_l = x_l'$ for all $l \neq i$ with $x_i = 0$ and $x_i' = 1$.

Bounding $|\mathbb{E}_{\boldsymbol{\theta},\boldsymbol{x}_{-j}}\, s_k(\boldsymbol{X}) - \mathbb{E}_{\boldsymbol{\theta},\boldsymbol{x}'_{-j}}\, s_k(\boldsymbol{X})|$. We take advantage of the coupling argument in Section 2.1 of Chazottes et al. [10] to bound deviations of conditional expectations.

Consider any $i \in \{1, \ldots, M\}$ and any $(\boldsymbol{x}, \boldsymbol{x}') \in \mathbb{X} \times \mathbb{X}$ such that $x_i = 0$ and $x_i' = 1$ while $x_l = x_l'$ for all $l \neq i$. Define

$$\mathbb{P}_{j,\boldsymbol{\theta},\boldsymbol{x}_{-j}}(a) \;\; \coloneqq \;\; \mathbb{P}_{\boldsymbol{\theta}}(X_j = a \,|\, \boldsymbol{X}_{-j} = \boldsymbol{x}_{-j}), \qquad a \in \{0, 1\}.$$

Let $(\boldsymbol{X}^{\star}, \boldsymbol{X}^{\star\star}) \in \{0,1\}^M \times \{0,1\}^M$ be an optimal coupling of the conditional probability mass functions $\mathbb{P}_{j,\boldsymbol{\theta},\boldsymbol{x}_{-j}}$ and $\mathbb{P}_{j,\boldsymbol{\theta},\boldsymbol{x}'_{-j}}$ such that

- the marginal probability mass function of $X_j^{\star}$ is $\mathbb{P}_{j,\boldsymbol{\theta},\boldsymbol{x}_{-j}}$;
- the marginal probability mass function of $X_j^{\star\star}$ is $\mathbb{P}_{j,\boldsymbol{\theta},\boldsymbol{x}'_{-j}}$;
- the following events occur with probability 1:
  - $\{X_i^{\star} = x_i = 0\}$,
  - $\{X_i^{\star\star} = x_i' = 1\}$,
  - $\{X_l^{\star} = X_l^{\star\star} = x_l\}$ for all $l \in \{1, \ldots, M\} \setminus \{i,\, j\}$.

An optimal coupling is guaranteed to exist, but it may not be unique [27, pp. 99–107]. That said, any optimal coupling will do. We denote the joint probability mass function of $(\boldsymbol{X}^{\star}, \boldsymbol{X}^{\star\star})$ by $\mathbb{T}_{j,\boldsymbol{\theta},\boldsymbol{x}_{-j},\boldsymbol{x}'_{-j}}$.

An important property of the coupling $\mathbb{T}_{j,\boldsymbol{\theta},\boldsymbol{x}_{-j},\boldsymbol{x}'_{-j}}$ is that, for all $j \in \{1, \ldots, M\} \setminus (\{i\} \cup \mathfrak{N}_i)$,

$$\mathbb{T}_{j,\boldsymbol{\theta},\boldsymbol{x}_{-j},\boldsymbol{x}'_{-j}}(X_j^{\star} \neq X_j^{\star\star}) \;\; = \;\; \|\mathbb{P}_{j,\boldsymbol{\theta},\boldsymbol{x}_{-j}} - \mathbb{P}_{j,\boldsymbol{\theta},\boldsymbol{x}'_{-j}}\|_{\mathrm{TV}} \;\; = \;\; 0, \tag{B.15}$$

because

- $\mathbb{T}_{j,\boldsymbol{\theta},\boldsymbol{x}_{-j},\boldsymbol{x}'_{-j}}$ is an optimal coupling, which implies that $\mathbb{T}_{j,\boldsymbol{\theta},\boldsymbol{x}_{-j},\boldsymbol{x}'_{-j}}(X_j^{\star} \neq X_j^{\star\star}) = \|\mathbb{P}_{j,\boldsymbol{\theta},\boldsymbol{x}_{-j}} - \mathbb{P}_{j,\boldsymbol{\theta},\boldsymbol{x}'_{-j}}\|_{\mathrm{TV}}$;
- the conditional independence of X_i and X_j implies that $\mathbb{P}_{j,\boldsymbol{\theta},\boldsymbol{x}_{-j}} = \mathbb{P}_{j,\boldsymbol{\theta},\boldsymbol{x}'_{-j}}$ for all $(\boldsymbol{x}, \boldsymbol{x}') \in \mathbb{X} \times \mathbb{X}$ such that $x_l = x_l'$ for all $l \neq i$, which implies that $\|\mathbb{P}_{j,\boldsymbol{\theta},\boldsymbol{x}_{-j}} - \mathbb{P}_{j,\boldsymbol{\theta},\boldsymbol{x}'_{-j}}\|_{\mathrm{TV}} = 0$.

By construction of the coupling $\mathbb{T}_{j,\boldsymbol{\theta},\boldsymbol{x}_{-j},\boldsymbol{x}'_{-j}}$, we can write

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta},\boldsymbol{x}_{-j}}\, s_k(\boldsymbol{X}) - \mathbb{E}_{\boldsymbol{\theta},\boldsymbol{x}'_{-j}}\, s_k(\boldsymbol{X}) \;&=\; \mathbb{E}_{\mathbb{T}_{j,\boldsymbol{\theta},\boldsymbol{x}_{-j},\boldsymbol{x}'_{-j}}}\, s_k(\boldsymbol{X}^{\star}) - \mathbb{E}_{\mathbb{T}_{j,\boldsymbol{\theta},\boldsymbol{x}_{-j},\boldsymbol{x}'_{-j}}}\, s_k(\boldsymbol{X}^{\star\star}) \\ &=\; \mathbb{E}_{\mathbb{T}_{j,\boldsymbol{\theta},\boldsymbol{x}_{-j},\boldsymbol{x}'_{-j}}}\, (s_k(\boldsymbol{X}^{\star}) - s_k(\boldsymbol{X}^{\star\star})).\end{aligned}$$

Taking advantage of the telescoping identity on page 205 and the bounding argument on page 206 of Chazottes et al. [10] gives rise to the bound

$$\begin{aligned}\left|\mathbb{E}_{\boldsymbol{\theta},\boldsymbol{x}_{-j}}\, s_k(\boldsymbol{X}) - \mathbb{E}_{\boldsymbol{\theta},\boldsymbol{x}'_{-j}}\, s_k(\boldsymbol{X})\right| &= \left|\mathbb{E}_{\mathbb{T}_{j,\boldsymbol{\theta},\boldsymbol{x}_{-j},\boldsymbol{x}'_{-j}}}(s_k(\boldsymbol{X}^\star) - s_k(\boldsymbol{X}^{\star\star}))\right| \\ &\leq \sum_{l=1}^{M} \Xi_{k,l}\; \mathbb{T}_{j,\boldsymbol{\theta},\boldsymbol{x}_{-j},\boldsymbol{x}'_{-j}}(X_l^\star \neq X_l^{\star\star}).\end{aligned}$$

The construction of the coupling $\mathbb{T}_{j,\boldsymbol{\theta},\boldsymbol{x}_{-j},\boldsymbol{x}'_{-j}}$ implies that

$$\begin{aligned}\left|\mathbb{E}_{\boldsymbol{\theta},\boldsymbol{x}_{-j}}\, s_k(\boldsymbol{X}) - \mathbb{E}_{\boldsymbol{\theta},\boldsymbol{x}'_{-j}}\, s_k(\boldsymbol{X})\right| &\leq \sum_{l=1}^{M} \Xi_{k,l}\; \mathbb{T}_{j,\boldsymbol{\theta},\boldsymbol{x}_{-j},\boldsymbol{x}'_{-j}}(X_l^\star \neq X_l^{\star\star}) \\ &\leq \begin{cases}\Xi_{k,i} & \text{if } i = j \\ \Xi_{k,i} + \Xi_{k,j} & \text{if } i \neq j \text{ and } j \in \mathfrak{N}_i \\ 0 & \text{if } i \neq j \text{ and } j \notin \mathfrak{N}_i.\end{cases}\end{aligned}$$

Collecting terms. Upon collecting terms, we obtain the bounds

$$\begin{aligned}&|g_k(\boldsymbol{\theta};\, \boldsymbol{x}) - g_k(\boldsymbol{\theta};\, \boldsymbol{x}')| \\ \leq\;& \sum_{j\,\in\,\{i\}\,\cup\,\mathfrak{N}_i} \left(\left|s_k(\boldsymbol{x}) - s_k(\boldsymbol{x}')\right| + \left|\mathbb{E}_{\boldsymbol{\theta},\boldsymbol{x}_{-j}}\, s_k(\boldsymbol{X}) - \mathbb{E}_{\boldsymbol{\theta},\boldsymbol{x}'_{-j}}\, s_k(\boldsymbol{X})\right|\right) \\ \leq\;& 2\,\Xi_{k,i} + \sum_{j\,\in\,\mathfrak{N}_i} (\Xi_{k,i} + (\Xi_{k,i} + \Xi_{k,j})) \\ \leq\;& 2\left(\Xi_{k,i} + \sum_{j\,\in\,\mathfrak{N}_i} (\Xi_{k,i} + \Xi_{k,j})\right)\end{aligned}$$

and

$$\begin{aligned}&\max_{(\boldsymbol{x},\boldsymbol{x}')\,\in\,\mathbb{X}\times\mathbb{X}:\; x_l = x'_l \text{ for all } l\neq i} |g_k(\boldsymbol{\theta};\, \boldsymbol{x}) - g_k(\boldsymbol{\theta};\, \boldsymbol{x}')| \\ \leq\;& 2\left(\Xi_{k,i} + \sum_{j\,\in\,\mathfrak{N}_i} (\Xi_{k,i} + \Xi_{k,j})\right).\end{aligned}$$

The Cauchy–Schwarz inequality implies that

$$\begin{aligned}
&\left(\max_{(\boldsymbol{x},\boldsymbol{x}')\,\in\,\mathbb{X}\times\mathbb{X}:\; x_l = x_l' \text{ for all } l \neq i} |g_k(\boldsymbol{\theta};\, \boldsymbol{x}) - g_k(\boldsymbol{\theta};\, \boldsymbol{x}')|\right)^2 \\
&\leq\; 4\left(\Xi_{k,i} + \sum_{j\,\in\,\mathfrak{N}_i} (\Xi_{k,i} + \Xi_{k,j})\right)^2 \\
&\leq\; 4\,(1 + 2\,D_N)\left(\Xi_{k,i}^2 + \sum_{j\,\in\,\mathfrak{N}_i} (\Xi_{k,i}^2 + \Xi_{k,j}^2)\right) \\
&\leq\; 8\,(1 + D_N)\left(\Xi_{k,i}^2 + \sum_{j\,\in\,\mathfrak{N}_i} (\Xi_{k,i}^2 + \Xi_{k,j}^2)\right) \\
&\leq\; 8\,(1 + D_N)\left((1 + D_N)\,\Xi_{k,i}^2 + \sum_{j\,\in\,\mathfrak{N}_i} \Xi_{k,j}^2\right),
\end{aligned}$$

using $D_N \coloneqq \max\{|\mathfrak{N}_1|, \ldots, |\mathfrak{N}_M|\}$. We hence obtain

$$\begin{aligned}
\|\boldsymbol{\Delta}_k\|_2^2 \;&\coloneqq\; \sum_{i=1}^{M}\left(\max_{(\boldsymbol{x},\boldsymbol{x}')\,\in\,\mathbb{X}\times\mathbb{X}:\; x_l = x_l' \text{ for all } l \neq i} |g_k(\boldsymbol{\theta}; \boldsymbol{x}) - g_k(\boldsymbol{\theta}; \boldsymbol{x}')|\right)^2 \\
&\leq\; \sum_{i=1}^{M} 8\,(1 + D_N)\left((1 + D_N)\,\Xi_{k,i}^2 + \sum_{j\,\in\,\mathfrak{N}_i} \Xi_{k,j}^2\right) \\
&=\; 8\,(1 + D_N)^2\, \|\boldsymbol{\Xi}_k\|_2^2 + 8\,(1 + D_N) \sum_{j=1}^{M} \Xi_{k,j}^2 \sum_{i=1}^{M} \mathbb{1}(j \,\in\, \mathfrak{N}_i).
\end{aligned}$$

To bound the second term on the right-hand side, note that edge variable X_j can be in the dependence neighborhoods $\mathfrak{N}_i$ $(i = 1, \ldots, M)$ of at most $D_N \coloneqq \max\{|\mathfrak{N}_1|, \ldots, |\mathfrak{N}_M|\}$ other edge variables X_i, which implies that

$$\begin{aligned}
8\,(1 + D_N) \sum_{j=1}^{M} \Xi_{k,j}^2 \sum_{i=1}^{M} \mathbb{1}(j \,\in\, \mathfrak{N}_i) \;&\leq\; 8\,(1 + D_N) \sum_{j=1}^{M} \Xi_{k,j}^2\, D_N \\
&\leq\; 8\,(1 + D_N)^2\, \|\boldsymbol{\Xi}_k\|_2^2.
\end{aligned}$$

We hence arrive at the following bound on $\|\boldsymbol{\Delta}_k\|_2^2$:

$$\begin{aligned}\|\boldsymbol{\Delta}_k\|_2^2 &\leq 8\,(1+D_N)^2\,\|\boldsymbol{\Xi}_k\|_2^2 + 8\,(1+D_N)\sum_{j=1}^{M}\Xi_{k,j}^2\sum_{i=1}^{M}\mathbb{1}(j\,\in\,\mathfrak{N}_i)\\ &\leq 16\,(1+D_N)^2\,\|\boldsymbol{\Xi}_k\|_2^2\\ &\leq 16\,(1+D_N)^2\,\Psi_N^2,\end{aligned}$$

where $\Psi_N \coloneqq \max_{1\leq k\leq p}\,\|\boldsymbol{\Xi}_k\|_2$.

Concentration result. By applying Theorem 1 of Chazottes et al. [10] to each coordinate $g_k(\boldsymbol{\theta};\,\boldsymbol{X}) - g_k(\boldsymbol{\theta})$ of $\boldsymbol{g}(\boldsymbol{\theta};\,\boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})$ $(k = 1,\ldots,p)$ using the above bound on $\|\boldsymbol{\Delta}_k\|_2$, we have, for all $t > 0$,

$$\begin{aligned}\mathbb{P}\left(|g_k(\boldsymbol{\theta};\,\boldsymbol{X}) - g_k(\boldsymbol{\theta})| \geq t\right) &\leq 2\,\exp\left(-\frac{2\,t^2}{16\,(1+D_N)^2\,\|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2^2\,\Psi_N^2}\right)\\ &= 2\,\exp\left(-\frac{t^2}{8\,(1+D_N)^2\,\|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2^2\,\Psi_N^2}\right),\end{aligned}$$

where $D_N \geq 0,\ \|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2 \geq 1,$ and $\Psi_N > 0$ provided N is large enough. A union bound over the p coordinates $g_k(\boldsymbol{\theta};\,\boldsymbol{X}) - g_k(\boldsymbol{\theta})$ of $\boldsymbol{g}(\boldsymbol{\theta};\,\boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})$ gives rise to the bound

$$\begin{aligned}&\mathbb{P}\left(\|\boldsymbol{g}(\boldsymbol{\theta};\,\boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty \geq t\right)\\ \leq\ & 2\,\exp\left(-\frac{t^2}{8\,(1+D_N)^2\,\|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2^2\,\Psi_N^2} + \log\,p\right).\end{aligned}$$

Since the above bound does not depend on $\boldsymbol{\theta}\in\boldsymbol{\Theta}$, we conclude that

$$\begin{aligned}&\mathbb{P}\left(\sup_{\boldsymbol{\theta}\,\in\,\boldsymbol{\Theta}}\,\|\boldsymbol{g}(\boldsymbol{\theta};\,\boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty\,\geq\,t\right)\\ \leq\ & 2\,\exp\left(-\frac{t^2}{8\,(1+D_N)^2\,\|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2^2\,\Psi_N^2} + \log\,p\right).\end{aligned}$$

□

Lemma 3. *The function* $\mathbb{E}\ \widetilde{\ell}(\boldsymbol{\theta};\,\boldsymbol{X})$ *is a strictly concave function on the convex set* $\boldsymbol{\Theta} = \mathbb{R}^p$*. In addition, the data-generating parameter vector* $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta}$ *maximizes the expected loglikelihood and pseudo-loglikelihood function:*

$$\boldsymbol{\theta}^\star \;=\; \underset{\boldsymbol{\theta}\,\in\,\boldsymbol{\Theta}}{\arg\max}\ \mathbb{E}\ \ell(\boldsymbol{\theta};\,\boldsymbol{X}) \;=\; \underset{\boldsymbol{\theta}\,\in\,\boldsymbol{\Theta}}{\arg\max}\ \mathbb{E}\ \widetilde{\ell}(\boldsymbol{\theta};\,\boldsymbol{X}).$$

Proof of Lemma 3. Section 2 of the manuscript shows that the family of densities $\{f_{\boldsymbol{\theta}},\, \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ parameterized by (2.1) and (2.2) is an exponential family of densities. We take advantage of the properties of exponential families [5] to prove Lemma 3, and divide the proof into three parts:

I. $\mathbb{E}\, \widetilde{\ell}(\boldsymbol{\theta};\, \boldsymbol{X})$ is a strictly concave function on the convex set $\boldsymbol{\Theta}$.

II. $\boldsymbol{\theta}^\star$ is the unique maximizer of $\mathbb{E}\, \ell(\boldsymbol{\theta};\, \boldsymbol{X})$.

III. $\boldsymbol{\theta}^\star$ is the unique maximizer of $\mathbb{E}\, \widetilde{\ell}(\boldsymbol{\theta};\, \boldsymbol{X})$.

I. $\mathbb{E}\, \widetilde{\ell}(\boldsymbol{\theta};\, \boldsymbol{X})$ is a strictly concave function on the convex set $\boldsymbol{\Theta}$. Let $\boldsymbol{x}$ be an observation of a random graph $\boldsymbol{X}$ with dependent edges. Then, by definition,

$$\widetilde{\ell}(\boldsymbol{\theta};\, \boldsymbol{x}) \;\; = \;\; \sum_{i=1}^{M} \widetilde{\ell}_i(\boldsymbol{\theta};\, \boldsymbol{x}),$$

where

$$\widetilde{\ell}_i(\boldsymbol{\theta};\, \boldsymbol{x}) \;\; = \;\; \langle \boldsymbol{\theta},\, s(\boldsymbol{x})\rangle - \psi_i(\boldsymbol{\theta};\, \boldsymbol{x}_{-i})$$

and

$$\psi_i(\boldsymbol{\theta};\, \boldsymbol{x}_{-i}) = \log\left(\exp(\langle \boldsymbol{\theta},\, s(\boldsymbol{x}_{-i}, x_i = 0)\rangle) + \exp(\langle \boldsymbol{\theta},\, s(\boldsymbol{x}_{-i}, x_i = 1)\rangle)\right).$$

We first show that $\mathbb{E}\, \widetilde{\ell}(\boldsymbol{\theta};\, \boldsymbol{X}) = \sum_{i=1}^{M} \mathbb{E}\, \widetilde{\ell}_i(\boldsymbol{\theta};\, \boldsymbol{X})$ is a concave function on the convex set $\boldsymbol{\Theta}$ by proving that the functions $\mathbb{E}\, \widetilde{\ell}_i(\boldsymbol{\theta};\, \boldsymbol{X})$ are concave on $\boldsymbol{\Theta}$. Observe that the functions $\mathbb{E}\, \widetilde{\ell}_i(\boldsymbol{\theta};\, \boldsymbol{X})$ are concave provided the functions $\widetilde{\ell}_i(\boldsymbol{\theta};\, \boldsymbol{x})$ are concave for all $\boldsymbol{x} \in \mathbb{X}$. To show that the functions $\widetilde{\ell}_i(\boldsymbol{\theta};\, \boldsymbol{x})$ are concave for all $\boldsymbol{x} \in \mathbb{X}$, consider any $i \in \{1, \ldots, M\}$, any $x_i \in \{0, 1\}$, and any $\boldsymbol{x}_{-i} \in \{0, 1\}^{M-1}$. Each $\widetilde{\ell}_i(\boldsymbol{\theta};\, \boldsymbol{x})$ consists of two terms. The first term, $\langle \boldsymbol{\theta},\, s(\boldsymbol{x})\rangle$, is a linear function of $\boldsymbol{\theta}$, so $\widetilde{\ell}_i(\boldsymbol{\theta};\, \boldsymbol{x})$ is a concave function of $\boldsymbol{\theta}$ if the second term, $\psi_i(\boldsymbol{\theta};\, \boldsymbol{x}_{-i})$, is a convex function of $\boldsymbol{\theta}$. Consider any $(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}) \in \boldsymbol{\Theta} \times \boldsymbol{\Theta}$ and any $\lambda \in (0, 1)$. Then, by Hölder's inequality,

$$\psi_i\left(\lambda\, \boldsymbol{\theta}^{(1)} + (1 - \lambda)\, \boldsymbol{\theta}^{(2)};\, \boldsymbol{x}_{-i}\right) \leq \lambda\, \psi_i\left(\boldsymbol{\theta}^{(1)};\, \boldsymbol{x}_{-i}\right) + (1 - \lambda)\, \psi_i\left(\boldsymbol{\theta}^{(2)};\, \boldsymbol{x}_{-i}\right).$$

As a consequence, for any $\boldsymbol{x}_{-i} \in \{0, 1\}^{M-1}$, $\psi_i(\boldsymbol{\theta};\, \boldsymbol{x}_{-i})$ is a convex function on $\boldsymbol{\Theta}$. Hence, for all $\boldsymbol{x} \in \mathbb{X}$, $\widetilde{\ell}(\boldsymbol{\theta};\, \boldsymbol{x})$ is a concave function on $\boldsymbol{\Theta}$, and so is $\mathbb{E}\, \widetilde{\ell}(\boldsymbol{\theta};\, \boldsymbol{X})$ as a finite sum of concave functions on $\boldsymbol{\Theta}$.

Second, we prove by contradiction that $\mathbb{E}\, \widetilde{\ell}(\boldsymbol{\theta};\, \boldsymbol{X})$ is a strictly concave function on $\boldsymbol{\Theta}$, by showing that there exists $i^\star \in \{1, \ldots, M\}$ such that $\mathbb{E}\, \psi_{i^\star}(\boldsymbol{\theta};\, \boldsymbol{X}_{-i^\star})$ is strictly convex on $\boldsymbol{\Theta}$, which implies that $\mathbb{E}\, \widetilde{\ell}_{i^\star}(\boldsymbol{\theta};\, \boldsymbol{X})$ is strictly concave on $\boldsymbol{\Theta}$. Suppose that there does not exist any $i^\star \in \{1, \ldots, M\}$

such that $\mathbb{E}\,\psi_{i^\star}(\boldsymbol{\theta};\,\boldsymbol{X}_{-i^\star})$ is strictly convex on $\boldsymbol{\Theta}$. Then, for all $i \in \{1,\ldots,M\}$, all $\boldsymbol{x}_{-i} \in \{0,1\}^{M-1}$, and all $x_i \in \{0,1\}$, there exists $(\boldsymbol{\theta}^{(1)},\boldsymbol{\theta}^{(2)}) \in \boldsymbol{\Theta} \times \boldsymbol{\Theta}$ such that

$$\exp\left(\langle \boldsymbol{\theta}^{(1)},\, s(\boldsymbol{x}_{-i},\, x_i)\rangle\right) \,\propto\, \exp\left(\langle \boldsymbol{\theta}^{(2)},\, s(\boldsymbol{x}_{-i},\, x_i)\rangle\right), \tag{B.16}$$

as Hölder's inequality reduces to an equality if and only if (B.16) holds, i.e.,

$$\psi_i\left(\lambda\,\boldsymbol{\theta}^{(1)} + (1-\lambda)\,\boldsymbol{\theta}^{(2)};\,\boldsymbol{x}_{-i}\right) = \lambda\,\psi_i\left(\boldsymbol{\theta}^{(1)};\,\boldsymbol{x}_{-i}\right) + (1-\lambda)\,\psi_i\left(\boldsymbol{\theta}^{(2)};\,\boldsymbol{x}_{-i}\right)$$

if and only if (B.16) holds. In other words, for all $\boldsymbol{x} \in \mathbb{X}$,

$$\exp\left(\langle \boldsymbol{\theta}^{(1)},\, s(\boldsymbol{x})\rangle\right) \,\propto\, \exp\left(\langle \boldsymbol{\theta}^{(2)},\, s(\boldsymbol{x})\rangle\right). \tag{B.17}$$

The conclusion (B.17) contradicts the assumption that the exponential family is minimal. Therefore, there exists $i^\star \in \{1,\ldots,M\}$ such that $\mathbb{E}\,\psi_{i^\star}(\boldsymbol{\theta};\,\boldsymbol{X}_{-i^\star})$ is strictly convex on $\boldsymbol{\Theta}$, which implies that $\mathbb{E}\,\widetilde{\ell}_{i^\star}(\boldsymbol{\theta};\,\boldsymbol{X})$ is strictly concave on $\boldsymbol{\Theta}$, and so is $\mathbb{E}\,\widetilde{\ell}(\boldsymbol{\theta};\,\boldsymbol{X}) = \sum_{i=1}^{M} \mathbb{E}\,\widetilde{\ell}_i(\boldsymbol{\theta};\,\boldsymbol{X})$.

II. $\boldsymbol{\theta}^\star$ is the unique maximizer of $\mathbb{E}\,\ell(\boldsymbol{\theta};\boldsymbol{X})$. Maximizing $\mathbb{E}\,\ell(\boldsymbol{\theta};\boldsymbol{X})$ is equivalent to solving

$$\nabla_{\boldsymbol{\theta}}\,\mathbb{E}\,\ell(\boldsymbol{\theta};\,\boldsymbol{X}) \;=\; \mathbb{E}\,s(\boldsymbol{X}) - \mathbb{E}_{\boldsymbol{\theta}}\,s(\boldsymbol{X}) \;=\; \boldsymbol{0}. \tag{B.18}$$

The unique solution of (B.18) is $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta} = \mathbb{R}^p$, because $\mathbb{E} \equiv \mathbb{E}_{\boldsymbol{\theta}^\star}$. The fact that the solution is unique follows from the fact the map $\boldsymbol{\mu} : \boldsymbol{\Theta} \mapsto \mathbb{M}$ defined by $\boldsymbol{\mu}(\boldsymbol{\theta}) \coloneqq \mathbb{E}_{\boldsymbol{\theta}}\,s(\boldsymbol{X})$ is one-to-one [5, Theorem 3.6, p. 74]. As a result, $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta} = \mathbb{R}^p$ is the unique maximizer of $\mathbb{E}\,\ell(\boldsymbol{\theta};\boldsymbol{X})$.

III. $\boldsymbol{\theta}^\star$ is the unique maximizer of $\mathbb{E}\,\widetilde{\ell}(\boldsymbol{\theta};\,\boldsymbol{X})$. Observe that, for any $\boldsymbol{x} \in \mathbb{X}$, $\widetilde{\ell}(\boldsymbol{\theta};\boldsymbol{x})$ is a sum of exponential-family loglikelihood functions, because the conditional distributions of edge variables X_i given $\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i}$ ($i = 1,\ldots,M$) are exponential-family distributions with sufficient statistic vector $s(\boldsymbol{x})$ and natural parameter vector $\boldsymbol{\theta}$. As a result, $\widetilde{\ell}(\boldsymbol{\theta};\boldsymbol{x})$ is continuously differentiable on $\boldsymbol{\Theta}$ for all $\boldsymbol{x} \in \mathbb{X}$ [5], and so is $\mathbb{E}\,\widetilde{\ell}(\boldsymbol{\theta};\boldsymbol{X})$. We then have

$$\boldsymbol{g}(\boldsymbol{\theta}) \;\coloneqq\; \mathbb{E}\,\nabla_{\boldsymbol{\theta}}\,\widetilde{\ell}(\boldsymbol{\theta};\boldsymbol{X}) \;=\; \mathbb{E}\sum_{i=1}^{M}\left(s(\boldsymbol{X}) - \mathbb{E}_{\boldsymbol{\theta},\boldsymbol{X}_{-i}}\,s(\boldsymbol{X})\right),$$

where $\mathbb{E}_{\boldsymbol{\theta},\boldsymbol{x}_{-i}}$ denotes the conditional expectation with respect to the conditional distribution of X_i given $\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i}$. By the law of total expectation and the fact that $\mathbb{E} \equiv \mathbb{E}_{\boldsymbol{\theta}^\star}$, we have $\mathbb{E}\,\mathbb{E}_{\boldsymbol{\theta}^\star,\boldsymbol{X}_{-i}}\,s(\boldsymbol{X}) = \mathbb{E}\,s(\boldsymbol{X})$, which implies

$$\boldsymbol{g}(\boldsymbol{\theta}^\star) \;=\; \mathbb{E}\sum_{i=1}^{M}\left(s(\boldsymbol{X}) - \mathbb{E}_{\boldsymbol{\theta}^\star,\boldsymbol{X}_{-i}}\,s(\boldsymbol{X})\right) \;=\; \boldsymbol{0}. \tag{B.19}$$

Thus, a root of $\boldsymbol{g}(\boldsymbol{\theta})$ exists, and $\boldsymbol{\theta}^\star$ is a root of $\boldsymbol{g}(\boldsymbol{\theta})$. In addition, $\mathbb{E}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$ is strictly concave on $\boldsymbol{\Theta}$, so $\boldsymbol{\theta}^\star$ is the unique root of $\boldsymbol{g}(\boldsymbol{\theta})$. As a consequence, the maximizer of $\mathbb{E}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$ as a function of $\boldsymbol{\theta} \in \boldsymbol{\Theta} = \mathbb{R}^p$ exists and is unique, and is given by $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta} = \mathbb{R}^p$. □

Lemma 4. *Let $\boldsymbol{g} : \boldsymbol{\Theta} \mapsto \mathbb{R}$ be any continuously differentiable function on the open and convex set $\boldsymbol{\Theta}$. If $\boldsymbol{g}(\boldsymbol{\theta})$ is strictly concave on $\boldsymbol{\Theta}$, then its gradient $\nabla_{\boldsymbol{\theta}}\, \boldsymbol{g}(\boldsymbol{\theta})$ exists, is continuous, and is one-to-one.*

Proof of Lemma 4. The existence and continuity of $\nabla_{\boldsymbol{\theta}}\, \boldsymbol{g}(\boldsymbol{\theta})$ on $\boldsymbol{\Theta}$ follow from the assumption that $\boldsymbol{g}(\boldsymbol{\theta})$ is continuously differentiable on the open and convex set $\boldsymbol{\Theta}$. We prove by contradiction that $\nabla_{\boldsymbol{\theta}}\, \boldsymbol{g}(\boldsymbol{\theta})$ is one-to-one on $\boldsymbol{\Theta}$. Suppose that $\nabla_{\boldsymbol{\theta}}\, \boldsymbol{g}(\boldsymbol{\theta})$ is not one-to-one on $\boldsymbol{\Theta}$, that is, there exists $(\boldsymbol{\theta}_1,\, \boldsymbol{\theta}_2) \in \boldsymbol{\Theta} \times \boldsymbol{\Theta}$ such that $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ and $\nabla_{\boldsymbol{\theta}}\, \boldsymbol{g}(\boldsymbol{\theta})\, |_{\boldsymbol{\theta}=\boldsymbol{\theta}_1} = \nabla_{\boldsymbol{\theta}}\, \boldsymbol{g}(\boldsymbol{\theta})\, |_{\boldsymbol{\theta}=\boldsymbol{\theta}_2}$. By the strict concavity of $\boldsymbol{g}(\boldsymbol{\theta})$ on $\boldsymbol{\Theta}$,

$$\langle \nabla_{\boldsymbol{\theta}}\, \boldsymbol{g}(\boldsymbol{\theta})\, |_{\boldsymbol{\theta}=\boldsymbol{\theta}_1},\, \boldsymbol{\theta}_2 - \boldsymbol{\theta}_1 \rangle \quad > \quad \boldsymbol{g}(\boldsymbol{\theta}_2) - \boldsymbol{g}(\boldsymbol{\theta}_1) \tag{B.20}$$

and

$$\langle \nabla_{\boldsymbol{\theta}}\, \boldsymbol{g}(\boldsymbol{\theta})\, |_{\boldsymbol{\theta}=\boldsymbol{\theta}_2},\, \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2 \rangle \quad > \quad \boldsymbol{g}(\boldsymbol{\theta}_1) - \boldsymbol{g}(\boldsymbol{\theta}_2). \tag{B.21}$$

By multiplying both sides of (B.21) by -1, we obtain

$$\langle \nabla_{\boldsymbol{\theta}}\, \boldsymbol{g}(\boldsymbol{\theta})\, |_{\boldsymbol{\theta}=\boldsymbol{\theta}_2},\, \boldsymbol{\theta}_2 - \boldsymbol{\theta}_1 \rangle \quad < \quad \boldsymbol{g}(\boldsymbol{\theta}_2) - \boldsymbol{g}(\boldsymbol{\theta}_1).$$

If $\nabla_{\boldsymbol{\theta}}\, g(\boldsymbol{\theta})\, |_{\boldsymbol{\theta}=\boldsymbol{\theta}_1} = \nabla_{\boldsymbol{\theta}}\, g(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_2}$, then

$$\langle \nabla_{\boldsymbol{\theta}}\, \boldsymbol{g}(\boldsymbol{\theta})\, |_{\boldsymbol{\theta}=\boldsymbol{\theta}_1},\, \boldsymbol{\theta}_2 - \boldsymbol{\theta}_1 \rangle \quad < \quad \boldsymbol{g}(\boldsymbol{\theta}_2) - \boldsymbol{g}(\boldsymbol{\theta}_1). \tag{B.22}$$

The conclusion (B.22) contradicts (B.20), so $\nabla_{\boldsymbol{\theta}}\, \boldsymbol{g}(\boldsymbol{\theta})$ is one-to-one on $\boldsymbol{\Theta}$. □

Lemma 5. *Under the assumptions of Theorem 2,*

$$\nabla_{\boldsymbol{\theta}}\, \mathbb{E}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}) \quad = \quad \mathbb{E}\, \nabla_{\boldsymbol{\theta}}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}) \quad \textit{for all} \quad \boldsymbol{\theta} \in \boldsymbol{\Theta} = \mathbb{R}^p.$$

Proof of Lemma 5. We start with two observations. First, the exponential family introduced in Section 2.1 of the manuscript is regular in the sense of Brown [5, p. 2], because

$$\boldsymbol{\Theta} := \{\boldsymbol{\theta} \in \mathbb{R}^p : \psi(\boldsymbol{\theta}) < \infty\} = \mathbb{R}^p \quad \text{and} \quad \boldsymbol{\Theta} = \mathbb{R}^p \text{ is open.}$$

Second, for any $\boldsymbol{x} \in \mathbb{X}$, $\widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{x})$ is a sum of exponential-family loglikelihood functions, because the conditional distribution of X_i given $\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i}$ is an exponential-family distribution with sufficient statistic vector $s(\boldsymbol{x})$ and natural parameter vector $\boldsymbol{\theta}$. Thus, for any $\boldsymbol{x} \in \mathbb{X}$, $\widetilde{\ell}(\,\cdot\,; \boldsymbol{x})$ is continuously differentiable on $\boldsymbol{\Theta}$.

Consider $\widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{x})$ as a function of $\boldsymbol{x} \in \mathbb{X}$ for fixed $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and define

$$\boldsymbol{g}(\boldsymbol{\theta}; \boldsymbol{x}) \;\; := \;\; \nabla_{\boldsymbol{\theta}}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{x}) \;\; = \;\; \sum_{i=1}^{M} \left(s(\boldsymbol{x}) - \mathbb{E}_{\boldsymbol{\theta},\boldsymbol{x}_{-i}}\, s(\boldsymbol{X})\right),$$

where $\mathbb{E}_{\boldsymbol{\theta},\boldsymbol{x}_{-i}}$ denotes the expectation with respect to the conditional distribution of X_i given $\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i}$. Here, $\boldsymbol{g}(\boldsymbol{\theta}; \boldsymbol{x})$ is considered as a function of $\boldsymbol{x} \in \mathbb{X}$ for fixed $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. By the triangle inequality, for each $k \in \{1, \ldots, p\}$,

$$|g_k(\boldsymbol{\theta}; \boldsymbol{x})| \;\; \leq \;\; M\, |s_k(\boldsymbol{x})| + \sum_{i=1}^{M} |\mathbb{E}_{\boldsymbol{\theta},\, \boldsymbol{x}_{-i}}\, s_k(\boldsymbol{X})| \;\; =: \;\; h_k(\boldsymbol{x}),$$

where the dependence of $h_k(\boldsymbol{x})$ on $\boldsymbol{\theta}$ is supressed. Since the exponential family is regular in the sense of Brown [5, p. 2], all moments of $s(\boldsymbol{X})$ exist [Theorem 2.2, p. 34, 5], implying that, for all $k \in \{1, \ldots, p\}$, $\mathbb{E}|s_k(\boldsymbol{X})| < \infty$ and $\mathbb{E}|\mathbb{E}_{\boldsymbol{\theta},\boldsymbol{X}_{-i}}\, s_k(\boldsymbol{X})| < \infty$. As a result,

$$\mathbb{E}\, h_k(\boldsymbol{X}) \;\; = \;\; M\, \mathbb{E}|s_k(\boldsymbol{X})| + \sum_{i=1}^{M} \mathbb{E}|\mathbb{E}_{\boldsymbol{\theta},\boldsymbol{X}_{-i}}\, s_k(\boldsymbol{X})| \;\; < \;\; \infty.$$

Since $|g_k(\boldsymbol{\theta}; \boldsymbol{x})| \leq h_k(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{X}$ and all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\mathbb{E}\, h_k(\boldsymbol{X}) < \infty$, Lebesgue's dominated convergence theorem implies that

$$\nabla_{\boldsymbol{\theta}}\, \mathbb{E}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}) \;\; = \;\; \mathbb{E}\, \nabla_{\boldsymbol{\theta}}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}) \quad \text{for all} \quad \boldsymbol{\theta} \in \boldsymbol{\Theta} = \mathbb{R}^p.$$

□

APPENDIX C: PROOFS OF COROLLARIES 1–3

We prove Corollaries 1–3 stated in Section 3.4 of the manuscript, using the auxiliary results proved in Appendices C.1 and C.2. To prove them, it is convenient to return to the notation used in Section 2 of the manuscript, denoting edge variables by $X_{i,j}$ ($\{i, j\} \subset \mathcal{N}$).

Proof of Corollaries 1–3. To prove Corollaries 1–3, we bound

$$\widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \;\; := \;\; \widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)\; (1 + D_N)\; \|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2\; \Psi_N\, \sqrt{\log \max\{N,\, p\}}.$$

We first bound Ψ_N and then prove Corollaries 1–3.

Bounding Ψ_N. Recall the definition of Ψ_N: For each $a \in \{1, \dots, p\}$ and each pair of nodes $\{i, j\} \subset \mathcal{N}$,

$$\Xi_{a,\{i,j\}} \;\; := \;\; \max_{(\boldsymbol{x}, \boldsymbol{x}') \,\in\, \mathbb{X} \times \mathbb{X}:\; x_{k,l} = x'_{k,l} \text{ for all } \{k,l\} \neq \{i,j\}} \; |s_a(\boldsymbol{x}) - s_a(\boldsymbol{x}')|$$

and

$$\Psi_N \;\; := \;\; \max_{1 \leq a \leq p} \; \|\boldsymbol{\Xi}_a\|_2.$$

We show that $\Psi_N \leq \sqrt{N}$ under Model 1 and $\Psi_N \leq \|s_{N+1}\|_{\text{Lip}} \sqrt{N}$ under Models 2 and 3 and bound $\|s_{N+1}\|_{\text{Lip}}$, where $\|s_{N+1}\|_{\text{Lip}}$ is the Lipschitz coefficient of $s_{N+1}(\boldsymbol{X})$ with respect to the Hamming metric on $\mathbb{X} \times \mathbb{X}$:

- Models 1, 2, and 3 have sufficient statistics $s_1(\boldsymbol{X}), \dots, s_N(\boldsymbol{X})$, the degrees of nodes $1, \dots, N$, respectively. Since the degrees of nodes are sums of $N - 1$ edge variables $X_{i,j} \in \{0, 1\}$, we have

$$\|\boldsymbol{\Xi}_a\|_2 \;\; = \;\; \sqrt{N-1} \;\; \leq \;\; \sqrt{N}, \qquad a = 1, \dots, N.$$

- Models 2 and 3 include the additional sufficient statistic for brokerage $s_{N+1}(\boldsymbol{x}) = \sum_{i<j}^{N} X_{i,j}\, I_{i,j}(\boldsymbol{X})$, where

$$I_{i,j}(\boldsymbol{x}) \;\; = \;\; \mathbb{1}\left(\sum_{h \,\in\, \mathcal{N}_i \,\cap\, \mathcal{N}_j} x_{i,h}\, x_{j,h} \;\geq\; 1\right), \qquad \{i,j\} \subset \mathcal{N}.$$

By the definition of $s_{N+1}(\boldsymbol{x})$, we have $\Xi_{N+1,\{i,j\}} = 0$ for all pairs of nodes $\{i, j\} \subset \mathcal{N}$ satisfying $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset$. The number of pairs of nodes $\{i, j\} \subset \mathcal{N}$ satisfying $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$ is bounded above by $N D_N^2$: For each of the N nodes $i \in \mathcal{N}$, there are at most D_N^2 distinct nodes $j \in \mathcal{N}_i$ such that $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$, a fact established by Lemma 11. In addition, Lemma 14 shows, for each $\{i, j\} \subset \mathcal{N}$, that $\Xi_{N+1,\{i,j\}} \leq 1 + D_N$. Thus,

$$\|\boldsymbol{\Xi}_{N+1}\|_2 \;\; \leq \;\; \sqrt{N\, D_N^2\, (1 + D_N)^2} \;\; \leq \;\; \sqrt{4\, N\, D_N^4} \;\; = \;\; 2\, D_N^2\, \sqrt{N}.$$

As a result, under Model 1,

$$\sqrt{N/2} \;\; \leq \;\; \Psi_N \;\; := \;\; \max_{1 \leq a \leq p} \|\boldsymbol{\Xi}_a\|_2 \;\; = \;\; \sqrt{N-1} \;\; \leq \;\; \sqrt{N},$$

whereas under Models 2 and 3,

$$\sqrt{N/2} \;\; \leq \;\; \Psi_N \;\; := \;\; \max_{1 \leq a \leq p} \|\boldsymbol{\Xi}_a\|_2 \;\; \leq \;\; 2\, D_N^2\, \sqrt{N},$$

noting that $D_N \geq 1$ under Models 2 and 3.

Convergence rates. We obtain the following convergence rates using the auxiliary results in Appendices C.1 and C.2. The following results hold for all large enough N. The constants vary from model to model.

- **Corollary 1:** The independence of edges under Model 1 implies that $D_N = 0$, $|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2 = 1$, and $\Lambda_N(\boldsymbol{\theta}^\star) = \widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)$, which in turn implies that $\Phi_N(\boldsymbol{\theta}^\star) = \widetilde{\Phi}_N(\boldsymbol{\theta}^\star)$. We have $p = N$ and $\Psi_N \leq \sqrt{N}$. By Lemma 6 with $\alpha = 0$ and $\vartheta \in [0, 7/2)$, there exist constants $B > 0$ and $N_0 \geq 3$, independent of N and p, such that, for all $N > N_0$,

  $$\widetilde{\Lambda}_N(\boldsymbol{\theta}^\star) \;\; \leq \;\; \frac{B}{N^{1-\vartheta/7}}.$$

  As a result, there exists a constant $C > 0$, independent N and p, such that, for all $N > N_0$,

  $$\Phi_N(\boldsymbol{\theta}^\star) \;\; = \;\; \widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \;\; \leq \;\; \frac{C\sqrt{N \log N}}{N^{1-\vartheta/7}} \;\; = \;\; C\sqrt{\frac{\log\, N}{N^{1-2\,\vartheta/7}}}.$$

- **Corollaries 2 and 3:** By assumption, $D_N \geq 1$ under Models 2 and 3, and Ψ_N is bounded as follows:

  $$\sqrt{N/2} \;\; \leq \;\; \Psi_N \;\; \leq \;\; 2\, D_N^2\, \sqrt{N}.$$

  To bound γ_N, recall that γ_N is given by

  $$\gamma_N \;\; = \;\; \sqrt{24}\;\, (1 + D_N)\; |||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2\; \Psi_N\; \sqrt{\log \max\{N,\, p\}}.$$

  Using $|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2 \geq 1$ along with $\Psi_N \geq \sqrt{N/2}$, we obtain the lower bound

  $$\gamma_N \;\; \geq \;\; \sqrt{12}\;\, D_N\, \sqrt{N \log N}$$

  and the upper bound

  $$\gamma_N \;\; \leq \;\; \sqrt{768}\, D_N^3\; |||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2\, \sqrt{N \log N},$$

  using $1 + D_N \leq 2\, D_N$ $(D_N \geq 1)$ along with $\Psi_N \;\leq\; 2\, D_N^2 \sqrt{N}$ and

  $$\log \max\{N,\, p\} \;\; = \;\; \log(N+1) \;\; \leq \;\; 2\, \log N.$$

  Thus, γ_N satisfies

  $$\sqrt{24}\;\, D_N\, \sqrt{N \log N} \;\leq\; \gamma_N \;\leq\; \sqrt{768}\;\, D_N^3\; |||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2\, \sqrt{N \log N}.$$

We now turn to bounding

$$\widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \;\; \coloneqq \;\; \widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)\; (1 + D_N)\; |||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2 \;\Psi_N\; \sqrt{\log \max\{N,\, p\}}.$$

By Lemma 6, there exist constants $C_1 > 0$ and $N_0 \geq 3$, independent of N and p, such that, for all $N > N_0$,

$$\widetilde{\Lambda}_N(\boldsymbol{\theta}^\star) \;\leq\; \frac{C_1\, D_N^9}{N^{1-(\alpha+\vartheta)}}, \quad \text{provided } \alpha \in [0,\, 1/2) \text{ and } \vartheta \in [0,\, 1/2 - \alpha).$$

As a result, there exists a constant $C_2 \coloneqq 4\, C_1 > 0$, independent of N and p, such that, for all $N > N_0$,

$$\begin{aligned} \widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \;\;&\leq\;\; \frac{C_2\, D_N^{9+1+2}\; |||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2\, \sqrt{N \log N}}{N^{1-(\alpha+\vartheta)}} \\ &=\;\; C_2\, D_N^{12}\; |||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2 \sqrt{\frac{\log N}{N^{1-2\,(\alpha+\vartheta)}}}, \end{aligned} \tag{C.1}$$

using the inequalities $1+D_N \leq 2\, D_N$ and $\Psi_N \;\leq\; 2\, D_N^2 \sqrt{N}$, noting that $D_N \geq 1$ under Models 2 and 3. By Lemma 12, there exist constants $C_3 > 0$ and $C_4 > 0$, independent of N and p, such that:

- Corollary 2 with $\vartheta \in [0,\, 1/2 - \alpha)$:

  $$|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2 \;\;\leq\;\; 1 + 4\, D_N^2 \;\;\leq\;\; 8\; D_N^2, \tag{C.2}$$

  using the fact that D_N satisfies $D_N \geq 1$ under Models 2 and 3.

- Corollary 3 with $\vartheta = 0$: If Assumption A is satisfied,

  $$\begin{aligned} |||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2 \;\;&\leq\;\; 1 + 4\, D_N^2 + \omega_1\, C_3\; \exp(C_4\, D_N^3) \\ &\leq\;\; 3\; \max\{4,\, \omega_1\, C_3\}\; D_N^2\; \exp(C_4\, D_N^3) \\ &\leq\;\; B_1\; \exp(2\; \log D_N + C_4\, D_N^3) \\ &\leq\;\; B_1\; \exp(A_1\, D_N^3), \end{aligned} \tag{C.3}$$

  using $D_N \;\geq\; 1$ along with $\log D_N \;\leq\; D_N \;\leq\; D_N^3$, and defining $A_1 \coloneqq 2 + C_4 > 0$ and $B_1 \coloneqq 3\, \max\{4,\, \omega_1\, C_3\} > 0$. Note that the constants $C_3 > 0$, $C_4 > 0$, and $\omega_1 \geq 0$ are independent of N and p, implying that A_1 and B_1 are likewise independent of N and p.

Upon collecting terms, we conclude that there exist constants $C_5 > 0$ and $C_6 > 0$, independent of N and p, such that, for all $N > N_0$:

- Corollary 2 with $\alpha \in [0,\, 1/2)$ and $\vartheta \in [0,\, 1/2 - \alpha)$: Equations (C.1) and (C.2) provide the bound
$$\widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \quad \leq \quad C_5\, D_N^{14} \sqrt{\frac{\log N}{N^{1-2\,(\alpha+\vartheta)}}},$$
where $C_5 \coloneqq 8\, C_2 > 0$ and γ_N satisfies (noting $\sqrt{768} \leq 28$)
$$4\, D_N \sqrt{N \log N} \quad \leq \quad \gamma_N \quad \leq \quad 28\, D_N^5 \sqrt{N \log N}.$$
- Corollary 3 with $\alpha \in [0,\, 1/2)$ and $\vartheta = 0$: If Assumption A is satisfied, Equations (C.1) and (C.3) provide the bound
$$\begin{aligned} \widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \quad &\leq \quad C_6\, D_N^{12} \exp(A_1\, D_N^3) \sqrt{\frac{\log N}{N^{1-2\,\alpha}}} \\ &\leq \quad C_6\, \exp(12 \log D_N + A_1\, D_N^3) \sqrt{\frac{\log N}{N^{1-2\,\alpha}}} \\ &\leq \quad C_6\, \exp(A_2\, D_N^3) \sqrt{\frac{\log N}{N^{1-2\,\alpha}}}, \end{aligned}$$
where γ_N satisfies
$$\begin{aligned} 4\, D_N \sqrt{N \log N} \quad &\leq \quad \gamma_N \\ &\leq \quad B_2\, \exp(A_3\, D_N^3) \sqrt{N \log N}, \end{aligned}$$
where $A_2 \coloneqq 12 + A_1 > 0$, $C_6 \coloneqq B_1\, C_2 > 0$, and $B_2 \coloneqq 28\, B_1 > 0$; note that the upper bound on γ_N leverages the fact that $D_N \geq 1$ under Models 2 and 3 along with $\log D_N \leq D_N \leq D_N^3$, so that
$$D_N^3 \exp(A_1\, D_N^3) = \exp\left(3 \log D_N + A_1\, D_N^3\right) \leq \exp\left(A_3\, D_N^3\right),$$
where $A_3 \coloneqq 3 + A_1 > 0$ is a constant, independent of N and p.

□

C.1. Bounding $\widetilde{\boldsymbol{\Lambda}}_{\boldsymbol{N}}(\boldsymbol{\theta}^\star)$.

We bound
$$\widetilde{\Lambda}_N(\boldsymbol{\theta}^\star) \quad \coloneqq \quad \sup_{\boldsymbol{\theta} \,\in\, \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)} |||(-\mathbb{E}\, \nabla_{\boldsymbol{\theta}}^2\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}))^{-1}|||_\infty$$
in Lemma 6, leveraging auxiliary results supplied by Lemmas 7–10. To do so, we first prepare the ground by introducing notation. The negative expected Hessian $-\mathbb{E}\, \nabla_{\boldsymbol{\theta}}^2\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$ corresponding to Models 2 and 3 is of the form
$$-\mathbb{E}\, \nabla_{\boldsymbol{\theta}}^2\, \widetilde{\ell}(\boldsymbol{\theta};\, \boldsymbol{X}) \quad = \quad \begin{pmatrix} \boldsymbol{A}(\boldsymbol{\theta}) & \boldsymbol{c}(\boldsymbol{\theta}) \\ \boldsymbol{c}(\boldsymbol{\theta})^\top & v(\boldsymbol{\theta}) \end{pmatrix}, \tag{C.4}$$
where

- the entries $A_{i,j}(\boldsymbol{\theta})$ of the matrix $\boldsymbol{A}(\boldsymbol{\theta}) \in \mathbb{R}^{N \times N}$ are given by

$$A_{i,j}(\boldsymbol{\theta}) \;\; = \;\; \sum_{a<b}^{N} \mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{a,b\}}}(s_i(\boldsymbol{X}),\, s_j(\boldsymbol{X})), \qquad i, j = 1, \dots, N;$$

- the entries $c_i(\boldsymbol{\theta})$ of the vector $\boldsymbol{c}(\boldsymbol{\theta}) \in \mathbb{R}^N$ are given by

$$c_i(\boldsymbol{\theta}) \;\; = \;\; \sum_{a<b}^{N} \mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{a,b\}}}(s_i(\boldsymbol{X}),\, s_{N+1}(\boldsymbol{X})), \qquad i = 1, \dots, N;$$

- $v(\boldsymbol{\theta}) \in \mathbb{R}^+$ is given by

$$v(\boldsymbol{\theta}) \;\; = \;\; \sum_{a<b}^{N} \mathbb{E}\, \mathbb{V}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{a,b\}}}\, s_{N+1}(\boldsymbol{X}),$$

where $\mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{a,b\}}}$ and $\mathbb{V}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{a,b\}}}$ are the conditional covariance and variance operators with respect to the conditional probability distribution of edge variable $X_{a,b}$ given all other edge variables $\boldsymbol{X}_{-\{a,b\}}$. The negative expected Hessian under Model 1 is $-\mathbb{E}\, \nabla^2_{\boldsymbol{\theta}}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}) = \boldsymbol{A}(\boldsymbol{\theta})$.

Lemma 6. *Assume that the data-generating parameter vector $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta} = \mathbb{R}^p$ satisfies*

$$\|\boldsymbol{\theta}^\star\|_\infty \;\; \leq \;\; \frac{L + \vartheta \, \log N}{14\,(3 + D_N)} - \epsilon^\star,$$

where $L \in [0,\, \infty)$, $\vartheta \in [0,\, \infty)$, and $\epsilon^\star \in (0,\, \infty)$ are constants, independent of N and p. Then $-\mathbb{E}\, \nabla^2_{\boldsymbol{\theta}}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$ is invertible on $\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$ and there exist constants $B > 0$, $C > 0$, and $N_0 \geq 3$, independent of N and p, such that, for all $N > N_0$, the ℓ_∞-induced matrix norm of $(-\mathbb{E}\, \nabla^2_{\boldsymbol{\theta}}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}))^{-1}$ satisfies the following upper bounds uniformly on $\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$:

- *Under Model 1,*

$$\sup_{\boldsymbol{\theta} \,\in\, \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)} \;\; |||(-\mathbb{E}\, \nabla^2_{\boldsymbol{\theta}}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}))^{-1}|||_\infty \;\; \leq \;\; \frac{B}{N^{1-(\alpha+\vartheta/7)}},$$

assuming $\alpha \in [0,\, 1/2)$ and $\vartheta \in [0,\, 7\,(1/2 - \alpha))$.

- *Under Models 2 and 3,*

$$\sup_{\boldsymbol{\theta} \,\in\, \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)} \;\; |||(-\mathbb{E}\, \nabla^2_{\boldsymbol{\theta}}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}))^{-1}|||_\infty \;\; \leq \;\; \frac{C\, D_N^9}{N^{1-(\alpha+\vartheta)}},$$

assuming $\alpha \in [0,\, 1/2)$, $\vartheta \in [0,\, 1/2 - \alpha)$, *and*

$$1 \;\leq\; D_N \;<\; \frac{L + \vartheta \log N}{14\, \epsilon^\star} - 3.$$

Remark. Under Model 1, edges are independent and $D_N = 0$, whereas under Models 2 and 3, edges are dependent and $D_N \geq 1$. The upper bound on D_N under Models 2 and 3 ensures that $\|\boldsymbol{\theta}^\star\|_\infty > 0$.

Proof of Lemma 6. Using (C.4), we can write the negative expected Hessian $-\mathbb{E}\, \nabla_{\boldsymbol{\theta}}^2\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$ corresponding to Models 2 and 3 as

$$-\mathbb{E}\, \nabla_{\boldsymbol{\theta}}^2\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}) \;=\; \begin{pmatrix} \boldsymbol{A}(\boldsymbol{\theta}) & \boldsymbol{c}(\boldsymbol{\theta}) \\ \boldsymbol{c}(\boldsymbol{\theta})^\top & v(\boldsymbol{\theta}) \end{pmatrix},$$

where $\boldsymbol{A}(\boldsymbol{\theta}) \in \mathbb{R}^{N \times N}$, $\boldsymbol{c}(\boldsymbol{\theta}) \in \mathbb{R}^N$, and $v(\boldsymbol{\theta}) \in \mathbb{R}^+$ are defined above.

Bounding $\|\boldsymbol{A}(\boldsymbol{\theta})^{-1}\|_\infty$. Lemma 7 proves that the smallest eigenvalue of $\boldsymbol{A}(\boldsymbol{\theta})$ is strictly positive on $\mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$, which implies that $\boldsymbol{A}(\boldsymbol{\theta})$ is invertible on $\mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$. Theorem 1.2 of Hillar and Wibisono [21] along with the bounds on the entries of $\boldsymbol{A}(\boldsymbol{\theta})$ given in (C.7) of Lemma 7 reveal that

$$\begin{aligned}
\|\boldsymbol{A}(\boldsymbol{\theta})^{-1}\|_\infty \;&\leq\; \frac{(3\, N - 4)\, (1 + \exp((3 + D_N)\, \|\boldsymbol{\theta}\|_\infty)))^2}{2\, N^{-\alpha}\, (N-2)\, (N-1)} \\
&\leq\; (1 + \exp((3 + D_N)\, \|\boldsymbol{\theta}\|_\infty)))^2\, \frac{9}{2\, N^{1-\alpha}} \\
&\leq\; (1 + \exp((3 + D_N)\, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)))^2\, \frac{9}{2\, N^{1-\alpha}} \\
&\leq\; (2\, \exp((3 + D_N)\, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)))^2\, \frac{9}{2\, N^{1-\alpha}} \\
&=\; \exp((3 + D_N)\, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star))^2\, \frac{18}{N^{1-\alpha}} \\
&=\; \frac{18\, \tau(\boldsymbol{\theta}^\star)^2}{N^{1-\alpha}},
\end{aligned}$$

where

$$\tau(\boldsymbol{\theta}^\star) \;\coloneqq\; \exp((3 + D_N)\, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)).$$

The above exploits the fact that $\|\boldsymbol{\theta}\|_\infty \leq \|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star$ for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$, along with the inequality

$$\frac{3\, N - 4}{2\, (N-2)\, (N-1)} \;=\; \frac{3\, (N-1) - 1}{2\, (N-2)\, (N-1)} \;\leq\; \frac{3\, (N-1)}{2\, (N-2)\, (N-1)} \;=\; \frac{3}{2\, (N-2)},$$

which is bounded above by

$$\frac{3}{2\,(N-2)} \quad \leq \quad \frac{9}{2\,N},$$

provided $N \geq 3$. Using the assumption

$$\|\boldsymbol{\theta}^\star\|_\infty \quad \leq \quad \frac{L + \vartheta\, \log N}{14\,(3 + D_N)} - \epsilon^\star,$$

we obtain

$$\begin{aligned}
\tau(\boldsymbol{\theta}^\star)^2 \quad &= \quad \exp(2\,(3 + D_N)\,(\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)) \;\leq\; \exp\left(2\,(3 + D_N)\,\frac{L + \vartheta\, \log N}{14\,(3 + D_N)}\right) \\
&= \quad \exp\left(\frac{L + \vartheta\, \log N}{7}\right) \;=\; \exp(L\,/\,7)\; N^{\vartheta\,/\,7}.
\end{aligned}$$

As a consequence, we find that

$$|\!|\!|\boldsymbol{A}(\boldsymbol{\theta})^{-1}|\!|\!|_\infty \quad \leq \quad \frac{18\;\tau(\boldsymbol{\theta}^\star)^2}{N^{1-\alpha}} \quad \leq \quad \frac{18\;\exp(L\,/\,7)\;N^{\vartheta\,/\,7}}{N^{1-\alpha}} \quad = \quad \frac{B}{N^{1-(\alpha+\vartheta/7)}}$$

where $B \coloneqq 18\;\exp(L\,/\,7) > 0$. Under Model 1, we hence obtain

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\,\in\,\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\,\epsilon^\star)} |\!|\!|(-\mathbb{E}\,\nabla_{\boldsymbol{\theta}}^2\,\widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}))^{-1}|\!|\!|_\infty \quad &= \quad \sup_{\boldsymbol{\theta}\,\in\,\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\,\epsilon^\star)} |\!|\!|\boldsymbol{A}(\boldsymbol{\theta})^{-1}|\!|\!|_\infty \\
&\leq \quad \frac{B}{N^{1-(\alpha+\vartheta/7)}},
\end{aligned}$$

assuming $\alpha \in [0,\, 1/2)$ and $\vartheta \in [0,\, 7\,(1/2 - \alpha))$.

Bounding $|\!|\!|(-\mathbb{E}\,\nabla_{\boldsymbol{\theta}}^2\,\widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}))^{-1}|\!|\!|_\infty$. Let

$$\xi(\boldsymbol{\theta}) \quad \coloneqq \quad v(\boldsymbol{\theta}) - \boldsymbol{c}(\boldsymbol{\theta})^\top\, \boldsymbol{A}(\boldsymbol{\theta})^{-1}\, \boldsymbol{c}(\boldsymbol{\theta}).$$

Theorem 8.5.11 of Harville [20, p. 98-99] implies that, if the inverse of $-\mathbb{E}\,\nabla_{\boldsymbol{\theta}}^2\,\widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$ exists, then it can be written as

$$(-\mathbb{E}\,\nabla_{\boldsymbol{\theta}}^2\,\widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}))^{-1} \quad = \quad \begin{pmatrix} \boldsymbol{A}(\boldsymbol{\theta}) & \boldsymbol{c}(\boldsymbol{\theta}) \\ \boldsymbol{c}(\boldsymbol{\theta})^\top & v(\boldsymbol{\theta}) \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} \boldsymbol{A}(\boldsymbol{\theta})^{-1} + \xi(\boldsymbol{\theta})^{-1}\,(\boldsymbol{A}(\boldsymbol{\theta})^{-1}\,\boldsymbol{c}(\boldsymbol{\theta}))\,(\boldsymbol{A}(\boldsymbol{\theta})^{-1}\,\boldsymbol{c}(\boldsymbol{\theta}))^\top & -\xi(\boldsymbol{\theta})^{-1}\,\boldsymbol{A}(\boldsymbol{\theta})^{-1}\,\boldsymbol{c}(\boldsymbol{\theta}) \\ -\xi(\boldsymbol{\theta})^{-1}\,(\boldsymbol{A}(\boldsymbol{\theta})^{-1}\,\boldsymbol{c}(\boldsymbol{\theta}))^\top & \xi(\boldsymbol{\theta})^{-1} \end{pmatrix}.$$

To establish that $-\mathbb{E}\,\nabla^2_{\boldsymbol{\theta}}\,\widetilde{\ell}(\boldsymbol{\theta};\boldsymbol{X})$ is invertible, note that $\boldsymbol{A}(\boldsymbol{\theta})$ is invertible on $\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\epsilon^\star)$ as its smallest eigenvalue is strictly positive by Lemma 7. In addition, we demonstrate below that $\xi(\boldsymbol{\theta}) > 0$. Thus, by Theorem 8.5.11 of Harville [20, p. 98-99], $-\mathbb{E}\,\nabla^2_{\boldsymbol{\theta}}\,\widetilde{\ell}(\boldsymbol{\theta};\boldsymbol{X})$ is invertible on $\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\epsilon^\star)$. We bound $|||(-\mathbb{E}\,\nabla^2_{\boldsymbol{\theta}}\,\widetilde{\ell}(\boldsymbol{\theta};\boldsymbol{X}))^{-1}|||_\infty$ under Models 2 and 3, assuming that $D_N \geq 1$.

To bound the ℓ_∞-induced matrix norm of $(-\mathbb{E}\,\nabla^2_{\boldsymbol{\theta}}\,\widetilde{\ell}(\boldsymbol{\theta};\boldsymbol{X}))^{-1}$, observe that

$$|||(-\mathbb{E}\,\nabla^2_{\boldsymbol{\theta}}\,\widetilde{\ell}(\boldsymbol{\theta};\boldsymbol{X}))^{-1}|||_\infty \quad \leq \quad \max\{T_1,\, T_2\},$$

where

$$\begin{aligned} T_1 &\coloneqq |||\boldsymbol{A}(\boldsymbol{\theta})^{-1} + \xi(\boldsymbol{\theta})^{-1}\,(\boldsymbol{A}(\boldsymbol{\theta})^{-1}\,\boldsymbol{c}(\boldsymbol{\theta}))\,(\boldsymbol{A}(\boldsymbol{\theta})^{-1}\,\boldsymbol{c}(\boldsymbol{\theta}))^\top|||_\infty \\ &+ \|\xi(\boldsymbol{\theta})^{-1}\,\boldsymbol{A}(\boldsymbol{\theta})^{-1}\,\boldsymbol{c}(\boldsymbol{\theta})\|_\infty \\ T_2 &\coloneqq \|\xi(\boldsymbol{\theta})^{-1}\,\boldsymbol{A}(\boldsymbol{\theta})^{-1}\,\boldsymbol{c}(\boldsymbol{\theta})\|_1 + |\xi(\boldsymbol{\theta})|^{-1}. \end{aligned}$$

We bound the terms T_1 and T_2 one by one.

Bounding T_1. The term T_1 is defined as

$$\begin{aligned} T_1 &\coloneqq |||\boldsymbol{A}(\boldsymbol{\theta})^{-1} + \xi(\boldsymbol{\theta})^{-1}\,(\boldsymbol{A}(\boldsymbol{\theta})^{-1}\,\boldsymbol{c}(\boldsymbol{\theta}))\,(\boldsymbol{A}(\boldsymbol{\theta})^{-1}\,\boldsymbol{c}(\boldsymbol{\theta}))^\top|||_\infty \\ &+ \|\xi(\boldsymbol{\theta})^{-1}\,\boldsymbol{A}(\boldsymbol{\theta})^{-1}\,\boldsymbol{c}(\boldsymbol{\theta})\|_\infty. \end{aligned}$$

We bound the first term of T_1 by using the triangle inequality:

$$\begin{aligned} & |||\boldsymbol{A}(\boldsymbol{\theta})^{-1} + \xi(\boldsymbol{\theta})^{-1}\,(\boldsymbol{A}(\boldsymbol{\theta})^{-1}\,\boldsymbol{c}(\boldsymbol{\theta}))\,(\boldsymbol{A}(\boldsymbol{\theta})^{-1}\,\boldsymbol{c}(\boldsymbol{\theta}))^\top|||_\infty \\ \leq\; & |||\boldsymbol{A}(\boldsymbol{\theta})^{-1}|||_\infty + |\xi(\boldsymbol{\theta})|^{-1}\,|||(\boldsymbol{A}(\boldsymbol{\theta})^{-1}\,\boldsymbol{c}(\boldsymbol{\theta}))\,(\boldsymbol{A}(\boldsymbol{\theta})^{-1}\,\boldsymbol{c}(\boldsymbol{\theta}))^\top|||_\infty \\ =\; & |||\boldsymbol{A}(\boldsymbol{\theta})^{-1}|||_\infty + |\xi(\boldsymbol{\theta})|^{-1}\,\|\boldsymbol{A}(\boldsymbol{\theta})^{-1}\,\boldsymbol{c}(\boldsymbol{\theta})\|_\infty\,\|\boldsymbol{A}(\boldsymbol{\theta})^{-1}\,\boldsymbol{c}(\boldsymbol{\theta})\|_1 \\ \leq\; & |||\boldsymbol{A}(\boldsymbol{\theta})^{-1}|||_\infty + N\,|\xi(\boldsymbol{\theta})|^{-1}\,\|\boldsymbol{A}(\boldsymbol{\theta})^{-1}\,\boldsymbol{c}(\boldsymbol{\theta})\|_\infty^2 \\ \leq\; & |||\boldsymbol{A}(\boldsymbol{\theta})^{-1}|||_\infty + N\,|\xi(\boldsymbol{\theta})|^{-1}\,\|\boldsymbol{c}(\boldsymbol{\theta})\|_\infty^2\,|||\boldsymbol{A}(\boldsymbol{\theta})^{-1}|||_\infty^2 \\ =\; & |||\boldsymbol{A}(\boldsymbol{\theta})^{-1}|||_\infty\,(1 + N\,|\xi(\boldsymbol{\theta})|^{-1}\,\|\boldsymbol{c}(\boldsymbol{\theta})\|_\infty^2\,|||\boldsymbol{A}(\boldsymbol{\theta})^{-1}|||_\infty), \end{aligned}$$

taking advantage of the identity

$$|||\boldsymbol{v}\,\boldsymbol{v}^\top|||_\infty \;=\; \max_{1\leq i\leq N}\sum_{j=1}^N |v_i\,v_j| \;=\; \max_{1\leq i\leq N}|v_i|\sum_{j=1}^N |v_j| \;=\; \|\boldsymbol{v}\|_\infty\,\|\boldsymbol{v}\|_1$$

applied to the vector

$$\boldsymbol{v} \;\coloneqq\; \boldsymbol{A}(\boldsymbol{\theta})^{-1}\,\boldsymbol{c}(\boldsymbol{\theta}),$$

along with the fact that $\|\!|\boldsymbol{A}(\boldsymbol{\theta})^{-1}|\!\|_1 = \|\!|\boldsymbol{A}(\boldsymbol{\theta})^{-1}|\!\|_\infty$, following from the symmetry of $\boldsymbol{A}(\boldsymbol{\theta})$. The second term of T_1 can be bounded as follows:

$$\|\xi(\boldsymbol{\theta})^{-1}\, \boldsymbol{A}(\boldsymbol{\theta})^{-1}\, \boldsymbol{c}(\boldsymbol{\theta})\|_\infty \quad\leq\quad |\xi(\boldsymbol{\theta})|^{-1}\, \|\boldsymbol{c}(\boldsymbol{\theta})\|_\infty\, \|\!|\boldsymbol{A}(\boldsymbol{\theta})^{-1}|\!\|_\infty.$$

Combining these results gives the following bound on T_1:

$$T_1 \quad\leq\quad \|\!|\boldsymbol{A}(\boldsymbol{\theta})^{-1}|\!\|_\infty \left(1 + N\, |\xi(\boldsymbol{\theta})|^{-1} \|\boldsymbol{c}(\boldsymbol{\theta})\|_\infty^2 \|\!|\boldsymbol{A}(\boldsymbol{\theta})^{-1}|\!\|_\infty + |\xi(\boldsymbol{\theta})|^{-1} \|\boldsymbol{c}(\boldsymbol{\theta})\|_\infty\right).$$

Bounding T_2. The term T_2 is defined as follows:

$$T_2 \quad\coloneqq\quad \|\xi(\boldsymbol{\theta})^{-1}\, \boldsymbol{A}(\boldsymbol{\theta})^{-1}\, \boldsymbol{c}(\boldsymbol{\theta})\|_1 + |\xi(\boldsymbol{\theta})|^{-1}.$$

We bound T_2 by noting that

$$\begin{aligned} T_2 &= \|\xi(\boldsymbol{\theta})^{-1}\, \boldsymbol{A}(\boldsymbol{\theta})^{-1}\, \boldsymbol{c}(\boldsymbol{\theta})\|_1 + |\xi(\boldsymbol{\theta})|^{-1} \\ &\leq |\xi(\boldsymbol{\theta})|^{-1}\, (1 + \|\boldsymbol{c}(\boldsymbol{\theta})\|_1\, \|\!|\boldsymbol{A}(\boldsymbol{\theta})^{-1}|\!\|_1) \\ &= |\xi(\boldsymbol{\theta})|^{-1}\, (1 + \|\boldsymbol{c}(\boldsymbol{\theta})\|_1\, \|\!|\boldsymbol{A}(\boldsymbol{\theta})^{-1}|\!\|_\infty) \\ &\leq |\xi(\boldsymbol{\theta})|^{-1}\, (1 + N\, \|\boldsymbol{c}(\boldsymbol{\theta})\|_\infty\, \|\!|\boldsymbol{A}(\boldsymbol{\theta})^{-1}|\!\|_\infty), \end{aligned}$$

using the inequality $\|\boldsymbol{v}\|_1 \leq N\, \|\boldsymbol{v}\|_\infty$, where the step from $\|\!|\boldsymbol{A}(\boldsymbol{\theta})^{-1}|\!\|_1$ to $\|\!|\boldsymbol{A}(\boldsymbol{\theta})^{-1}|\!\|_\infty$ follows from the symmetry of $\boldsymbol{A}(\boldsymbol{\theta})$. We bound the terms in T_1 and T_2 one by one. The resulting bounds hold for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$.

Bounding $\|\!|\boldsymbol{A}(\boldsymbol{\theta})^{-1}|\!\|_\infty$. We have shown above that

$$\|\!|\boldsymbol{A}(\boldsymbol{\theta})^{-1}|\!\|_\infty \quad\leq\quad \frac{18\, \tau(\boldsymbol{\theta}^\star)^2}{N^{1-\alpha}}.$$

Bounding $\|\boldsymbol{c}(\boldsymbol{\theta})\|_\infty$. Lemma 10 proves that $\|\boldsymbol{c}(\boldsymbol{\theta})\|_\infty \leq 3\, D_N^3$.

Bounding $|\xi(\boldsymbol{\theta})|^{-1}$. The term $\xi(\boldsymbol{\theta})$ is defined as

$$\begin{aligned} \xi(\boldsymbol{\theta}) &\coloneqq v(\boldsymbol{\theta}) - \boldsymbol{c}(\boldsymbol{\theta})^\top\, \boldsymbol{A}(\boldsymbol{\theta})^{-1}\, \boldsymbol{c}(\boldsymbol{\theta}) \\ &= (\boldsymbol{c}(\boldsymbol{\theta})^\top\, \boldsymbol{c}(\boldsymbol{\theta})) \left(\frac{v(\boldsymbol{\theta})}{\boldsymbol{c}(\boldsymbol{\theta})^\top\, \boldsymbol{c}(\boldsymbol{\theta})} - \frac{\boldsymbol{c}(\boldsymbol{\theta})^\top\, \boldsymbol{A}(\boldsymbol{\theta})^{-1}\, \boldsymbol{c}(\boldsymbol{\theta})}{\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta})}\right). \end{aligned}$$

To bound $\xi(\boldsymbol{\theta})$, we leverage bounds established in Lemmas 8–10:

$$\begin{aligned} \boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta}) &\geq \frac{N}{256\, \tau(\boldsymbol{\theta}^\star)^8} \quad \text{by Lemma 9}, \\ \frac{v(\boldsymbol{\theta})}{\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta})} &\geq \frac{1}{144\, D_N^6\, \tau(\boldsymbol{\theta}^\star)^4} \quad \text{by Lemma 10}, \\ \frac{\boldsymbol{c}(\boldsymbol{\theta})^\top\, \boldsymbol{A}(\boldsymbol{\theta})^{-1}\, \boldsymbol{c}(\boldsymbol{\theta})}{\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta})} &\leq \frac{12\, \tau(\boldsymbol{\theta}^\star)^2}{N^{1-\alpha}} \quad \text{by Lemma 8}. \end{aligned}$$

All of the above quantities are well-defined, because $N \geq 3,\ D_N \geq 1,$ and $\tau(\boldsymbol{\theta}^\star) > 0$ under Models 2 and 3. These results help bound $\xi(\boldsymbol{\theta})$ as follows:

$$\begin{aligned}
\xi(\boldsymbol{\theta}) &= (\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta})) \left(\frac{v(\boldsymbol{\theta})}{\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta})} - \frac{\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{A}(\boldsymbol{\theta})^{-1} \boldsymbol{c}(\boldsymbol{\theta})}{\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta})}\right) \\
&\geq \frac{N}{256\, \tau(\boldsymbol{\theta}^\star)^8} \left(\frac{1}{144\, D_N^6\, \tau(\boldsymbol{\theta}^\star)^4} - \frac{12\, \tau(\boldsymbol{\theta}^\star)^2}{N^{1-\alpha}}\right) \\
&= \frac{N}{(256)\,(144)\, D_N^6\, \tau(\boldsymbol{\theta}^\star)^{12}} \left(1 - \frac{(12)\,(144)\, D_N^6\, \tau(\boldsymbol{\theta}^\star)^6}{N^{1-\alpha}}\right).
\end{aligned}$$

To bound the term $1 - (12)\,(144)\ D_N^6\, \tau(\boldsymbol{\theta}^\star)^6 \,/\, N^{1-\alpha}$, observe that the assumption

$$\|\boldsymbol{\theta}^\star\|_\infty \;\leq\; \frac{L + \vartheta\, \log N}{14\,(3 + D_N)} - \epsilon^\star$$

implies that the term $\tau(\boldsymbol{\theta}^\star)^6$ is bounded above by

$$\begin{aligned}
\tau(\boldsymbol{\theta}^\star)^6 &= \exp(6\,(3 + D_N)\,(\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)) \\
&\leq \exp\left(\frac{6}{14}\,(L + \vartheta\, \log N)\right) \\
&= \exp((3/7)\, L)\ N^{3\,\vartheta/7},
\end{aligned}$$

which in turn implies that

$$\frac{D_N^6\, \tau(\boldsymbol{\theta}^\star)^6}{N^{1-\alpha}} \;\leq\; \frac{\exp((3/7)\, L)\ D_N^6\ N^{3\,\vartheta/7}}{N^{1-\alpha}} \;\leq\; \frac{\exp((3/7)\, L)\ D_N^6}{N^{1/2}},$$

using the assumption that $\alpha \in [0,\, 1/2)$ and $\vartheta \in [0,\, 1/2 - \alpha)$. Since D_N must satisfy $D_N = O(\log N)$ to ensure $\|\boldsymbol{\theta}^\star\|_\infty > 0$, there exist constants $C_1 \in (0,\, 1)$ and $N_1 \geq 3$, independent of N and p, such that, for all $N > N_1$,

$$1 - \frac{(12)\,(144)\ D_N^6\ \tau(\boldsymbol{\theta}^\star)^6}{N^{1-\alpha}} \;\geq\; 1 - \frac{(12)\,(144)\ \exp((3/7)\, L)\ D_N^6}{N^{1/2}} \;\geq\; C_1 \;>\; 0.$$

We then obtain, for all $N > N_1$, that

$$\begin{aligned}
\xi(\boldsymbol{\theta}) &= \frac{N}{(256)\,(144)\ D_N^6\ \tau(\boldsymbol{\theta}^\star)^{12}} \left(1 - \frac{(12)\,(144)\ D_N^6\ \tau(\boldsymbol{\theta}^\star)^6}{N^{1-\alpha}}\right) \\
&\geq \frac{C_1\ N}{(256)\,(144)\ D_N^6\ \tau(\boldsymbol{\theta}^\star)^{12}},
\end{aligned}$$

which shows that $\xi(\boldsymbol{\theta}) > 0$ and hence

$$|\xi(\boldsymbol{\theta})|^{-1} \;=\; \frac{1}{\xi(\boldsymbol{\theta})} \;\leq\; \frac{C_2\, D_N^6\, \tau(\boldsymbol{\theta}^\star)^{12}}{N},$$

defining $C_2 \coloneqq (256)\,(144)\,/\,C_1 > 0$.

Bounding $\max\{T_1,\, T_2\}$. We have shown that

$$\begin{aligned}
T_1 &\leq |\!|\!|\boldsymbol{A}(\boldsymbol{\theta})^{-1}|\!|\!|_\infty \left(1 + N\, |\xi(\boldsymbol{\theta})|^{-1} \|\boldsymbol{c}(\boldsymbol{\theta})\|_\infty^2\, |\!|\!|\boldsymbol{A}(\boldsymbol{\theta})^{-1}|\!|\!|_\infty + |\xi(\boldsymbol{\theta})|^{-1} \|\boldsymbol{c}(\boldsymbol{\theta})\|_\infty\right)\\
T_2 &\leq |\xi(\boldsymbol{\theta})|^{-1}\, (1 + N\, \|\boldsymbol{c}(\boldsymbol{\theta})\|_\infty\, |\!|\!|\boldsymbol{A}(\boldsymbol{\theta})^{-1}|\!|\!|_\infty).
\end{aligned}$$

Using the bounds derived above, we obtain

$$\begin{aligned}
T_1 &\leq |\!|\!|\boldsymbol{A}(\boldsymbol{\theta})^{-1}|\!|\!|_\infty \left(1 + N\, |\xi(\boldsymbol{\theta})|^{-1}\, \|\boldsymbol{c}(\boldsymbol{\theta})\|_\infty^2\, |\!|\!|\boldsymbol{A}(\boldsymbol{\theta})^{-1}|\!|\!|_\infty + |\xi(\boldsymbol{\theta})|^{-1}\, \|\boldsymbol{c}(\boldsymbol{\theta})\|_\infty\right)\\
&\leq \frac{18\, \tau(\boldsymbol{\theta}^\star)^2}{N^{1-\alpha}} \left(1 + N \left(\frac{C_2\, D_N^6\, \tau(\boldsymbol{\theta}^\star)^{12}}{N}\right) (3\, D_N^3)^2 \left(\frac{18\, \tau(\boldsymbol{\theta}^\star)^2}{N^{1-\alpha}}\right) + \left(\frac{C_2\, D_N^6\, \tau(\boldsymbol{\theta}^\star)^{12}}{N}\right) (3\, D_N^3)\right)\\
&\leq \frac{18\, \tau(\boldsymbol{\theta}^\star)^2}{N^{1-\alpha}} \left(1 + \frac{(18)\,(3)^2\, C_2\, D_N^{12}\, \tau(\boldsymbol{\theta}^\star)^{14}}{N^{1-\alpha}} + \frac{3\, C_2\, D_N^9\, \tau(\boldsymbol{\theta}^\star)^{12}}{N}\right)\\
&= \frac{(18)^2\,(3)^2\, C_2\, D_N^{12}\, \tau(\boldsymbol{\theta}^\star)^{16}}{N^{1-\alpha}} \left(\frac{1}{(18)\,(3)^2\, C_2\, D_N^{12}\, \tau(\boldsymbol{\theta}^\star)^{14}} + \frac{1}{N^{1-\alpha}} + \frac{1}{(18)\,(3)\, D_N^3\, \tau(\boldsymbol{\theta}^\star)^2\, N}\right)\\
&\leq \frac{(18)^2\,(3)^2\, C_2\, D_N^{12}\, \tau(\boldsymbol{\theta}^\star)^{16}}{N^{1-\alpha}} \left(\frac{1}{C_2\, D_N^{12}\, \tau(\boldsymbol{\theta}^\star)^{14}} + \frac{1}{N^{1-\alpha}} + \frac{1}{D_N^3\, \tau(\boldsymbol{\theta}^\star)^2\, N}\right)\\
&\leq \frac{(3)^3\,(18)^2\, C_2\, D_N^{12}\, \tau(\boldsymbol{\theta}^\star)^{16}}{N^{1-\alpha}}\; \frac{1}{\min\{C_2\, D_N^{12}\, \tau(\boldsymbol{\theta}^\star)^{14},\; N^{1-\alpha},\; D_N^3\, \tau(\boldsymbol{\theta}^\star)^2\, N\}}
\end{aligned}$$

and

$$\begin{aligned}
T_2 &\leq |\xi(\boldsymbol{\theta})|^{-1}\, (1 + N\, \|\boldsymbol{c}(\boldsymbol{\theta})\|_\infty\, |\!|\!|\boldsymbol{A}(\boldsymbol{\theta})^{-1}|\!|\!|_\infty)\\
&\leq \frac{C_2\, D_N^6\, \tau(\boldsymbol{\theta}^\star)^{12}}{N} \left(1 + N D_N^3\, \frac{(3)\,(18)\, \tau(\boldsymbol{\theta}^\star)^2}{N^{1-\alpha}}\right)\\
&= \frac{C_2\, D_N^6\, \tau(\boldsymbol{\theta}^\star)^{12}}{N} \left(1 + (3)\,(18)\, D_N^3\, N^\alpha\, \tau(\boldsymbol{\theta}^\star)^2\right)\\
&\leq \frac{C_3\, D_N^9\, \tau(\boldsymbol{\theta}^\star)^{14}}{N^{1-\alpha}},
\end{aligned}$$

where $C_3 \coloneqq (2)\,(3)\,(18)\,C_2 > 0$, and using the fact that $D_N^3\, N^{\alpha}\, \tau(\boldsymbol{\theta}^\star)^2 \geq 1$ under Models 2 and 3. Define

$$\begin{aligned} U &\coloneqq \frac{C_3\, D_N^9\, \tau(\boldsymbol{\theta}^\star)^{14}}{N^{1-\alpha}} \\ V &\coloneqq \frac{C_4\, D_N^3\, \tau(\boldsymbol{\theta}^\star)^2}{\min\{C_2\, D_N^{12}\, \tau(\boldsymbol{\theta}^\star)^{14},\; N^{1-\alpha},\; D_N^3\, \tau(\boldsymbol{\theta}^\star)^2\, N\}}, \end{aligned}$$

so that the bounds on T_1 and T_2 can be stated in terms of U and V:

$$\begin{aligned} T_1 &\leq \frac{C_3\, C_4\, D_N^{12}\, \tau(\boldsymbol{\theta}^\star)^{16}}{N^{1-\alpha}\, \min\{C_2\, D_N^{12}\, \tau(\boldsymbol{\theta}^\star)^{14},\; N^{1-\alpha},\; D_N^3\, \tau(\boldsymbol{\theta}^\star)^2\, N\}} \;=\; U\, V \\ T_2 &\leq U, \end{aligned}$$

where $C_4 \coloneqq (3)^3\,(18)^2\, C_2 \,/\, C_3 > 0$. Thus,

$$\sup_{\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)} \|\!|(-\mathbb{E}\, \nabla_{\boldsymbol{\theta}}^2\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}))^{-1}\|\!|_\infty \;\leq\; \max\{T_1,\, T_2\} \;\leq\; \max\{U\, V,\; U\}.$$

To make the bound on $\max\{T_1,\, T_2\}$ as tight as possible, we need constants $C_5 \geq 1$ and $N_2 \geq N_1$, independent of N and p, such that, for all $N > N_2$,

$$V \;\coloneqq\; \frac{C_4\, D_N^3\, \tau(\boldsymbol{\theta}^\star)^2}{\min\{C_2\, D_N^{12}\, \tau(\boldsymbol{\theta}^\star)^{14},\; N^{1-\alpha},\; D_N^3\, \tau(\boldsymbol{\theta}^\star)^2\, N\}} \;\leq\; C_5.$$

Upon inspecting the denominator of V,

$$\min\{C_2\, D_N^{12}\, \tau(\boldsymbol{\theta}^\star)^{14},\; N^{1-\alpha},\; D_N^3\, \tau(\boldsymbol{\theta}^\star)^2\, N\},$$

and observing that $\alpha \in [0, 1/2)$, it is evident that

- the first term $C_2\, D_N^{12}\, \tau(\boldsymbol{\theta}^\star)^{14}$ grows either slower or faster than $N^{1/2}$ depending on the growth of D_N and $\tau(\boldsymbol{\theta}^\star)$;
- the second term $N^{1-\alpha}$ grows faster than $N^{1/2}$, because $1 - \alpha > 1/2$ for all $\alpha \in [0, 1/2)$;
- the third term $D_N^3\, \tau(\boldsymbol{\theta}^\star)^2\, N$ grows faster than $N^{1/2}$, because $D_N \geq 1$ and $\tau(\boldsymbol{\theta}^\star) \geq 1$ under Models 2 and 3.

To bound the first term, observe that the assumption

$$\|\boldsymbol{\theta}^\star\|_\infty \;\leq\; \frac{L + \vartheta\, \log N}{14\,(3 + D_N)} - \epsilon^\star$$

implies that the term $\tau(\boldsymbol{\theta}^\star)^{14}$ is bounded above by

$$\tau(\boldsymbol{\theta}^\star)^{14} \;=\; \exp(14\,(3+D_N)\,(\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)) \;\leq\; \exp(L + \vartheta\,\log N) \;=\; \exp(L)\; N^{\vartheta},$$

which in turn implies that the first term is bounded above by

$$C_2\; D_N^{12}\; \tau(\boldsymbol{\theta}^\star)^{14} \quad\leq\quad C_2\,\exp(L)\; D_N^{12}\; N^{\vartheta}.$$

Since $\alpha \in [0,\, 1/2)$ and $\vartheta \in [0,\, 1/2 - \alpha)$ under Models 2 and 3, the constant ϑ satisfies $\vartheta < 1/2$ while D_N needs to satisfy $D_N = O(\log N)$ to ensure $\|\boldsymbol{\theta}^\star\|_\infty > 0$. As a result, the first term grows slower than $N^{1/2}$, while the second and third term grow at least as fast as $N^{1/2}$. Thus, there exist constants $C_5 > 0$ and $N_2 \geq N_1$, independent of N and p, such that, for all $N > N_2$,

$$\begin{aligned} V \;&\coloneqq\; \frac{C_4\; D_N^3\; \tau(\boldsymbol{\theta}^\star)^2}{\min\{C_2\; D_N^{12}\; \tau(\boldsymbol{\theta}^\star)^{14},\; N^{1-\alpha},\; D_N^3\; \tau(\boldsymbol{\theta}^\star)^2\; N\}} \\ &=\; \frac{C_4\; D_N^3\; \tau(\boldsymbol{\theta}^\star)^2}{C_2\; D_N^{12}\; \tau(\boldsymbol{\theta}^\star)^{14}} \;=\; \frac{C_4}{C_2\; D_N^{9}\; \tau(\boldsymbol{\theta}^\star)^{12}} \;\leq\; C_5. \end{aligned}$$

It is worth noting that D_N and $\tau(\boldsymbol{\theta}^\star)$ may or may not increase as a function of N, but both quantities are bounded below by 1 under Models 2 and 3.

In conclusion, for all $N > N_2 \geq 3$,

$$\begin{aligned} \sup_{\boldsymbol{\theta}\,\in\,\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\,\epsilon^\star)} \|\!|(-\mathbb{E}\,\nabla_{\boldsymbol{\theta}}^2\; \widetilde{\ell}(\boldsymbol{\theta};\boldsymbol{X}))^{-1}|\!\|_\infty \;&\leq\; \max\{T_1,\, T_2\} \\ &\leq\; \max\{U\,V,\; U\} \\ &\leq\; \max\{1,\, C_5\}\; U \\ &=\; \frac{C_3\;\max\{1,\,C_5\}\; D_N^{9}\; \tau(\boldsymbol{\theta}^\star)^{14}}{N^{1-\alpha}} \\ &\leq\; \frac{C_3\;\max\{1,\,C_5\}\;\exp(L)\; D_N^{9}\; N^{\vartheta}}{N^{1-\alpha}} \\ &=\; \frac{C\,D_N^9}{N^{1-(\alpha+\vartheta)}}, \end{aligned}$$

assuming $\alpha \in [0,\, 1/2)$ and $\vartheta \in [0,\, 1/2 - \alpha)$, where the constant $C \coloneqq C_3\,\max\{1,\, C_5\}\,\exp(L) > 0$ is independent of N and p. □

Lemma 7. *Consider Models 1, 2, and 3 with* $\alpha \in [0,\, 1/2)$. *Then*

$$\inf_{\boldsymbol{\theta}\,\in\,\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\epsilon^\star)} \lambda_{\min}(\boldsymbol{A}(\boldsymbol{\theta})) \;\;\geq\;\; \frac{N^{1-\alpha}}{12\,\exp((3+D_N)\,(\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star))^2} \;\;>\;\; 0,$$

where $\lambda_{\min}(\boldsymbol{A}(\boldsymbol{\theta}))$ *is the smallest eigenvalue of* $\boldsymbol{A}(\boldsymbol{\theta})$.

Proof of Lemma 7. By definition,

$$\widetilde{\ell}(\boldsymbol{\theta};\, \boldsymbol{x}) \;\; = \;\; \sum_{i<j}^{N} \log \mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = x_{i,j} \mid \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}), \qquad \boldsymbol{x} \in \mathbb{X}.$$

Note that the conditional distribution of edge variable $X_{i,j}$ conditional on the event $\boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}$ is an exponential-family distribution with sufficient statistic vector $s(\boldsymbol{x})$ and natural parameter vector $\boldsymbol{\theta}$. Using standard properties of exponential families, it is straightforward to calculate, for each pair of nodes $\{i,j\} \subset \mathcal{N}$ and coordinates $(t,l) \in \{1,\dots,N\}^2$:

$$\begin{aligned} & -\sum_{i<j}^{N} \frac{\partial}{\partial\theta_t\, \partial\theta_l} \log \mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = x_{i,j} \mid \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}) \\ = \;\; & \sum_{i<j}^{N} \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{x}_{-\{i,j\}}}(s_t(\boldsymbol{X}), s_l(\boldsymbol{X})), \end{aligned}$$

where $\mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{x}_{-\{i,j\}}}(s_t(\boldsymbol{X}), s_l(\boldsymbol{X}))$ denotes the conditional covariance of $s_t(\boldsymbol{X})$ and $s_l(\boldsymbol{X})$, computed with respect to the conditional distribution of $X_{i,j}$ given $\boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}$. We have, for all $\{i,j\} \subset \mathcal{N}$ and $\boldsymbol{x}_{-\{i,j\}} \in \{0,1\}^{\binom{N}{2}-1}$,

$$\mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{x}_{-\{i,j\}}}(s_t(\boldsymbol{X}), s_l(\boldsymbol{X})) \;\; = \;\; \sum_{h_1 \,\in \mathcal{N} \setminus \{t\}} \; \sum_{h_2 \,\in \mathcal{N} \setminus \{l\}} \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{x}_{-\{i,j\}}}(X_{t,h_1},\, X_{l,h_2}),$$

as

$$s_t(\boldsymbol{X}) \;\; = \;\; \sum_{h \in \mathcal{N} \setminus \{t\}} X_{t,h}, \qquad t \in \{1,\dots,N\}.$$

For each pair of nodes $\{i,j\} \subset \mathcal{N}$, we distinguish two cases:

1. If either $t \notin \{i,j\}$ or $l \notin \{i,j\}$, then $s_t(\boldsymbol{X})$ and $s_l(\boldsymbol{X})$ cannot both be a function of $X_{i,j}$. It then follows that, conditional on $\boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}$,

$$\mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{x}_{-\{i,j\}}}(s_t(\boldsymbol{X}), s_l(\boldsymbol{X})) \;\; = \;\; 0,$$

as in this case either $s_t(\boldsymbol{X})$ or $s_l(\boldsymbol{X})$ will be almost surely constant.

2. If either $\{t,l\} = \{i,j\}$ or $t = l \in \{i,j\}$, then both $s_t(\boldsymbol{X})$ and $s_l(\boldsymbol{X})$ are functions of $X_{i,j}$. Conditional on $\boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}$, edge variables $X_{a,b}$ corresponding to pairs of nodes $\{a,b\} \neq \{i,j\}$ are almost surely constant, implying

$$\mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{x}_{-\{i,j\}}}(X_{t,h_1},\, X_{l,h_2}) \;\; = \;\; 0,$$

for all $\{t, h_1\} \neq \{i, j\}$ and all $\{l, h_2\} \neq \{i, j\}$. We then have, in the case $\{t, l\} = \{i, j\}$ $(t \neq l)$, that

$$\mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{x}_{-\{i,j\}}}(s_t(\boldsymbol{X}), s_l(\boldsymbol{X})) \;=\; \mathbb{V}_{\boldsymbol{\theta}, \boldsymbol{x}_{-\{i,j\}}}\, X_{i,j},$$

and in the case when $t = l \in \{i, j\}$,

$$\mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{x}_{-\{i,j\}}}(s_t(\boldsymbol{X}), s_l(\boldsymbol{X})) \;=\; \mathbb{V}_{\boldsymbol{\theta}, \boldsymbol{x}_{-\{i,j\}}}\, s_t(\boldsymbol{X}) \;=\; \mathbb{V}_{\boldsymbol{\theta}, \boldsymbol{x}_{-\{i,j\}}}\, X_{i,j}.$$

As a result, for all $t \neq l \in \{1, \ldots, N\}$,

$$\sum_{i<j}^{N} \mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{i,j\}}}(s_t(\boldsymbol{X}), s_l(\boldsymbol{X})) \;=\; \mathbb{E}\, \mathbb{V}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,l\}}}\, X_{t,l} \tag{C.5}$$

and all $t \in \{1, \ldots, N\}$,

$$\sum_{i<j}^{N} \mathbb{E}\, \mathbb{V}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{i,j\}}}\, s_t(\boldsymbol{X}) \;=\; \sum_{l \,\in\, \mathcal{N} \setminus \{t\}} \mathbb{E}\, \mathbb{V}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,l\}}}\, X_{t,l}. \tag{C.6}$$

An important consequence of (C.5) and (C.6) is that the matrix $\boldsymbol{A}(\boldsymbol{\theta})$ given in (C.4) is diagonally balanced, in the sense of Hillar and Wibisono [21]. Observe that

$$\begin{aligned} \mathbb{V}_{\boldsymbol{\theta}, \boldsymbol{x}_{-\{i,j\}}}\, X_{i,j} \;&=\; \mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = 1 \,|\, \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}) \\ &\times\; (1 - \mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = 1 \,|\, \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}})). \end{aligned}$$

Applying Lemma 13, for all $\boldsymbol{x}_{-\{i,j\}} \in \{0, 1\}^{\binom{N}{1} - 1}$,

$$\mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = 1 \,|\, \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}) \;\geq\; \frac{N^{-\alpha}}{1 + \exp((3 + D_N) \|\boldsymbol{\theta}\|_\infty)},$$

$$\mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = 1 \,|\, \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}) \;\leq\; \frac{1}{1 + \exp(-(3 + D_N) \|\boldsymbol{\theta}\|_\infty)},$$

noting that $D_N = 0$ under Model 1, which implies that

$$1 - \mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = 1 \,|\, \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}) \;\geq\; \frac{1}{1 + \exp((3 + D_N)\, \|\boldsymbol{\theta}\|_\infty)}.$$

Thus, for all pairs of nodes $\{i, j\} \subset \mathcal{N}$ and all $\boldsymbol{x}_{-\{i,j\}} \in \{0, 1\}^{\binom{N}{2} - 1}$,

$$\mathbb{V}_{\boldsymbol{\theta}, \boldsymbol{x}_{-\{i,j\}}}\, X_{i,j} \;\geq\; \frac{N^{-\alpha}}{(1 + \exp((3 + D_N)\, \|\boldsymbol{\theta}\|_\infty))^2},$$

which implies

$$\mathbb{E}\,\mathbb{V}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{i,j\}}} X_{i,j} \;\;\geq\;\; \frac{N^{-\alpha}}{(1+\exp((3+D_N)\,\|\boldsymbol{\theta}\|_\infty))^2}.$$

As a result, each element $A_{t,l}(\boldsymbol{\theta})$ of $\boldsymbol{A}(\boldsymbol{\theta})$ is bounded from below by

$$A_{t,l}(\boldsymbol{\theta}) \;\;\geq\;\; \frac{N^{-\alpha}}{(1+\exp((3+D_N)\,\|\boldsymbol{\theta}\|_\infty))^2} \;\;>\;\; 0. \tag{C.7}$$

By invoking Lemma 2.1 of Hillar and Wibisono [21] using the above bounds, the smallest eigenvalue $\lambda_{\min}(\boldsymbol{A}(\boldsymbol{\theta}))$ of the matrix $\boldsymbol{A}(\boldsymbol{\theta})$ satisfies

$$\lambda_{\min}(\boldsymbol{A}(\boldsymbol{\theta})) \;\;\geq\;\; \frac{N^{-\alpha}\,(N-2)}{(1+\exp((3+D_N)\,\|\boldsymbol{\theta}\|_\infty))^2} \;\;\geq\;\; \frac{N^{-\alpha}\,(N-2)}{4\,\exp((3+D_N)\,\|\boldsymbol{\theta}\|_\infty))^2}.$$

Using the inequality $N-2 \;\geq\; N\,/\,3$ (for $N \geq 3$), we obtain

$$\lambda_{\min}(\boldsymbol{A}(\boldsymbol{\theta})) \;\geq\; \frac{N^{-\alpha}\,(N-2)}{4\,\exp((3+D_N)\,\|\boldsymbol{\theta}\|_\infty))^2} \;\geq\; \frac{N^{1-\alpha}}{12\,\exp((3+D_N)\,\|\boldsymbol{\theta}\|_\infty))^2}.$$

Finally, for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$, we have $\|\boldsymbol{\theta}\|_\infty \leq \|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star$, implying

$$\inf_{\boldsymbol{\theta}\,\in\,\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\epsilon^\star)} \lambda_{\min}(\boldsymbol{A}(\boldsymbol{\theta})) \;\;\geq\;\; \frac{N^{1-\alpha}}{12\,\exp((3+D_N)\,(\|\boldsymbol{\theta}^\star\|_\infty+\epsilon^\star))^2}.$$

□

Lemma 8. *Consider Models 2 and 3 with* $\alpha \in [0,\,1/2)$. *Then*

$$\sup_{\boldsymbol{\theta}\,\in\,\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\epsilon^\star)} \frac{\boldsymbol{c}(\boldsymbol{\theta})^\top\,\boldsymbol{A}(\boldsymbol{\theta})^{-1}\,\boldsymbol{c}(\boldsymbol{\theta})}{\boldsymbol{c}(\boldsymbol{\theta})^\top\boldsymbol{c}(\boldsymbol{\theta})} \;\;\leq\;\; \frac{12\,\exp((3+D_N)\,(\|\boldsymbol{\theta}^\star\|_\infty+\epsilon^\star))^2}{N^{1-\alpha}}.$$

Proof of Lemma 8. Based on (C.4), define

$$R(\boldsymbol{A}(\boldsymbol{\theta})^{-1},\,\boldsymbol{c}(\boldsymbol{\theta})) \;\;\coloneqq\;\; \frac{\boldsymbol{c}(\boldsymbol{\theta})^\top\,\boldsymbol{A}(\boldsymbol{\theta})^{-1}\,\boldsymbol{c}(\boldsymbol{\theta})}{\boldsymbol{c}(\boldsymbol{\theta})^\top\,\boldsymbol{c}(\boldsymbol{\theta})},$$

recognizing $R(\boldsymbol{A}(\boldsymbol{\theta})^{-1},\,\boldsymbol{c}(\boldsymbol{\theta}))$ to be the Rayleigh quotient of $\boldsymbol{A}(\boldsymbol{\theta})^{-1} \in \mathbb{R}^{N\times N}$, assuming $\boldsymbol{c}(\boldsymbol{\theta}) \in \mathbb{R}^N \setminus \boldsymbol{0}$, where $\boldsymbol{0} \in \mathbb{R}^N$ denotes the N-dimensional zero vector. To bound $R(\boldsymbol{A}(\boldsymbol{\theta})^{-1},\,\boldsymbol{c}(\boldsymbol{\theta}))$, note that if $\lambda_1,\ldots,\lambda_N$ are the eigenvalues of $\boldsymbol{A}(\boldsymbol{\theta})$, then $1/\lambda_1$, …, $1/\lambda_N$ are the eigenvalues of $\boldsymbol{A}(\boldsymbol{\theta})^{-1}$. Let $\lambda_{\min}(\boldsymbol{A}(\boldsymbol{\theta}))$ denote the smallest eigenvalue of $\boldsymbol{A}(\boldsymbol{\theta})$, so that $1/\lambda_{\min}(\boldsymbol{A}(\boldsymbol{\theta}))$ is the largest eigenvalue of $\boldsymbol{A}(\boldsymbol{\theta})^{-1}$. Since the Rayleigh quotient of a matrix is bounded above by the largest eigenvalue of that matrix, we obtain, using Lemma 7,

$$R(\boldsymbol{A}(\boldsymbol{\theta})^{-1},\, \boldsymbol{c}(\boldsymbol{\theta})) \quad \leq \quad \frac{1}{\lambda_{\min}(\boldsymbol{A}(\boldsymbol{\theta}))} \quad \leq \quad \frac{12\, \exp((3 + D_N)\, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star))^2}{N^{1-\alpha}},$$

for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$. As a result,

$$\sup_{\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)} \frac{\boldsymbol{c}(\boldsymbol{\theta})^\top\, \boldsymbol{A}(\boldsymbol{\theta})^{-1}\, \boldsymbol{c}(\boldsymbol{\theta})}{\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta})} \quad \leq \quad \frac{12\, \exp((3 + D_N)\, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star))^2}{N^{1-\alpha}}.$$

□

Lemma 9. *Consider Models 2 and 3 with* $\alpha \in [0,\, 1/2)$*. Then*

$$\inf_{\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)} \boldsymbol{c}(\boldsymbol{\theta})^\top\, \boldsymbol{c}(\boldsymbol{\theta}) \quad \geq \quad \frac{N}{256\, \exp((3 + D_N)\, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star))^8}.$$

Proof of Lemma 9. From (C.4), the coordinates of $\boldsymbol{c}(\boldsymbol{\theta})$ are given by

$$c_t(\boldsymbol{\theta}) \quad = \quad \sum_{i<j}^{N} \mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{i,j\}}}(s_t(\boldsymbol{X}), s_{N+1}(\boldsymbol{X})), \qquad t \in \{1, \ldots, N\}.$$

Recall that

$$s_t(\boldsymbol{X}) \quad \coloneqq \quad \sum_{a \in \mathcal{N} \setminus \{t\}} X_{t,a}, \qquad t \in \{1, \ldots, N\}.$$

Then

$$\begin{aligned} c_t(\boldsymbol{\theta}) \quad &= \quad \sum_{i<j}^{N} \mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{i,j\}}}(s_t(\boldsymbol{X}), s_{N+1}(\boldsymbol{X})) \\ &= \quad \sum_{i<j}^{N} \mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{i,j\}}} \left(\sum_{a \in \mathcal{N} \setminus \{t\}} X_{t,a},\ s_{N+1}(\boldsymbol{X}) \right) \\ &= \quad \sum_{i<j}^{N} \sum_{a \in \mathcal{N} \setminus \{t\}} \mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{i,j\}}}(X_{t,a},\ s_{N+1}(\boldsymbol{X})) \\ &= \quad \sum_{a \in \mathcal{N} \setminus \{t\}} \mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\ s_{N+1}(\boldsymbol{X})). \end{aligned}$$

The last equality follows from that fact that $\mathbb{C}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{i,j\}}}(X_{t,a},\, s_{N+1}(\boldsymbol{X})) = 0$ almost surely for all $\{i,j\} \neq \{t,a\}$, as $\mathbb{C}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{i,j\}}}$ is the conditional covariance operator with respect to the conditional distribution of $X_{i,j}$ given $\boldsymbol{X}_{-\{i,j\}}$, implying $X_{t,a}$ is almost surely constant whenever $\{i,j\} \neq \{t,a\}$. Recall that

$$s_{N+1}(\boldsymbol{X}) \;\; \coloneqq \;\; \sum_{i<j}^{N} X_{i,j}\, I_{i,j}(\boldsymbol{X}),$$

where

$$I_{i,j}(\boldsymbol{X}) \;\; = \;\; \mathbb{1}\left(\sum_{h\,\in\,\mathcal{N}_i\,\cap\,\mathcal{N}_j} X_{i,h}\, X_{j,h} \;\geq\; 1\right), \qquad \{i,j\} \subset \mathcal{N}.$$

Thus, we have

$$\mathbb{E}\,\mathbb{C}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, s_{N+1}(\boldsymbol{X})) \;\; = \;\; \sum_{i<j}^{N} \mathbb{E}\,\mathbb{C}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, X_{i,j}\, I_{i,j}(\boldsymbol{X})),$$

implying

$$\begin{aligned} &\sum_{a\,\in\,\mathcal{N}\setminus\{t\}} \mathbb{E}\,\mathbb{C}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, s_{N+1}(\boldsymbol{X})) \\ &\quad = \;\; \sum_{a\,\in\,\mathcal{N}\setminus\{t\}} \sum_{i<j}^{N} \mathbb{E}\,\mathbb{C}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, X_{i,j}\, I_{i,j}(\boldsymbol{X})). \end{aligned}$$

The FKG inequality implies that

$$\mathbb{C}_{\boldsymbol{\theta},\boldsymbol{x}_{-\{t,a\}}}(X_{t,a},\, X_{i,j}\, I_{i,j}(\boldsymbol{X})) \;\; \geq \;\; 0 \quad \text{for all} \quad \boldsymbol{x}_{-\{t,a\}} \in \{0,1\}^{\binom{N}{2}-1},$$

because the conditional covariance is computed with respect to the conditional distribution of $X_{t,a}$ and both $X_{t,a}$ and $X_{i,j}\, I_{i,j}(\boldsymbol{X})$ ($\{i,j\} \subset \mathcal{N}$) are monotone non-decreasing functions of $X_{t,a}$. As a result,

$$\sum_{a\,\in\,\mathcal{N}\setminus\{t\}} \sum_{i<j}^{N} \mathbb{E}\,\mathbb{C}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{t,a\}}}(X_{t,a}, X_{i,j} I_{i,j}(\boldsymbol{X})) \;\geq\; \mathbb{E}\,\mathbb{C}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{t,a\}}}(X_{t,a}, X_{t,a} I_{t,a}(\boldsymbol{X})),$$

for some $a \in \mathcal{N}$ satisfying $\mathcal{N}_a \cap \mathcal{N}_t \neq \emptyset$. Such a node $a \in \mathcal{N}$ exists because each node $t \in \mathcal{N}$ belongs to one or more subpopulations $\mathcal{A}_k$ ($k \in \{1,\ldots,K\}$) and $|\mathcal{A}_k| \geq 3$ for all $k \in \{1,\ldots,K\}$. We then obtain

$$\boldsymbol{c}(\boldsymbol{\theta})^{\top}\, \boldsymbol{c}(\boldsymbol{\theta}) \;\; \geq \;\; \sum_{t=1}^{N} \left(\mathbb{E}\,\mathbb{C}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, X_{t,a}\, I_{t,a}(\boldsymbol{X}))\right)^{2}.$$

It is therefore enough to demonstrate that

$$\left(\mathbb{E}\,\mathbb{C}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, X_{t,a}\, I_{t,a}(\boldsymbol{X}))\right)^2 \;\geq\; \frac{1}{(1+\exp((3+D_N)\,(\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)))^8}.$$

Each node $t \in \mathcal{N}$ belongs to one or more subpopulations $\mathcal{A}_k$ for some $k \in \{1,\dots,K\}$. Since $|\mathcal{A}_l| \geq 3$ for all $l \in \{1,\dots,K\}$, there exists a node $b \in \mathcal{N}_t \cap \mathcal{N}_a$ so that $\mathbb{P}_{\boldsymbol{\theta}}(I_{t,a}(\boldsymbol{X}) = 1) \geq \mathbb{P}_{\boldsymbol{\theta}}(X_{t,b}\, X_{a,b} = 1)$, because the event $X_{t,b}\, X_{a,b} = 1$ implies the event $I_{t,a}(\boldsymbol{X}) = 1$. By Lemma 13, for pairs $\{i,j\} \subset \mathcal{N}$ with $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$, we have the bounds

$$\mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = 1 \mid \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}) \;\;\geq\;\; \frac{1}{1+\exp((3+D_N)\,\|\boldsymbol{\theta}\|_\infty)}$$

and

$$\mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = 0 \mid \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}) \;\;\geq\;\; \frac{1}{1+\exp((3+D_N)\,\|\boldsymbol{\theta}\|_\infty)}.$$

We can partition the sample space of $\boldsymbol{X}_{-\{t,a\}}$ based on whether $I_{t,a}(\boldsymbol{X}) = 0$ or $I_{t,a}(\boldsymbol{X}) = 1$. When $I_{t,a}(\boldsymbol{X}) = 0$,

$$\mathbb{C}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, X_{t,a}\, I_{t,a}(\boldsymbol{X})) \;\;=\;\; \mathbb{C}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, 0) \;\;=\;\; 0$$

and when $I_{t,a}(\boldsymbol{X}) = 1$,

$$\mathbb{C}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, X_{t,a}\, I_{t,a}(\boldsymbol{X})) \;\;=\;\; \mathbb{V}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{t,a\}}}\, X_{t,a}.$$

Using the above bounds, we obtain

$$\mathbb{V}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{t,a\}}}\, X_{t,a} \;\;\geq\;\; \left(\frac{1}{1+\exp((3+D_N)\,\|\boldsymbol{\theta}\|_\infty)}\right)^2.$$

Next, using the law of total expectation, we obtain

$$\begin{aligned}
&\mathbb{E}\,\mathbb{C}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, X_{t,a}\, I_{t,a}(\boldsymbol{X}))\\
=\;\; &\mathbb{P}(I_{t,a}(\boldsymbol{X}) = 1)\, \mathbb{V}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{t,a\}}}\, X_{t,a}\\
\geq\;\; &\mathbb{P}(X_{t,b}\; X_{a,b} = 1)\; \left(\frac{1}{1+\exp((3+D_N)\,\|\boldsymbol{\theta}\|_\infty)}\right)^2\\
\geq\;\; &\left(\frac{1}{1+\exp((3+D_N)\,(\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star))}\right)^2 \left(\frac{1}{1+\exp((3+D_N)\,\|\boldsymbol{\theta}\|_\infty)}\right)^2,
\end{aligned}$$

where the last inequality follows by writing

$$\begin{aligned}\mathbb{P}(X_{t,b}\, X_{a,b} = 1) &= \mathbb{P}(X_{t,b} = 1 \,|\, X_{a,b} = 1)\, \mathbb{P}(X_{a,b} = 1) \\ &\geq \left(\inf_{\boldsymbol{x}_{-\{i,j\}} \in \{0,1\}^{\binom{N}{2}-1}} \mathbb{P}(X_{i,j} = 1 \,|\, \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}})\right)^2 \\ &\geq \left(\frac{1}{1 + \exp((3 + D_N)\, \|\boldsymbol{\theta}^\star\|_\infty)}\right)^2 \\ &\geq \left(\frac{1}{1 + \exp((3 + D_N)\, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star))}\right)^2.\end{aligned}$$

For all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$, $\|\boldsymbol{\theta}\|_\infty \leq \|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star$, which implies that

$$\frac{1}{1 + \exp((3 + D_N)\, \|\boldsymbol{\theta}\|_\infty)} \quad \geq \quad \frac{1}{1 + \exp((3 + D_N)\, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star))}.$$

Thus, we have shown, for all $t \in \{1, \ldots, N\}$ and all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$, that

$$c_t(\boldsymbol{\theta})^2 \quad \geq \quad \frac{1}{(1 + \exp((3 + D_N)\, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)))^8},$$

which in turn implies, for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$, that

$$\begin{aligned}\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta}) &\geq \frac{N}{(1 + \exp((3 + D_N)\, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)))^8} \\ &\geq \frac{N}{(2\, \exp((3 + D_N)\, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)))^8} \\ &= \frac{N}{256\, \exp((3 + D_N)\, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star))^8}.\end{aligned}$$

As a result,

$$\inf_{\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)} \boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta}) \quad \geq \quad \frac{N}{256\, \exp((3 + D_N)\, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star))^8}.$$

□

Lemma 10. *Consider Models 2 and 3 with $\alpha \in [0,\, 1/2)$. Then*

$$\sup_{\boldsymbol{\theta} \,\in\, \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)} \|\boldsymbol{c}(\boldsymbol{\theta})\|_\infty \;\;\leq\;\; 3\, D_N^3$$

and

$$\inf_{\boldsymbol{\theta}\in\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)} \dfrac{v(\boldsymbol{\theta})}{\boldsymbol{c}(\boldsymbol{\theta})^\top\, \boldsymbol{c}(\boldsymbol{\theta})} \;\;\geq\;\; \dfrac{1}{144\; D_N^6\; \exp((3 + D_N)\, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star))^4},$$

recalling that $D_N \geq 1$ under Models 2 and 3.

Proof of Lemma 10. Recall that $s_{N+1}(\boldsymbol{X})$ is defined by

$$s_{N+1}(\boldsymbol{X}) \;\;\coloneqq\;\; \sum_{i<j}^{N} X_{i,j}\, I_{i,j}(\boldsymbol{X}),$$

where

$$I_{i,j}(\boldsymbol{X}) \;=\; \mathbb{1}\left(\sum_{h \,\in\, \mathcal{N}_i \,\cap\, \mathcal{N}_j} X_{i,h}\, X_{j,h} \;\geq\; 1\right), \qquad \{i,j\} \subset \mathcal{N}.$$

According to (C.4), $v(\boldsymbol{\theta})$ is given by

$$\begin{aligned} v(\boldsymbol{\theta}) \;\;&\coloneqq\;\; \sum_{i<j}^{N} \mathbb{E}\, \mathbb{V}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{i,j\}}}\, s_{N+1}(\boldsymbol{X}) \\ &=\;\; \sum_{i<j}^{N} \mathbb{E}\, \mathbb{V}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{i,j\}}} \left(\sum_{a<b}^{N} X_{a,b}\, I_{a,b}(\boldsymbol{X})\right). \end{aligned}$$

Given any pair of nodes $\{i,j\} \subset \mathcal{N}$, the FKG inequality implies, for all pairs of nodes $\{a,b\} \subset \mathcal{N}$ and $\{r,t\} \subset \mathcal{N}$, that

$$\mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{x}_{-\{i,j\}}}(X_{a,b}\, I_{a,b}(\boldsymbol{X}),\; X_{r,t}\, I_{r,t}(\boldsymbol{X})) \;\;\geq\;\; 0,$$

because the conditional covariance is computed with respect to the conditional distribution of $X_{i,j}$ and each $X_{a,b}\, I_{a,b}(\boldsymbol{X})$ ($\{a,b\} \subset \mathcal{N}$) is a monotone non-decreasing function of $X_{i,j}$. Thus,

$$\begin{aligned} v(\boldsymbol{\theta}) &= \sum_{i<j}^{N} \mathbb{E}\,\mathbb{V}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{i,j\}}} \left(\sum_{a<b}^{N} X_{a,b}\, I_{a,b}(\boldsymbol{X})\right) \\ &\geq \sum_{i<j}^{N} \sum_{a<b}^{N} \mathbb{E}\,\mathbb{V}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{i,j\}}} \left(X_{a,b}\, I_{a,b}(\boldsymbol{X})\right) \\ &\geq \sum_{i<j}^{N} \mathbb{E}\,\mathbb{V}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{i,j\}}} \left(X_{i,j}\, I_{i,j}(\boldsymbol{X})\right) \\ &= \sum_{i<j\,:\,\mathcal{N}_i\,\cap\,\mathcal{N}_j\neq\emptyset} \mathbb{E}\,\mathbb{V}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{i,j\}}} \left(X_{i,j}\, I_{i,j}(\boldsymbol{X})\right), \end{aligned}$$

noting that $I_{i,j}(\boldsymbol{X}) = 0$ almost surely when $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset$. We can then partition the sample space of $\boldsymbol{X}_{-\{i,j\}}$ based on whether $I_{i,j}(\boldsymbol{X}) = 0$ or $I_{i,j}(\boldsymbol{X}) = 1$. Using the law of total expectation,

$$\begin{aligned} &\mathbb{E}\,\mathbb{V}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{i,j\}}} \left(X_{i,j}\, I_{i,j}(\boldsymbol{X})\right) \\ =\;& \mathbb{E}\,(\mathbb{V}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{i,j\}}}\, X_{i,j}\, I_{i,j}(\boldsymbol{X}) \mid I_{i,j}(\boldsymbol{X}) = 1)\; \mathbb{P}(I_{i,j}(\boldsymbol{X}) = 1), \end{aligned}$$

noting $\mathbb{V}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{i,j\}}}\, X_{i,j}\, I_{i,j}(\boldsymbol{X}) = 0$ almost surely when $I_{i,j}(\boldsymbol{X}) = 0$. Hence,

$$v(\boldsymbol{\theta}) \;\geq\; \sum_{i<j\,:\,\mathcal{N}_i\,\cap\,\mathcal{N}_j\,\neq\,\emptyset}^{N} \mathbb{E}\,(\mathbb{V}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{i,j\}}}\, X_{i,j} \mid I_{i,j}(\boldsymbol{X}) = 1)\; \mathbb{P}(I_{i,j}(\boldsymbol{X}) = 1).$$

We bound

$$\mathbb{E}\,(\mathbb{V}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{i,j\}}}\, X_{i,j} \mid I_{i,j}(\boldsymbol{X}) = 1)$$

from below, for any $\boldsymbol{x} \in \{0,1\}^{\binom{N}{2}}$ with $I_{i,j}(\boldsymbol{x}) = 1$, by

$$\mathbb{V}_{\boldsymbol{\theta},\boldsymbol{x}_{-\{i,j\}}}\, X_{i,j} \;\geq\; \frac{1}{(1+\exp((3+D_N)\,\|\boldsymbol{\theta}\|_\infty))^2}. \tag{C.8}$$

The lower bound in (C.8) follows from Lemma 13, which shows that, for all $\boldsymbol{x}_{-\{i,j\}} \in \{0,1\}^{\binom{N}{2}-1}$ and $\{i,j\} \subset \mathcal{N}$ satisfying $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$,

$$\mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = 1 \mid \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}) \;\geq\; \frac{1}{1+\exp((3+D_N)\,\|\boldsymbol{\theta}\|_\infty)}$$

and

$$\mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = 0 \mid \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}) \;\; \geq \;\; \frac{1}{1 + \exp((3 + D_N)\, \|\boldsymbol{\theta}\|_\infty)}.$$

Next, given $h \in \mathcal{N}_i \,\cap\, \mathcal{N}_j \neq \emptyset$, the event $\{X_{i,h}\, X_{j,h} = 1\}$ implies the event $\{I_{i,j}(\boldsymbol{X}) = 1\}$, so that

$$\begin{aligned}\mathbb{P}(I_{i,j}(\boldsymbol{X}) = 1) \;\; &\geq \;\; \mathbb{P}(X_{i,h}\, X_{j,h} = 1) \\ &= \;\; \mathbb{P}(X_{i,h} = 1 \mid X_{j,h} = 1)\;\, \mathbb{P}(X_{j,h} = 1) \\ &\geq \;\; \left(\inf_{\boldsymbol{x}_{-\{i,j\}} \in \{0,1\}^{\binom{N}{2}-1}} \mathbb{P}(X_{i,j} = 1 \mid \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}})\right)^2 \\ &\geq \;\; \left(\frac{1}{1 + \exp((3 + D_N)\, \|\boldsymbol{\theta}\|_\infty)}\right)^2.\end{aligned}$$

Hence,

$$v(\boldsymbol{\theta}) \;\; \geq \;\; \sum_{i<j:\; \mathcal{N}_i \,\cap \mathcal{N}_j \,\neq \emptyset}^{N} \frac{1}{(1 + \exp((3 + D_N)\, \|\boldsymbol{\theta}\|_\infty))^4}.$$

Since $|\mathcal{A}_k| \geq 3$ $(k = 1, \dots, K)$, given a node $i \in \mathcal{N}$, there exists at least one other node $j \in \mathcal{N}$ such that $\mathcal{N}_i \,\cap\, \mathcal{N}_j \neq \emptyset$. Thus,

$$v(\boldsymbol{\theta}) \;\; \geq \;\; \frac{N}{(1 + \exp((3 + D_N)\, \|\boldsymbol{\theta}\|_\infty))^4}.$$

Observe that $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$ implies $\|\boldsymbol{\theta}\|_\infty \leq \|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star$, which in turn implies that

$$v(\boldsymbol{\theta}) \;\; \geq \;\; \frac{N}{(1 + \exp((3 + D_N)\, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)))^4} \quad \text{for all } \boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star).$$

We proceed to bound $\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta})$. We proved in Lemma 9 that

$$\begin{aligned}c_t(\boldsymbol{\theta}) \;\; &= \;\; \sum_{a \,\in \mathcal{N} \setminus \{t\}} \mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, s_{N+1}(\boldsymbol{X})) \\ &= \;\; \sum_{a \,\in \mathcal{N} \setminus \{t\}} \sum_{i<j}^{N} \mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, X_{i,j}\, I_{i,j}(\boldsymbol{X})) \\ &= \;\; \sum_{a \,\in \mathcal{N} \setminus \{t\} \,:\, \mathcal{N}_a \,\cap \mathcal{N}_t \,\neq \emptyset} \;\; \sum_{i<j}^{N} \mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, X_{i,j}\, I_{i,j}(\boldsymbol{X})),\end{aligned}$$

noting that, by Proposition 2, $X_{t,a}$ is independent of all other edge variables when $\mathcal{N}_t \cap \mathcal{N}_a = \emptyset$, in which case $\mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, X_{i,j}\, I_{i,j}(\boldsymbol{X})) = 0$. Hence,

$$\begin{aligned} & \sum_{a \,\in\, \mathcal{N} \setminus \{t\} \,:\, \mathcal{N}_a \,\cap\, \mathcal{N}_t \,\neq\, \emptyset} \; \sum_{i<j}^{N} \mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, X_{i,j}\, I_{i,j}(\boldsymbol{X})) \\ \leq \;\; & D_N^2 \left(\sup_{a \,\in\, \mathcal{N} \setminus \{t\}} \; \sum_{i<j}^{N} \mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, X_{i,j}\, I_{i,j}(\boldsymbol{X})) \right). \end{aligned}$$

This bound follows from Lemma 11, which shows that, for all $t \in \mathcal{N}$, that

$$|\{a \in \mathcal{N} \setminus \{t\} \,:\, \mathcal{N}_a \cap \mathcal{N}_t \neq \emptyset\}| \;\; \leq \;\; D_N^2.$$

If $\mathcal{N}_t \cap \mathcal{N}_a \neq \emptyset$, then $\mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, X_{i,j}\, I_{i,j}(\boldsymbol{X})) = 0$ if

1. $\{i, j\} \neq \{t, a\}$, in which case $X_{i,j}$ is constant almost surely, and
2. $I_{i,j}(\boldsymbol{X})$ is constant in $X_{t,a}$, implying $I_{i,j}(\boldsymbol{X})$ is constant almost surely.

The justification for the above statements regarding constancy is the fact that the conditional covariance $\mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, X_{i,j}\, I_{i,j}(\boldsymbol{X}))$ is computed with respect to the conditional distribution of $X_{t,a}$ conditional on $\boldsymbol{X}_{-\{t,a\}}$. It is therefore enough to bound

- the number of pairs $\{i, j\} \subset \mathcal{N}$ which do not satisfy either point 1. or 2. above, for a given $\{t, a\} \subset \mathcal{N}$, and
- the quantity $\mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, X_{i,j}\, I_{i,j}(\boldsymbol{X}))$.

Since $\mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, X_{i,j}\, I_{i,j}(\boldsymbol{X})) \leq 1$, we focus on the bounding the number of pairs $\{i, j\} \subset \mathcal{N}$ for which $X_{i,j}\, I_{i,j}(\boldsymbol{X})$ is a function of $X_{t,a}$:

- First, $\{i, j\} = \{t, a\}$ for only one pair $\{i, j\} \subset \mathcal{N}$.
- Second,

  $$I_{i,j}(\boldsymbol{X}) \;\; = \;\; \mathbb{1}\left(\sum_{h \in \mathcal{N}_i \,\cap\, \mathcal{N}_j} X_{i,h}\, X_{j,h} \;\geq\; 1 \right), \qquad \{i, j\} \neq \{t, a\},$$

  is a function of $X_{t,a}$ if and only if one of the following holds:

  1. $\{i, j\} \cap \{t, a\} = a$ and $t \in \mathcal{N}_i \cap \mathcal{N}_j$, or
  2. $\{i, j\} \cap \{t, a\} = t$ and $a \in \mathcal{N}_i \cap \mathcal{N}_j$.

  In either case, the number of possible pairs $\{i, j\} \neq \{t, a\}$ is bounded above by $\max\{|\mathcal{N}_i|,\, |\mathcal{N}_j|\} \leq D_N$ by Lemma 11, and hence is bounded above by $2\, D_N$ in either case.

Recalling $D_N \geq 1$ under Models 2 and 3, we have the bound

$$\sum_{a \in \mathcal{N} \setminus \{t\}} \sum_{i<j}^{N} \mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\; X_{i,j}\, I_{i,j}(\boldsymbol{X})) \;\; \leq \;\; D_N^2\,(1 + 2\, D_N) \;\; \leq \;\; 3\, D_N^3,$$

which shows, for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$, that

$$\|\boldsymbol{c}(\boldsymbol{\theta})\|_\infty \;\; \leq \;\; 3\, D_N^3$$

and

$$\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta}) \;\; \leq \;\; N\,(3\, D_N^3)^2 \;\; = \;\; 9\, D_N^6\, N.$$

Collecting terms reveals that, for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$,

$$\begin{aligned}
\frac{v(\boldsymbol{\theta})}{\boldsymbol{c}(\boldsymbol{\theta})^\top\, \boldsymbol{c}(\boldsymbol{\theta})} \;\; &\geq \;\; \frac{N}{(1 + \exp((3 + D_N)\,(\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star))^4} \left(\frac{1}{9\, D_N^6\, N}\right) \\
&\geq \;\; \frac{1}{9\, D_N^6\,(1 + \exp((3 + D_N)\,(\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)))^4} \\
&\geq \;\; \frac{1}{9\, D_N^6\,(2\, \exp((3 + D_N)\,(\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)))^4} \\
&= \;\; \frac{1}{144\, D_N^6\, \exp((3 + D_N)\,(\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star))^4}.
\end{aligned}$$

As a result,

$$\sup_{\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)} \|\boldsymbol{c}(\boldsymbol{\theta})\|_\infty \;\; \leq \;\; 3\, D_N^3$$

and

$$\inf_{\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)} \frac{v(\boldsymbol{\theta})}{\boldsymbol{c}(\boldsymbol{\theta})^\top\, \boldsymbol{c}(\boldsymbol{\theta})} \;\; \geq \;\; \frac{1}{144\; D_N^6\; \exp((3 + D_N)\,(\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star))^4}.$$

□

C.2. Bounding $\|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2$. To bound the spectral norm $\|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2$ of the coupling matrix $\mathcal{D}_N(\boldsymbol{\theta}^\star)$, we first review undirected graphical models encoding the conditional independence properties of generalized β-models with dependent edges in Appendices C.2.1 and C.2.2. We then bound $\|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2$ by using these conditional independence properties in Appendix C.2.3. Auxiliary results can be found in Appendix C.2.4.

C.2.1. *Undirected graphical models of random graphs.* Let $\mathcal{G}(\mathcal{V}, \mathcal{E})$ be an undirected graph with set of vertices $\mathcal{V}$ and set of edges

$$\mathcal{E} \;\; \subseteq \;\; \big\{\{v, w\} \, : \, v \in \mathcal{V},\, w \in \mathcal{V} \setminus \{v\}\big\}.$$

An undirected graphical model of a random graph [25] is a family of probability measures $\{\mathbb{P}_{\boldsymbol{\theta}},\, \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ dominated by a σ-finite measure ν, with factorization and conditional independence properties [14] of the form

$$f_{\boldsymbol{\theta}}\,(\boldsymbol{x}) \;\; \coloneqq \;\; \frac{\mathrm{d}\,\mathbb{P}_{\boldsymbol{\theta}}}{\mathrm{d}\,\nu}(\boldsymbol{x}) \;\; \propto \;\; \prod_{\mathcal{C}\,\in\,\mathfrak{C}} g_{\mathcal{C}}(\boldsymbol{x}_{\mathcal{C}};\, \boldsymbol{\theta}), \qquad \boldsymbol{x} \in \mathbb{X}, \tag{C.9}$$

where $\mathfrak{C}$ is the set of all maximal complete subsets of the conditional independence graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ with set of vertices $\mathcal{V} = \{X_1, \ldots, X_M\}$ and set of edges $\mathcal{E} \subset \{\{v, w\} \, : \, v \in \mathcal{V},\, w \in \mathcal{V} \setminus \{v\}\}$. The functions $g_{\mathcal{C}} : \mathbb{X} \times \boldsymbol{\Theta} \mapsto \mathbb{R}^{+} \cup \{0\}$ are non-negative functions defined on the maximal complete subsets $\mathcal{C} \in \mathfrak{C}$ of the conditional independence graph $\mathcal{G}$. A complete subset of the conditional independence graph $\mathcal{G}$ is a subset of vertices such that each pair of vertices is connected by an edge, and a complete subset is maximal complete if no vertices can be added without losing the property of completeness.

The probability density functions introduced in Section 2 are of the form

$$f_{\boldsymbol{\theta}}\,(\boldsymbol{x}) \;\; \propto \;\; \prod_{i\,<\,j}^{N} \varphi_{i,j}(x_{i,j},\, \boldsymbol{x}_{\mathcal{S}_{i,j}};\, \boldsymbol{\theta}), \qquad \boldsymbol{x} \in \mathbb{X}, \tag{C.10}$$

where $\mathcal{S}_{i,j} \subset \{\{v, w\} \, : \, v \in \mathcal{N},\, w \in \mathcal{N} \setminus \{w\}\} \setminus \{i, j\}$ for all $\{i, j\} \subset \mathcal{N}$. Probability density functions of the form (C.10) can be represented as probability density functions of the form (C.9) by grouping the functions $\varphi_{i,j}$ in accordance with the maximal complete subsets of conditional independence graph $\mathcal{G}$. The conditional independence graph $\mathcal{G}$ depends on the model: e.g., the conditional independence graph of Model 1 has no edges, because all edge variables are independent. By contrast, the conditional independence graph of Models 2 and 3 shown in Figure 2 has edges, which indicate the absence of conditional independence among edge variables due to brokerage in overlapping subpopulations.

To distinguish the random graph (representing data structure) from the conditional independence graph $\mathcal{G}$ (representing conditional independence structure, i.e., model structure), we call elements of $\mathcal{V}$ vertices rather than nodes, and elements of $\mathcal{E}$ edges rather than edge variables.

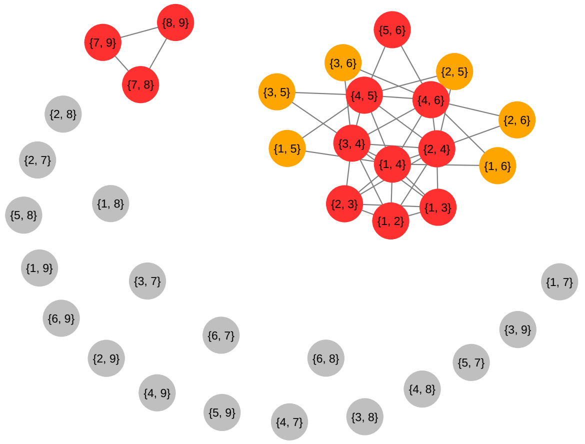


FIG 2. *The conditional independence graph of Models 2 and 3 with population of nodes* $\mathcal{N} := \{1, \ldots, 9\}$*, consisting of overlapping subpopulations* $\mathcal{A}_1 := \{1, 2, 3, 4\}$*,* $\mathcal{A}_2 := \{4, 5, 6\}$*, and* $\mathcal{A}_3 := \{7, 8, 9\}$*. Edge variables* $X_{i,j}$ *are represented by circles with labels* $\{i, j\}$*. If nodes* i *and* j *share a subpopulation,* $X_{i,j}$ *is colored red. If nodes* i *and* j *do not share a subpopulation but belong to overlapping subpopulations,* $X_{i,j}$ *is colored orange. Otherwise,* $X_{i,j}$ *is colored gray.*

C.2.2. *Conditional independence properties.* We prove selected conditional independence properties that help establish consistency results and convergence rates for generalized β-models with dependent edges.

By Equations (2.1) and (2.2), for each $\{i, j\} \subset \mathcal{N}$,

$$\varphi_{i,j}(x_{i,j},\, \boldsymbol{x}_{\mathcal{S}_{i,j}};\, \boldsymbol{\theta}) \;:=\; a_{i,j}(x_{i,j}) \exp\left((\theta_i + \theta_j)\, x_{i,j} + \theta_{N+1}\, b_{i,j}(x_{i,j},\, \boldsymbol{x}_{\mathcal{S}_{i,j}})\right),$$

where

$$b_{i,j}(x_{i,j}, \boldsymbol{x}_{\mathcal{S}_{i,j}}) \quad := \quad \begin{cases} 0 & \text{if } \mathcal{N}_i \cap \mathcal{N}_j = \emptyset \\ x_{i,j}\, \mathbb{1}\left(\displaystyle\sum_{h \,\in\, \mathcal{N}_i \,\cap\, \mathcal{N}_j} x_{i,h}\, x_{j,h} \geq 1\right) & \text{if } \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset. \end{cases}$$

Definition 1. Neighborhood intersection property. *Consider a random graph model with a probability density function parameterized by* (2.1) *and* (2.2)*. If* $\mathcal{S}_{i,j} = \{\{a, b\} \subset \mathcal{N} : (a, b) \in \{i, j\} \times \{\mathcal{N}_i \cap \mathcal{N}_j\}\}$ *for all pairs of nodes* $\{i, j\} \subset \mathcal{N}$*, then the random graph is said to satisfy the neighborhood intersection property.*

By construction, generalized β-models with dependent edges satisfy the neighborhood intersection property, which implies conditional independence properties, including—but not limited to—the conditional independence properties established in Proposition 2 below. Define

$$\mathfrak{D}_N \quad := \quad \big\{\{a, b\} \,:\, a \in \mathcal{N},\; b \in \mathcal{N} \setminus \{a\}\big\}.$$

We will utilize $\mathfrak{D}_N$ as the index set of all possible edge variables, which will be useful in constructing statements which exclude certain edge variables.

Proposition 2. *A random graph with overlapping subpopulations $\mathcal{A}_k$ of sizes $|\mathcal{A}_k| \geq 3$ $(k = 1, \ldots, K)$ satisfying the neighborhood intersection property possesses the following conditional independence properties:*

1. *For all pairs of nodes $\{i, j\} \subset \mathcal{N}$ such that $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset$:*

$$X_{i,j} \quad \perp\!\!\!\perp \quad \boldsymbol{X}_{\mathfrak{D}_N \setminus \{i,j\}}.$$

2. *For all pairs of nodes $\{i, j\} \subset \mathcal{N}$ such that $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$ and there exists $k \in \{1, \ldots, K\}$ such that $\{i, j\} \subset \mathcal{A}_k$:*

$$X_{i,j} \quad \perp\!\!\!\perp \quad \boldsymbol{X}_{\mathfrak{D}_N \setminus (\{i,j\} \cup \mathfrak{N}_{i,j})} \mid \boldsymbol{X}_{\mathfrak{N}_{i,j}},$$

where $\mathfrak{N}_{i,j} = \mathcal{J}_i^{(j)} \cup \mathcal{J}_j^{(i)}$, using the definition

$$\mathcal{J}_t^{(v)} \quad := \quad \left[\bigcup_{b \in \mathcal{N}_t \setminus \{v\}} \{v, b\}\right] \cup \left[\bigcup_{b \in \mathcal{N}_t} \big\{\{a, b\} : (a, b) \in \{v, b\} \times \mathcal{N}_v \cap \mathcal{N}_b\big\}\right],$$

and with the property that

$$\mathfrak{N}_{i,j} \quad \subseteq \quad \big\{\{a, b\} \,:\, a \in \mathcal{N}_i \cup \mathcal{N}_j,\, b \in \mathcal{N}_i \cup \mathcal{N}_j \setminus \{a\}\big\}.$$

3. *For all pairs of nodes $\{i, j\} \subset \mathcal{N}$ such that $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$ and there exists no $k \in \{1, \ldots, K\}$ such that $\{i, j\} \subset \mathcal{A}_k$:*

$$X_{i,j} \quad \perp\!\!\!\perp \quad \boldsymbol{X}_{\mathfrak{D}_N \setminus (\{i,j\} \cup \mathcal{S}_{i,j})} \mid \boldsymbol{X}_{\mathcal{S}_{i,j}}.$$

where $\mathcal{S}_{i,j} = \{\{a, b\} \subset \mathcal{N} \,:\, (a, b) \in \{i, j\} \times \{\mathcal{N}_i \cap \mathcal{N}_j\}\}$.

Proof of Proposition 2. In the following, we use the characterizations of conditional independence due to Dawid [14], which relate factorization properties of probability density functions to conditional independence properties. Using these characterizations of conditional independence, we establish the conditional independence properties of Proposition 2 by showing that, for each pair of nodes $\{i, j\} \subset \mathcal{N}$, there exists a subset of edge indices $\mathfrak{N}_{i,j} \subseteq \mathfrak{D}_N \setminus \{i,j\}$ and non-negative functions g and h such that the probability density function can be written as

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) \;\; \propto \;\; g(x_{i,j},\, \boldsymbol{x}_{\mathfrak{N}_{i,j}})\; h(\boldsymbol{x}_{\mathfrak{D}_N \setminus \{i,j\}}),$$

where $\mathfrak{D}_N \coloneqq \{\{a,b\} \,:\, a \in \mathcal{N},\, b \in \mathcal{N} \setminus \{a\}\}$, implying that

$$X_{i,j} \;\; \perp\!\!\!\perp \;\; \boldsymbol{X}_{\mathfrak{D}_N \setminus (\{i,j\} \,\cup\, \mathfrak{N}_{i,j})} \mid \boldsymbol{X}_{\mathfrak{N}_{i,j}}.$$

Proposition 2 assumes that the neighborhood intersection property is satisfied, allowing us to write

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) \;\; \propto \;\; \prod_{i\,<\,j}^{N} \varphi_{i,j}(x_{i,j},\, \boldsymbol{x}_{\mathcal{S}_{i,j}}),$$

where

$$\mathcal{S}_{i,j} \;\; = \;\; \{\{v,w\} : (v,w) \in \{i,j\} \times \mathcal{N}_i \cap \mathcal{N}_j\}\,,$$

recalling the definition of the node neighborhood sets $\mathcal{N}_i$ $(i \in \mathcal{N})$:

$$\mathcal{N}_i \;\; = \;\; \{h \in \mathcal{N} \setminus \{i\} \;:\; \text{ exists } k \in \{1,\ldots,K\} \text{ such that } \{i,h\} \subset \mathcal{A}_k\}.$$

Condition 1: Consider any pair of nodes $\{i,j\} \subset \mathcal{N}$ with $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset$, that is, nodes i and j neither belong to a common subpopulations nor belong to distinct subpopulations that overlap. Since $\{i,j\} \times \mathcal{N}_i \cap \mathcal{N}_j = \emptyset$,

$$\varphi_{i,j}(x_{i,j},\, \boldsymbol{x}_{\mathcal{S}_{i,j}}) \;\; \equiv \;\; \varphi_{i,j}(x_{i,j}).$$

It remains to check whether any $\varphi_{a,b}$ with $\{a,b\} \neq \{i,j\}$ can be a function of $x_{i,j}$, which can happen in one of two different ways:

- $i \in \{a,b\}$ and $j \in \mathcal{N}_a \cap \mathcal{N}_b$, in which case, by the definition of the node neighborhood sets $\mathcal{N}_v$ $(v \in \mathcal{N})$, it must be that $j \in \mathcal{N}_i$; or
- $j \in \{a,b\}$ and $i \in \mathcal{N}_a \cap \mathcal{N}_b$, in which case, similarly, $i \in \mathcal{N}_j$.

We prove that there exists no $\varphi_{a,b}$ with $\{a,b\} \neq \{i,j\}$ which is a function of $x_{i,j}$ by contradiction. Note $i \in \mathcal{N}_j$ implies the existence of a $k \in \{1,\ldots,K\}$ such that $\{i,j\} \subset \mathcal{A}_k$. By assumption, $|\mathcal{A}_k| \geq 3$ for all $k \in \{1,\ldots,K\}$, implying there exists at least one other node $h \in \mathcal{A}_k \setminus \{i,j\}$. Thus, if $\{i,j,h\} \subseteq \mathcal{A}_k$, then both $h \in \mathcal{N}_i$ and $h \in \mathcal{N}_j$ must hold, violating the assumption that $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset$. The case when $j \in \{a,b\}$ and $i \in \mathcal{N}_a \cap \mathcal{N}_b$ is proved similarly. Therefore, there cannot exist a pair of nodes $\{a,b\} \neq \{i,j\}$ such that $\varphi_{a,b}$ is a function of $x_{i,j}$. As a consequence, taking

$$g(x_{i,j}) \;\; = \;\; \varphi_{i,j}(x_{i,j})$$

and

$$h(\boldsymbol{x}_{\mathfrak{D}_N \setminus \{i,j\}}) \quad = \prod_{a<b:\ \{a,b\} \neq \{i,j\}}^{N} \varphi_{a,b}(x_{a,b},\, \boldsymbol{x}_{\mathcal{S}_{a,b}})$$

shows that $f_{\boldsymbol{\theta}}(\boldsymbol{x})$ can be written as

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) \quad \propto \quad g(x_{i,j})\; h(\boldsymbol{x}_{\mathfrak{D}_N \setminus \{i,j\}}),$$

which implies $X_{i,j} \perp\!\!\!\perp \boldsymbol{X}_{\mathfrak{D}_N \setminus \{i,j\}}$, i.e., $X_{i,j}$ is independent of all others edges.

Condition 2: Consider any pair of nodes $\{i,j\} \subset \mathcal{N}$ with $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$ and such that there exists a $k \in \{1, \ldots, K\}$ such that $\{i,j\} \subset \mathcal{A}_k$. By definition, $\varphi_{i,j}(x_{i,j}, \boldsymbol{x}_{\mathcal{S}_{i,j}})$ is a function of $x_{i,j}$. Recall the key condition for the neighborhood intersection assumption, which was that $\mathcal{S}_{a,b}$ satisfies

$$\mathcal{S}_{a,b} \quad = \quad \{\{v,w\} : (v,w) \in \{a,b\} \times \mathcal{N}_a \cap \mathcal{N}_b\}\,, \qquad \{a,b\} \subset \mathcal{N}.$$

For any $\varphi_{a,b}(x_{a,b}, \boldsymbol{x}_{\mathcal{S}_{a,b}})$ with $\{a,b\} \neq \{i,j\}$ to be a function of $x_{i,j}$, one of the following must hold:

1. $i \in \{a,b\}$ and $j \in \mathcal{N}_a \cap \mathcal{N}_b$, in which case, by the definition of the node neighborhood sets $\mathcal{N}_v$ $(v \in \mathcal{N})$, $j \in \mathcal{N}_i$; or
2. $j \in \{a,b\}$ and $i \in \mathcal{N}_a \cap \mathcal{N}_b$, in which case, similarly, $i \in \mathcal{N}_j$.

Consider the first case: $i \in \{a,b\}$ and $j \in \mathcal{N}_a \cap \mathcal{N}_b$, and without loss, take $a = i$. The condition for this case implies that $j \in \mathcal{N}_i \cap \mathcal{N}_b$. By assumption, $\{i,j\} \subset \mathcal{A}_k$ for some $k \in \{1, \ldots, K\}$, which implies $j \in \mathcal{N}_i$. If $j \in \mathcal{N}_b$, then $b \in \mathcal{N}_j$, implying $\varphi_{i,b}$ is a function of $x_{i,j}$ for all $\{i,b\} \subset \mathcal{N}$ with $b \in \mathcal{N}_j$. Applying the same argument to the second case where $j \in \{a,b\}$ and $i \in \mathcal{N}_a \cap \mathcal{N}_b$ reveals that $\varphi_{j,b}$ is a function of $x_{i,j}$ for all $\{j,b\}$ with $b \in \mathcal{N}_i$.

Summarily, $\varphi_{a,b}$ is a function of $x_{i,j}$ if it is in the following list:

- $\varphi_{i,j}(x_{i,j}, \boldsymbol{x}_{\mathcal{S}_{i,j}})$, where $\{v,w\} \in \mathcal{S}_{i,j}$ if $(v,w) \in \{i,j\} \times \mathcal{N}_i \cap \mathcal{N}_j$.
- $\varphi_{i,b}(x_{i,b}, \boldsymbol{x}_{\mathcal{S}_{i,b}})$ $(b \in \mathcal{N}_j)$, where $\{v,w\} \in \mathcal{S}_{i,b}$ if $(v,w) \in \{i,b\} \times \mathcal{N}_i \cap \mathcal{N}_b$.
- $\varphi_{j,b}(x_{j,b}, \boldsymbol{x}_{\mathcal{S}_{j,b}})$ $(b \in \mathcal{N}_i)$, where $\{v,w\} \in \mathcal{S}_{j,b}$ if $(v,w) \in \{j,b\} \times \mathcal{N}_j \cap \mathcal{N}_b$.

This collection of functions is a function of all edge variables $X_{a,b}$ with indices $\{a,b\}$ in $\mathcal{J}_i^{(j)} \cup \mathcal{J}_j^{(i)} \cup \{i,j\}$, where

$$\mathcal{J}_t^{(v)} \quad \coloneqq \quad \left[\bigcup_{b \in \mathcal{N}_t \setminus \{v\}} \{v,b\}\right] \;\cup\; \left[\bigcup_{b \in \mathcal{N}_t} \big\{\{a,b\} : (a,b) \in \{v,b\} \times \mathcal{N}_v \cap \mathcal{N}_b\big\}\right].$$

Thus, there exist non-negative functions g and h such that the probability density function can be written as follows:

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) \quad \propto \quad g(x_{i,j}, \boldsymbol{x}_{\mathfrak{N}_{i,j}})\; h(\boldsymbol{x}_{\mathfrak{D}_N \setminus \{i,j\}}),$$

where $\mathfrak{N}_{i,j} = \mathcal{J}_i^{(j)} \cup \mathcal{J}_j^{(i)}$, implying that

$$X_{i,j} \quad \perp\!\!\!\perp \quad \boldsymbol{X}_{\mathfrak{D}_N \setminus (\{i,j\} \cup \mathfrak{N}_{i,j})} \mid \boldsymbol{X}_{\mathfrak{N}_{i,j}}.$$

As $\{i,j\} \subset \mathcal{N}_i \cup \mathcal{N}_j$,

$$\mathfrak{N}_{i,j} \quad \subseteq \quad \big\{\{a,b\} \,:\, a \in \mathcal{N}_i \cup \mathcal{N}_j,\, b \in \mathcal{N}_i \cup \mathcal{N}_j \setminus \{a\}\big\}.$$

Condition 3: Consider any pair of nodes $\{i,j\} \subset \mathcal{N}$ with $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$ and such that there exists no $k \in \{1,\ldots,K\}$ such that $\{i,j\} \subset \mathcal{A}_k$. It is clear that $\varphi_{i,j}$ is a function of $x_{i,j}$. For any $\varphi_{a,b}(x_{a,b}, \boldsymbol{x}_{\mathcal{S}_{a,b}})$ with $\{a,b\} \neq \{i,j\}$ to be a function of $x_{i,j}$, one of the following must hold:

- $i \in \{a,b\}$ and $j \in \mathcal{N}_a \cap \mathcal{N}_b$, in which case, by the definition of the node neighborhood sets $\mathcal{N}_v$ ($v \in \mathcal{N}$), it must be that $j \in \mathcal{N}_i$; or
- $j \in \{a,b\}$ and $i \in \mathcal{N}_a \cap \mathcal{N}_b$, in which case, similarly, $i \in \mathcal{N}_j$.

In both conditions, $i \in \mathcal{N}_j$ and $j \in \mathcal{N}_i$, which implies that $\{i,j\} \subset \mathcal{A}_k$ for some $k \in \{1,\ldots,K\}$, violating the assumption that no such k exists. Thus, $\varphi_{a,b}(x_{a,b}, \boldsymbol{x}_{\mathcal{S}_{a,b}})$ is a function of $x_{i,j}$ if and only if $\{a,b\} = \{i,j\}$. As a result, there exist non-negative functions g and h such that

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) \quad \propto \quad g(x_{i,j},\, \boldsymbol{x}_{\mathcal{S}_{i,j}})\; h(\boldsymbol{x}_{\mathfrak{D}_N \setminus \{i,j\}})$$

which implies $X_{i,j} \perp\!\!\!\perp \boldsymbol{X}_{\mathfrak{D}_N \setminus (\{i,j\} \cup \mathcal{S}_{i,j})} \mid \boldsymbol{X}_{\mathcal{S}_{i,j}}$. □

Lemma 11. *Consider Models 2 and 3. Then* $\max_{t \in \mathcal{N}} |\mathcal{N}_t| \leq D_N$ *and*

$$\max_{t \in \mathcal{N}} |\{a \in \mathcal{N} \setminus \{t\} \,:\, \mathcal{N}_a \cap \mathcal{N}_t \neq \emptyset\}| \quad \leq \quad D_N^2,$$

where $D_N \coloneqq \max_{\{i,j\} \subset \mathcal{N}} |\mathfrak{N}_{i,j}|$.

Proof of Lemma 11. By Proposition 2, for any $\{i,j\} \subset \mathcal{N}$ satisfying $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$ and for which there exists $k \in \{1,\ldots,K\}$ such that $\{i,j\} \subset \mathcal{A}_k$, we have $\mathfrak{N}_{i,j} = \mathcal{J}_i^{(j)} \cup \mathcal{J}_j^{(i)}$, where, for all $\{t,v\} \subset \mathcal{N}$,

$$\mathcal{J}_t^{(v)} \quad \coloneqq \quad \left[\bigcup_{b \in \mathcal{N}_t \setminus \{v\}} \{v,b\}\right] \;\cup\; \left[\bigcup_{b \in \mathcal{N}_t} \big\{\{a,b\} : (a,b) \in \{v,b\} \times \mathcal{N}_v \cap \mathcal{N}_b\big\}\right].$$

Then, for each $t \in \mathcal{N}$, there exists $v \in \mathcal{N} \setminus \{t\}$ and $k \in \{1, \dots, K\}$ such that $\{t, v\} \subset \mathcal{A}_k$, due to the assumption that $|\mathcal{A}_k| \geq 3$ for all $k \in \{1, \dots, K\}$ under Models 2 and 3, implying that

$$|\mathfrak{N}_{v,t}| \;=\; |\mathcal{J}_t^{(v)} \cup \mathcal{J}_v^{(t)}| \;\geq\; \left| \left[\bigcup_{b \in \mathcal{N}_t \setminus \{v\}} \{v, b\} \right] \cup \left[\bigcup_{b \in \mathcal{N}_v \setminus \{t\}} \{t, b\} \right] \right| \;\geq\; |\mathcal{N}_t|.$$

Thus $D_N \coloneqq \max_{\{i,j\} \subset \mathcal{N}} |\mathfrak{N}_{i,j}| \geq |\mathcal{N}_t|$ for all $t \in \mathcal{N}$. Next, for all $t \in \mathcal{N}$,

$$|\{a \in \mathcal{N} \setminus \{t\} \,:\, \mathcal{N}_a \cap \mathcal{N}_t \neq \emptyset\}| \;\leq\; \left| \bigcup_{r \in \mathcal{N}_t} \mathcal{N}_r \right| \;\leq\; |\mathcal{N}_t| \left(\max_{r \in \mathcal{N}} |\mathcal{N}_r| \right) \;\leq\; D_N^2,$$

using the above-proven fact that $\max_{t \in \mathcal{N}} |\mathcal{N}_t| \leq D_N$. □

C.2.3. *Bounding the spectral norm of the coupling matrix.* We bound the spectral norm $\|\!|\mathcal{D}_N(\boldsymbol{\theta}^\star)|\!\|_2$ of the coupling matrix $\mathcal{D}_N(\boldsymbol{\theta}^\star)$. Throughout, we adopt the notation used in Section 3 of the manuscript and denote the number of edge variables by $M = \binom{N}{2}$ and edge variables by $X_1, \dots, X_M$.

Lemma 12. *Consider Models 2 and 3. Assume that Assumption A is satisfied and that the data-generating parameter vector $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta} = \mathbb{R}^p$ satisfies*

$$\|\boldsymbol{\theta}^\star\|_\infty \;\;\leq\;\; \frac{L + \vartheta \log N}{14\,(3 + D_N)} - \epsilon^\star, \tag{C.11}$$

where $L \in [0,\, \infty)$, $\vartheta \in [0,\, \infty)$, and $\epsilon^\star \in (0,\, \infty)$ are constants, independent of N and p, and $\epsilon^\star$ is the same as in the definition of $\widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)$ and $\widetilde{\Phi}_N(\boldsymbol{\theta}^\star)$.

1. *If the subpopulations do not intersect ($\omega_1 = \omega_2 = 0$) and $\boldsymbol{\theta}^\star \in \mathbb{R}^{N+1}$ satisfies condition (C.11) with $\vartheta \in [0,\, 1/2 - \alpha)$, then*
$$\|\!|\mathcal{D}_N(\boldsymbol{\theta}^\star)|\!\|_2 \;\;\leq\;\; 1 + 4\, D_N^2.$$
2. *If the subpopulations do intersect ($\omega_1 > 0$) and $\boldsymbol{\theta}^\star \in \mathbb{R}^{N+1}$ satisfies condition (C.11) with $\vartheta = 0$, then there exists finite constants $C_1 > 0$ and $C_2 > 0$, independent of N and p, such that*
$$\|\!|\mathcal{D}_N(\boldsymbol{\theta}^\star)|\!\|_2 \;\;\leq\;\; 1 + 4\, D_N^2 + \omega_1\, C_1\, \exp(C_2\, D_N^3).$$

Condition (C.11) is identical to condition (3.15) on page 20 of the main manuscript.

Proof of Lemma 12. We adapt the coupling approach of van den Berg and Maes [39, pp. 759–760] from the literature on Gibbs measures and Markov random fields to coupling conditional distributions of subgraphs of random graphs. Let $i \in \mathcal{V}$ be any vertex of the conditional independence graph $\mathcal{G}$, corresponding to edge variable X_i, and consider any $\boldsymbol{x}_{1:i-1} \in \{0,1\}^{i-1}$. Define

$$\mathbb{P}_{i,\boldsymbol{x}_{1:i-1},0}(\boldsymbol{X}_{i+1:M} = \boldsymbol{a}) \;\;\coloneqq\;\; \mathbb{P}(\boldsymbol{X}_{i+1:M} = \boldsymbol{a} \mid \boldsymbol{X}_{1:i-1} = \boldsymbol{x}_{1:i-1}, X_i = 0)$$

and

$$\mathbb{P}_{i,\boldsymbol{x}_{1:i-1},1}(\boldsymbol{X}_{i+1:M} = \boldsymbol{a}) \;\;\coloneqq\;\; \mathbb{P}(\boldsymbol{X}_{i+1:M} = \boldsymbol{a} \mid \boldsymbol{X}_{1:i-1} = \boldsymbol{x}_{1:i-1}, X_i = 1),$$

where $\boldsymbol{X}_{1:i-1} = (X_1, \ldots, X_{i-1})$, $\boldsymbol{X}_{i+1:M} = (X_{i+1}, \ldots, X_M)$, and $\boldsymbol{a} \in \{0,1\}^{M-i}$.

We divide the proof into three parts:

I. Coupling conditional distributions of subgraphs.

II. Bounding the elements of the coupling matrix $\mathcal{D}_N(\boldsymbol{\theta}^\star)$.

III. Bounding the spectral norm $|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2$ of the coupling matrix $\mathcal{D}_N(\boldsymbol{\theta}^\star)$.

I. Coupling conditional distributions of subgraphs. Given any vertex $i \in \mathcal{V}$ of the conditional independence graph $\mathcal{G}$ and any $\boldsymbol{x}_{1:i-1} \in \{0,1\}^{i-1}$, we construct a coupling $(\boldsymbol{X}^\star, \boldsymbol{X}^{\star\star})$ of the conditional probability distributions $\mathbb{P}_{i,\boldsymbol{x}_{1:i-1},0}$ and $\mathbb{P}_{i,\boldsymbol{x}_{1:i-1},1}$. Some background on coupling can be found in Lindvall [27].

It will be convenient to assume that the coupling $(\boldsymbol{X}^\star, \boldsymbol{X}^{\star\star})$ takes on values in the set $\{0,1\}^M \times \{0,1\}^M$ rather than the set $\{0,1\}^{M-i} \times \{0,1\}^{M-i}$, where we set $(\boldsymbol{X}^\star_{1:i-1},\, X^\star_i) = (\boldsymbol{x}_{1:i-1},\, 0)$ and $(\boldsymbol{X}^{\star\star}_{1:i-1},\, X^{\star\star}_i) = (\boldsymbol{x}_{1:i-1},\, 1)$ with probability 1. As a consequence, the random vectors $\boldsymbol{X}^\star \in \{0,1\}^M$ and $\boldsymbol{X}^{\star\star} \in \{0,1\}^M$ have the same dimension as random vector $\boldsymbol{X} \in \{0,1\}^M$. We construct a coupling of the conditional probability distributions $\mathbb{P}_{i,\boldsymbol{x}_{1:i-1},0}$ and $\mathbb{P}_{i,\boldsymbol{x}_{1:i-1},1}$ as follows:

1. Initialize the subset of vertices $\mathfrak{V} = \{1, \ldots, i\}$.

2. Check whether there exists a vertex $j \in \mathcal{V} \setminus \mathfrak{V}$ connected to a vertex $v \in \mathfrak{V}$ in the conditional independence graph $\mathcal{G}$ such that the coupling disagrees at vertex $v \in \mathfrak{V}$, in the sense that $X^\star_v \neq X^{\star\star}_v$.

   (a) If such a vertex j exists, pick the smallest such vertex, and let $(X^\star_j,\, X^{\star\star}_j)$ be distributed according to an optimal coupling of $\mathbb{P}(X_j = \,\cdot\, \mid \boldsymbol{X}_{\mathfrak{V}} = \boldsymbol{x}^\star_{\mathfrak{V}})$ and $\mathbb{P}(X_j = \,\cdot\, \mid \boldsymbol{X}_{\mathfrak{V}} = \boldsymbol{x}^{\star\star}_{\mathfrak{V}})$.

(b) If no such vertex j exists, select the smallest $j \in \mathcal{V} \setminus \mathfrak{V}$ and let $(X_j^\star, X_j^{\star\star})$ be distributed according to an optimal coupling of $\mathbb{P}(X_j = \cdot \mid \boldsymbol{X}_{\mathfrak{V}} = \boldsymbol{x}_{\mathfrak{V}}^\star)$ and $\mathbb{P}(X_j = \cdot \mid \boldsymbol{X}_{\mathfrak{V}} = \boldsymbol{x}_{\mathfrak{V}}^{\star\star})$. In this case, an optimal coupling will ensure $X_j^\star = X_j^{\star\star}$ with probability 1, as conditional independence properties and the equality of edge variables in the conditioning statement in this case will imply

$$\mathbb{P}(X_j = a \mid \boldsymbol{X}_{\mathfrak{V}} = \boldsymbol{x}_{\mathfrak{V}}^\star) \quad = \quad \mathbb{P}(X_j = a \mid \boldsymbol{X}_{\mathfrak{V}} = \boldsymbol{x}_{\mathfrak{V}}^{\star\star}), \quad a \in \{0, 1\},$$

resulting in a total variation distance of 0.

In both steps, an optimal coupling exists [27, Theorem 5.2, p. 19], but it may not be unique. Any optimal coupling will do.

3. Replace $\mathfrak{V}$ by $\mathfrak{V} \cup \{j\}$ and repeat Step 2 until $\mathcal{V} \setminus \mathfrak{V} = \emptyset$.

Denote the resulting coupling distribution by $\mathbb{Q}_{i,\boldsymbol{x}_{1:i-1}}$. Lemma 15 verifies that the above algorithm constructs a valid coupling of the conditional distributions $\mathbb{P}_{i,\boldsymbol{x}_{1:i-1},0}$ and $\mathbb{P}_{i,\boldsymbol{x}_{1:i-1},1}$, in the sense that the marginal distributions of $\boldsymbol{X}^\star$ and $\boldsymbol{X}^{\star\star}$ are $\mathbb{P}_{i,\boldsymbol{x}_{1:i-1},0}$ and $\mathbb{P}_{i,\boldsymbol{x}_{1:i-1},1}$, respectively.

For any two distinct vertices $i \in \mathcal{V}$ and $j \in \{i+1, \ldots, M\}$ of the conditional independence graph $\mathcal{G}$, define the event $i \not\longleftrightarrow j$ to be the event that there exists a *path of disagreement* between i and j in $\mathcal{G}$. A path of disagreement $i \not\longleftrightarrow j$ between vertices X_i and X_j is a sequence of two or more distinct vertices $(X_i, \ldots, X_j)$ in the conditional independence graph $\mathcal{G}$ starting at vertex X_i and ending at vertex X_j, such that

- each subsequent pair of vertices (X_v, X_w) in the sequence is connected by an edge in the conditional independence graph $\mathcal{G}$, which indicates the absence of conditional independence of vertices X_v and X_w;
- the coupling $(\boldsymbol{X}_{i+1:M}^\star, \boldsymbol{X}_{i+1:M}^{\star\star}) \in \{0,1\}^{M-i} \times \{0,1\}^{M-i}$ with joint probability mass function $\mathbb{Q}_{\boldsymbol{\theta}^\star,i,\boldsymbol{x}_{1:i-1}}$ disagrees at each vertex X_v in the sequence, in the sense that $X_v^\star \neq X_v^{\star\star}$.

Theorem 1 of van den Berg and Maes [39, p. 753] shows that

$$\mathbb{Q}_{i,\boldsymbol{x}_{1:i-1}}(X_j^\star \neq X_j^{\star\star}) \quad = \quad \mathbb{Q}_{i,\boldsymbol{x}_{1:i-1}}(i \not\longleftrightarrow j) \quad \leq \quad \mathbb{B}_{\boldsymbol{\pi}}(i \not\longleftrightarrow j), \tag{C.12}$$

where $\mathbb{B}_{\boldsymbol{\pi}}$ is a Bernoulli product measure on $\{0,1\}^M$ with probability vector $\boldsymbol{\pi} \in [0,1]^M$. The coordinates π_v of $\boldsymbol{\pi}$ are given by

$$\pi_v \quad := \quad \begin{cases} 0 & \text{if } v \in \{1, \ldots, i-1\} \\ 1 & \text{if } v = i \\ \displaystyle\max_{(\boldsymbol{x}_{-v},\, \boldsymbol{x}'_{-v}) \,\in\, \{0,1\}^{M-1} \times \{0,1\}^{M-1}} \pi_{v,\, \boldsymbol{x}_{-v},\, \boldsymbol{x}'_{-v}} & \text{if } v \in \{i+1, \ldots, M\}, \end{cases}$$

where

$$\pi_{v,\boldsymbol{x}_{-v},\boldsymbol{x}'_{-v}} \;\; := \;\; \|\mathbb{P}(\,\cdot \mid \boldsymbol{X}_{-v} = \boldsymbol{x}_{-v}) - \mathbb{P}(\,\cdot \mid \boldsymbol{X}'_{-v} = \boldsymbol{x}'_{-v})\|_{\mathrm{TV}}.$$

Observe that the total variation distance

$$\|\mathbb{P}(\,\cdot \mid \boldsymbol{X}_{-v} = \boldsymbol{x}_{-v}) - \mathbb{P}(\,\cdot \mid \boldsymbol{X}'_{-v} = \boldsymbol{x}'_{-v})\|_{\mathrm{TV}}$$

is equal to

$$\sup_{x_v \,\in\, \{0,1\}} \left|\mathbb{P}(X_v = x_v \mid \boldsymbol{X}_{-v} = \boldsymbol{x}_{-v}) - \mathbb{P}(X_v = x_v \mid \boldsymbol{X}'_{-v} = \boldsymbol{x}'_{-v})\right|.$$

The Bernoulli product measure $\mathbb{B}_{\boldsymbol{\pi}}$ assumes that independent Bernoulli experiments are carried out at vertices $v \in \{1, \ldots, M\}$. The Bernoulli experiment at vertex $v \in \{i+1, \ldots, M\}$ has two possible outcomes: Either vertex v is *open,* in the sense that the event $\{X_v^\star \neq X_v^{\star\star}\}$ occurs and hence vertex v allows a path of disagreement from i to j to pass through, or vertex v is *closed.* A vertex v is open with probability π_v, and closed with probability $1-\pi_v$. By construction, vertices $v \in \{1, \ldots, i-1\}$ are closed with probability 1, and vertex i is open with probability 1.

The coupling argument of van den Berg and Maes [39] is useful, in that it translates the hard problem of bounding probabilities of events involving dependent random variables into the more convenient problem of bounding probabilities of events involving independent random variables. Indeed, we can bound the above-diagonal elements $\mathcal{D}_{i,j}(\boldsymbol{\theta}^\star)$ of $\mathcal{D}_N(\boldsymbol{\theta}^\star)$ by

$$\mathcal{D}_{i,j}(\boldsymbol{\theta}^\star) \;:=\; \sup_{\boldsymbol{x}_{1:i-1} \in \{0,1\}^{i-1}} \mathbb{Q}_{i,\boldsymbol{x}_{1:i-1}}(X_j^\star \neq X_j^{\star\star}) \;\leq\; \mathbb{B}_{\boldsymbol{\pi}}(i \not\leftrightarrow j). \tag{C.13}$$

By the construction of $\mathcal{D}_N(\boldsymbol{\theta}^\star)$, the below-diagonal and diagonal elements of $\mathcal{D}_N(\boldsymbol{\theta}^\star)$ are 0 and 1, respectively. We define $\pi^\star \in (0,1)$ by

$$\pi^\star \;\; := \;\; \max_{1 \leq v \leq M} \pi_v,$$

and note that Lemma 16, together with the assumption that $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta}_N = \mathbb{R}^p$ satisfies (C.11), implies that

$$\pi^\star \;\; \leq \;\; \frac{1}{1 + \exp(-L - \vartheta \log N)} \;\; < \;\; 1,$$

where $L \in [0, \infty)$ and $\vartheta \in [0, \infty)$ are the same constants as in (C.11), assumed to be independent of N and p. Let

$$U \;\; := \;\; \frac{1}{1 + \exp(-L - \vartheta \log N)}$$

and define the vector $\boldsymbol{\xi} \in [0,1]^M$ by

$$\xi_i \;\; \coloneqq \;\; \begin{cases} 0 & \text{if } v \in \{1, \ldots, i-1\} \\ 1 & \text{if } v = i \\ U & \text{if } v \in \{i+1, \ldots, M\} \end{cases} .$$

Observe that the probabilities $\mathbb{B}_{\boldsymbol{\pi}}(i \not\longleftrightarrow j)$ of the events $\{i \not\longleftrightarrow j\}$ are non-decreasing in the coordinates of $\boldsymbol{\pi}$, so that

$$\mathbb{B}_{\boldsymbol{\pi}}(i \not\longleftrightarrow j) \;\; \leq \;\; \mathbb{B}_{\boldsymbol{\xi}}(i \not\longleftrightarrow j).$$

We bound the elements $\mathcal{D}_{i,j}(\boldsymbol{\theta}^\star)$ of $\mathcal{D}_N(\boldsymbol{\theta}^\star)$ by bounding the probabilities $\mathbb{B}_{\boldsymbol{\xi}}(i \not\longleftrightarrow j)$ of the events $\{i \not\longleftrightarrow j\}$.

II. Bounding the elements of the coupling matrix $\mathcal{D}_N(\boldsymbol{\theta}^\star)$. To bound the elements $\mathcal{D}_{i,j}(\boldsymbol{\theta}^\star)$ of $\mathcal{D}_N(\boldsymbol{\theta}^\star)$, we bound the probabilities $\mathbb{B}_{\boldsymbol{\xi}}(i \not\longleftrightarrow j)$ of the events $\{i \not\longleftrightarrow j\}$ using Assumption A. To do so, define

$$\mathcal{S}_{\mathcal{G},i,k} \;\; \coloneqq \;\; \{v \in \mathcal{V} \setminus \{i\} \,:\, d_{\mathcal{G}}(i, v) = k\}\,, \qquad k = 1, \ldots, M-1,$$

where $d_{\mathcal{G}}(i, v)$ is the graph distance (i.e., the length of the shortest path) between vertices $i \in \mathcal{V}$ and $v \in \mathcal{V}$ in the conditional independence graph $\mathcal{G}$. The set $\mathcal{S}_{\mathcal{G},i,k} \subseteq \mathcal{V}$ represents the subset of vertices in the conditional independence graph $\mathcal{G}$ at graph distance k from vertex i in $\mathcal{G}$.

We bound $\mathbb{B}_{\boldsymbol{\xi}}(i \not\longleftrightarrow j)$ by placing restrictions on the subpopulation structure, which determines which edges are present in $\mathcal{G}$. To do so, define the *subpopulation graph* $\mathcal{G}_{\mathcal{A}}$ to be the graph with the set of subpopulations $\{\mathcal{A}_1, \ldots, \mathcal{A}_K\}$ as vertices and edges between vertices $\mathcal{A}_r$ and $\mathcal{A}_l$ if and only if $\mathcal{A}_r \cap \mathcal{A}_l \neq \emptyset$. In $\mathcal{G}_{\mathcal{A}}$, two vertices corresponding to subpopulations $\mathcal{A}_r$ and $\mathcal{A}_l$ are connected by an edge if and only if they overlap. Let $d_{\mathcal{G}_{\mathcal{A}}}(\mathcal{A}_r, \mathcal{A}_l)$ denote the graph distance (i.e., the length of the shortest path) between vertices $\mathcal{A}_r$ and $\mathcal{A}_l$ in $\mathcal{G}_{\mathcal{A}}$. Define, for all $\mathcal{A}_r \in \{\mathcal{A}_1, \ldots, \mathcal{A}_K\}$ and $k \in \{1, \ldots, K-1\}$,

$$\mathcal{V}_{\mathcal{A}_r,k} \;\; \coloneqq \;\; \big\{\mathcal{A}_l \in \{\mathcal{A}_1, \ldots, \mathcal{A}_K\} \setminus \{\mathcal{A}_r\} \,:\, d_{\mathcal{G}_{\mathcal{A}}}(\mathcal{A}_r, \mathcal{A}_l) = k\big\}.$$

Let $g : \{1, 2, \ldots\} \mapsto [0, \infty)$ be such that, for all $K \in \{1, 2, \ldots\}$,

$$|\mathcal{V}_{\mathcal{A}_r,k}| \;\; \leq \;\; g(k), \quad k \in \{1, \ldots, K-1\}, \qquad \text{for all } \mathcal{A}_r \in \{\mathcal{A}_1, \ldots, \mathcal{A}_K\}.$$

In words, $g(k)$ bounds the number of subpopulations at graph distance k from any given subpopulation in $\mathcal{G}_{\mathcal{A}}$ for all conceivable subpopulations and thus all conceivable subpopulation structures, i.e., for all $\mathcal{G}_{\mathcal{A}}$ defined for subpopulations $\mathcal{A}_1, \ldots, \mathcal{A}_K$ at all values of $K \in \{1, 2, \ldots\}$.

Models 2 and 3 satisfy Definition 1 and posses the neighborhood intersection property. By Proposition 2, the dependence neighborhood of any edge variable X_i between nodes $\{a, b\} \subset \mathcal{N}$ is not larger than the subset of edge indices contained in the set $\mathcal{M}_{a,b} \coloneqq \big\{\{c, d\} \,:\, c \in \mathcal{N}_a \cup \mathcal{N}_b,\, d \in \mathcal{N}_a \cup \mathcal{N}_b \setminus \{c\}\big\}$, i.e., the edge variables contained in the set $\mathfrak{N}_i$ will correspond to edge variables between pairs of nodes in $\mathcal{M}_{a,b}$. We construct a graph covering $\mathcal{G}^\star$ of the conditional independence graph $\mathcal{G}$ as follows:

1. Initialize $\mathcal{G}^\star$ with the same set of vertices and edges as $\mathcal{G}$.
2. For each vertex X_i in $\mathcal{G}$ corresponding to an edge variable between nodes $\{a, b\} \subset \mathcal{N}$ with degree greater than 0 in $\mathcal{G}$, add edges between X_i and any other edge variables X_j contained in the subgraph $\boldsymbol{X}_{\mathcal{M}_{a,b}}$ which are not already present in $\mathcal{G}$.

The construction of $\mathcal{G}^\star$ ensures that the dependence neighborhood of any given vertex X_i in $\mathcal{G}^\star$ corresponding to the edge variable between pair of nodes $\{a, b\} \subset \mathcal{N}$ is either empty or is equal to the set of vertices corresponding to edge variables contained in the subgraph $\boldsymbol{X}_{\mathcal{M}_{a,b}}$. Moreover, the fact that $\mathcal{G} \subseteq \mathcal{G}^\star$ implies

$$\mathbb{B}_{\boldsymbol{\xi}}(i \not\longleftrightarrow j \text{ in } \mathcal{G}) \quad \leq \quad \mathbb{B}_{\boldsymbol{\xi}}(i \not\longleftrightarrow j \text{ in } \mathcal{G}^\star).$$

In words, the probability of the existence of a path of disagreement does not decrease through the addition of edges in the graph. Henceforth and for ease of presentation, the event $i \not\longleftrightarrow j$ will represent a path of disagreement in the graph covering $\mathcal{G}^\star$ of $\mathcal{G}$ and we will assume $\mathcal{S}_{i,k} \equiv \mathcal{S}_{\mathcal{G}^\star,i,k}$.

We bound each $|\mathcal{S}_{i,k}|$ ($k \in \{1, 2, \ldots\}$) for arbitrary $i \in \mathcal{V}$ with non-zero degree in $\mathcal{G}^\star$:

- *Bounding* $|\mathcal{S}_{i,1}|$. Let X_i denote the edge variable between pair of nodes $\{a, b\} \subset \mathcal{N}$. By definition, $\mathcal{S}_{i,1}$ contains all vertices in $\mathcal{G}^\star$ corresponding to edge variables X_j which lie in the dependence neighborhood of edge variable X_i in $\mathcal{G}^\star$. By the construction of $\mathcal{G}^\star$, the dependence neighborhood of edge variable X_i in $\mathcal{G}^\star$ is equal to the set of edge variables contained in the subgraph $\boldsymbol{X}_{\mathcal{M}_{a,b}}$, the number of which is bounded above by $4\, D_N^2$:

  $$|\mathcal{M}_{a,b}| \leq |\mathcal{N}_a \cup \mathcal{N}_b|^2 \leq (|\mathcal{N}_a| + |\mathcal{N}_b|)^2 \leq (2\, \max\{|\mathcal{N}_a|,\, |\mathcal{N}_b|\})^2 \leq 4\, D_N^2,$$

  where by Lemma 11, $D_N \geq \max_{t \in \mathcal{N}} |\mathcal{N}_t|$. Hence, $|\mathcal{S}_{i,1}| \leq 4\, D_N^2$.

- *Bounding* $|\mathcal{S}_{i,2}|$. Consider any $j \in \mathcal{S}_{i,2}$ and let X_i denote the edge variable between pair of nodes $\{a, b\} \subset \mathcal{N}$. The shortest path between edge variables X_i and X_j in $\mathcal{G}^\star$ is of length 2, implying the following:

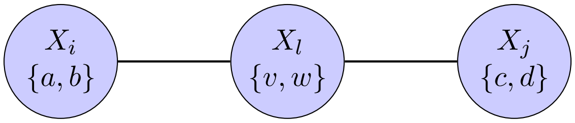
  (F.1) X_j is not in the dependence neighborhood of X_i in $\mathcal{G}^\star$.

  (F.2) In $\mathcal{G}^\star$, there is at least one edge variable X_l in the dependence neighborhood of X_i such that X_j is likewise in the dependence neighborhood of X_l.

  By the construction of $\mathcal{G}^\star$, facts (F.1) and (F.2) imply there exist

  – a pair of nodes $\{v, w\} \subseteq \mathcal{N}_a \cup \mathcal{N}_b$ such that X_l is the edge variable between $\{v, w\}$, and

  – a pair of nodes $\{c, d\} \subseteq \mathcal{N}_v \cup \mathcal{N}_w$ such that X_j is the edge variable between $\{c, d\}$ and $\{c, d\} \not\subseteq \mathcal{N}_a \cup \mathcal{N}_b$, otherwise X_j would be in the dependence neighborhood of X_i, violating the assumption that $j \in \mathcal{S}_{i,2}$.



  Recall the definition, for all $v \in \mathcal{N}$,

$$\mathcal{N}_v \quad := \quad \{w \in \mathcal{N} \setminus \{v\} \,:\, \text{exists } r \in \{1, \ldots, K\} \text{ such that } \{v, w\} \subset \mathcal{A}_r\}.$$

  As $\{c, d\} \subseteq \mathcal{N}_v \cup \mathcal{N}_w$, there must exist $r, t \in \{1, \ldots, K\}$ such that:

  – either $\{c, v\} \subset \mathcal{A}_r$ or $\{c, w\} \subset \mathcal{A}_r$, and

  – either $\{d, v\} \subset \mathcal{A}_t$ or $\{d, w\} \subset \mathcal{A}_t$.

  Since $\{c, d\} \not\subseteq \mathcal{N}_a \cup \mathcal{N}_b$, therefore $\{a, b\} \not\subseteq \mathcal{A}_r \cup \mathcal{A}_t \subseteq \mathcal{N}_c \cup \mathcal{N}_d$. Finally, $\{v, w\} \subseteq \mathcal{N}_a \cup \mathcal{N}_b$ implies that there exists $n, m \in \{1, \ldots, K\}$ such that

  – either $\{v, a\} \subset \mathcal{A}_n$ or $\{v, b\} \subset \mathcal{A}_n$, and

  – either $\{w, a\} \subset \mathcal{A}_m$ or $\{w, b\} \subset \mathcal{A}_m$.

  As a result, $(\mathcal{A}_n \cup \mathcal{A}_m) \cap (\mathcal{A}_r \cup \mathcal{A}_t) \neq \emptyset$, implying either v or w belong to a subpopulation $\mathcal{A}_z \not\subseteq \mathcal{N}_a \cup \mathcal{N}_b$ ($z \in \{1, \ldots, K\}$) for which $d_{\mathcal{G}_\mathcal{A}}(\mathcal{A}_z, \mathcal{A}_y) = 1$ for some $y \in \{1, \ldots, K\}$ with $\mathcal{A}_y \subseteq \mathcal{N}_a \cup \mathcal{N}_b$, i.e., a subpopulation with graph distance at least 1 in $\mathcal{G}_\mathcal{A}$ from all subpopulations represented in $\mathcal{N}_a \cup \mathcal{N}_b$ and equal to 1 for at least one such subpopulation. The same holds for either c or d. Thus,

$$|\mathcal{S}_{i,2}| \quad \leq \quad 2\, D_N^3\, (g(1) + 1)\, g(1),$$

  which follows from the following argument:

- First, the number of subpopulations contained in $\mathcal{N}_a \cup \mathcal{N}_b$ is bounded above by $2(g(1)+1)$, because $g(1)$ bounds the number of subpopulations which overlap with any other subpopulation, so that $g(1)+1$ bounds the number of subpopulations to which any node $a \in \mathcal{N}$ or $b \in \mathcal{N}$ may belong;

- Second, the number of subpopulations with graph distance 1 in the subpopulation graph $\mathcal{G}_{\mathcal{A}}$ to any subpopulation represented in $\mathcal{N}_a \cup \mathcal{N}_b$ is bounded above by $2\,(g(1)+1)\,g(1)$;

- Third, note that either c or d must be in one of the subpopulations with graph distance 1 in $\mathcal{G}_{\mathcal{A}}$ to at least one of the subpopulations represented in $\mathcal{N}_a \cup \mathcal{N}_b$. Without loss, let this be c. Then the total number of such nodes c which are contained in one of the aforementioned subpopulations at graph distance 1 is bounded above by $2\,D_N\,(g(1)+1)\,g(1)$, using the bound $|\mathcal{A}_k| \leq |\mathcal{N}_i| \leq D_N$ from Lemma 11 which holds for all $k \in \{1,\dots,K\}$ and $i \in \mathcal{A}_k$.

- Finally, we bound the number of possible d that may be paired with c. Note X_j has non-zero degree in $\mathcal{G}^\star$. By Proposition 2, X_j is independent of all other edges if $\mathcal{N}_c \cap \mathcal{N}_d = \emptyset$. Thus, we bound the number of edge variables X_j between node $c \in \mathcal{N}$ and nodes $d \in \mathcal{N} \setminus \{c\}$ for which $\mathcal{N}_c \cap \mathcal{N}_d \neq \emptyset$ using Lemma 11:

$$|\{d \in \mathcal{N} \setminus \{c\} \,:\, \mathcal{N}_c \cap \mathcal{N}_d \neq \emptyset\}| \quad \leq \quad D_N^2.$$

  Hence, the number of such d (for a given c) numbers no more than D_N^2, the total of which is bounded above by $2\,D_N^3\,(g(1)+1)\,g(1)$.

- *Bounding* $|\mathcal{S}_{i,k}|$ *for* $k \in \{3,4,\dots\}$. Consider any $k \in \{3,4,\dots\}$ and any $j \in \mathcal{S}_{i,k}$. Let X_i be the edge variable between nodes $\{a,b\} \subset \mathcal{N}$ and X_j be the edge variable between pair of nodes $\{c,d\} \subset \mathcal{N}$. For $j \in \mathcal{S}_{i,k}$, there must exist an $l \in \mathcal{S}_{i,k-1}$ such that $j \in \mathcal{S}_{l,1}$. Let X_l be the edge variable between nodes $\{v,w\} \subset \mathcal{N}$. Leveraging arguments from the case bounding $|\mathcal{S}_{i,1}|$ above, $\{c,d\} \subseteq \mathcal{N}_v \cup \mathcal{N}_w$, implying both c and d belong to at least one subpopulation to which either v or w also belong. Assume the following:

  (A.1) $h \in \mathcal{S}_{i,k-1}$ if and only if X_h is an edge variable between nodes $\{r,t\} \subset \mathcal{N}$ and either r or t belongs to a subpopulation at graph distance $k-2$ in $\mathcal{G}_{\mathcal{A}}$ to a subpopulation represented in $\mathcal{N}_a \cup \mathcal{N}_b$ and not less than $k-2$ to all others in $\mathcal{N}_a \cup \mathcal{N}_b$.

(A.2) $h \in \mathcal{S}_{i,k-2}$ if and only if X_h is an edge variable between nodes $\{r, t\} \subset \mathcal{N}$ and either r or t belongs to a subpopulation at graph distance $k-3$ in $\mathcal{G}_{\mathcal{A}}$ to a subpopulation represented in $\mathcal{N}_a \cup \mathcal{N}_b$ and not less than $k-3$ to all others in $\mathcal{N}_a \cup \mathcal{N}_b$.

We have shown above that (A.1) and (A.2) are satisfied when $k = 3$:

- $h \in \mathcal{S}_{i,1}$ corresponds to edge variables X_h between nodes $\{r, t\} \subset \mathcal{N}_a \cup \mathcal{N}_b$, in which case both r and t belong to a subpopulation represented in $\mathcal{N}_a \cup \mathcal{N}_b$ and therefore have graph distance 0 in $\mathcal{G}_{\mathcal{A}}$ to a subpopulation represented in $\mathcal{N}_a \cup \mathcal{N}_b$.
- $h \in \mathcal{S}_{i,2}$ corresponds to edge variables X_h between nodes $\{r, t\} \subset \mathcal{N}$ for which either r or t belongs to a subpopulation which is not represented in $\mathcal{N}_a \cup \mathcal{N}_b$, but which is at graph distance 1 in $\mathcal{G}_{\mathcal{A}}$ to a subpopulation represented in $\mathcal{N}_a \cup \mathcal{N}_b$, and at graph distance no less than 1 to all others in $\mathcal{N}_a \cup \mathcal{N}_b$.

Assumptions (A.1) and (A.2) require that neither v nor w belong to a subpopulation at graph distance less than or equal to $k-2$ in $\mathcal{G}_{\mathcal{A}}$ from any subpopulation represented in $\mathcal{N}_a \cup \mathcal{N}_b$. Leveraging the argument used in the case bounding $|\mathcal{S}_{i,2}|$: For $\{c, d\} \subseteq \mathcal{N}_v \cup \mathcal{N}_w$, either c or d must belong to a subpopulation $\mathcal{A}_r$ ($r \in \{1, \ldots, K\}$) jointly with either v or w which is at graph distance at least $k-1$ in in $\mathcal{G}_{\mathcal{A}}$ from any subpopulation represented in $\mathcal{N}_i \cup \mathcal{N}_j$. Thus, there must exist some $q \in \{1, \ldots, K\}$ such that $\mathcal{A}_r \cap \mathcal{A}_q \neq \emptyset$ and for which both of the following are satisfied:

- $d_{\mathcal{G}_{\mathcal{A}}}(\mathcal{A}_q, \mathcal{A}_y) \geq k-2$ for all $y \in \{1, \ldots, K\}$ satisfying $\mathcal{A}_y \subseteq \mathcal{N}_a \cup \mathcal{N}_b$;
- $d_{\mathcal{G}_{\mathcal{A}}}(\mathcal{A}_q, \mathcal{A}_y) = k-2$ for at least one $y \in \{1, \ldots, K\}$ satisfying $\mathcal{A}_y \subseteq \mathcal{N}_a \cup \mathcal{N}_b$.

Hence, there exists at least one $y \in \{1, \ldots, K\}$ satisfying $\mathcal{A}_y \subseteq \mathcal{N}_a \cup \mathcal{N}_b$ for which $d_{\mathcal{G}_{\mathcal{A}}}(\mathcal{A}_r, \mathcal{A}_y) = k-1$. Repeating the counting argument from the case bounding $|\mathcal{S}_{i,2}|$ above shows that the number of such pairs $\{a, b\} \subset \mathcal{N}$ is bounded above by

$$|\mathcal{S}_{i,k}| \quad \leq \quad 2\, D_N^3\, (g(1) + 1)\, g(k-1),$$

and the result is proved by induction.

From the above, we have the bounds $|\mathcal{S}_{i,1}| \leq 4\, D_N^2$ and

$$|\mathcal{S}_{i,k}| \quad \leq \quad 2\, D_N^3\, (g(1) + 1)\, g(k-1), \qquad k \in \{2, 3, \ldots\}. \tag{C.14}$$

We proceed with bounding $\mathcal{D}_{i,j}(\boldsymbol{\theta}^\star)$ under Assumption A. Define the function $g : \{1, 2, \dots\} \mapsto [0, \infty)$ by

$$g(k) \;=\; \omega_1 + \frac{\omega_2}{2\, D_N^3} \log k, \qquad k \in \{1, 2, \dots\}, \tag{C.15}$$

where $\omega_1 \geq 0$ and $\omega_2 \in [0, \omega_1]$ are independent of N and p by Assumption A, and $\omega_2 \in [0, \omega_1]$ additionally satisfies

$$\omega_2 \;<\; \frac{1}{(\omega_1 + 1)\, |\log(1 - U)|}.$$

Using (C.15) and (C.14), we obtain the bounds $|\mathcal{S}_{i,1}| \leq 4\, D_N^2$ and

$$|\mathcal{S}_{i,k}| \;\leq\; (\omega_1 + 1)\, (2\, D_N^3\, \omega_1 + \omega_2\, \log(k - 1)), \qquad k \in \{2, 3, \dots\}.$$

By construction,

$$\mathbb{B}_{\boldsymbol{\xi}}(v \text{ is open in } \mathcal{G}^\star) \;\leq\; U \;<\; 1, \qquad v \in \{i + 1, \dots, M\}.$$

For there to be a path of disagreement $i \not\leftrightarrow j$ in $\mathcal{G}^\star$ between a vertex $i \in \{1, \dots, M\}$ and $j \in \mathcal{S}_{i,k}$, there must be at least one open vertex in each of the sets $\mathcal{S}_{i,1}, \dots, \mathcal{S}_{i,k-1}$ and j must be open (placing no restrictions on the connectedness of vertices within sets or between two sequential sets); note that i is open with probability 1. The probability that there exists at least one open vertex $v \in \mathcal{S}_{i,1}$ is bounded above by

$$1 - (1 - U)^{|\mathcal{S}_{i,1}|} \;\leq\; 1 - (1 - U)^{4\, D_N^2} \;\leq\; 1,$$

and for $k \in \{2, 3, \dots\}$, the same is bounded above by

$$\begin{aligned} 1 - (1 - U)^{|\mathcal{S}_{i,k}|} \;&\leq\; 1 - (1 - U)^{(\omega_1+1)\, (2\, D_N^3\, \omega_1 + \omega_2\, \log(k-1))} \\ &=\; 1 - (1 - U)^{C_1 D_N^3 + C_2 \log(k-1)}, \end{aligned}$$

defining $C_1 \coloneqq 2\, \omega_1\, (\omega_1 + 1) \in [0, \infty)$ and $C_2 \coloneqq \omega_2\, (\omega_1 + 1) \in [0, \infty)$. Since the events that vertices are open are independent under the Bernoulli product measure $\mathbb{B}_{\boldsymbol{\xi}}$, we obtain

$$\begin{aligned} \mathbb{B}_{\boldsymbol{\xi}}(i \not\leftrightarrow j) \;&\leq\; U \left[1 - (1 - U)^{4 D_N^2}\right] \prod_{l=2}^{k-1} \left[1 - (1 - U)^{C_1 D_N^3 + C_2 \log(l-1)}\right] \\ &\leq\; U \left[1 - (1 - U)^{4 D_N^2}\right] \left[1 - (1 - U)^{C_1 D_N^3 + C_2 \log(k-2)}\right]^{k-2} \\ &\leq\; \left[1 - (1 - U)^{C_1 D_N^3 + C_2 \log(k-2)}\right]^{k-2}, \end{aligned}$$

by the monotonicity of logarithms. We then bound

$$1-(1-U)^{C_1 D_N^3+C_2\log(k-2)} \quad\leq\quad \exp\left(-(1-U)^{C_1 D_N^3+C_2\log(k-2)}\right),$$

using the inequality $1-z\leq\exp(-z)$ $(z\in(0,1))$. We proceed by writing

$$\begin{aligned}
&\exp\left(-(1-U)^{C_1 D_N^3+C_2\log(k-2)}\right)\\
&= \exp\left(-\exp\left(\left[C_1\,D_N^3+C_2\log(k-2)\right]\,\log(1-U)\right)\right)\\
&= \exp\left(-\exp\left(-\left[C_1\,D_N^3+C_2\log(k-2)\right]\,|\log(1-U)|\right)\right)\\
&= \exp\left(-\exp(-C_1\,D_N^3\,|\log(1-U)|)\,(k-2)^{-C_2\,|\log(1-U)|}\right),
\end{aligned}$$

where in the above we used the fact that $\log(1-U)<0$. Define

$$A \quad\coloneqq\quad \exp(-C_1\,D_N^3\,|\log(1-U)|) \quad\in\quad (0,1),$$

noting that $D_N\geq 1$ under Models 2 and 3. Then

$$\begin{aligned}
\left[1-(1-U)^{C_1 D_N^3+C_2\log(k-2)}\right]^{k-2} &\leq \left[\exp\left(-A\,(k-2)^{-C_2\,|\log(1-U)|}\right)\right]^{k-2}\\
&= \exp\left(-A\,(k-2)^{1-C_2\,|\log(1-U)|}\right),
\end{aligned}$$

demonstrating the bound (for $i\in\{1,\dots,M\}$ and $j\in\mathcal{S}_{i,k}$)

$$\mathbb{B}_{\boldsymbol{\pi}}(i \not\leftrightarrow j) \quad\leq\quad \mathbb{B}_{\boldsymbol{\xi}}(i \not\leftrightarrow j) \quad\leq\quad \exp\left(-A\,(k-2)^{1-C_2\,|\log(1-U)|}\right).$$

We hence obtain (for $i\in\{1,\dots,M\}$ and $j\in\mathcal{S}_{i,k}$)

$$\mathcal{D}_{i,j}(\boldsymbol{\theta}^\star) \quad\leq\quad \exp\left(-A\,(k-2)^{1-C_2\,|\log(1-U)|}\right). \tag{C.16}$$

III. Bounding the spectral norm $|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2$ of the coupling matrix $\mathcal{D}_N(\boldsymbol{\theta}^\star)$. To bound the spectral norm $|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2$ of the coupling matrix $\mathcal{D}_N(\boldsymbol{\theta}^\star)$, we first use Hölder's inequality to obtain

$$|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2 \quad\leq\quad \sqrt{|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_1\;|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_\infty}.$$

Next, we form a symmetric $M\times M$ matrix $\mathcal{T}$ by defining

$$\mathcal{T} \quad\coloneqq\quad \mathcal{D}_N(\boldsymbol{\theta}^\star)+\mathcal{D}_N(\boldsymbol{\theta}^\star)^\top-\operatorname{diag}(\mathcal{D}_N(\boldsymbol{\theta}^\star)),$$

where $\mathcal{D}_N(\boldsymbol{\theta}^\star)^\top$ is the $M \times M$ transpose of $\mathcal{D}_N(\boldsymbol{\theta}^\star)$ and $\operatorname{diag}(\mathcal{D}_N(\boldsymbol{\theta}^\star))$ is the $M \times M$ diagonal matrix with elements $\mathcal{D}_{1,1}, \ldots, \mathcal{D}_{M,M}$ on the main diagonal. By the construction of $\mathcal{T}$, the elements $\mathcal{T}_{i,j}$ of $\mathcal{T}$ are given by

$$\mathcal{T}_{i,j} \;=\; \begin{cases} \mathcal{D}_{i,j}(\boldsymbol{\theta}^\star) & \text{if } j > i \\[1em] \mathcal{D}_{i,i}(\boldsymbol{\theta}^\star) & \text{if } i = j\ , \\[1em] \mathcal{D}_{j,i}(\boldsymbol{\theta}^\star) & \text{if } j < i \end{cases}$$

where $\mathcal{D}_{i,i}(\boldsymbol{\theta}^\star) = 1$ $(i = 1, \ldots, M)$ by the definition of $\mathcal{D}_N(\boldsymbol{\theta}^\star)$. Using the fact that $\mathcal{T}_{i,j} = \max(\mathcal{D}_{i,j}(\boldsymbol{\theta}^\star),\, \mathcal{D}_{j,i}(\boldsymbol{\theta}^\star))$ $(i, j = 1, \ldots, M)$, we obtain

$$|\!|\!| \mathcal{D}_N(\boldsymbol{\theta}^\star) |\!|\!|_1 \;=\; \max_{1 \leq j \leq M} \sum_{i=1}^{M} |\mathcal{D}_{i,j}(\boldsymbol{\theta}^\star)| \;\leq\; \max_{1 \leq j \leq M} \sum_{i=1}^{M} |\mathcal{T}_{i,j}| \;=\; |\!|\!| \mathcal{T} |\!|\!|_1$$

and

$$|\!|\!| \mathcal{D}_N(\boldsymbol{\theta}^\star) |\!|\!|_\infty \;=\; \max_{1 \leq i \leq M} \sum_{j=1}^{M} |\mathcal{D}_{i,j}(\boldsymbol{\theta}^\star)| \;\leq\; \max_{1 \leq i \leq M} \sum_{j=1}^{M} |\mathcal{T}_{i,j}| \;=\; |\!|\!| \mathcal{T} |\!|\!|_\infty.$$

In addition, we know that $\mathcal{T}_{i,j} = \mathcal{T}_{j,i}$ $(i, j = 1, \ldots, M)$, which implies that

$$|\!|\!| \mathcal{T} |\!|\!|_1 \;=\; |\!|\!| \mathcal{T}^\top |\!|\!|_\infty \;=\; |\!|\!| \mathcal{T} |\!|\!|_\infty.$$

As a consequence, we obtain

$$|\!|\!| \mathcal{D}_N(\boldsymbol{\theta}^\star) |\!|\!|_2 \;\leq\; \sqrt{|\!|\!| \mathcal{D}_N(\boldsymbol{\theta}^\star) |\!|\!|_1 \, |\!|\!| \mathcal{D}_N(\boldsymbol{\theta}^\star) |\!|\!|_\infty} \;\leq\; \sqrt{|\!|\!| \mathcal{T} |\!|\!|_1 \, |\!|\!| \mathcal{T} |\!|\!|_\infty} \;=\; |\!|\!| \mathcal{T} |\!|\!|_\infty,$$

where $|\!|\!| \mathcal{T} |\!|\!|_\infty$ can be bounded above by using (C.13):

$$|\!|\!| \mathcal{T} |\!|\!|_\infty \;=\; \max_{1 \leq i \leq M} \sum_{j=1}^{M} |\mathcal{T}_{i,j}| \;\leq\; 1 + \max_{1 \leq i \leq M} \sum_{j=1:\, j \neq i}^{M} \mathbb{B}_{\boldsymbol{\xi}}(i \not\longleftrightarrow j).$$

Hence using (C.16) along with Assumption A,

$$|\!|\!| \mathcal{D}_N(\boldsymbol{\theta}^\star) |\!|\!|_2 \;\leq\; 1 + \max_{1 \leq i \leq M} \sum_{j=1:\, j \neq i}^{M} \mathbb{B}_{\boldsymbol{\xi}}(i \not\longleftrightarrow j)$$

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

$$\begin{aligned}
&\sum_{j=1:\, j \neq i}^{M} \mathbb{B}_{\boldsymbol{\xi}}(i \not\leftrightarrow j) \\
\leq\;& |\mathcal{S}_{i,1}| + \sum_{k=2}^{\infty} |\mathcal{S}_{i,k}| \, \exp\left(-A\,(k-2)^{1 - C_2\,|\log(1-U)|}\right) \\
\leq\;& 4\, D_N^2 + \sum_{k=2}^{\infty} (C_1\, D_N^3 + C_2\, \log(k-1)) \exp\left(-A\,(k-2)^{1 - C_2\,|\log(1-U)|}\right),
\end{aligned}$$

using the above bounds on the number of vertices $|\mathcal{S}_{i,k}|$ which are at graph distance in k to any given vertex $i \in \mathcal{V}$ in $\mathcal{G}^\star$. We focus on bounding the infinite series

$$\begin{aligned}
&\sum_{k=2}^{\infty} (C_1\, D_N^3 + C_2\, \log(k-1)) \exp\left(-A\,(k-2)^{1 - C_2\,|\log(1-U)|}\right) \\
=\;& C_1\, D_N^3 \sum_{k=2}^{\infty} \exp\left(-A\,(k-2)^{1 - C_2\,|\log(1-U)|}\right) \\
+\;& C_2 \sum_{k=2}^{\infty} \log(k-1)\, \exp\left(-A\,(k-2)^{1 - C_2\,|\log(1-U)|}\right).
\end{aligned}$$

For the first series,

$$\sum_{k=2}^{\infty} \exp\left(-A\,(k-2)^{1 - C_2\,|\log(1-U)|}\right) \;=\; 1 + \sum_{k=1}^{\infty} \exp\left(-A\, k^{1 - C_2\,|\log(1-U)|}\right),$$

noting that $\exp\left(-A\,(k-2)^{1 - C_2\,|\log(1-U)|}\right) = 1$ when $k = 2$. By a Taylor expansion of $\exp(z)$, we can establish the inequality $\exp(z) > z^u \,/\, u!$ for any $z > 0$ and any $u > 0$, which further implies the inequality $\exp(-z) < u! \,/\, z^u$ for the same. Using this inequality,

$$\exp\left(-A\, k^{1 - C_2\,|\log(1-U)|}\right) \;<\; \frac{u!}{A^u\, k^{u\,(1 - C_2\,|\log(1-U)|)}}. \tag{C.17}$$

Assume that $1 - C_2\,|\log(1-U)| > 0$, which is satisfied when

$$\omega_2 \;<\; \frac{1}{(\omega_1 + 1)\,|\log(1-U)|},$$

recalling $C_2 \coloneqq \omega_2\,(\omega_1 + 1) > 0$. Taking $u = \lceil 2\,/\,(1 - C_2\,|\log(1-U)|)\rceil > 0$,

$$\sum_{k=1}^{\infty} \exp\left(-A\,k^{1-C_2\,|\log(1-U)|}\right) \;\;\leq\;\; \frac{u!}{A^u}\sum_{k=1}^{\infty}\frac{1}{k^2} \;\;=\;\; \frac{u!}{A^u}\left(\frac{\pi^2}{6}\right).$$

Thus, the first infinite series is bounded above by

$$C_1\,D_N^3\left(1 + \frac{u!}{A^u}\left(\frac{\pi^2}{6}\right)\right). \tag{C.18}$$

For the second infinite series, we write

$$\begin{aligned} & C_2\sum_{k=2}^{\infty}\log(k-1)\,\exp\left(-A\,(k-2)^{1-C_2\,|\log(1-U)|}\right) \\ = \;\; & C_2\sum_{k=3}^{\infty}\log(k-1)\,\exp\left(-A\,(k-2)^{1-C_2\,|\log(1-U)|}\right) \\ = \;\; & C_2\sum_{k=1}^{\infty}\log(k+1)\,\exp\left(-A\,k^{1-C_2\,|\log(1-U)|}\right). \end{aligned}$$

We employ (C.17) once more to show that

$$\exp\left(-A\,k^{1-C_2\,|\log(1-U)|}\right) \;\;\leq\;\; \frac{u!}{A^u\,k^3},$$

taking $u = \lceil 3\,/\,(1 - C_2\,|\log(1-U)|)\rceil > 0$ this time. Thus,

$$\begin{aligned} \sum_{k=1}^{\infty}\log(k+1)\,\exp\left(-A\,k^{1-C_2\,|\log(1-U)|}\right) \;\;&\leq\;\; \frac{u!}{A^u}\sum_{k=1}^{\infty}\frac{\log(k+1)}{k^3} \\ \leq\;\; \frac{u!}{A^u}\sum_{k=1}^{\infty}\frac{k}{k^3} \;\;=\;\; \frac{u!}{A^u}\sum_{k=1}^{\infty}\frac{1}{k^2} \;\;&=\;\; \frac{u!}{A^u}\left(\frac{\pi^2}{6}\right), \end{aligned}$$

using the inequality $\log(z+1) \leq z$ for $z \in (0,\infty)$. As a result, the second infinite series is bounded above by

$$C_2\,\frac{u!}{A^u}\left(\frac{\pi^2}{6}\right). \tag{C.19}$$

Combining (C.18), (C.19), and the bound $2 \geq \pi^2\,/\,6$,

$$\begin{aligned} \|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2 \;\;&\leq\;\; 1 + 4\,D_N^2 + C_1\,D_N^3\left(1 + 2\,\frac{u!}{A^u}\right) + 2\,C_2\,\frac{u!}{A^u} \\ &\leq\;\; 1 + 4\,D_N^2 + \max\{C_1, C_2\}\,A^{-u}\left((A^u + 2\,u!)\,D_N^3 + 2\,u!\right) \\ &\leq\;\; 1 + 4\,D_N^2 + \max\{C_1, C_2\}\,A^{-u}\left((1 + 2\,u!)\,D_N^3 + 1 + 2\,u!\right) \\ &\leq\;\; 1 + 4\,D_N^2 + \max\{C_1, C_2\}\,A^{-u}\,(1 + 2\,u!)\left(D_N^3 + 1\right), \end{aligned}$$

recalling that $A \coloneqq \exp(-C_1\, D_N^3\, |\log(1-U)|)$, which implies $A^u \in (0,1)$. Next, using the definitions of C_1 and C_2, and the assumption that $\omega_2 \leq \omega_1$,

$$\max\{C_1,\, C_2\} \;=\; \max\{2\,\omega_1\,(\omega_1+1),\, \omega_2\,(\omega_1+1)\} \;\leq\; 2\,\omega_1\,(\omega_1+1).$$

Then, there exist finite constants

$$C_3 \;\coloneqq\; 2\,(\omega_1+1)\,(1+2\,u!) \;>\; 0$$

and $C_4 \coloneqq C_1\,|\log(1-U)| > 0$, independent of N and p, such that

$$|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2 \;\leq\; 1 + 4\,D_N^2 + \omega_1\,C_3\,\exp(C_4\,D_N^3)\,(1+D_N^3).$$

We complete the proof by noticing two key facts:

- If $\omega_1 = 0$, then $\omega_1\,(1+D_N^3) \leq 2\,\omega_1\,D_N^3$ for all D_N^3.
- Since $D_N \geq 1$ under Models 2 and 3, $\omega_1\,(1+D_N^3) \leq 2\,\omega_1\,D_N^3$.

Thus, there exists $C_5 \coloneqq 2\,C_3 > 0$, independent of N and p, such that

$$\begin{aligned} |||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2 \;&\leq\; 1 + 4\,D_N^2 + \omega_1\,C_5\,D_N^3\,\exp(C_4\,D_N^3) \\ &\leq\; 1 + 4\,D_N^2 + \omega_1\,C_5\,\exp(C_4\,D_N^3 + 3\,\log D_N) \\ &\leq\; 1 + 4\,D_N^2 + \omega_1\,C_5\,\exp(C_6\,D_N^3), \end{aligned}$$

taking $C_6 \coloneqq (3 + C_4) > 0$ and since $\log D_N \leq D_N^3$. Recall that

$$U \;\coloneqq\; \frac{1}{1+\exp(-L-\vartheta\,\log N)}$$

is bounded away from 1 when $\vartheta = 0$.

We then have the following cases:

- If subpopulations overlap, i.e., $\omega_1 > 0$ and $\omega_2 \in [0, \omega_1]$, provided

  $$\omega_2 \;<\; \frac{1}{(1+\omega_1)\,|\log(1-U)|},$$

  then

  $$|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2 \;\leq\; 1 + 4\,D_N^2 + \omega_1\,C_5\,\exp(C_6\,D_N^3),$$

  provided $\vartheta = 0$ to ensure the constants are independent of N and p.
- If subpopulations do not overlap, i.e., $\omega_1 = \omega_2 = 0$, then

  $$|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2 \;\leq\; 1 + 4\,D_N^2.$$

  Since D_N does not depend on U, we allow $\vartheta > 0$. □

C.2.4. *Auxiliary results.* We prove Lemmas 13–16, which establish auxiliary results utilized in the proof of Lemma 12.

Lemma 13. *Consider Models 1, 2, and 3 with $\boldsymbol{\theta} \in \mathbb{R}^p$ and $\alpha \in [0,\, 1/2)$. Then there exist functions $L_k : \mathbb{R}^p \mapsto (0,1)$ and $U_k : \mathbb{R}^p \mapsto (0,1)$ $(k = 0,1)$ such that, for all $\{i,j\} \subset \mathcal{N}$ and $\boldsymbol{x}_{-\{i,j\}} \in \{0,1\}^{M-1}$,*

$$0 \;<\; L_k(\boldsymbol{\theta}) \;\leq\; \mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = k \mid \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}) \;\leq\; U_k(\boldsymbol{\theta}) \;<\; 1.$$

The functions $L_k(\boldsymbol{\theta})$ and $U_k(\boldsymbol{\theta})$ $(k = 0,1)$ are given by

$$L_1(\boldsymbol{\theta}) \;\coloneqq\; \begin{cases} \dfrac{1}{1 + \exp((3 + D_N)\, \|\boldsymbol{\theta}\|_\infty)} & \text{if } \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset \\[2ex] \dfrac{N^{-\alpha}}{1 + \exp(3\, \|\boldsymbol{\theta}\|_\infty)} & \text{if } \mathcal{N}_i \cap \mathcal{N}_j = \emptyset \end{cases}$$

$$U_1(\boldsymbol{\theta}) \;\coloneqq\; \begin{cases} \dfrac{1}{1 + \exp(-(3 + D_N)\, \|\boldsymbol{\theta}\|_\infty)} & \text{if } \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset \\[2ex] \dfrac{1}{1 + \exp(-3\, \|\boldsymbol{\theta}\|_\infty)\, N^{\alpha}} & \text{if } \mathcal{N}_i \cap \mathcal{N}_j = \emptyset \end{cases}$$

$$L_0(\boldsymbol{\theta}) \;\coloneqq\; \begin{cases} \dfrac{1}{1 + \exp((3 + D_N)\, \|\boldsymbol{\theta}\|_\infty)} & \text{if } \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset \\[2ex] \dfrac{1}{1 + \exp(3\, \|\boldsymbol{\theta}\|_\infty)\, N^{-\alpha}} & \text{if } \mathcal{N}_i \cap \mathcal{N}_j = \emptyset \end{cases}$$

$$U_0(\boldsymbol{\theta}) \;\coloneqq\; \begin{cases} \dfrac{1}{1 + \exp(-(3 + D_N)\, \|\boldsymbol{\theta}\|_\infty)} & \text{if } \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset \\[2ex] \dfrac{1}{1 + \exp(-3\, \|\boldsymbol{\theta}\|_\infty)\, N^{-\alpha}} & \text{if } \mathcal{N}_i \cap \mathcal{N}_j = \emptyset. \end{cases},$$

Proof of Lemma 13. Consider any pair of nodes $\{i,j\} \subset \mathcal{N}$ and any $\boldsymbol{x}_{-\{i,j\}} \in \{0,1\}^{\binom{N}{2}-1}$. We can express the full conditional probability

$$\mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = x_{i,j} \mid \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}})$$

two different ways depending whether $\mathcal{N}_i$ and $\mathcal{N}_j$ are disjoint.

First, if $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset$,

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = x_{i,j} \mid \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}) \\
= \;& \frac{\exp(\langle \boldsymbol{\theta},\, s(\boldsymbol{x}_{-\{i,j\}},\, x_{i,j}) \rangle)\, N^{-\alpha\, x_{i,j}}}{\exp(\langle \boldsymbol{\theta},\, s(\boldsymbol{x}_{-\{i,j\}},\, x_{i,j} = 0)\rangle) + \exp(\langle \boldsymbol{\theta},\, s(\boldsymbol{x}_{-\{i,j\}},\, x_{i,j} = 1)\rangle)\, N^{-\alpha}} \\
= \;& \frac{1}{g(0;\, \boldsymbol{x}_{-\{i,j\}},\, x_{i,j},\, \boldsymbol{\theta})\, N^{\alpha\, x_{i,j}} + g(1;\, \boldsymbol{x}_{-\{i,j\}},\, x_{i,j},\, \boldsymbol{\theta})\, N^{-\alpha\,(1 - x_{i,j})}},
\end{aligned}$$

defining, for $y \in \{0, 1\}$,

$$g(y;\, \boldsymbol{x}_{-\{i,j\}},\, x_{i,j},\, \boldsymbol{\theta}) \;\; \coloneqq \;\; \exp(\langle \boldsymbol{\theta},\, s(\boldsymbol{x}_{-\{i,j\}},\, y) - s(\boldsymbol{x}_{-\{i,j\}},\, x_{i,j})\rangle).$$

Note $g(x_{i,j};\, \boldsymbol{x}_{-\{i,j\}},\, x_{i,j},\, \boldsymbol{\theta}) = 1$ for all $\boldsymbol{x}_{-\{i,j\}} \in \{0,1\}^{\binom{N}{2}-1}$ and all $\boldsymbol{\theta} \in \mathbb{R}^p$.

Second, if $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$,

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = x_{i,j} \mid \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}) \\
= \;& \frac{\exp(\langle \boldsymbol{\theta},\, s(\boldsymbol{x}_{-\{i,j\}},\, x_{i,j}) \rangle)}{\exp(\langle \boldsymbol{\theta},\, s(\boldsymbol{x}_{-\{i,j\}},\, x_{i,j} = 0)\rangle) + \exp(\langle \boldsymbol{\theta},\, s(\boldsymbol{x}_{-\{i,j\}},\, x_{i,j} = 1)\rangle)} \\
= \;& \frac{1}{1 + g(1 - x_{i,j};\, \boldsymbol{x}_{-\{i,j\}},\, x_{i,j},\, \boldsymbol{\theta})}.
\end{aligned}$$

Next, observe that

$$\begin{aligned}
&\max_{\boldsymbol{x}_{-\{i,j\}} \in \{0,1\}^{M-1}} \left| s_l(\boldsymbol{x}_{-\{i,j\}},\, x_{i,j} = 0) - s_l(\boldsymbol{x}_{-\{i,j\}},\, x_{i,j} = 1) \right| \\
\leq \;& \begin{cases} 0 & \text{if } l \in \{1, \ldots, N\} \setminus \{i, j\} \\ 1 & \text{if } l \in \{i, j\} \\ 1 + D_N & \text{if } l = N+1 \text{ and } \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset \\ 0 & \text{if } l = N+1 \text{ and } \mathcal{N}_i \cap \mathcal{N}_j = \emptyset \end{cases}.
\end{aligned}$$

The bound on s_{N+1} follows from Lemma 14, whereas the conditions follow from Proposition 2: s_{N+1} is a function of only dependent edge variables and $X_{i,j}$ is independent of all other edges in the graph when $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset$. The bound on $s_l(\boldsymbol{x})$ ($l \in \mathcal{N}$) follows because $s_l(\boldsymbol{x}) = \sum_{h \in \mathcal{N} \setminus \{l\}} x_{l,h}$ is a function of $x_{i,j}$ if and only if $l \in \{i, j\}$. As a result, the triangle inequality shows that

$$\left|\left\langle \boldsymbol{\theta},\, s(\boldsymbol{x}_{-\{i,j\}},\, x_{i,j} = 1)\right\rangle - \left\langle \boldsymbol{\theta},\, s(\boldsymbol{x}_{-\{i,j\}},\, x_{i,j} = 0)\right\rangle\right|$$

$$\leq \begin{cases} (3 + D_N)\; \|\boldsymbol{\theta}\|_\infty & \text{if } \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset \\ 3\, \|\boldsymbol{\theta}\|_\infty & \text{if } \mathcal{N}_i \cap \mathcal{N}_j = \emptyset \end{cases},$$

implying, for $\{i, j\} \subseteq \mathcal{N}$ with $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$,

$$\exp(-(3 + D_N)\, \|\boldsymbol{\theta}\|_\infty) \;\leq\; g(1 - x_{i,j};\, \boldsymbol{x}_{-\{i,j\}},\, x_{i,j},\, \boldsymbol{\theta}) \;\leq\; \exp((3 + D_N)\, \|\boldsymbol{\theta}\|_\infty),$$

and for $\{i, j\} \subseteq \mathcal{N}$ with $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset$,

$$\exp(-3\, \|\boldsymbol{\theta}\|_\infty) \;\leq\; g(1 - x_{i,j};\, \boldsymbol{x}_{-\{i,j\}},\, x_{i,j},\, \boldsymbol{\theta}) \;\leq\; \exp(3\, \|\boldsymbol{\theta}\|_\infty).$$

As a result, for all $k \in \{0, 1\}$,

$$0 \quad < \quad L_k(\boldsymbol{\theta}) \quad \leq \quad \mathbb{P}(X_{i,j} = k \,|\, \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}) \quad \leq \quad U_k(\boldsymbol{\theta}) \quad < \quad 1,$$

where

$$L_1(\boldsymbol{\theta}) \quad \coloneqq \quad \begin{cases} \dfrac{1}{1 + \exp((3 + D_N)\, \|\boldsymbol{\theta}\|_\infty)} & \text{if } \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset \\[2ex] \dfrac{N^{-\alpha}}{1 + \exp(3\, \|\boldsymbol{\theta}\|_\infty)} & \text{if } \mathcal{N}_i \cap \mathcal{N}_j = \emptyset \end{cases}$$

and

$$U_1(\boldsymbol{\theta}) \quad \coloneqq \quad \begin{cases} \dfrac{1}{1 + \exp(-(3 + D_N)\, \|\boldsymbol{\theta}\|_\infty)} & \text{if } \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset \\[2ex] \dfrac{1}{1 + \exp(-3\, \|\boldsymbol{\theta}\|_\infty)\, N^{\alpha}} & \text{if } \mathcal{N}_i \cap \mathcal{N}_j = \emptyset \end{cases}.$$

We obtain $L_0(\boldsymbol{\theta})$ and $U_0(\boldsymbol{\theta})$ by noting

$$\mathbb{P}(X_{i,j} = 0 \,|\, \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}) \quad = \quad 1 - \mathbb{P}(X_{i,j} = 1 \,|\, \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}),$$

implying

$$1 - U_1(\boldsymbol{\theta}) \quad \leq \quad \mathbb{P}(X_{i,j} = 0 \,|\, \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}) \quad \leq \quad 1 - L_1(\boldsymbol{\theta}),$$

which allows us to obtain

$$L_0(\boldsymbol{\theta}) \quad \coloneqq \quad \begin{cases} \dfrac{1}{1+\exp((3+D_N)\,\|\boldsymbol{\theta}\|_\infty)} & \text{if } \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset \\[2ex] \dfrac{1}{1+\exp(3\,\|\boldsymbol{\theta}\|_\infty)\,N^{-\alpha}} & \text{if } \mathcal{N}_i \cap \mathcal{N}_j = \emptyset \end{cases}$$

and

$$U_0(\boldsymbol{\theta}) \quad \coloneqq \quad \begin{cases} \dfrac{1}{1+\exp(-(3+D_N)\,\|\boldsymbol{\theta}\|_\infty)} & \text{if } \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset \\[2ex] \dfrac{1}{1+\exp(-3\,\|\boldsymbol{\theta}\|_\infty)\,N^{-\alpha}} & \text{if } \mathcal{N}_i \cap \mathcal{N}_j = \emptyset \end{cases}.$$

□

Lemma 14. *Consider*

$$s_{N+1}(\boldsymbol{x}) \quad = \quad \sum_{i<j}^{N} x_{i,j}\, I_{i,j}(\boldsymbol{x}),$$

where

$$I_{i,j}(\boldsymbol{x}) \quad = \quad \begin{cases} 0 & \textit{if } \mathcal{N}_i \cap \mathcal{N}_j = \emptyset \\[2ex] \mathbb{1}\left(\displaystyle\sum_{h \in \mathcal{N}_i \cap \mathcal{N}_j} x_{i,h}\, x_{j,h} \;\geq\; 1\right) & \textit{if } \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset. \end{cases}$$

Then, for all $\{i,j\} \subset \mathcal{N}$*,*

$$\max_{(\boldsymbol{x},\boldsymbol{x}') \,\in\, \mathbb{X}\times\mathbb{X}:\; x_{v,w}=x'_{v,w},\; \{v,w\}\neq\{i,j\}} |s_{N+1}(\boldsymbol{x}) - s_{N+1}(\boldsymbol{x}')| \quad \leq \quad 1 + D_N,$$

where $D_N \coloneqq \max_{\{i,j\}\subset\mathcal{N}} |\mathfrak{N}_{i,j}|$.

Proof of Lemma 14. Consider any pair of nodes $\{i,j\} \subset \mathcal{N}$. The number of $x_{a,b}\, I_{a,b}(\boldsymbol{x})$ ($\{a,b\} \neq \{i,j\}$) which are a function of $x_{i,j}$ includes

- $\{a,b\} = \{i,b\}$ ($b \in \mathcal{N} \setminus \{i,j\}$) satisfying $j \in \mathcal{N}_i \cap \mathcal{N}_b \subseteq \mathcal{N}_i$, and
- $\{a,b\} = \{j,b\}$ ($b \in \mathcal{N} \setminus \{i,j\}$) satisfying $i \in \mathcal{N}_j \cap \mathcal{N}_b \subseteq \mathcal{N}_j$.

As a result, the number of summands $x_{a,b}\, I_{a,b}(\boldsymbol{x})$ $(\{a,b\} \neq \{i,j\})$ which can change value due to changing the value of $x_{i,j}$ is bounded above by $|(\{i\} \times \mathcal{N}_i) \cup (\{j\} \times \mathcal{N}_j)|$. Proposition 2 establishes that, for any $k \in \{1, \dots, K\}$ and $\{i,j\} \subset \mathcal{A}_k$,

$$\big\{\{a,b\} \,:\, (a,b) \in \{i\} \times \mathcal{N}_i \text{ or } (a,b) \in \{j\} \times \mathcal{N}_j\big\} \quad \subseteq \quad \mathfrak{N}_{i,j}.$$

Hence $|(\{i\} \times \mathcal{N}_i) \cup (\{j\} \times \mathcal{N}_j)| \leq D_N$, noting $D_N := \max_{\{i,j\} \subset \mathcal{N}} |\mathfrak{N}_{i,j}|$. As a result, the number of total summands $x_{a,b}\, I_{a,b}(\boldsymbol{x})$ which are a function of $x_{i,j}$ is bounded above by $1 + D_N$, now counting the case when $\{a,b\} = \{i,j\}$. Consider any $(\boldsymbol{x},\, \boldsymbol{x}') \in \mathbb{X} \times \mathbb{X}$ such that $x_{v,w} = x'_{v,w}$ for all $\{v,w\} \neq \{i,j\}$. Then, by the triangle inequality,

$$|s_{N+1}(\boldsymbol{x}) - s_{N+1}(\boldsymbol{x}')| \quad \leq \quad \sum_{\{a,b\} \subset \mathcal{N}} |x_{a,b}\, I_{a,b}(\boldsymbol{x}) - x'_{a,b}\, I_{a,b}(\boldsymbol{x}')| \quad \leq \quad 1 + D_N,$$

using $x_{a,b}\, I_{a,b}(\boldsymbol{x}) \in \{0,1\}$ for all $\{a,b\} \subset \mathcal{N}$ and all $\boldsymbol{x} \in \mathbb{X}$. □

Lemma 15. *Choose any $i \in \{1, \dots, M\}$ and any $\boldsymbol{x}_{1:i-1} \in \{0,1\}^{i-1}$. Then the coupling of the conditional distributions*

$$\mathbb{P}(\,\cdot \mid \boldsymbol{X}_{1:i-1} = \boldsymbol{x}_{1:i-1},\; X_i = 0) \quad \textit{and} \quad \mathbb{P}(\,\cdot \mid \boldsymbol{X}_{1:i-1} = \boldsymbol{x}_{1:i-1},\; X_i = 1)$$

of $\boldsymbol{X}_{(i+1):M}$ constructed in Lemma 12 is a valid coupling.

Proof of Lemma 15. Denote the coupling distribution generated by the algorithm in Lemma 12 by $\mathbb{Q}_{i,\boldsymbol{x}_{1:i-1}}$ and let $v_1, \dots, v_{M-i}$ be the vertices added to the set $\mathfrak{V}$ at iteration $1, \dots, M-i$ of the algorithm. To reduce the notational burden, define

$$\begin{aligned} & q(\boldsymbol{x}^\star_{a:b},\; \boldsymbol{x}^{\star\star}_{a:b} \mid \boldsymbol{x}^\star_{c:d},\; \boldsymbol{x}^{\star\star}_{c:d}) \\ := \quad & \mathbb{Q}_{i,\boldsymbol{x}_{1:i-1}}(\boldsymbol{X}^\star_{a:b} = \boldsymbol{x}^\star_{a:b},\; \boldsymbol{X}^{\star\star}_{a:b} = \boldsymbol{x}^{\star\star}_{a:b} \mid \boldsymbol{X}^\star_{c:d} = \boldsymbol{x}^\star_{c:d},\; \boldsymbol{X}^{\star\star}_{c:d} = \boldsymbol{x}^{\star\star}_{c:d}), \end{aligned}$$

where $a, b, c, d \in \{1, \dots, M\}$ are distinct and $\{a, \dots, b\} \cap \{c, \dots, d\} = \emptyset$. By construction,

$$q(\boldsymbol{x}^\star_{i+1:M},\; \boldsymbol{x}^{\star\star}_{i+1:M}) \quad = \quad q(x^\star_{v_1},\; x^{\star\star}_{v_1}) \prod_{l=2}^{M-i} q(x^\star_{v_l},\; x^{\star\star}_{v_l} \mid \boldsymbol{x}^\star_{v_1,\dots,v_{l-1}},\; \boldsymbol{x}^{\star\star}_{v_1,\dots,v_{l-1}}).$$

Observe that

$$\begin{aligned} & \sum_{x^\star_{v_{M-i}} \in \{0,1\}} q(x^\star_{v_{M-i}},\; x^{\star\star}_{v_{M-i}} \mid \boldsymbol{x}^\star_{v_1,\dots,v_{M-i-1}},\; \boldsymbol{x}^{\star\star}_{v_1,\dots,v_{M-i-1}}) \\ = \; & \mathbb{P}(X_{v_{M-i}} = x^{\star\star}_{v_{M-i}} \mid \boldsymbol{X}_{1:i-1} = \boldsymbol{x}_{1:i-1},\; X_i = 1,\; \boldsymbol{X}_{v_1,\dots,v_{M-i-1}} = \boldsymbol{x}^{\star\star}_{v_1,\dots,v_{M-i-1}}) \end{aligned}$$

and

$$\sum_{x^{\star\star}_{v_{M-i}} \in \{0,1\}} q(x^{\star}_{v_{M-i}},\, x^{\star\star}_{v_{M-i}} \mid \boldsymbol{x}^{\star}_{v_1,\dots,v_{M-i-1}},\, \boldsymbol{x}^{\star\star}_{v_1,\dots,v_{M-i-1}})$$

$$= \mathbb{P}(X_{v_{M-i}} = x^{\star}_{v_{M-i}} \mid \boldsymbol{X}_{1:i-1} = \boldsymbol{x}_{1:i-1},\ X_i = 0,\ \boldsymbol{X}_{v_1,\dots,v_{M-i-1}} = \boldsymbol{x}^{\star}_{v_1,\dots,v_{M-i-1}}),$$

owing to the fact that $(X^{\star}_{v_{M-i}},\, X^{\star\star}_{v_{M-i}})$ is distributed according to the optimal coupling of the conditional distributions

$$\mathbb{P}(X_{v_{M-i}} = \,\cdot \mid \boldsymbol{X}_{1:i-1} = \boldsymbol{x}_{1:i-1},\ X_i = 0,\ \boldsymbol{X}_{v_1,\dots,v_{M-i-1}} = \boldsymbol{x}^{\star}_{v_1,\dots,v_{M-i-1}})$$

and

$$\mathbb{P}(X_{v_{M-i}} = \,\cdot \mid \boldsymbol{X}_{1:i-1} = \boldsymbol{x}_{1:i-1},\ X_i = 1,\ \boldsymbol{X}_{v_1,\dots,v_{M-i-1}} = \boldsymbol{x}^{\star\star}_{v_1,\dots,v_{M-i-1}}).$$

We can repeat the same argument to show that

$$\begin{aligned} &\sum_{x^{\star}_{v_1} \in \{0,1\}} \cdots \sum_{x^{\star}_{v_{M-i}} \in \{0,1\}} q(\boldsymbol{x}^{\star}_{v_1,\dots,v_{M-i}},\, \boldsymbol{x}^{\star\star}_{i+1:M}) \\ = \;& \mathbb{P}(\boldsymbol{X}_{i+1:M} = \boldsymbol{x}^{\star\star}_{1+i:M} \mid \boldsymbol{X}_{1:i-1} = \boldsymbol{x}_{1:i-1},\ X_i = 1) \end{aligned}$$

and

$$\begin{aligned} &\sum_{x^{\star\star}_{v_1} \in \{0,1\}} \cdots \sum_{x_{v^{\star\star}_{M-i}} \in \{0,1\}} q(\boldsymbol{x}^{\star}_{i+1:M},\, \boldsymbol{x}^{\star\star}_{v_1,\dots,v_{M-i}}) \\ = \;& \mathbb{P}(\boldsymbol{X}_{i+1:M} = \boldsymbol{x}^{\star}_{1+i:M} \mid \boldsymbol{X}_{1:i-1} = \boldsymbol{x}_{1:i-1},\ X_i = 0), \end{aligned}$$

so the coupling is indeed a valid coupling of the conditional distributions

$$\mathbb{P}(\boldsymbol{X}_{i+1:M} = \,\cdot \mid \boldsymbol{X}_{1:i-1} = \boldsymbol{x}_{1:i-1},\ X_i = 0)$$

and

$$\mathbb{P}(\boldsymbol{X}_{i+1:M} = \,\cdot \mid \boldsymbol{X}_{1:i-1} = \boldsymbol{x}_{1:i-1},\ X_i = 1).$$

□

Lemma 16. *Consider Models 1, 2, and 3, any* $v \in \{1,\dots,M\}$, *and any* $(\boldsymbol{x}_{-v},\, \boldsymbol{x}'_{-v}) \in \{0,1\}^{M-1} \times \{0,1\}^{M-1}$. *Define*

$$\pi_{v,\,\boldsymbol{x}_{-v},\,\boldsymbol{x}'_{-v}} \;:=\; \|\mathbb{P}(\,\cdot \mid \boldsymbol{X}_{-v} = \boldsymbol{x}_{-v}) - \mathbb{P}(\,\cdot \mid \boldsymbol{X}_{-v} = \boldsymbol{x}'_{-v})\|_{TV}$$

and

$$\pi^{\star} \;:=\; \max_{1 \le v \le M} \; \max_{(\boldsymbol{x}_{-v},\,\boldsymbol{x}'_{-v}) \in \{0,1\}^{M-1} \times \{0,1\}^{M-1}} \pi_{v,\,\boldsymbol{x}_{-v},\,\boldsymbol{x}'_{-v}},$$

and define $D_N \coloneqq \max_{\{i,j\} \subset \mathcal{N}} |\mathfrak{N}_{i,j}|$. *Then*

$$\pi^\star \;\; \leq \;\; \begin{cases} 0 & \textit{under Model 1} \\ \dfrac{1}{1 + \exp(-(3 + D_N)\, \|\boldsymbol{\theta}^\star\|_\infty)} & \textit{under Models 2 and 3.} \end{cases}$$

Proof of Lemma 16. Under Model 1, edge variables X_v are independent, which implies that $\pi_{v,\, \boldsymbol{x}_{-v},\, \boldsymbol{x}'_{-v}} = 0$ for all $v \in \{1, \ldots, M\}$ and all $(\boldsymbol{x}_{-v},\, \boldsymbol{x}'_{-v}) \in \{0,1\}^{M-1} \times \{0,1\}^{M-1}$, which in turn implies that $\pi^\star = 0$. To bound $\pi^\star$ under Models 2 and 3, we distinguish two cases:

(a) If edge variable X_v corresponds to a pair of nodes with non-intersecting node neighborhoods, i.e., a pair $\{i, j\} \subset \mathcal{N}$ with $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset$, then X_v is independent of all other edge variables by Proposition 2. As a result, $\pi_{v,\, \boldsymbol{x}_{-v},\, \boldsymbol{x}'_{-v}} = 0$ for all $(\boldsymbol{x}_{-v},\, \boldsymbol{x}'_{-v}) \in \{0,1\}^{M-1} \times \{0,1\}^{M-1}$.

(b) If edge variable X_v corresponds to a pair of nodes with intersecting node neighborhoods, i.e., a pair $\{i, j\} \subset \mathcal{N}$ with $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$, then X_v is not independent of all other edges, implying $\pi_{v,\, \boldsymbol{x}_{-v},\, \boldsymbol{x}'_{-v}} > 0$ for some or all $(\boldsymbol{x}_{-v},\, \boldsymbol{x}'_{-v}) \in \{0,1\}^{M-1} \times \{0,1\}^{M-1}$.

We focus henceforth on case (b). Consider any $v \in \{1, \ldots, M\}$ such that $\pi_{v,\, \boldsymbol{x}_{-v},\, \boldsymbol{x}'_{-v}} > 0$ for some $(\boldsymbol{x}_{-v},\, \boldsymbol{x}'_{-v}) \in \{0,1\}^{M-1} \times \{0,1\}^{M-1}$ and define

$$a_0 \;=\; \mathbb{P}(X_v = 0 \mid \boldsymbol{X}_{-v} = \boldsymbol{x}_{-v}) \quad \text{and} \quad a_1 \;=\; \mathbb{P}(X_v = 1 \mid \boldsymbol{X}_{-v} = \boldsymbol{x}_{-v})$$

$$b_0 \;=\; \mathbb{P}(X_v = 0 \mid \boldsymbol{X}_{-v} = \boldsymbol{x}'_{-v}) \quad \text{and} \quad b_1 \;=\; \mathbb{P}(X_v = 1 \mid \boldsymbol{X}_{-v} = \boldsymbol{x}'_{-v}).$$

Then

$$\pi_{v,\, \boldsymbol{x}_{-v},\, \boldsymbol{x}'_{-v}} = \frac{1}{2}\left(|(1 - a_1) - (1 - b_1)| + |a_1 - b_1|\right) = |a_1 - b_1| \leq \max\{a_1,\, b_1\}.$$

By symmetry,

$$\pi_{v,\, \boldsymbol{x}_{-v},\, \boldsymbol{x}'_{-v}} \;\; \leq \;\; \max\{a_0,\, b_0\},$$

which implies that

$$\pi_{v,\, \boldsymbol{x}_{-v},\, \boldsymbol{x}'_{-v}} \;\; \leq \;\; \min\left\{\max\{a_0,\, b_0\},\; \max\{a_1,\, b_1\}\right\}.$$

Lemma 13 shows that, under Models 2 and 3,

$$\mathbb{P}(X_v = 0 \mid \boldsymbol{X}_{-v} = \boldsymbol{x}_{-v}) \;\; \leq \;\; \frac{1}{1 + \exp(-(3 + D_N)\, \|\boldsymbol{\theta}^\star\|_\infty)}$$

and

$$\mathbb{P}(X_v = 1 \mid \boldsymbol{X}_{-v} = \boldsymbol{x}_{-v}) \quad \leq \quad \frac{1}{1 + \exp(-(3 + D_N)\, \|\boldsymbol{\theta}^\star\|_\infty)}.$$

We therefore conclude that, under Models 2 and 3,

$$\begin{aligned} \pi^\star \quad &\leq \quad \min\{\max\{a_0,\, b_0\},\; \max\{a_1,\, b_1\}\} \\ &\leq \quad \frac{1}{1 + \exp(-(3 + D_N)\, \|\boldsymbol{\theta}^\star\|_\infty)}. \end{aligned}$$

□

APPENDIX D: PROOF OF PROPOSITION 1

We prove Proposition 1 stated in Section 2.4 of the manuscript.

Proof of Proposition 1. The expected degree of any node $i \in \mathcal{N}$ under Model 3 with $\alpha \in (0,\, 1]$ is given by

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\theta}} \left(\sum_{j \neq i}^{N} X_{i,j} \right) \quad &= \quad \sum_{j \,\in\, \mathfrak{A}_{i,1}} \mathbb{E}_{\boldsymbol{\theta}}\, X_{i,j} + \sum_{j \,\in\, \mathfrak{A}_{i,2}} \mathbb{E}_{\boldsymbol{\theta}}\, X_{i,j} \\ &\leq \quad |\mathfrak{A}_{i,1}| \max_{j \,\in\, \mathfrak{A}_{i,1}} \mathbb{E}_{\boldsymbol{\theta}}\, X_{i,j} + |\mathfrak{A}_{i,2}| \max_{j \,\in\, \mathfrak{A}_{i,2}} \mathbb{E}_{\boldsymbol{\theta}}\, X_{i,j}, \end{aligned}$$

where

- $\mathfrak{A}_{i,1} = \{j \in \mathcal{N} \setminus \{i\} : \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset\}$;
- $\mathfrak{A}_{i,2} = \{j \in \mathcal{N} \setminus \{i\} : \mathcal{N}_i \cap \mathcal{N}_j = \emptyset\}$.

We bound the expectations of edges $\mathbb{E}_{\boldsymbol{\theta}}\, X_{i,j}$ by using the bound

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\theta}}\, X_{i,j} \quad &= \quad \mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = 1) \\ &\leq \quad \max_{\boldsymbol{x}_{-\{i,\,j\}} \,\in\, \{0,1\}^{\binom{N}{2}-1}} \mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = 1 \mid \boldsymbol{X}_{-\{i,\,j\}} = \boldsymbol{x}_{-\{i,\,j\}}). \end{aligned}$$

For any $j \in \mathfrak{A}_{i,1}$, $\mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = 1 \,|\, \boldsymbol{X}_{-\{i,\,j\}} = \boldsymbol{x}_{-\{i,\,j\}}) \leq 1 \leq \exp(3\, \|\boldsymbol{\theta}\|_\infty)$ for all $\boldsymbol{x}_{-\{i,\,j\}} \in \{0,1\}^{\binom{N}{2}-1}$. In addition, for any $j \in \mathfrak{A}_{i,2}$, Lemma 13 in Appendix C.2.3 shows that

$$\mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = 1 \mid \boldsymbol{X}_{-\{i,\,j\}} = \boldsymbol{x}_{-\{i,\,j\}}) \leq \frac{1}{1 + \exp(-3\, \|\boldsymbol{\theta}\|_\infty)\, N^\alpha} < \frac{\exp(3\, \|\boldsymbol{\theta}\|_\infty)}{N^\alpha},$$

for all $\boldsymbol{x}_{-\{i,j\}} \in \{0,1\}^{\binom{N}{2}-1}$. Hence,

$$\mathbb{E}_{\boldsymbol{\theta}} \left(\sum_{j \neq i}^{N} X_{i,j} \right) \;\leq\; \exp(3\, \|\boldsymbol{\theta}\|_\infty)\, (|\mathfrak{A}_{i,1}| + |\mathfrak{A}_{i,2}|\, N^{-\alpha}).$$

Bounding $|\mathfrak{A}_{i,1}|$. To bound $|\mathfrak{A}_{i,1}|$, we distinguish two cases:

- $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$ and $j \in \mathcal{N}_i$, which implies that there exists $k \in \{1, \dots, K\}$ such that $\{i,\, j\} \subset \mathcal{A}_k$, in which case $j \in \mathcal{N}_i$.
- $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$ and $j \notin \mathcal{N}_i$, in which case there exists $h \in \mathcal{N}_i \cap \mathcal{N}_j$, which implies that $h \in \mathcal{N}_i$ and $h \in \mathcal{N}_j$, which further implies $j \in \mathcal{N}_h$.

The number of nodes $j \in \mathcal{N}$ satisfying the first case is bounded above by $|\mathcal{N}_i| \leq \max_{1 \leq r \leq N} |\mathcal{N}_r| \leq (\max_{1 \leq r \leq N} |\mathcal{N}_r|)^2$, since $|\mathcal{N}_r| \in \{0, 1, \dots, N-1\}$, and the number of $j \in \mathcal{N}$ satisfying the second case is bounded above by

$$\left| \bigcup_{h \in \mathcal{N}_i} \mathcal{N}_h \right| \;\leq\; |\mathcal{N}_i| \max_{h \in \mathcal{N}_i} |\mathcal{N}_h| \;\leq\; \left(\max_{1 \leq h \leq N} |\mathcal{N}_h| \right)^2.$$

In conclusion,

$$|\mathfrak{A}_{i,1}| \;\leq\; 2 \left(\max_{1 \leq h \leq N} |\mathcal{N}_h| \right)^2.$$

Bounding $|\mathfrak{A}_{i,2}|$. For each node $i \in \mathcal{N}$, there are at most $N - 1 < N$ other nodes $j \in \mathcal{N} \setminus \mathcal{N}_i$, hence $|\mathfrak{A}_{i,2}| \leq N \leq 2\, N$.

Conclusion. By collecting terms, for all nodes $i \in \mathcal{N}$,

$$\mathbb{E}_{\boldsymbol{\theta}} \left(\sum_{j \neq i}^{N} X_{i,j} \right) \;\leq\; 2\, \exp(3\, \|\boldsymbol{\theta}\|_\infty) \left(\left(\max_{1 \leq h \leq N} |\mathcal{N}_h| \right)^2 + N^{1-\alpha} \right).$$

□

APPENDIX E: SIMULATION RESULTS

We study the performance of maximum pseudo-likelihood estimators by considering populations with $N = 125$, 250, 500, and 1,000 nodes. We focus on maximum pseudo-likelihood estimators, because computing maximum likelihood and Monte Carlo maximum likelihood estimators is too time-consuming when N is large (e.g., when $N = 500$ and $N = 1{,}000$). For each value of N, we generate 1,000 populations with overlapping subpopulations as follows:

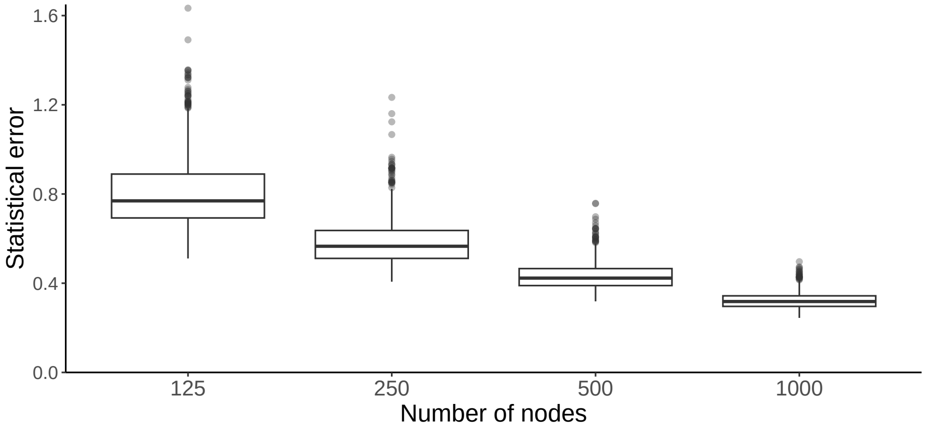


FIG 3. *The statistical error $\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star\|_\infty$ of maximum pseudo-likelihood estimator $\widetilde{\boldsymbol{\theta}}$ as an estimator of $\boldsymbol{\theta}^\star \in \mathbb{R}^{N+1}$ plotted against the number of nodes N.*

- The number of subpopulations K is $N \,/\, 25$.
- Each node $i \in \mathcal{N}$ belongs to $1 + Y_i$ subpopulations, where $Y_i \overset{\text{iid}}{\sim} \text{Binomial}(K - 1,\; 1 \,/\, K)$ $(i = 1, \ldots, N)$.
- For node $i \in \{1, \ldots, N\}$, the $1 + Y_i$ subpopulation memberships are sampled from the Multinomial$(p_1^{(i)}, \ldots, p_K^{(i)})$ distribution with

$$p_k^{(i)} \;=\; \begin{cases} \dfrac{1}{K} & \text{if } i = 1 \\[2ex] \dfrac{1}{K-1}\left(1 - \dfrac{N_k^{(i-1)}}{N_1^{(i-1)} + \ldots + N_K^{(i-1)}}\right) & \text{if } i \in \{2, \ldots, N\}, \end{cases}$$

  where $N_k^{(i-1)}$ is the number of nodes in $\{1, \ldots, i-1\}$ that belong to subpopulation $\mathcal{A}_k$ $(k = 1, \ldots, K)$ at the current time.

We consider Model 2 with degree parameters $\theta_1^\star, \ldots, \theta_N^\star$ drawn from Uniform$(-1.25, -.75)$ and brokerage parameter $\theta_{N+1}^\star = .25$. For each population size $N \in \{125, 250, 500, 1000\}$, we generate a graph from Model 2 and compute the maximum pseudo-likelihood estimator from the generated graph. For each value of N, the gradient ascent algorithm used to compute the maximum pseudo-likelihood estimator converged for at least 95% of the simulated data sets, and the following simulation results are based on the simulated data sets for which the gradient ascent algorithm converged.

Figure 3 demonstrates that the statistical error $\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star\|_\infty$ of $\widetilde{\boldsymbol{\theta}}$ as an estimator of the data-generating parameter vector $\boldsymbol{\theta}^\star \in \mathbb{R}^{N+1}$ decreases as

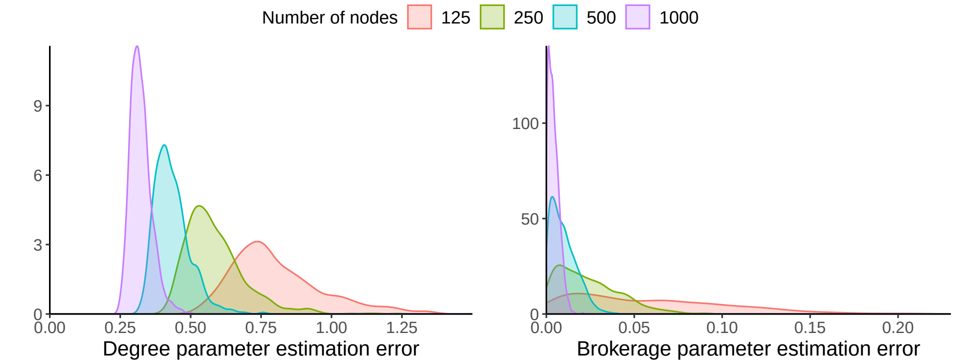


Fig 4. *The maximum deviation* $\max_{1 \leq i \leq N} |\widetilde{\theta}_i - \theta_i^\star|$ *of the maximum pseudo-likelihood estimators* $\widetilde{\theta}_i$ *from the data-generating degree parameters* $\theta_i^\star$ *(*$i = 1, \ldots, N$*) (left) and the deviation* $|\widetilde{\theta}_{N+1} - \theta_{N+1}^\star|$ *of the maximum pseudo-likelihood estimator* $\widetilde{\theta}_{N+1}$ *from the data-generating brokerage parameter* $\theta_{N+1}^\star$ *(right).*

the number of nodes N increases. Figure 4 decomposes the statistical error of $\widetilde{\boldsymbol{\theta}}$ into the statistical error of the degree parameter estimators $\widetilde{\theta}_1, \ldots, \widetilde{\theta}_N$ and the statistical error of the brokerage parameter estimator $\widetilde{\theta}_{N+1}$. Figure 4 reveals that the brokerage parameter is estimated with greater accuracy than the degree parameters, which makes sense as the degree parameters are greater in absolute value than the brokerage parameter and there are N estimated degree parameters $\widetilde{\theta}_1, \ldots, \widetilde{\theta}_N$, compared with the single estimated brokerage parameter $\widetilde{\theta}_{N+1}$.

REFERENCES

[1] Amini, A. A., Chen, A., Bickel, P. J., and Levina, E. (2013), "Pseudo-likelihood methods for community detection in large sparse networks," *The Annals of Statistics*, 41, 2097–2122.

[2] Bhattacharya, B. B., and Mukherjee, S. (2018), "Inference in Ising models," *Bernoulli*, 24, 493–525.

[3] Bickel, P. J., and Chen, A. (2009), "A nonparametric view of network models and Newman-Girvan and other modularities," in *Proceedings of the National Academy of Sciences*, Vol. 106, pp. 21068–21073.

[4] Bollobás, B., Riordan, O., Spencer, J., and Tusnády, G. (2001), "The degree sequence of a scale-free random graph process," *Random Structures & Algorithms*, 18, 279–290.

[5] Brown, L. (1986), *Fundamentals of Statistical Exponential Families: With Applications in Statistical Decision Theory*, Hayworth, CA, USA: Institute of Mathematical Statistics.

[6] Caron, F., and Fox, E. B. (2017), "Sparse graphs using exchangeable random measures," *Journal of the Royal Statistical Society, Series B (with discussion)*, 79, 1–44.

[7] Chatterjee, S. (2007), "Estimation in spin glasses: A first step," *The Annals of Statistics*, 35, 1931–1946.

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

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

[10] 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.

[11] 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.

[12] Comets, F. (1992), “On consistency of a class of estimators for exponential families of Markov random fields on the lattice,” *The Annals of Statistics*, 20, 455–468.

[13] Crane, H., and Dempsey, W. (2018), “Edge exchangeable models for interaction networks,” *Journal of the American Statistical Association*, 113, 1311–1326.

[14] Dawid, A. P. (1979), “Conditional independence in statistical theory,” *Journal of the Royal Statistical Society, Series B*, 41, 1–31.

[15] Frank, O., and Strauss, D. (1986), “Markov graphs,” *Journal of the American Statistical Association*, 81, 832–842.

[16] Furi, M., and Martelli, M. (1991), “On the mean value theorem, inequality, and inclusion,” *The American Mathematical Monthly*, 98, 840–846.

[17] Gao, C., Ma, Z., Zhang, A. Y., and Zhou, H. H. (2018), “Community detection in degree-corrected block models,” *The Annals of Statistics*, 46, 2153–2185.

[18] Ghosal, P., and Mukherjee, S. (2020), “Joint estimation of parameters in Ising model,” *The Annals of Statistics*, 48, 785–810.

[19] Handcock, M. S. (2003), “Statistical Models for Social Networks: Inference and Degeneracy,” in *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, eds. Breiger, R., Carley, K., and Pattison, P., Washington, D.C.: National Academies Press, pp. 1–12.

[20] Harville, D. A. (1997), *Matrix algebra from a statistician’s perspective*, New York: Springer.

[21] Hillar, C. J., and Wibisono, A. (2015), “A Hadamard-type lower bound for symmetric diagonally dominant positive matrices,” *Linear Algebra and its Applications*, 472, 135–141.

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

[23] Holland, P. W., and Leinhardt, S. (1981), “An exponential family of probability distributions for directed graphs,” *Journal of the American Statistical Association*, 76, 33–65.

[24] Karwa, V., and Slavković, A. B. (2016), “Inference using noisy degrees: Differentially private β-model and synthetic graphs,” *The Annals of Statistics*, 44, 87–112.

[25] 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.

[26] Lehmann, E. L., and Romano, J. P. (2005), *Testing Statistical Hypotheses*, New York: Springer-Verlag, 3rd ed.

[27] Lindvall, T. (2002), *Lectures On The Coupling Method*, Courier Corporation.

[28] Mukherjee, R., Mukherjee, S., and Sen, S. (2018), “Detection thresholds for the β-model on sparse graphs,” *The Annals of Statistics*, 46, 1288–1317.

[29] Mukherjee, S. (2020), “Degeneracy in sparse ERGMs with functions of degrees as

sufficient statistics," *Bernoulli*, 26, 1016–1043.

[30] 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.

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

[32] Rinaldo, A., Fienberg, S. E., and Zhou, Y. (2009), "On the geometry of discrete exponential families with application to exponential random graph models," *Electronic Journal of Statistics*, 3, 446–484.

[33] Rinaldo, A., Petrović, S., and Fienberg, S. E. (2013), "Maximum likelihood estimation in the β-model," *The Annals of Statistics*, 41, 1085–1110.

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

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

[36] 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.

[37] Schweinberger, M., and Stewart, J. R. (2020), "Concentration and consistency results for canonical and curved exponential-family models of random graphs," *The Annals of Statistics*, 48, 374–396.

[38] Stewart, J. R., and Schweinberger, M. (2023), "Supplement to: Pseudo-likelihood-based M-estimators for random graphs with dependent edges and parameter vectors of increasing dimension," *Department of Statistics, Florida State University.*

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

[40] Watts, D. J. (2003), *Six Degrees. The Science of a Connected Age*, Norton.

[41] Yan, T., Leng, C., and Zhu, J. (2016), "Asymptotics in directed exponential random graph models with an increasing bi-degree sequence," *The Annals of Statistics*, 44, 31–57.

[42] Yan, T., and Xu, J. (2013), "A central limit theorem in the β-model for undirected random graphs with a diverging number of vertices," *Biometrika*, 100, 519–524.

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

Michael Schweinberger
Department of Statistics
Penn State University
326 Thomas Building
University Park, PA 16802
E-mail: mus47@psu.edu