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., 27, 13, 52, 41, 30, 36, 15]; exchangeable random graph models [e.g., 10, 17]; stochastic block models [e.g., 1, 5, 6, 42, 34, 2, 21]; latent space models [e.g., 26, 47]; and exponential-family models of random graphs [e.g., 23, 3, 43, 12, 37, 45]. Other models are small-world networks [50] and scale-free networks with power law degree distributions [7]. 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 [27, 19], three questions have loomed large:

I. How can one construct models that allow the propensities of nodes to form edges and other subgraphs 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 [8], 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 [27, 13], generalized linear models of random graphs, and undirected graphical models of random graphs [19, 32]. An alternative route, not considered here, is provided by the Hoover-Aldous representation theorem [5] via exchangeable random graphs [10, 17], 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 more often than not 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., 13, 52, 41, 30, 51, 36, 15] – are scarce in the literature on exponential-family random graph models, with two recent exceptions. Mukherjee [37] 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 [45] 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. [13] – studied by Rinaldo et al. [41], Yan and Xu [52], Karwa and Slavković [30], Mukherjee et al. [36], Chen et al. [15], 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 with dependent edges have in common that the number of parameters increases with the number of nodes.

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 [29, 16, 11, 4, 22], 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. [39] 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 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. Simulation results can be found in the supplement [46].

1.5. *Notation.* Let $\mathcal{N} \coloneqq \{1, \ldots, N\}$ $(N \geq 2)$ 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. To denote subgraphs induced by subsets of nodes $\mathcal{C} \subseteq \mathcal{N}$ and $\mathcal{D} \subseteq \mathcal{N}$, we write $\boldsymbol{X}_{\mathcal{C},\,\mathcal{D}} \coloneqq (X_{i,j} :\; i \neq j,\; i \in \mathcal{C},\; j \in \mathcal{D})$. If $\mathcal{C} = \mathcal{D}$, we write $\boldsymbol{X}_{\mathcal{C}}$ instead of $\boldsymbol{X}_{\mathcal{C},\mathcal{C}}$. 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 determinant of $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ is written as $\det(\boldsymbol{A})$. The closed hypercube in $\mathbb{R}^d$ centered at $\boldsymbol{c} \in \mathbb{R}^d$ with radius $\rho > 0$ is denoted by $\mathcal{B}_\infty(\boldsymbol{c},\, \rho) \coloneqq \{\boldsymbol{a} \in \mathbb{R}^d : \|\boldsymbol{a} - \boldsymbol{c}\|_\infty \leq \rho\}$. For any subset $\mathcal{S} \subset \mathbb{R}^d$, $\operatorname{int}(\mathcal{S})$ and $\operatorname{bd}(\mathcal{S})$ denote the interior and boundary of $\mathcal{S}$, respectively. The total variation distance between two probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ defined on a common measurable space is denoted by $\|\mathbb{P}_1 - \mathbb{P}_2\|_{\mathrm{TV}}$. 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 $A, B, C, \ldots$ 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. Consider a family of probability measures $\{\mathbb{P}_{\boldsymbol{\theta}},\, \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ 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 $\varphi_{i,j} : \{0,1\}^{|\mathcal{S}_{i,j}|+1} \times \boldsymbol{\Theta} \mapsto \mathbb{R}^+ \cup \{0\}$ 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 $\{a, b\} \subset \mathcal{N}$. We allow the dimension p of parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$ to increase as a function of the number of nodes N. A natural choice of reference measure ν is 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 [18], and the resulting models can be viewed as undirected graphical models of random graphs [19, 32]. In contrast to the undirected graphical models of random graphs by Frank and Strauss [19], which allow edges to depend on many other edges and can give rise to undesirable behavior [e.g., model near-degeneracy, 23, 40, 3, 43, 9, 12], 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. Other possible models are sketched in Section 2.5.

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. [32] and Wainwright and Jordan [49]. We therefore assume that the functions of edges $\varphi_{i,j}$ are of the form

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

where $a_{i,j} : \{0,1\}^{|\mathbb{S}_{i,j}|+1} \mapsto \mathbb{R}^+ \cup \{0\}$ is a function of $x_{i,j}$ and $\boldsymbol{x}_{\mathbb{S}_{i,j}}$, which can be used to induce sparsity by penalizing edges, and $\langle \boldsymbol{\theta},\, s_{i,j}(x_{i,j},\, \boldsymbol{x}_{\mathbb{S}_{i,j}}) \rangle$ is the inner product of a vector of parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$ and a vector of statistics $s_{i,j} : \{0,1\}^{|\mathbb{S}_{i,j}|+1} \mapsto \mathbb{R}^p$ ($\{i,j\} \subset \mathcal{N}$). 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 $a : \mathbb{X} \mapsto \mathbb{R}^+ \cup \{0\}$ is given by $a(\boldsymbol{x}) \coloneqq \prod_{i<j}^N a_{i,j}(x_{i,j},\, \boldsymbol{x}_{\mathbb{S}_{i,j}})$ and $s : \mathbb{X} \mapsto \mathbb{R}^p$ is given by

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

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

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

The parameter space is $\boldsymbol{\Theta} \coloneqq \{\boldsymbol{\theta} \in \mathbb{R}^p \;:\; \psi(\boldsymbol{\theta}) < \infty\} = \mathbb{R}^p$, because the family of densities is an exponential family of densities with respect to a σ-finite measure with a finite support [8]. To ensure that $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is identifiable, we assume that the exponential family is minimal in the sense of Brown [8, 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 [8, 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 [13]. 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 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

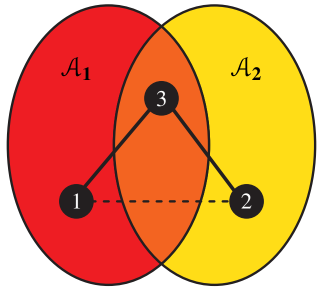

FIG 1. *A graphical representation of the dependencies among edges induced by brokerage. Consider two overlapping subpopulations $\mathcal{A}_1$ and $\mathcal{A}_2$. The nodes $1 \in \mathcal{A}_1 \setminus \mathcal{A}_2$ and $2 \in \mathcal{A}_2 \setminus \mathcal{A}_1$ do not belong to the same subpopulation, 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.*

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 [5]. 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 $i \in \mathcal{N}$:

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

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 $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{N+1}) \in \mathbb{R}^{N+1}$.

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

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

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 [8, 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., 41, 31, 36, 37, 15]. 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

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

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*

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

Proposition 1 reveals that when the neighborhoods $\mathcal{N}_i$ of nodes $i \in \mathcal{N}$ 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 the sizes of neighborhoods and $\|\boldsymbol{\theta}\|_\infty$ are bounded above, the expected degrees are $O(N^{1-\alpha})$.

2.5. *Other possible models.* The probabilistic framework is neither restricted to generalized β-models nor the interaction terms used by generalized β-models. Such interaction terms have turned out to be useful in practice [e.g., 28], but other interaction terms can be used to capture brokerage and other network-mediated phenomena [e.g., 44]. Examples include

- generalized β-models of communication chains, which include indicators of whether edges between nodes belong to a path of length $k \in \{2, 3, \dots\}$ passing through shared neighborhoods;
- models of random attributes and random connections allowing dependence between attributes and connections;
- spatio-temporal models of random attributes and random connections indexed by space and time.

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. Let $\boldsymbol{X}_{a:b} \coloneqq (X_a, \dots, X_b) \in \{0, 1\}^{b-a+1}$, where $a \leq b$ and $a, b \in \{1, \dots, M\}$. For each $i \in \{1, \dots, M\}$, we denote the conditional 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 $\mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{x}_{1:i-1}, x_i}$:

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

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 techniques [35]. Let $(\boldsymbol{X}^\star_{i+1:M}, \boldsymbol{X}^{\star\star}_{i+1:M})$ be a coupling of $\mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{x}_{1:i-1}, 0}$ and $\mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{x}_{1:i-1}, 1}$ and denote the joint probability mass function of $(\boldsymbol{X}^\star_{i+1:M}, \boldsymbol{X}^{\star\star}_{i+1:M})$ by $\mathbb{Q}_{\boldsymbol{\theta}^\star, i, \boldsymbol{x}_{1:i-1}}$. A coupling is constructed in the supplement [46]. Based on the coupling, we construct an $M \times M$ coupling matrix $\mathcal{D}_N(\boldsymbol{\theta}^\star)$ with elements

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

The elements $\mathcal{D}_{i,j}(\boldsymbol{\theta}^\star)$ represent bounds on the total variation distance between conditional probability mass functions of edge variables X_j $(j > i)$ given subgraph $(\boldsymbol{X}_{1:i-1}, X_i) = (\boldsymbol{x}_{1:i-1}, x_i)$ due to changing $x_i = 0$ to $x_i = 1$. While the definition of $\mathcal{D}_N(\boldsymbol{\theta}^\star)$ depends on the ordering of edge variables, it is possible to obtain bounds on $|\!|\!|\mathcal{D}_N(\boldsymbol{\theta}^\star)|\!|\!|_2$ that hold for all orderings. In Section 3.3.2, we take advantage of coupling techniques from the literature on percolation theory to bound the spectral norm $|\!|\!|\mathcal{D}_N(\boldsymbol{\theta}^\star)|\!|\!|_2$ of $\mathcal{D}_N(\boldsymbol{\theta}^\star)$.

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

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

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

$$\Psi_N \quad \coloneqq \quad \max_{1 \leq i \leq p} \|\boldsymbol{\Xi}_i\|_2.$$

To exclude the trivial case $\Psi_N = 0$, we assume that there exist finite constants $C_0 > 0$ and $N_0 \geq 2$ such that $\Psi_N > C_0 > 0$ 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 $\ell(\boldsymbol{\theta};\, \boldsymbol{x}) \coloneqq \log f_{\boldsymbol{\theta}}(\boldsymbol{x})$ and

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

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 $\epsilon^\star \in (0,\, \infty)$ be a constant, independent of N and p. Define $\mathcal{I}(\boldsymbol{\theta}) \coloneqq \nabla_{\boldsymbol{\theta}}^2\, \psi(\boldsymbol{\theta}) = \mathbb{C}_{\boldsymbol{\theta}}\, s(\boldsymbol{X}) = -\mathbb{E}_{\boldsymbol{\theta}}\, \nabla_{\boldsymbol{\theta}}^2\, \ell(\boldsymbol{\theta};\, \boldsymbol{X})$ [33, Theorem 2.7.1, p. 49] and

$$\begin{aligned}
\Lambda_N(\boldsymbol{\theta}^\star) \;\; &\coloneqq \;\; \inf_{\boldsymbol{\theta} \,\in\, \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)} \; \sqrt[p]{\det(\mathcal{I}(\boldsymbol{\theta}))} \\
\mathcal{C}_N(\boldsymbol{\theta}^\star) \;\; &\coloneqq \;\; \sup_{\boldsymbol{\theta} \,\in\, \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)} \; |||\mathcal{I}(\boldsymbol{\theta})|||_\infty \; |||(\mathcal{I}(\boldsymbol{\theta}))^{-1}|||_\infty \\
\Phi_N(\boldsymbol{\theta}^\star) \;\; &\coloneqq \;\; \frac{\Lambda_N(\boldsymbol{\theta}^\star)}{\mathcal{C}_N(\boldsymbol{\theta}^\star)\; |||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2\; \Psi_N\, \sqrt{\log \max\{N,\, p\}}},
\end{aligned}$$

where $\Lambda_N(\boldsymbol{\theta}^\star) \geq 0$, $\;\mathcal{C}_N(\boldsymbol{\theta}^\star) \,\geq\, 1,$ and $|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2 \,\geq\, 1$ by definition, while $\Psi_N > C_0 > 0$ by assumption, provided N is large enough.

Theorem 1. *Consider a single observation of a random graph with N nodes and $M = \binom{N}{2}$ dependent edges. Assume that $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta} = \mathbb{R}^p$, where $p \to \infty$ as $N \to \infty$ is allowed. If $\Phi_N(\boldsymbol{\theta}^\star) \to \infty$ as $N \to \infty$, there exists a finite constant $N_0 \geq 2$ such that, for all $N > N_0$, the random set $\widehat{\boldsymbol{\Theta}}$ is non-empty and its unique element $\widehat{\boldsymbol{\theta}}$ satisfies*

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

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

The assumption $\Phi_N(\boldsymbol{\theta}^\star) \,\to\, \infty$ ensures that $\mathcal{I}(\boldsymbol{\theta})$ is invertible for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ in a neighborhood $\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$ of the data-generating parameter vector $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta} = \mathbb{R}^p$, provided N is large enough. 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 Fisher information matrix $\mathcal{I}(\boldsymbol{\theta})$, via the geometric mean of its eigenvalues $\Lambda_N(\boldsymbol{\theta}^\star)$ and its ℓ_∞-induced condition number $\mathcal{C}_N(\boldsymbol{\theta}^\star)$ in a neighborhood of the data-generating parameter vector $\boldsymbol{\theta}^\star \in \mathbb{R}^p$;
- the dependence induced by the model, quantified by $|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2$;
- 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 [23, 40, 3, 43, 9, 12, 44]. It is known that some random graph models with dependent edges [e.g., the ill-posed triangle model, 23, 40, 3, 43, 9, 12, 44] 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 in the sense that the ℓ_∞-induced condition number of $\mathcal{I}(\boldsymbol{\theta})$ can be large for some or all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$, reducing the convergence rate via $\mathcal{C}_N(\boldsymbol{\theta}^\star)$.

Model near-degeneracy. On subsets of $\boldsymbol{\Theta}$ inducing near-degenerate distributions, the variances of sufficient statistics (e.g., the number of edges) can be small, so that $\det(\mathcal{I}(\boldsymbol{\theta})) = \det(\mathbb{C}_{\boldsymbol{\theta}}\, s(\boldsymbol{X}))$ 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 [43, 9], depressing the convergence rate via $|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2$ and Ψ_N. An example is the ill-posed triangle model [23, 40, 3, 43, 9, 12, 44]. 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.

Conclusion. On subsets of $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta}$ with $\Phi_N(\boldsymbol{\theta}^\star) \to 0$, Theorem 1 does not establish consistency. Theorem 1 does establish consistency on subsets of $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta}$ with $\Phi_N(\boldsymbol{\theta}^\star) \to \infty$. It is worth noting that phase transitions are of interest in physics, but are often viewed as a nuisance in the statistical analysis of relations involving living organisms. Thus, Theorem 1 establishes consistency on those subsets of $\boldsymbol{\Theta}$ that are capable of generating data resembling networks arising in the life sciences and the social sciences.

PROOF OF THEOREM 1. The mapping $\boldsymbol{\mu} : \boldsymbol{\Theta} \mapsto \mathbb{M}$ from the natural parameter space $\boldsymbol{\Theta} = \mathbb{R}^p$ to the mean-value parameter space $\mathbb{M} \coloneqq \boldsymbol{\mu}(\boldsymbol{\Theta})$ defined by $\boldsymbol{\mu}(\boldsymbol{\theta}) \coloneqq \nabla_{\boldsymbol{\theta}}\, \psi(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}\, s(\boldsymbol{X})$ is a homeomorphism [8, Theorem 3.6, p. 74], which implies that its inverse $\boldsymbol{\mu}^{-1} : \mathbb{M} \mapsto \boldsymbol{\Theta}$ exists and $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^{-1}$ are continuous, one-to-one, and onto. Consider any $\epsilon \in (0, \epsilon^\star)$ small enough so that $\boldsymbol{\mu}(\mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon)) \subseteq \operatorname{int}(\mathbb{M})$ and define

$$\delta(\epsilon) \;\; \coloneqq \inf_{\boldsymbol{\mu}' \,\in\, \operatorname{bd}(\boldsymbol{\mu}(\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon)))} \|\boldsymbol{\mu}' - \boldsymbol{\mu}(\boldsymbol{\theta}^\star)\|_\infty.$$

By the definition of $\delta(\epsilon)$, $\mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, \delta(\epsilon))$ is the largest hypercube centered at $\boldsymbol{\mu}(\boldsymbol{\theta}^\star)$ and contained in $\boldsymbol{\mu}(\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon))$, so that

$$\mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, \delta(\epsilon)) \;\; \subseteq \;\; \boldsymbol{\mu}(\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon)) \;\; \subseteq \;\; \operatorname{int}(\mathbb{M}). \tag{3.1}$$

In the event $s(\boldsymbol{X}) \in \boldsymbol{\mu}(\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon)) \subseteq \operatorname{int}(\mathbb{M})$, the random set $\widehat{\boldsymbol{\Theta}}$ is non-empty and its unique element $\widehat{\boldsymbol{\theta}}$ solves $\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}}) = s(\boldsymbol{X})$ [8, Theorem 5.5, p. 148]. The mapping $\boldsymbol{\mu} : \boldsymbol{\Theta} \mapsto \mathbb{M}$ is one-to-one, hence

$$\begin{aligned}
&\mathbb{P}\left(\widehat{\boldsymbol{\theta}} \,\in\, \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon)\right) \;\; = \;\; \mathbb{P}\left(\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}}) \,\in\, \boldsymbol{\mu}(\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon))\right) \\
\geq \;\; &\mathbb{P}\left(\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}}) \,\in\, \mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, \delta(\epsilon))\right) \;\; \geq \;\; 1 - 2\, \tau(\delta(\epsilon)),
\end{aligned} \tag{3.2}$$

where the first lower bound follows from (3.1) while the second lower bound follows from Lemma 1 in the supplement [46], with $\tau(t)$ defined by

$$\tau(t) \;\; \coloneqq \;\; \exp\left(-\frac{2\, t^2}{|\!|\!|\mathcal{D}_N(\boldsymbol{\theta}^\star)|\!|\!|_2^2\;\, \Psi_N^2} + \log p\right), \qquad t > 0. \tag{3.3}$$

To convert convergence rates of $\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}})$ into convergence rates of $\widehat{\boldsymbol{\theta}}$ based on (3.2) and (3.3), we bound $\delta(\epsilon)$ from below by using the continuity of $\boldsymbol{\mu} : \boldsymbol{\Theta} \mapsto \mathbb{M}$, which implies that there exists $\mathfrak{V}(\delta(\epsilon)) \in (0,\, \epsilon]$ such that

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

First, we show that $\delta(\epsilon)$ is bounded below by $\mathfrak{V}(\delta(\epsilon))\, \Lambda_N(\boldsymbol{\theta}^\star)$, by bounding the p-dimensional Lebesgue measure $\mathcal{L}$ of the hypercube $\mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, \delta(\epsilon))$:

$$\begin{aligned}
&(2\, \delta(\epsilon))^p \;=\; \mathcal{L}(\mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, \delta(\epsilon))) \;\geq\; \int\limits_{\mathbb{R}^p} \mathbb{1}(\boldsymbol{\mu}' \in \boldsymbol{\mu}(\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \mathfrak{V}(\delta(\epsilon)))))\; \mathrm{d}\,\mathcal{L}(\boldsymbol{\mu}') \\
&=\; \int\limits_{\mathbb{R}^p} \mathbb{1}(\boldsymbol{\theta}' \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \mathfrak{V}(\delta(\epsilon))))\; \det(\mathcal{I}(\boldsymbol{\theta}'))\; \mathrm{d}\,\mathcal{L}(\boldsymbol{\theta}') \;\geq\; (2\, \mathfrak{V}(\delta(\epsilon))\; \Lambda_N(\boldsymbol{\theta}^\star))^p.
\end{aligned}$$

Second, we show that a stronger result can be obtained: $\delta(\epsilon)$ is bounded below by a multiple of $\epsilon\,\Lambda_N(\boldsymbol{\theta}^\star)$, which provides the key to converting convergence rates of $\boldsymbol{\mu}(\widehat{\boldsymbol{\theta}})$ into convergence rates of $\widehat{\boldsymbol{\theta}}$ based on (3.2) and (3.3). The main insight is that all $\rho > 0$ for which $\boldsymbol{\mu}^{-1}(\mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, \rho)) \subseteq \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon)$ satisfy $\rho \leq \delta(\epsilon)$ by the definition of $\delta(\epsilon)$. Leveraging it, we show that $\delta(\epsilon)$ is bounded below by a multiple of $\epsilon\,\Lambda_N(\boldsymbol{\theta}^\star)$ by proving that

$$\boldsymbol{\mu}^{-1}\left(\mathcal{B}_\infty\left(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, \frac{\epsilon\,\Lambda_N(\boldsymbol{\theta}^\star)}{\mathcal{C}_N(\boldsymbol{\theta}^\star)}\right)\right) \;\subseteq\; \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon) \;\subset\; \boldsymbol{\Theta} \;=\; \mathbb{R}^p, \tag{3.4}$$

which implies that $\epsilon\,\Lambda_N(\boldsymbol{\theta}^\star) \,/\, \mathcal{C}_N(\boldsymbol{\theta}^\star) \leq \delta(\epsilon)$. It is worth noting that $\mathcal{C}_N(\boldsymbol{\theta}^\star)$

- is either bounded above by a finite constant, independent of N and p;
- or increases as a function of N, which is permissible but comes at a cost in terms of the convergence rate: see Corollary 1 in Section 3.4.

To establish (3.4) in both scenarios, let

- $\boldsymbol{\mu}' \in \mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, \epsilon\,\Lambda_N(\boldsymbol{\theta}^\star) \,/\, \mathcal{C}_N(\boldsymbol{\theta}^\star))$ be an arbitrary mean-value parameter vector;
- $\boldsymbol{\theta}' \in \boldsymbol{\mu}^{-1}(\mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, \epsilon\,\Lambda_N(\boldsymbol{\theta}^\star) \,/\, \mathcal{C}_N(\boldsymbol{\theta}^\star)))$ be the natural parameter vector corresponding to $\boldsymbol{\mu}' \in \mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, \epsilon\,\Lambda_N(\boldsymbol{\theta}^\star) \,/\, \mathcal{C}_N(\boldsymbol{\theta}^\star))$;

and note that $\boldsymbol{\mu}^{-1} : \mathbb{M} \mapsto \boldsymbol{\Theta}$ is one-to-one and $\boldsymbol{\mu}'$ is identical to $\boldsymbol{\mu}(\boldsymbol{\theta}')$. By the multivariate mean-value theorem [20, Theorem 5], there exists $\boldsymbol{\theta}'' \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$ such that $\boldsymbol{\mu}(\boldsymbol{\theta}') - \boldsymbol{\mu}(\boldsymbol{\theta}^\star) = \mathcal{I}(\boldsymbol{\theta}'')\,(\boldsymbol{\theta}' - \boldsymbol{\theta}^\star)$. Observe that $\mathcal{I}(\boldsymbol{\theta}'')$ is invertible provided N is large enough, because the assumption $\Phi_N(\boldsymbol{\theta}^\star) \to \infty$ as $N \to \infty$ implies that there exist finite constants $C_1 > 0$, $C_2 > 0$, and $N_1 \geq 2$ such that $\Phi_N(\boldsymbol{\theta}^\star) \geq C_1$ and $\Lambda_N(\boldsymbol{\theta}^\star) \geq C_2$ for all $N > N_1$. It follows that $\boldsymbol{\theta}' \in \boldsymbol{\mu}^{-1}(\mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, \epsilon\,\Lambda_N(\boldsymbol{\theta}^\star) \,/\, \mathcal{C}_N(\boldsymbol{\theta}^\star)))$ satisfies

$$\|\boldsymbol{\theta}' - \boldsymbol{\theta}^\star\|_\infty \;\leq\; |||(\mathcal{I}(\boldsymbol{\theta}''))^{-1}|||_\infty\, \|\boldsymbol{\mu}(\boldsymbol{\theta}') - \boldsymbol{\mu}(\boldsymbol{\theta}^\star)\|_\infty \;\leq\; |||(\mathcal{I}(\boldsymbol{\theta}''))^{-1}|||_\infty\, \frac{\epsilon\,\Lambda_N(\boldsymbol{\theta}^\star)}{\mathcal{C}_N(\boldsymbol{\theta}^\star)}.$$

We bound $\Lambda_N(\boldsymbol{\theta}^\star)$ by noting that $\sqrt[p]{\det(\mathcal{I}(\boldsymbol{\theta}))} \leq \lambda_{\max}(\boldsymbol{\theta}) \leq |||\mathcal{I}(\boldsymbol{\theta})|||_\infty$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\lambda_{\max}(\boldsymbol{\theta})$ is the largest eigenvalue of $\mathcal{I}(\boldsymbol{\theta})$. We obtain

$$\begin{aligned}
\|\boldsymbol{\theta}' - \boldsymbol{\theta}^\star\|_\infty \;&\leq\; \frac{\epsilon}{\mathcal{C}_N(\boldsymbol{\theta}^\star)}\, |||(\mathcal{I}(\boldsymbol{\theta}''))^{-1}|||_\infty \inf_{\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)} |||\mathcal{I}(\boldsymbol{\theta})|||_\infty \\
&\leq\; \frac{\epsilon}{\mathcal{C}_N(\boldsymbol{\theta}^\star)}\, |||(\mathcal{I}(\boldsymbol{\theta}''))^{-1}|||_\infty\, |||\mathcal{I}(\boldsymbol{\theta}'')|||_\infty \;\leq\; \epsilon,
\end{aligned}$$

where the last inequality follows from the definition of $\mathcal{C}_N(\boldsymbol{\theta}^\star)$. We conclude that (3.4) holds, which implies $\delta(\epsilon) \geq \epsilon\,\Lambda_N(\boldsymbol{\theta}^\star) \,/\, \mathcal{C}_N(\boldsymbol{\theta}^\star)$ and hence

$$\mathbb{P}\left(\widehat{\boldsymbol{\theta}} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon)\right) \;\geq\; 1 - 2\,\tau(\delta(\epsilon)) \;\geq\; 1 - 2\,\tau\left(\frac{\epsilon\,\Lambda_N(\boldsymbol{\theta}^\star)}{\mathcal{C}_N(\boldsymbol{\theta}^\star)}\right).$$

To make the probability of event $\widehat{\boldsymbol{\theta}} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon)$ at least $1 - 2\,/\max\{N,\, p\}^2$, set $\epsilon = \sqrt{3/2}\,/\,\Phi_N(\boldsymbol{\theta}^\star)$. The assumption $\Phi_N(\boldsymbol{\theta}^\star) \to \infty$ as $N \to \infty$ implies that there exists a finite constant $N_2 \geq 2$ such that $\epsilon = \sqrt{3/2}\,/\,\Phi_N(\boldsymbol{\theta}^\star) \in (0,\, \epsilon^\star)$ for all $N > N_2$. Thus, for all $N > \max\{N_1,\, N_2\} \geq 2$, the random set $\widehat{\boldsymbol{\Theta}}$ is non-empty and its unique element $\widehat{\boldsymbol{\theta}}$ satisfies

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star\|_\infty \;\; \leq \;\; \frac{\sqrt{3/2}}{\Phi_N(\boldsymbol{\theta}^\star)} \tag{3.5}$$

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

3.3. *Pseudo-likelihood-based M-estimators.* Maximum likelihood estimators are unappealing on computational grounds, because evaluating $\ell(\boldsymbol{\theta};\, \boldsymbol{x})$ requires 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 $M = \binom{N}{2}$ is small or the model makes restrictive independence assumptions. As a scalable alternative, consider M-estimators

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

based on the pseudo-loglikelihood function

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

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

To state the main result, let $i \in \{1, \ldots, M\}$ and $\mathfrak{N}_i \subseteq \{1, \ldots, M\} \setminus \{i\}$ be the smallest subset of $\{1, \ldots, M\} \setminus \{i\}$ such that

$$X_i \;\;\perp\!\!\!\perp\;\; \boldsymbol{X} \setminus (X_i,\, \boldsymbol{X}_{\mathfrak{N}_i}) \mid \boldsymbol{X}_{\mathfrak{N}_i}.$$

Let $\epsilon^\star \in (0,\, \infty)$ be a constant, independent of N and p, and define

$$\begin{aligned}
\widetilde{\Lambda}_N(\boldsymbol{\theta}^\star) &\coloneqq \inf_{\boldsymbol{\theta}\,\in\,\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\,\epsilon^\star)} \sqrt[p]{\det(-\mathbb{E}\, \nabla_{\boldsymbol{\theta}}^2\; \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}))} \\
\widetilde{\mathcal{C}}_N(\boldsymbol{\theta}^\star) &\coloneqq \sup_{\boldsymbol{\theta}\,\in\,\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\,\epsilon^\star)} \|-\mathbb{E}\, \nabla_{\boldsymbol{\theta}}^2\; \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})\|_\infty\, \|(-\mathbb{E}\, \nabla_{\boldsymbol{\theta}}^2\; \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}))^{-1}\|_\infty \\
\widetilde{\Phi}_N(\boldsymbol{\theta}^\star) &\coloneqq \frac{\widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)}{\widetilde{\mathcal{C}}_N(\boldsymbol{\theta}^\star)\, \max\left\{1,\, \max\limits_{1 \leq i \leq M} |\mathfrak{N}_i|\right\} \|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2\; \Psi_N\; \sqrt{\log \max\{N,\, p\}}},
\end{aligned}$$

where $\widetilde{\Lambda}_N(\boldsymbol{\theta}^\star) \geq 0$, $\widetilde{\mathcal{C}}_N(\boldsymbol{\theta}^\star) \geq 1$, and $|||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2 \geq 1$ by definition, while $\Psi_N > C_0 > 0$ by assumption, provided N is large enough.

We then obtain the following consistency results and convergence rates of pseudo-likelihood-based M-estimators. In contrast to the single-observation literature on Ising models and discrete Markov random fields [e.g., 29, 16, 11, 4, 22], which has focused on a fixed number of parameters p, we accommodate scenarios with $p \to \infty$ parameters.

Theorem 2. *Consider a single observation of a random graph with N nodes and $M = \binom{N}{2}$ dependent edges. Assume that $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta} = \mathbb{R}^p$, where $p \to \infty$ as $N \to \infty$ is allowed. If $\widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \to \infty$ as $N \to \infty$, there exists a finite constant $N_0 \geq 2$ 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*

$$\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star\|_\infty \;\;\leq\;\; \frac{3}{\widetilde{\Phi}_N(\boldsymbol{\theta}^\star)}$$

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

$$\gamma_N \;\;=\;\; \sqrt{3/2}\;\max\left\{1,\; \max_{1 \leq i \leq M} |\mathfrak{N}_i|\right\}\; |||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2\; \Psi_N\; \sqrt{\log \max\{N,\, p\}}.$$

A proof of Theorem 2 is provided in the supplement [46]. 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. It is worth noting that the maximum neighborhood size $\max_{1 \leq i \leq M} |\mathfrak{N}_i|$ in Theorem 2 can increase as a function of N, as demonstrated by Corollaries 2 and 3 in Section 3.4.

We present applications to generalized β-models with dependent edges and $p \geq N \to \infty$ parameters in Section 3.4, but first provide a simple application of Theorems 1 and 2 and explore how fast p can grow as a function of N. We then explain how $\widetilde{\Phi}_N(\boldsymbol{\theta}^\star)$ can be bounded.

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, where the edge probabilities $\mu_i = \mathbb{E}\, X_i$ satisfy $0 < C_1 < \mu_i < C_2 < 1$ 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 parameters $\theta_k = \operatorname{logit}(\pi_k)$ 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

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

Then there exists a constant $C_5 > 0$, independent of N and p, such that

$$\inf_{\boldsymbol{\theta} \,\in\, \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)} \sqrt[p]{\det(\mathcal{I}(\boldsymbol{\theta}))} \;\;\geq\;\; \inf_{\boldsymbol{\theta} \,\in\, \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)} \; \min_{1 \leq k \leq p} \mathbb{V}_{\boldsymbol{\theta}}\, s_k(\boldsymbol{X}) \;\;\geq\;\; \frac{C_5\, N^2}{p}.$$

As a result, $\Lambda_N(\boldsymbol{\theta}^\star) = \widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)$ is bounded below by $C_5\; N^2 / \, p$. In addition, $|||\mathcal{I}(\boldsymbol{\theta})|||_\infty$ is bounded above by a constant multiple of $N^2 \,/\, p$ while $|||(\mathcal{I}(\boldsymbol{\theta}))^{-1}|||_\infty$ is bounded above by a constant multiple of $p \,/\, N^2$, so that $\mathcal{C}_N(\boldsymbol{\theta}^\star) = \widetilde{\mathcal{C}}_N(\boldsymbol{\theta}^\star)$ is bounded above by a constant, independent of N and p. By the independence of edges, $\max_{1 \leq i \leq M} |\mathfrak{N}_i| = 0$ 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 $\Psi_N \coloneqq \max_{1 \leq k \leq p} \|\boldsymbol{\Xi}_k\|_2$ 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 $\|\boldsymbol{\Xi}_k\|_2 \leq \sqrt{C_4\, N^2 /\, p}$ for all k and hence $\Psi_N \leq \sqrt{C_4\, N^2 /\, p}$. Thus, there exists a constant $C > 0$, independent of N and p, such that

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

If $p = o(N^2 / \log N)$, then $\Phi_N(\boldsymbol{\theta}^\star) \to \infty$ 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 [38, 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 $\Lambda_N(\boldsymbol{\theta}^\star)$.

3.3.2. *Bounding the spectral norm of the coupling matrix.* If edges are dependent, the spectral norm $\|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2$ of the coupling matrix $\mathcal{D}_N(\boldsymbol{\theta}^\star)$ can be bounded by using the inequality

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

The elements $\mathcal{D}_{i,j}(\boldsymbol{\theta}^\star)$ of $\mathcal{D}_N(\boldsymbol{\theta}^\star)$ can be bounded by leveraging coupling techniques from percolation theory [48], applied to a conditional independence graph $\mathcal{G}$ with a set of vertices $\mathcal{V} = \{X_1, \ldots, X_M\}$ and a set of edges representing the conditional independence structure of a random graph. It is worth repeating that models with densities of the form (2.1) possess factorization properties and hence conditional independence properties [18], which can be represented by a conditional independence graph $\mathcal{G}$ [32].

For any vertex $X_i \in \mathcal{V}$ of the conditional independence graph $\mathcal{G}$ and any $\boldsymbol{x}_{1:i-1} \in \{0,1\}^{i-1}$, we construct a coupling $(\boldsymbol{X}^\star_{i+1:M}, \boldsymbol{X}^{\star\star}_{i+1:M})$ of the conditional probability mass functions $\mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{x}_{1:i-1}, 0}$ and $\mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{x}_{1:i-1}, 1}$ defined in Section 3.1. The coupling is described in the supplement [46] and helps translate the hard problem of bounding probabilities of events involving dependent edges into the more convenient problem of bounding probabilities of events involving independent edges. To demonstrate, consider any vertex $X_i \in \mathcal{V}$ and any vertex $X_j \in \{X_{i+1}, \ldots, X_M\}$ and let $i \nleftrightarrow j$ be the event that there exists a path of disagreement between vertices X_i and X_j in $\mathcal{G}$, in the sense that $X_v^\star \neq X_v^{\star\star}$ for all vertices X_v along the path. Theorem 1 of van den Berg and Maes [48] implies that

$$\mathbb{Q}_{\boldsymbol{\theta}^\star, i, \boldsymbol{x}_{1:i-1}}(X_j^\star \neq X_j^{\star\star}) = \mathbb{Q}_{\boldsymbol{\theta}^\star, i, \boldsymbol{x}_{1:i-1}}(i \nleftrightarrow j) \leq \mathbb{B}_{\boldsymbol{\pi}(\boldsymbol{\theta}^\star)}(i \nleftrightarrow j), \tag{3.6}$$

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

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

where

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

is the total variation distance between the conditional probability mass functions of vertex X_v given $\boldsymbol{X}_{-v} = \boldsymbol{x}_{-v}$ and $\boldsymbol{X}_{-v} = \boldsymbol{x}'_{-v}$. The Bernoulli product measure $\mathbb{B}_{\boldsymbol{\pi}(\boldsymbol{\theta}^\star)}$ assumes that independent Bernoulli experiments are carried out at vertices X_v of $\mathcal{G}$, with two possible outcomes: Either a vertex X_v is "open," in the sense that $X_v^\star \neq X_v^{\star\star}$ and hence a path of disagreement $i \not\leftrightarrow j$ can pass through X_v, or X_v is "closed." A vertex X_v is "open" with probability $\pi_v(\boldsymbol{\theta}^\star)$. By construction, vertices $X_v \in \{1, \ldots, i-1\}$ are "closed" and vertex X_i is "open" with probability one.

Using (3.6), we can bound the above-diagonal elements $\mathcal{D}_{i,j}(\boldsymbol{\theta}^\star)$ of $\mathcal{D}_N(\boldsymbol{\theta}^\star)$:

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

while the below-diagonal and diagonal elements of $\mathcal{D}_N(\boldsymbol{\theta}^\star)$ are 0 and 1, respectively. Specific bounds will depend on the model and its data-generating 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 [46].

3.3.3. *Bounding the p-th root determinant of the negative expected Hessian.* To establish convergence rates, we need to bound $\Lambda_N(\boldsymbol{\theta}^\star)$ and $\widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)$. Consider $\widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)$, the infimum of the p-th root determinant of $-\mathbb{E}\, \nabla^2_{\boldsymbol{\theta}}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$ on $\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$. In applications to generalized β-models, we bound $\widetilde{\Lambda}_N(\boldsymbol{\theta}^\star)$ by exploiting the properties of determinants along with the conditional independence properties of generalized β-models. We first bound the determinant of the $N \times N$ submatrix of $-\mathbb{E}\, \nabla^2_{\boldsymbol{\theta}}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$ corresponding to the degree parameters $\theta_1, \ldots, \theta_N$ by bounding its smallest eigenvalue. We then bound the determinant of $-\mathbb{E}\, \nabla^2_{\boldsymbol{\theta}}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$ by relating it to the determinant of its submatrix. Details can be found in the supplement [46].

3.3.4. *Bounding the smoothness of the sufficient statistics.* The quantity $\Psi_N := \max_{1 \leq i \leq p} \|\boldsymbol{\Xi}_i\|_2$ 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, which is – more often than not – a simple combinatorical exercise.

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

$$\|\boldsymbol{\theta}^\star\|_\infty \;\leq\; \begin{cases} \dfrac{L + (1-\vartheta)\,\log N}{8} & \text{under Model 1} \\[2ex] \dfrac{L + (1-\vartheta)\,\log N}{16\,(3 + D_N)} & \text{under Models 2 and 3,} \end{cases} \tag{3.7}$$

where $L \geq 0$ and $\vartheta \in (4/5,\, 1]$ are finite constants, independent of N and p, and $D_N \coloneqq \max_{1 \leq i \leq M} |\mathfrak{N}_i| \geq 0$ can increase as a function of N.

We start with the β-model [13], because its theoretical properties have been studied and it is therefore a convenient benchmark.

Corollary 1. *Consider Model 1, the β-model with independent edges, with $\boldsymbol{\theta}^\star \in \mathbb{R}^N$ satisfying (3.7). Then*

$$\gamma_N \;=\; \sqrt{3/2}\,\sqrt{N\,\log N}$$

and there exist finite constants $B \geq 1$, $C > 0$, and $N_0 \geq 2$, independent of N and p, such that, for all $N > N_0$, $\mathcal{C}_N(\boldsymbol{\theta}^\star) = \widetilde{\mathcal{C}}_N(\boldsymbol{\theta}^\star) \;\leq\; B\, N^{3\,(1-\vartheta)/2}$ and

$$\Phi_N(\boldsymbol{\theta}^\star) \;=\; \widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \;\geq\; C\,\sqrt{\frac{N^{5\,\vartheta-4}}{\log N}}, \qquad \vartheta \in (4/5,\, 1].$$

Corollary 1 reveals that the highest convergence rate is obtained when $\|\boldsymbol{\theta}^\star\|_\infty$ is bounded above (case $\vartheta = 1$). If $\|\boldsymbol{\theta}^\star\|_\infty$ increases as a function of N (case $\vartheta \in (4/5,\, 1)$), the ℓ_∞-induced condition number $\widetilde{\mathcal{C}}_N(\boldsymbol{\theta}^\star)$ increases, which reduces the convergence rate via ϑ. Condition (3.7) is the weakest known condition on $\|\boldsymbol{\theta}^\star\|_\infty$: Chatterjee et al. [13, Theorem 1.3] report a non-asymptotic error bound of the form $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star\|_\infty \leq C\,\sqrt{N\,/\,\log N}$ assuming that $\|\boldsymbol{\theta}^\star\|_\infty$ is bounded above (case $\vartheta = 1$), while Yan and Xu [52, Theorem 1] report asymptotic consistency and normality results assuming that $\|\boldsymbol{\theta}^\star\|_\infty = o(\log\log N)$. By contrast, condition (3.7) assumes that $\|\boldsymbol{\theta}^\star\|_\infty < (1/40)\,\log N$ ($\boldsymbol{\theta}^\star \in \mathbb{R}^N,\; L = 0$), 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. [51, 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 best 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.

To demonstrate that Theorem 2 covers random graph models with dependent edges, we turn to generalized β-models with dependent edges.

Corollary 2. *Consider Model 2, the generalized β-model with dependent edges, with $\boldsymbol{\theta}^\star \in \mathbb{R}^{N+1}$ satisfying (3.7). Then there exist finite constants $C > 0$ and $N_0 \geq 2$, independent of N and p, such that, for all $N > N_0$,*

$$\sqrt{3/2}\ \|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2 \sqrt{N \log N} \leq \gamma_N \leq 6\ D_N^2\ \|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2 \sqrt{N \log N} \tag{3.8}$$

and

$$\widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \quad \geq \quad \frac{C}{\widetilde{\mathcal{C}}_N(\boldsymbol{\theta}^\star)\ D_N^3\ \|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2} \sqrt{\frac{N^{2\,\vartheta-1}}{\log N}},$$

where edges are dependent when $\theta_{N+1} \neq 0$, $D_N \coloneqq \max_{1 \leq i \leq M} |\mathfrak{N}_i| \geq 1$ can increase as a function of N when $\theta_{N+1} \neq 0$, and $\min_{1 \leq k \leq K} |\mathcal{A}_k| \geq 3$.

Corollary 2 reveals that the convergence rate depends on the dependence among edges in terms of $\|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2$ and the maximum neighborhood size D_N, which can increase as a function of N. It is worth noting that condition (3.7) implies that D_N must satisfy $D_N = O(\log N)$ to prevent $\|\boldsymbol{\theta}^\star\|_\infty$ from approaching 0 (case $\vartheta \in (4/5,\, 1)$). If $\|\boldsymbol{\theta}^\star\|_\infty$ is allowed to approach 0, D_N can increase faster than $\log N$. Conditions under which $\|\boldsymbol{\theta}^\star\|_\infty$ is allowed to approach 0 are known as mean-field conditions [e.g., 11, 22] and are used to study Ising models and other models in physics.

To bound $\|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2$, we introduce a subpopulation graph $\mathcal{G}_{\mathcal{A}}$ with a set of vertices $\mathcal{V}_{\mathcal{A}} = \{\mathcal{A}_1, \ldots, \mathcal{A}_K\}$, where a pair of distinct subpopulations $\mathcal{A}_k$ and $\mathcal{A}_l$ is connected by an edge if and only 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 $d_{\mathcal{G}_{\mathcal{A}}}(\mathcal{A}_k,\, \mathcal{A}_k) = 0$ and $d_{\mathcal{G}_{\mathcal{A}}}(\mathcal{A}_k,\, \mathcal{A}_l) = \infty$ 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 $\xi \coloneqq 1/(1 + \exp(-L\,/\,16))$, where $L \geq 0$ is identical to the finite constant L in (3.7) and is independent of N and p. Let $D_N \coloneqq \max_{1 \leq i \leq M} |\mathfrak{N}_i| \geq 1$ and assume that there exist finite constants $\omega_1 \geq 0$ and $\omega_2 \in [0,\, 1/|\log(1 - \xi)|)$, independent of N and p, such that*

$$\max_{1 \leq k \leq K} |\mathcal{V}_{\mathcal{A}_k,l}| \quad \leq \quad \omega_1 + \omega_2\, \frac{\log l}{8\, D_N^2}, \qquad l \in \{1, \ldots, K-1\}.$$

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. *Consider Model 3, the generalized β-model with dependent edges and a known level of sparsity $\alpha \in (0, 1/2)$, with $\boldsymbol{\theta}^\star \in \mathbb{R}^{N+1}$ satisfying (3.7) and γ_N satisfying (3.8). Let $C > 0$, $C_1 > 0$, $C_2 > 0$, $C_3 > 0$, and $N_0 \geq 2$ be finite constants, independent of N and p. Then, for all $N > N_0$,*

$$\widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \;\; \geq \;\; \frac{C}{\widetilde{\mathcal{C}}_N(\boldsymbol{\theta}^\star)\; D_N^3 \; |||\mathcal{D}_N(\boldsymbol{\theta}^\star)|||_2} \sqrt{\frac{N^{2\,(\vartheta-\alpha)-1}}{\log N}},$$

where the random graph is sparse when $\alpha \in (0, 1/2)$, edges are dependent when $\theta_{N+1} \neq 0$, $D_N \coloneqq \max_{1 \leq i \leq M} |\mathfrak{N}_i| \geq 1$ can increase as a function of N when $\theta_{N+1} \neq 0$, and $\min_{1 \leq k \leq K} |\mathcal{A}_k| \geq 3$. If Assumption A holds, then:

- *Scenario 1: If the subpopulations do not intersect ($\omega_1 = \omega_2 = 0$) and $\boldsymbol{\theta}^\star \in \mathbb{R}^{N+1}$ satisfies (3.7) with $\vartheta \in (\max\{4/5,\, 1/2 + \alpha\},\, 1]$, then*

$$\widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \;\; \geq \;\; \frac{C_1}{\widetilde{\mathcal{C}}_N(\boldsymbol{\theta}^\star)\; D_N^5} \sqrt{\frac{N^{2\,(\vartheta-\alpha)-1}}{\log N}}.$$

- *Scenario 2: If the subpopulations do intersect and $\boldsymbol{\theta}^\star \in \mathbb{R}^{N+1}$ satisfies (3.7) with $\vartheta = 1$, then*

$$\widetilde{\Phi}_N(\boldsymbol{\theta}^\star) \;\; \geq \;\; \frac{C_2}{\widetilde{\mathcal{C}}_N(\boldsymbol{\theta}^\star)\; D_N^5\; \exp(C_3\, D_N^2)} \sqrt{\frac{N^{1-2\,\alpha}}{\log N}}.$$

Corollary 3 shows that sparsity reduces the convergence rate. The maximum neighborhood size D_N can increase as a function of N, regardless of whether the subpopulations do not overlap (first scenario) or do overlap (second scenario). That said, overlap comes at a cost in terms of the convergence rate: e.g., in the second scenario, the convergence rate is reduced by a factor of $\exp(C_3\, D_N^2)$, compared with the first scenario (case $\vartheta = 1$). It is worth noting that results on related statistical exponential-family models for dependent random variables in single-observation scenarios indicate that neighborhoods cannot be too large. For example, the results of Chatterjee and Diaconis [12] along with earlier results [23, 40, 3, 43] suggest that exponential-family random graphs with neighborhoods of sizes $O(N)$ may not be well-posed models, and statistical inference for ill-posed models is questionable. In addition, the recent results of Ghosal and Mukherjee [22] on two-parameter Ising models, which are likewise exponential-family models, suggest that consistency results may not be obtainable unless

(a) either mean-field conditions are satisfied, which have a long tradition in physics but are less common in other areas;

(b) or the graph that represents the conditional independence structure of the model satisfies bounded neighborhood assumptions [see condition (1.2) on p. 787 and Theorem 1.8 on p. 789 of 22].

A final remark is in order with regard to the ℓ_∞-induced condition number $\widetilde{\mathcal{C}}_N(\boldsymbol{\theta}^\star)$. In contrast to the ℓ_2-induced condition number of $-\mathbb{E}\, \nabla^2_{\boldsymbol{\theta}}\ \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$, which is the ratio of the largest and smallest eigenvalue of $-\mathbb{E}\, \nabla^2_{\boldsymbol{\theta}}\ \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$ and can be bounded by restricting eigenvalues, bounding the ℓ_∞-induced condition number $\widetilde{\mathcal{C}}_N(\boldsymbol{\theta}^\star)$ is less simple. In the special case where $-\mathbb{E}\, \nabla^2_{\boldsymbol{\theta}}\ \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$ is diagonally dominant in the sense of Hillar and Wibisono [25], the ℓ_∞-induced condition number $\widetilde{\mathcal{C}}_N(\boldsymbol{\theta}^\star)$ can be bounded, as demonstrated by Corollary 1. In practice, the ℓ_∞-induced condition number $\widetilde{\mathcal{C}}_N(\boldsymbol{\theta}^\star)$ can be approximated by numerical methods, provided $\boldsymbol{\theta}^\star$ is replaced by $\widetilde{\boldsymbol{\theta}}$.

Supplementary materials. Due to space restrictions, we present proofs of theoretical results along with simulation results in the supplement [46].

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

References.

[1] Airoldi, E., Blei, D., Fienberg, S., and Xing, E. (2008), "Mixed membership stochastic blockmodels," *Journal of Machine Learning Research*, 9, 1981–2014.

[2] 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.

[3] Bhamidi, S., Bresler, G., and Sly, A. (2011), "Mixing time of exponential random graphs," *The Annals of Applied Probability*, 21, 2146–2170.

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

[5] 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.

[6] Bickel, P. J., Chen, A., and Levina, E. (2011), "The method of moments and degree distributions for network models," *The Annals of Statistics*, 39, 2280–2301.

[7] 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.

[8] Brown, L. (1986), *Fundamentals of Statistical Exponential Families*, Hayworth, CA, USA: Institute of Mathematical Statistics.

[9] Butts, C. T. (2011), "Bernoulli graph bounds for general random graph models," *Sociological Methodology*, 41, 299–345.

[10] 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.

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

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

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

[14] 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.

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

[16] 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.

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

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

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

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

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

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

[23] 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.

[24] Hillar, C. J., Lin, S., and Wibisono, A. (2012), "Inverses of symmetric, diagonally dominant positive matrices and applications," *arXiv preprint arXiv:1203.6812*.

[25] 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.

[26] 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.

[27] 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.

[28] Hunter, D. R., Krivitsky, P. N., and Schweinberger, M. (2012), "Computational statistical methods for social network models," *Journal of Computational and Graphical Statistics*, 21, 856–882.

[29] Jensen, J. L., and Møller, J. (1991), "Pseudolikelihood for exponential family models of spatial point processes," *The Annals of Applied Probability*, 1, 445–461.

[30] 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.

[31] Krivitsky, P. N., and Kolaczyk, E. D. (2015), "On the question of effective sample size in network modeling: An asymptotic inquiry," *Statistical Science*, 30, 184–198.

[32] Lauritzen, S., Rinaldo, A., and Sadeghi, K. (2018), "Random networks, graphical

models and exchangeability," *Journal of the Royal Statistical Society: Series B*, 80, 481–508.

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

[34] Lei, J., and Rinaldo, A. (2015), "Consistency of spectral clustering in stochastic block models," *The Annals of Statistics*, 43, 215–237.

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

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

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

[38] 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.

[39] 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.

[40] 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.

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

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

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

[44] 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.

[45] 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.

[46] Stewart, J. R., and Schweinberger, M. (2022), "Pseudo-likelihood-based M-estimators for random graphs with dependent edges and parameter vectors of increasing dimension," Tech. rep., Florida State University, arXiv:2012.07167.

[47] Tang, M., Sussman, D. L., and Priebe, C. E. (2013), "Universally consistent vertex classification for latent positions graphs," *The Annals of Statistics*, 41, 1406–1430.

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

[49] Wainwright, M. J., and Jordan, M. I. (2008), "Graphical models, exponential families, and variational inference," *Foundations and Trends in Machine Learning*, 1, 1–305.

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

[51] 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.

[52] 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: 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, and C, it is convenient to adopt the notation used in Section 3 of the manuscript, denoting the number of edge variables by $M = \binom{N}{2}$ and edge variables by $X_1, \ldots, X_M$, where N is the number of nodes. In addition, we denote the data-generating parameter vector by $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta} = \mathbb{R}^p$, the data-generating probability measure and expectation by $\mathbb{P} \equiv \mathbb{P}_{\boldsymbol{\theta}^\star}$ and $\mathbb{E} \equiv \mathbb{E}_{\boldsymbol{\theta}^\star}$, and the probability density function of $\mathbb{P} \equiv \mathbb{P}_{\boldsymbol{\theta}^\star}$ with respect to a σ-finite measure ν by $f \equiv f_{\boldsymbol{\theta}^\star}$. Throughout, we assume that $\min_{1 \leq k \leq K} |\mathcal{A}_k| \geq 3$. To simplify the proofs, we suppress the argument $\boldsymbol{\theta}^\star$ of all other quantities that depend on the data-generating parameter vector $\boldsymbol{\theta}^\star$, and suppress the subscript N of all quantities that depend on the number of nodes N.

APPENDIX A: ADDITIONAL RESULTS

We present an additional version of generalized β-models with size-dependent parameterizations, called Model 4.

Model 2, introduced in Section 2.3 of the manuscript, assumes that the weight of the brokerage statistic $b_{i,j}$ does not depend on the sizes of neighborhood intersections. While convenient on mathematical grounds, the assumption that small and large neighborhood intersections have the same brokerage parameter may be unwarranted. To allow small and large neighborhood intersections to have distinct brokerage parameters, we consider functions of edges $\varphi_{i,j}$ of the form

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

where $a_{i,j}$ is 1 if $x_{i,j} \in \{0, 1\}$ and is 0 otherwise, $b_{i,j}$ is defined by (2.5), and

$$\eta_{i,j}(\theta_{N+1}) \;\;=\;\; \theta_{N+1} \log\left(1 + \frac{\log |\mathcal{N}_i \cap \mathcal{N}_j|}{|\mathcal{N}_i \cap \mathcal{N}_j|}\right), \qquad \theta_{N+1} \in \mathbb{R}.$$

Corollary 4. *Consider Model 4, the generalized β-model with dependent edges and a size-dependent parameterization, with $\boldsymbol{\theta}^\star \in \mathbb{R}^{N+1}$ and γ satisfying the same conditions as in Corollary 3. Let $C > 0$, $C_1 > 0$, $C_2 > 0$, $C_3 > 0$, and $N_0 \geq 2$ be finite constants, independent of N and p. Then, for all $N > N_0$,*

$$\widetilde{\Phi} \;\;\geq\;\; \frac{C}{\widetilde{\mathcal{C}}\; D^3\; |||\mathcal{D}|||_2} \sqrt{\frac{N^{2\,\vartheta-1}}{\log N}},$$

where edges are dependent provided $\theta_{N+1} \neq 0$, $D \coloneqq \max_{1 \leq i \leq M} |\mathfrak{N}_i| \geq 1$ can increase as a function of N provided $\theta_{N+1} \neq 0$, and $\min_{1 \leq k \leq K} |\mathcal{A}_k| \geq 3$. If Assumption A is satisfied, then:

- *Scenario 1: If the subpopulations do not intersect ($\omega_1 = \omega_2 = 0$) and $\boldsymbol{\theta}^\star \in \mathbb{R}^{N+1}$ satisfies (3.7) with $\vartheta \in (4/5,\, 1]$, then*

$$\widetilde{\Phi} \;\;\geq\;\; \frac{C_1}{\widetilde{\mathcal{C}}\; D^5} \sqrt{\frac{N^{2\,\vartheta-1}}{\log N}}.$$

- *Scenario 2: If the subpopulations do intersect and $\boldsymbol{\theta}^\star \in \mathbb{R}^{N+1}$ satisfies (3.7) with $\vartheta = 1$, then*

$$\widetilde{\Phi} \;\;\geq\;\; \frac{C_2}{\widetilde{\mathcal{C}}\; D^5\; \exp(C_3\, D^2)} \sqrt{\frac{N}{\log N}}.$$

Corollary 4 demonstrates that theoretical guarantees can be obtained for models with size-dependent parameterizations, which allow parameters to depend on the sizes of neighborhood intersections.

APPENDIX B: PROOF OF 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 $t > 0$,*

$$\mathbb{P}\left(s(\boldsymbol{X}) \;\in\; \mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, t)\right) \quad \geq \quad 1 - 2\,\tau(t),$$

with $\tau(t)$ defined by

$$\tau(t) \quad \coloneqq \quad \exp\left(-\frac{2\,t^2}{|||\mathcal{D}|||_2^2\;\Psi^2} + \log p\right),$$

where $|||\mathcal{D}|||_2 \geq 1$, $p \geq 1$, and $\Psi > C_0 > 0$, provided N is large enough.

Proof of Lemma 1. By Theorem 1 of Chazottes et al. [14, p. 207], for all $t > 0$,

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

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

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

where $\Psi = \max_{1 \leq i \leq p}\, \|\boldsymbol{\Xi}_i\|_2$. As a result,

$$\mathbb{P}\left(s(\boldsymbol{X}) \;\in\; \mathcal{B}_\infty(\boldsymbol{\mu}(\boldsymbol{\theta}^\star),\, t)\right) \quad \geq \quad 1 - 2\,\tau(t),$$

where we used the fact that $\boldsymbol{\mu}(\boldsymbol{\theta}^\star) \coloneqq \mathbb{E}_{\boldsymbol{\theta}^\star}\, s(\boldsymbol{X})$.

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. [14] 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. [14] 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 [8].

APPENDIX C: PROOF OF THEOREM 2

We prove Theorem 2 stated in Section 3.3 of the manuscript. Auxiliary results are proved in Appendix C.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 2, 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 3, its gradient

$$\boldsymbol{g}(\boldsymbol{\theta}) \;=\; \nabla_{\boldsymbol{\theta}}\, \mathbb{E}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}) \;=\; \mathbb{E}\, \nabla_{\boldsymbol{\theta}}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$$

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 is permitted by Lebesgue's dominated convergence theorem [33, Theorem 2.7.1, p. 49].

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

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

We divide the remainder of the proof into three parts:

I. Existence.

II. Convergence rate.

III. Uniform convergence of $\boldsymbol{g}(\,\cdot\,;\, \boldsymbol{X})$ on $\boldsymbol{\Theta}$.

I. Existence. Consider any $\gamma \in (0,\, \delta(\epsilon)/2]$ and define

$$\mathbb{G}(\gamma) \;\coloneqq\; \left\{\boldsymbol{x} \in \mathbb{X} :\; \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty \;\leq\; \gamma\right\} \;\subseteq\; \mathbb{X}.$$

In the event $\boldsymbol{x} \in \mathbb{G}(\gamma)$,

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

because $\boldsymbol{g}(\boldsymbol{\theta}^\star) = \mathbf{0}$ by Lemma 2. As a consequence, the set

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

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

II. Convergence rate. Suppose that the event $\boldsymbol{x} \in \mathbb{G}(\gamma)$ occurs, so that the set $\widetilde{\boldsymbol{\Theta}}(\gamma)$ is non-empty. Then, for any element $\widetilde{\boldsymbol{\theta}}$ of $\widetilde{\boldsymbol{\Theta}}(\gamma)$,

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

because $\boldsymbol{g}(\boldsymbol{\theta}^\star) = \mathbf{0}$ by Lemma 2 and $\|\boldsymbol{g}(\widetilde{\boldsymbol{\theta}};\, \boldsymbol{x})\|_\infty \leq \gamma$ for any element $\widetilde{\boldsymbol{\theta}}$ of $\widetilde{\boldsymbol{\Theta}}(\gamma)$ by construction of the set $\widetilde{\boldsymbol{\Theta}}(\gamma)$. In addition, by construction of the set $\mathbb{G}(\gamma)$ and the choice $\gamma \in (0,\, \delta(\epsilon)\,/\,2]$, we obtain

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

In other words, any element $\widetilde{\boldsymbol{\theta}}$ of $\widetilde{\boldsymbol{\Theta}}(\gamma)$ satisfies

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

which implies $\widetilde{\boldsymbol{\theta}} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon)$ and hence

$$\widetilde{\boldsymbol{\Theta}}(\gamma) \;\subseteq\; \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon) \;\subset\; \boldsymbol{\Theta} \;=\; \mathbb{R}^p.$$

To establish convergence rates, we take advantage of the continuity argument of Theorem 1, applied to $\boldsymbol{g}(\boldsymbol{\theta})$ rather than $\boldsymbol{\mu}(\boldsymbol{\theta})$, which implies that

$$\delta(\epsilon) \;\;\geq\;\; \frac{\epsilon\, \widetilde{\Lambda}}{\widetilde{\mathcal{C}}}.$$

We establish convergence rates by choosing ϵ so that the event $\widetilde{\boldsymbol{\Theta}}(\gamma) \subseteq \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon)$ occurs with high probability. To do so, we need to establish uniform convergence of $\boldsymbol{g}(\,\cdot\,;\, \boldsymbol{X})$ on $\boldsymbol{\Theta}$.

III. Uniform convergence of $\boldsymbol{g}(\,\cdot\,;\, \boldsymbol{X})$ on $\boldsymbol{\Theta}$. We have seen that, for any $\boldsymbol{x} \in \mathbb{X}$ such that

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty + \gamma \;\;\leq\;\; \delta(\epsilon),$$

the set $\widetilde{\mathbf{\Theta}}(\gamma)$ is non-empty and satisfies

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

Thus, we have

$$\mathbb{P}\left(\widetilde{\mathbf{\Theta}}(\gamma) \;\subseteq\; \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon)\right) \;\geq\; \mathbb{P}\left(\sup_{\boldsymbol{\theta}\,\in\,\mathbf{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty + \gamma \;\leq\; \delta(\epsilon)\right).$$

The fact that $\delta(\epsilon) \geq \epsilon\, \widetilde{\Lambda} \,/\, \widetilde{\mathcal{C}}$ implies that

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\boldsymbol{\theta}\,\in\,\mathbf{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty + \gamma \;\leq\; \delta(\epsilon)\right) \\
\geq\;\; &\mathbb{P}\left(\sup_{\boldsymbol{\theta}\,\in\,\mathbf{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty + \gamma \;\leq\; \frac{\epsilon\, \widetilde{\Lambda}}{\widetilde{\mathcal{C}}}\right).
\end{aligned}$$

We bound the probability of the complement of the event on the right-hand side by bounding

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\boldsymbol{\theta}\,\in\,\mathbf{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty + \gamma \;>\; \frac{\epsilon\, \widetilde{\Lambda}}{\widetilde{\mathcal{C}}}\right) \\
\leq\; &\mathbb{P}\left(\sup_{\boldsymbol{\theta}\,\in\,\mathbf{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty \;>\; \frac{\epsilon\, \widetilde{\Lambda}}{2\, \widetilde{\mathcal{C}}}\right) + \mathbb{P}\left(\gamma > \frac{\epsilon\, \widetilde{\Lambda}}{2\, \widetilde{\mathcal{C}}}\right).
\end{aligned}$$

The first term on the right-hand side can be bounded by using Lemma 4, which shows that

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\boldsymbol{\theta}\,\in\,\mathbf{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty \;>\; \frac{\epsilon\, \widetilde{\Lambda}}{2\, \widetilde{\mathcal{C}}}\right) \\
\leq\;\; &2\, \exp\left(-\frac{\epsilon^2\, \widetilde{\Lambda}^2}{2\;\widetilde{\mathcal{C}}^2\, \max\{1,\; \max_{1 \leq i \leq M} |\mathfrak{N}_i|\}^2\; |||\mathcal{D}|||_2^2\; \Psi^2} + \log\, p\right).
\end{aligned}$$

To ensure that the event $\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta}; \boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty > \epsilon\, \widetilde{\Lambda} \,/\, (2\;\, \widetilde{\mathcal{C}})$ occurs with low probability, choose

$$\epsilon \;\;=\;\; \sqrt{6}\;\; \frac{\widetilde{\mathcal{C}}\;\max\{1,\; \max_{1 \leq i \leq M} |\mathfrak{N}_i|\}\; |||\mathcal{D}|||_2\; \Psi\; \sqrt{\log \max\{N,\, p\}}}{\widetilde{\Lambda}}.$$

The choice of ϵ implies that

$$\mathbb{P}\left(\sup_{\boldsymbol{\theta}\,\in\,\boldsymbol{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty \;>\; \frac{\epsilon\, \widetilde{\Lambda}}{2\, \widetilde{\mathcal{C}}}\right) \;\leq\; \frac{2}{\max\{N,\, p\}^2}.$$

The second term, $\mathbb{P}(\gamma > \epsilon\, \widetilde{\Lambda} \,/\, (2\, \widetilde{\mathcal{C}}))$, can be bounded above as follows. The choice of ϵ implies that

$$\frac{\epsilon\, \widetilde{\Lambda}}{\widetilde{\mathcal{C}}} \;=\; \sqrt{6}\;\max\left\{1,\, \max_{1\leq i\leq M} |\mathfrak{N}_i|\right\}\; |||\mathcal{D}|||_2\; \Psi\; \sqrt{\log \max\{N,\, p\}}.$$

As a consequence, choosing

$$\begin{aligned}\gamma \;=\; \frac{\epsilon\, \widetilde{\Lambda}}{2\, \widetilde{\mathcal{C}}} \;&=\; \sqrt{3/2}\;\max\left\{1,\, \max_{1\leq i\leq M} |\mathfrak{N}_i|\right\}\; |||\mathcal{D}|||_2\; \Psi\; \sqrt{\log \max\{N,\, p\}} \\ &\leq\; \frac{\delta(\epsilon)}{2}\end{aligned}$$

ensures that

$$\mathbb{P}\left(\gamma > \frac{\epsilon\, \widetilde{\Lambda}}{2\, \widetilde{\mathcal{C}}}\right) \;=\; 0$$

and, at the same time, ensures that the random set $\widetilde{\boldsymbol{\Theta}}(\gamma)$ is non-empty with at least probability $1 - 2 \,/\, \max\{N,\, p\}^2$.

Upon collecting terms, we obtain

$$\mathbb{P}\left(\widetilde{\boldsymbol{\Theta}}(\gamma) \;\subseteq\; \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon)\right) \;\geq\; 1 - \frac{2}{\max\{N,\, p\}^2}. \tag{C.1}$$

Last, but not least, recall that $\widetilde{\Phi}$ is defined by

$$\widetilde{\Phi} \;\coloneqq\; \frac{\widetilde{\Lambda}}{\widetilde{\mathcal{C}}\;\max\{1,\, \max_{1\leq i\leq M} |\mathfrak{N}_i|\}\; |||\mathcal{D}|||_2\; \Psi\; \sqrt{\log \max\{N,\, p\}}}.$$

By the choice of ϵ and the definition of $\widetilde{\Phi}$, we have $\epsilon < 3 \,/\, \widetilde{\Phi}$. The assumption $\widetilde{\Phi} \to \infty$ as $N \to \infty$ implies that there exists a constant $N_0 \geq 2$ such that $\epsilon < 3 \,/\, \widetilde{\Phi} \,\in\, (0,\, \epsilon^\star)$ for all $N > N_0$. Thus, for all $N > N_0$, the random set $\widetilde{\boldsymbol{\Theta}}(\gamma)$ is non-empty and any element $\widetilde{\boldsymbol{\theta}}$ of $\widetilde{\boldsymbol{\Theta}}(\gamma)$ satisfies

$$\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star\|_\infty \;\leq\; \frac{3}{\widetilde{\Phi}}$$

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

Remark. Extensions to dependent random variables with countable and uncountable sample spaces. Neither Theorem 1 nor Theorem 2 are restricted to random graphs with dependent edges. Both theorems can be extended to dependent random variables with countable and uncountable sample spaces: see the remark in Appendix B.

C.1. Auxiliary results.

Lemma 2. *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 2. 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 [8] to prove Lemma 2, 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 $i \in \{1, \dots, M\}$, 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 $i^\star \in \{1, \dots, M\}$ 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 $i^\star \in \{1, \dots, M\}$ such that $\mathbb{E}\, \psi_{i^\star}(\boldsymbol{\theta};\, \boldsymbol{X}_{-i^\star})$ is strictly convex on $\boldsymbol{\Theta}$. Then, for all $i \in \{1, \dots, M\}$, 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

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

In other words, for all $\boldsymbol{x} \in \mathbb{X}$,

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

The conclusion (C.2) contradicts the assumption that the exponential family is minimal. Therefore, there exists $i^\star \in \{1, \dots, M\}$ 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

$$\nabla_{\boldsymbol{\theta}}\, \mathbb{E}\, \ell(\boldsymbol{\theta};\, \boldsymbol{X}) \;=\; \mathbb{E}\, s(\boldsymbol{X}) - \mathbb{E}_{\boldsymbol{\theta}}\, s(\boldsymbol{X}) \;=\; \boldsymbol{0}. \tag{C.3}$$

The unique solution of (C.3) is $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta} = \mathbb{R}^p$, because $\mathbb{E}\, s(\boldsymbol{X}) \equiv \mathbb{E}_{\boldsymbol{\theta}^\star}\, s(\boldsymbol{X})$. The fact that the solution is unique follows from the fact the map $\boldsymbol{\mu} : \boldsymbol{\Theta} \mapsto \mathbb{M}$ defined by $\boldsymbol{\mu}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}\, s(\boldsymbol{X})$ is one-to-one [8, 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 $\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i}$

are Bernoulli distributions ($i = 1, \ldots, M$), and Bernoulli distributions are exponential-family distributions. As a result, $\widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{x})$ is continuously differentiable on $\boldsymbol{\Theta}$ for all $\boldsymbol{x} \in \mathbb{X}$ [8], and so is $\mathbb{E}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$. We have

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

where $\mathbb{E}_{\boldsymbol{\theta}, \boldsymbol{x}_{-i}}$ denotes the conditional expectation, computed 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 that

$$\boldsymbol{g}(\boldsymbol{\theta}^\star) \;\; = \;\; \mathbb{E} \sum_{i=1}^{M} \left(s(\boldsymbol{X}) - \mathbb{E}_{\boldsymbol{\theta}^\star, \boldsymbol{X}_{-i}}\, s(\boldsymbol{X})\right) \;\; = \;\; \mathbf{0}. \tag{C.4}$$

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 3. *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 3. 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 $\boldsymbol{\Theta}$,

$$\langle \nabla_{\boldsymbol{\theta}}\, \boldsymbol{g}(\boldsymbol{\theta})\,|_{\boldsymbol{\theta}=\boldsymbol{\theta}_1},\; \boldsymbol{\theta}_2 - \boldsymbol{\theta}_1 \rangle \;\; > \;\; \boldsymbol{g}(\boldsymbol{\theta}_2) - \boldsymbol{g}(\boldsymbol{\theta}_1) \tag{C.5}$$

and

$$\langle \nabla_{\boldsymbol{\theta}}\, \boldsymbol{g}(\boldsymbol{\theta})\,|_{\boldsymbol{\theta}=\boldsymbol{\theta}_2},\; \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2 \rangle \;\; > \;\; \boldsymbol{g}(\boldsymbol{\theta}_1) - \boldsymbol{g}(\boldsymbol{\theta}_2). \tag{C.6}$$

By multiplying both sides of (C.6) by -1, we obtain

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

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

$$\langle \nabla_{\boldsymbol{\theta}}\, \boldsymbol{g}(\boldsymbol{\theta})\,|_{\boldsymbol{\theta}=\boldsymbol{\theta}_1},\; \boldsymbol{\theta}_2 - \boldsymbol{\theta}_1 \rangle \;\; < \;\; \boldsymbol{g}(\boldsymbol{\theta}_2) - \boldsymbol{g}(\boldsymbol{\theta}_1). \tag{C.7}$$

The conclusion (C.7) contradicts (C.5), so $\nabla_{\boldsymbol{\theta}}\, \boldsymbol{g}(\boldsymbol{\theta})$ is one-to-one on $\boldsymbol{\Theta}$.

Lemma 4. *For all $t > 0$,*

$$\mathbb{P}\left(\sup_{\boldsymbol{\theta}\,\in\,\boldsymbol{\Theta}} \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty \,>\, t\right)$$

$$\leq\quad 2\,\exp\left(-\frac{2\,t^2}{\max\{1,\ \max_{1\leq i\leq M} |\mathfrak{N}_i|\}^2\ \left|\left|\left|\mathcal{D}\right|\right|\right|_2^2\ \Psi^2} + \log\, p\right),$$

where $|||\mathcal{D}|||_2 \geq 1,\ p \geq 1,$ and $\Psi > C_0 > 0,$ provided N is large enough.

Proof of Lemma 4. First, consider any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and recall that

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

where $\mathbb{E}_{\boldsymbol{\theta},\boldsymbol{x}_{-i}}$ denotes the conditional expectation, computed with respect to the conditional distribution of X_i given $\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i}$. Consider any $i \in \{1,\dots,M\}$ and any $(\boldsymbol{x},\boldsymbol{x}') \in \mathbb{X}\times\mathbb{X}$ such that $x_j = x_j'$ for all $j \neq i$. Then, by the triangle inequality, we obtain, for each $k \in \{1,\dots,p\}$,

$$\begin{aligned}
|g_k(\boldsymbol{\theta};\, \boldsymbol{x}) - g_k(\boldsymbol{\theta};\, \boldsymbol{x}')| \quad &= \quad \left|\sum_{j=1}^{M} \left(\mathbb{E}_{\boldsymbol{\theta},\boldsymbol{x}_{-j}}\, s_k(\boldsymbol{X}) - \mathbb{E}_{\boldsymbol{\theta},\boldsymbol{x}'_{-j}}\, s_k(\boldsymbol{X})\right)\right| \\
&\leq \quad \sum_{j=1}^{M} \left|\mathbb{E}_{\boldsymbol{\theta},\boldsymbol{x}_{-j}}\, s_k(\boldsymbol{X}) - \mathbb{E}_{\boldsymbol{\theta},\boldsymbol{x}'_{-j}}\, s_k(\boldsymbol{X})\right| \\
&\leq \quad \max\left\{1,\ \max_{1\leq i\leq M} |\mathfrak{N}_i|\right\}\ \Xi_{i,k}.
\end{aligned}$$

The last inequality follows from the fact that, given any $i \in \{1,\dots,M\}$, $\mathfrak{N}_i \subseteq \{1,\dots,M\}\setminus\{i\}$ is defined as the smallest subset of indices such that

$$X_i \quad \perp\!\!\!\perp \quad \boldsymbol{X} \setminus (X_i,\, \boldsymbol{X}_{\mathfrak{N}_i}) \mid \boldsymbol{X}_{\mathfrak{N}_i},$$

which implies that a change of a single edge variable can affect the conditional distributions of at most $\max_{1\leq i\leq M} |\mathfrak{N}_i|$ other edge variables. As a result,

$$\sup_{(\boldsymbol{x},\boldsymbol{x}')\in\mathbb{X}\times\mathbb{X}:\, x_j = x_j' \text{ for all } j\neq i} |g_k(\boldsymbol{\theta};\boldsymbol{x}) - g_k(\boldsymbol{\theta};\boldsymbol{x}')| \leq \max\left\{1,\ \max_{1\leq i\leq M} |\mathfrak{N}_i|\right\}\ \Xi_{i,k}.$$

By applying Theorem 1 of Chazottes et al. [14] to each coordinate $g_k(\boldsymbol{\theta};\, \boldsymbol{X}) - g_k(\boldsymbol{\theta})$ of $\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})$ $(k = 1, \dots, p)$, we have, for all $t > 0$,

$$\begin{aligned} & \mathbb{P}\left(|g_k(\boldsymbol{\theta};\, \boldsymbol{X}) - g_k(\boldsymbol{\theta})| > t\right) \\ \leq\; & 2\, \exp\left(-\frac{2\, t^2}{\max\{1,\, \max_{1 \leq i \leq M} |\mathfrak{N}_i|\}^2\;\; |||\mathcal{D}|||_2^2\;\; \Psi^2}\right), \end{aligned}$$

provided $|||\mathcal{D}|||_2 > 0$ and $\Psi > 0$. A union bound shows that

$$\begin{aligned} & \mathbb{P}\left(\|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty \,>\, t\right) \\ \leq\; & 2\, \exp\left(-\frac{2\, t^2}{\max\{1,\, \max_{1 \leq i \leq M} |\mathfrak{N}_i|\}^2\;\; |||\mathcal{D}|||_2^2\;\; \Psi^2} + \log\, p\right) \end{aligned}$$

and

$$\begin{aligned} & \mathbb{P}\left(\sup_{\boldsymbol{\theta} \,\in\, \boldsymbol{\Theta}}\, \|\boldsymbol{g}(\boldsymbol{\theta};\, \boldsymbol{X}) - \boldsymbol{g}(\boldsymbol{\theta})\|_\infty \,>\, t\right) \\ \leq\; & 2\, \exp\left(-\frac{2\, t^2}{\max\{1,\, \max_{1 \leq i \leq M} |\mathfrak{N}_i|\}^2\;\; |||\mathcal{D}|||_2^2\;\; \Psi^2} + \log\, p\right). \end{aligned}$$

APPENDIX D: PROOFS OF COROLLARIES 1–4

We prove Corollaries 1–3 stated in Section 3.4 of the manuscript and Corollary 4 stated in Appendix A, using the auxiliary results proved in Appendices D.1 and D.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 1–4. To prove Corollaries 1–4, we bound

$$\widetilde{\Phi} \;\;:=\;\; \frac{\widetilde{\Lambda}}{\widetilde{\mathcal{C}}\;\, \max\left\{1,\; \max_{1 \leq i < j \leq N} |\mathfrak{N}_{i,j}|\right\}\;\, |||\mathcal{D}|||_2\;\, \Psi\; \sqrt{\log \max\{N,\, p\}}}.$$

Throughout, we assume that $\epsilon^\star$ is a constant, independent of N and p, and satisfies $\epsilon^\star \in (0,\, \|\boldsymbol{\theta}^\star\|_\infty]$, while $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta} = \mathbb{R}^p$ satisfies

$$\|\boldsymbol{\theta}^\star\|_\infty \;\;\leq\;\; \begin{cases} \dfrac{L + (1 - \vartheta)\, \log N}{8} & \text{under Model 1} \\[2ex] \dfrac{L + (1 - \vartheta)\, \log N}{16\, (3 + D)} & \text{under Models 2–4,} \end{cases}$$

where $D \coloneqq \max_{1 \leq i < j \leq N} |\mathfrak{N}_{i,j}| \geq 1$. To prepare the ground for the proofs of Corollaries 1–4, we first bound Ψ.

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

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

and

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

We show that $\Psi \leq \sqrt{N}$ under Model 1 and $\Psi \leq \|s_{N+1}\|_{\mathrm{Lip}} \sqrt{N}$ under Models 2–4 and bound $\|s_{N+1}\|_{\mathrm{Lip}}$, where $\|s_{N+1}\|_{\mathrm{Lip}}$ is the Lipschitz coefficient of $s_{N+1}(\boldsymbol{X})$ with respect to the Hamming metric on $\mathbb{X} \times \mathbb{X}$:

- Models 1–4 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 $\|\boldsymbol{\Xi}_a\|_2 = \sqrt{N-1} \leq \sqrt{N}$ $(a = 1, \dots, N)$.
- Models 2–4 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

  $$I_{i,j}(\boldsymbol{x}) \;=\; \mathbb{1}\left(\sum_{h \,\in\, \mathcal{N}_i \,\cap\, \mathcal{N}_j} x_{i,h}\, x_{j,h} \;>\; 0\right), \qquad \{i, j\} \subset \mathcal{N}.$$

  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 ND: For each of the N nodes $i \in \mathcal{N}$, there are at most D distinct nodes $j \in \mathcal{N}_i$ such that $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$, because Proposition 2 reveals

  $$|\mathcal{N}_i \cup \mathcal{N}_j| \;\leq\; 2\, |\mathcal{N}_i \cup \mathcal{N}_j| \;\leq\; \max_{\{i,j\} \subset \mathcal{N}} |\mathfrak{N}_{i,j}| \;=\; D.$$

  In addition, Lemma 13 shows that $\Xi_{a,\{i,j\}} \leq 1 + D$. Taken together,

  $$\|\boldsymbol{\Xi}_{N+1}\|_2 \;\leq\; \sqrt{N D (1 + D)^2} \;\leq\; \sqrt{4 N D^3} \;\leq\; 3\, D^2 \sqrt{N}.$$

As a result, Models 1–4 have

$$\Psi \;=\; \max_{1 \leq a \leq p} \|\boldsymbol{\Xi}_a\|_2 \;\leq\; \max\{1,\; 3 D^2\} \sqrt{N}.$$

Convergence rates. By collecting terms, we obtain the following convergence rates; note that the constants may be recycled from line to line and may vary from model to model. All results hold for large enough N.

- **Corollary 1:** We have $\alpha = 0$, $p = N$, and $\Psi = \sqrt{N}$. The independence of edges under the β-model implies that $\Lambda = \widetilde{\Lambda}$, $\mathcal{C} = \widetilde{\mathcal{C}}$, $\max_{1 \leq i < j \leq N} |\mathfrak{N}_{i,j}| = 0$, and $|||\mathcal{D}|||_2 = 1$, which in turn implies that $\Phi = \widetilde{\Phi}$. As a result, γ reduces to

$$\begin{aligned}\gamma &= \sqrt{3/2}\, \max\left\{1,\, \max_{1 \leq i < j \leq N} |\mathfrak{N}_{i,j}|\right\}\, |||\mathcal{D}|||_2\, \Psi\, \sqrt{\log \max\{N,\, p\}} \\ &= \sqrt{3/2}\, \sqrt{N\, \log N}.\end{aligned}$$

  By Lemma 5 with $\alpha = 0$, there exist constants $A > 0$ and $B \geq 1$, independent of N and p, such that $\Lambda = \widetilde{\Lambda} \geq A\, N^{\vartheta}$ and $\mathcal{C} = \widetilde{\mathcal{C}} \leq B\, N^{3(1-\vartheta)/2}$. Thus, there exists a constant $C > 0$, independent N and p, such that

$$\Phi = \widetilde{\Phi} \geq \frac{C\, N^{\vartheta}}{N^{3(1-\vartheta)/2}\, \sqrt{N}\, \sqrt{\log N}} = C\, \sqrt{\frac{N^{5\,\vartheta - 4}}{\log N}}, \quad \vartheta \in (4/5,\, 1].$$

- **Corollary 2:** We have $\alpha = 0$, $p = N + 1$, and

$$1 \leq \max\left\{1,\, \max_{1 \leq i < j \leq N} |\mathfrak{N}_{i,j}|\right\} \leq D.$$

  Therefore, using the above, we have shown that Ψ satisfies

$$\sqrt{N} \leq \Psi \leq \max\{1,\, 3\, D^2\}\, \sqrt{N} \leq 3\, D^2\, \sqrt{N}.$$

  To bound γ, recall that γ is given by

$$\gamma = \sqrt{3/2}\, \max\left\{1,\, \max_{1 \leq i < j \leq N} |\mathfrak{N}_{i,j}|\right\}\, |||\mathcal{D}|||_2\, \Psi\, \sqrt{\log \max\{N,\, p\}}.$$

  Using $\Psi \geq \sqrt{N}$ along with $p = N + 1 \geq N$, we obtain the lower bound

$$\gamma \geq \sqrt{3/2}\, |||\mathcal{D}|||_2\, \sqrt{N\, \log N}$$

  and, using $\Psi \leq 3\, D^2\, \sqrt{N}$ along with

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

we obtain the upper bound

$$\gamma \;\; \leq \;\; 6\, D^3\, |||\mathcal{D}|||_2\, \sqrt{N\, \log N}.$$

Thus, γ satisfies

$$\sqrt{3/2}\;\, |||\mathcal{D}|||_2\, \sqrt{N\, \log\, N} \;\;\; \leq \;\;\; \gamma \;\;\; \leq \;\;\; 6\;\, D^2\, |||\mathcal{D}|||_2\, \sqrt{N \log N}.$$

By Lemma 6 with $\alpha = 0$ and $\vartheta \in (4/5,\, 1]$, there exists a constant $C_1 > 0$, independent of N and p, such that $\widetilde{\Lambda} \;\geq\; C_1\, N^{\vartheta}$, provided $\min_{1 \leq k \leq K} |\mathcal{A}_k| \geq 3$. As a consequence, there exists a constant $C_2 > 0$, independent of N and p, such that

$$\begin{aligned} \widetilde{\Phi} \;\; &\geq \;\; \frac{C_1\, N^{\vartheta}}{\widetilde{\mathcal{C}}\;\, D\; |||\mathcal{D}|||_2\, (3\, D^2\, \sqrt{N})\, \sqrt{2\, \log N}} \\ &\geq \;\; \frac{C_2}{\widetilde{\mathcal{C}}\;\, D^3\; |||\mathcal{D}|||_2}\, \sqrt{\frac{N^{2\,\vartheta - 1}}{\log N}}, \end{aligned}$$

where edges are dependent provided $\theta_{N+1} \neq 0$ and $D = \max_{1 \leq i < j \leq N} |\mathfrak{N}_{i,j}| \geq 1$ provided $\theta_{N+1} \neq 0$.

- **Corollary 3:** By Lemma 6 with $\alpha \in (0,\, 1/2)$, there exists a constant $C_1 > 0$, independent of N and p, such that $\widetilde{\Lambda} \;\geq\; C_1\; N^{\vartheta - \alpha}$, provided $\min_{1 \leq k \leq K} |\mathcal{A}_k| \geq 3$. As a result, there exists a constant $C_2 > 0$, independent of N and p, such that

$$\begin{aligned} \widetilde{\Phi} \;\; &\geq \;\; \frac{C_1\; N^{\vartheta - \alpha}}{\widetilde{\mathcal{C}}\;\, D\;\; |||\mathcal{D}|||_2\, (3\, D^2\, \sqrt{N})\, \sqrt{2\, \log N}} \\ &\geq \;\; \frac{C_2}{\widetilde{\mathcal{C}}\;\, D^3\; |||\mathcal{D}|||_2}\, \sqrt{\frac{N^{2\,(\vartheta - \alpha) - 1}}{\log N}}, \end{aligned}$$

where edges are dependent provided $\theta_{N+1} \neq 0$ and $D = \max_{1 \leq i < j \leq N} |\mathfrak{N}_{i,j}| \geq 1$ provided $\theta_{N+1} \neq 0$. Observe that ϑ must satisfy $\vartheta \in (4/5,\, 1]$ and $2\,(\vartheta - \alpha) - 1 > 0$ with $\alpha \in (0,\, 1/2)$, which implies that ϑ must satisfy

$$\vartheta \;\; \in \;\; (\max\{4/5,\; 1/2 + \alpha\},\;\, 1]\,.$$

We bound $|||\mathcal{D}|||_2$ in scenarios 1 and 2:

1. Assume that the subpopulations do not intersect ($\omega_1 = \omega_2 = 0$) and that $\boldsymbol{\theta}^\star \in \mathbb{R}^{N+1}$ satisfies (3.7) with $\vartheta \in (\max\{4/5,\, 1/2+\alpha\},\, 1]$. Then Lemma 11 implies that

$$|||\mathcal{D}|||_2 \;\;\leq\;\; 1 + D^2.$$

Thus, there exists a constant $C_3 > 0$, independent of N and p, such that

$$\widetilde{\Phi} \;\;\geq\;\; \frac{C_3}{\widetilde{\mathcal{C}}\, D^5} \sqrt{\frac{N^{2\,(\vartheta-\alpha)-1}}{\log N}}.$$

2. The subpopulations intersect and $\boldsymbol{\theta}^\star \in \mathbb{R}^{N+1}$ satisfies (3.7) with $\vartheta = 1$. Then, by Lemma 11, there exist constants $C_4 > 0$ and $C_5 > 0$, independent of N and p, such that

$$|||\mathcal{D}|||_2 \;\;\leq\;\; C_4\, D^2 \exp(C_5\, D^2).$$

Thus, there exists a constant $C_6 > 0$, independent of N and p, such that

$$\widetilde{\Phi} \;\;\geq\;\; \frac{C_6}{\widetilde{\mathcal{C}}\;\, D^5\, \exp(C_5\, D^2)} \sqrt{\frac{N^{1-2\,\alpha}}{\log N}}.$$

- **Corollary 4:** The proof of Corollary 4 resembles the proof of Corollary 3 with $\alpha = 0$.

D.1. Bounding $\widetilde{\boldsymbol{\Lambda}}$. To bound $\widetilde{\Lambda}$, we prove Lemmas 5–10. To do so, we first introduce the sparse β-model with independent edges, with probability mass function

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) \;\;=\;\; \prod_{i<j}^{N} \frac{\exp((\theta_i + \theta_j)\, x_{i,j})\, N^{-\alpha\, x_{i,j}}}{1 + \exp(\theta_i + \theta_j)\, N^{-\alpha}}, \tag{D.1}$$

where $\alpha \in [0,\, 1/2)$ is a known constant, which may be interpreted as the level of sparsity of the random graph.

Lemma 5 bounds the smallest eigenvalue of the expected Hessian $-\mathbb{E}\, \nabla^2_{\boldsymbol{\theta}}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$ under the sparse β-model; note that Model 1 is the special case of the sparse β-model with $\alpha = 0$. Under Models 2 and 3, the expected Hessian $-\mathbb{E}\, \nabla^2_{\boldsymbol{\theta}}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$ is of the form

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

where

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

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

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

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

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

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

Under Model 4, the expected Hessian $-\mathbb{E}\, \nabla_{\boldsymbol{\theta}}^2\, \widetilde{\ell}(\boldsymbol{\theta};\, \boldsymbol{X})$ involves derivatives of the size-dependent parameters $\eta_{i,j}(\theta_{N+1})$, defined by

$$\eta_{i,j}(\theta_{N+1}) \;\;=\;\; \theta_{N+1}\, \log\left(1 + \frac{\log |\mathcal{N}_i \,\cap\, \mathcal{N}_j|}{|\mathcal{N}_i \,\cap\, \mathcal{N}_j|}\right), \qquad \theta_{N+1} \in \mathbb{R}.$$

That said, the canonical form of exponential families is not unique, so the terms $\log(1 + \log |\mathcal{N}_i \,\cap\, \mathcal{N}_j| \,/\, |\mathcal{N}_i \,\cap\, \mathcal{N}_j|)$ can be absorbed into the sufficient statistic $s_{N+1}(\boldsymbol{X})$. The advantage of doing so is that the expected expected Hessian $-\mathbb{E}\, \nabla_{\boldsymbol{\theta}}^2\, \widetilde{\ell}(\boldsymbol{\theta};\, \boldsymbol{X})$ then has the same form under Models 2, 3, and 4. We therefore absorb the terms $\log(1+\log |\mathcal{N}_i \cap \mathcal{N}_j| \,/\, |\mathcal{N}_i \cap \mathcal{N}_j|)$ into $s_{N+1}(\boldsymbol{X})$ to streamline the proofs of Lemmas 5–10. As a consequence, we can ignore the terms $\log(1 + \log |\mathcal{N}_i \,\cap\, \mathcal{N}_j| \,/\, |\mathcal{N}_i \,\cap\, \mathcal{N}_j|)$, because the terms are bounded above and do not affect the rate of growth of $\widetilde{\Lambda}$.

Lemma 5. *Consider Models 1–4 with a known level of sparsity $\alpha \in [0,\, 1/2)$. Assume that the data-generating parameter vector $\boldsymbol{\theta}^\star \in \mathbb{R}^p$ satisfies*

$$\|\boldsymbol{\theta}^\star\|_\infty \;\;\leq\;\; \frac{L + (1 - \vartheta)\, \log N}{4\,(3 + D)},$$

where $L \geq 0$ and $\vartheta \in (4/5,\, 1]$ are finite constants, independent of N and p, and $D \coloneqq \max_{1 \leq i < j \leq N} |\mathfrak{N}_{i,j}| \geq 0$. Then, under Models 1–4, there exists a finite constant $C \coloneqq \exp(-L) \,/\, 3 \in (0,\, 1/3]$, independent of N and p, such that, for all $\epsilon^\star \in (0,\, \|\boldsymbol{\theta}^\star\|_\infty]$ and all $N \geq 3$,

$$\inf_{\boldsymbol{\theta} \,\in\, \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)} \sqrt[N]{|\det(\boldsymbol{A}(\boldsymbol{\theta}))|} \;\;\geq\;\; C\, N^{\vartheta - \alpha}.$$

In addition, under Model 1, there exist finite constants $C \coloneqq \exp(-L) \,/\, 12 \in (0,\, 1/12]$ *and* $U \coloneqq 40 \exp(3\,L \,/\, 2) \in [40,\, \infty)$*, independent of* N *and* p*, such that, for all* $\epsilon^\star \in (0,\, \|\boldsymbol{\theta}^\star\|_\infty]$ *and all* $N \geq 3$*,*

$$\begin{aligned} \widetilde{\Lambda} &\geq C\, N^{\vartheta-\alpha} \\ \widetilde{\mathcal{C}} &\leq U\, N^{3(1-\vartheta)/2}. \end{aligned}$$

Proof of Lemma 5. By definition,

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

It is straightforward to calculate, for each pair of nodes $\{i, j\} \subset \mathcal{N}$ and each pair of coordinates $(t, l) \in \{1, \ldots, N\} \times \{1, \ldots, N\}$ of $-\nabla^2_{\boldsymbol{\theta}}\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{x})$,

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

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

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

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

1. If $t \notin \{i, j\}$ or $l \notin \{i, j\}$, then $s_t(\boldsymbol{X})$ and $s_l(\boldsymbol{X})$ cannot be both a function of $X_{i,j}$. It then follows that, conditional on $\boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}$,
$$\mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{x}_{-\{i,j\}}}(s_t(\boldsymbol{X}), s_l(\boldsymbol{X})) \;=\; 0.$$
2. If $\{t, l\} \subseteq \{i, j\}$, then either $\{t, l\} = \{i, j\}$ or $t = l \in \{i, j\}$. In both cases, $s_t(\boldsymbol{X})$ 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

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

In the special case when $t = l$,

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

As a result, for all $t \neq l$,

$$\sum_{i<j}^{N} \mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{i,j\}}}(s_t(\boldsymbol{X}), s_l(\boldsymbol{X})) \;\;=\;\; \mathbb{E}\, \mathbb{V}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{t,l\}}}\, X_{t,l} \tag{D.3}$$

and

$$\sum_{i<j}^{N} \mathbb{E}\, \mathbb{V}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{i,j\}}}\, s_t(\boldsymbol{X}) \;\;=\;\; \sum_{l \,\in\, \mathcal{N} \setminus \{t\}} \mathbb{E}\, \mathbb{V}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{t,l\}}}\, X_{t,l}. \tag{D.4}$$

An important consequence of (D.3) and (D.4) is that the matrix $-\mathbb{E}\, \nabla_{\boldsymbol{\theta}}^2\, \tilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$ is diagonally balanced, in the sense of Hillar and Wibisono [25]. 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 12 and since $N^{-\alpha} \leq 1$ $(N \geq 1)$, for all $\boldsymbol{x}_{-\{i,j\}} \in \{0,1\}^{\binom{N}{1}-1}$,

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

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

and

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

where $D \geq 0$. We thus obtain, for all pairs of nodes $\{i, j\} \subset \mathcal{N}$ and all $\boldsymbol{x}_{-\{i,j\}} \in \{0,1\}^{\binom{N}{2}-1}$,

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

which implies

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

By invoking Lemma 2.1 of Hillar and Wibisono [25] and the above bounds, the smallest eigenvalue $\lambda_{\min}(\boldsymbol{\theta})$ of the matrix $\boldsymbol{A}(\boldsymbol{\theta}) \in \mathbb{R}^{N \times N}$ at $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$ satisfies

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

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

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

By invoking Corollary 2.1 of Hillar et al. [24][1] and using the inequality $(3\,N-4)\,/\,(N-2) \leq 5$ ($N \geq 3$),

$$\begin{aligned}\kappa_\infty(\boldsymbol{\theta}) \;&\coloneqq\; \|\boldsymbol{A}(\boldsymbol{\theta})\|_\infty\; \|\,(\boldsymbol{A}(\boldsymbol{\theta}))^{-1}\|_\infty \\ &\leq\; \frac{(1+\exp((3+D)\,\|\boldsymbol{\theta}\|_\infty))^3\,(3\,N-4)}{N-2} \\ &\leq\; 5\,(1+\exp((3+D)\,\|\boldsymbol{\theta}\|_\infty))^3.\end{aligned}$$

By the reverse triangle inequality and the assumption that $\epsilon^\star \in (0,\, \|\boldsymbol{\theta}^\star\|_\infty]$, we have $\|\boldsymbol{\theta}\|_\infty \leq \|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star \leq 2\,\|\boldsymbol{\theta}^\star\|_\infty$, which implies that

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

and

$$\sup_{\boldsymbol{\theta}\,\in\,\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\,\epsilon^\star)} \kappa_\infty(\boldsymbol{\theta}) \;\;\leq\;\; 5\,(1+\exp(2\,(3+D)\,\|\boldsymbol{\theta}^\star\|_\infty))^3.$$

We want to show that there exists a finite constant $C > 0$, independent of N and p, such that

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

[1]The proof of Corollary 2.1 of Hillar et al. [24] follows from Theorem 1.2 of Hillar and Wibisono [25] (Theorem 1.1 in Hillar et al. [24]), as noted by Hillar et al. [24].

Upon rearranging terms, we obtain

$$\begin{aligned}\left(\frac{N^{1-\vartheta}}{3\,C}\right)^{1/2} &\geq 1 + \exp(2\,(3+D)\,\|\boldsymbol{\theta}^\star\|_\infty)) \\ &\geq \exp(2\,(3+D)\,\|\boldsymbol{\theta}^\star\|_\infty)).\end{aligned}$$

Upon taking the natural logarithm on both sides, we obtain

$$\frac{(1-\vartheta)\,\log N - \log 3\,C}{4\,(3+D)} \;\geq\; \|\boldsymbol{\theta}^\star\|_\infty.$$

Let $C \coloneqq \exp(-L)\,/\,3 \in (0,\,1/3]$ and note that C is independent of N and p, because $L \geq 0$ is independent of N and p. Then, for all $\epsilon^\star \in (0,\,\|\boldsymbol{\theta}^\star\|_\infty]$ and all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\,\epsilon^\star)$,

$$\inf_{\boldsymbol{\theta}\,\in\,\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\,\epsilon^\star)} \lambda_{\min}(\boldsymbol{\theta}) \;\geq\; C\,N^{\vartheta-\alpha},$$

provided that

$$\|\boldsymbol{\theta}^\star\|_\infty \;\leq\; \frac{L + (1-\vartheta)\,\log N}{4\,(3+D)}. \tag{D.5}$$

Using the properties of determinants and the above bound on $\lambda_{\min}(\boldsymbol{\theta})$,

$$|\det(-\mathbb{E}\,\nabla^2_{\boldsymbol{\theta}}\,\tilde{\ell}(\boldsymbol{\theta};\boldsymbol{X}))| \;=\; |\det(\boldsymbol{A}(\boldsymbol{\theta}))| \;\geq\; |\lambda_{\min}(\boldsymbol{\theta})|^N \;\geq\; (C\,N^{\vartheta-\alpha})^N$$

implying

$$\inf_{\boldsymbol{\theta}\,\in\,\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\,\epsilon^\star)} \sqrt[N]{|\det(\boldsymbol{A}(\boldsymbol{\theta}))|} \;\geq\; C\,N^{\vartheta-\alpha},$$

provided that (D.5) is satisfied and $N \geq 3$.

Last, but not least, we consider Model 1. Under Model 1 with $D = 0$,

$$\widetilde{\Lambda} \;=\; \inf_{\boldsymbol{\theta}\,\in\,\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\,\epsilon^\star)} \sqrt[N]{|\det(\boldsymbol{A}(\boldsymbol{\theta}))|} \;\geq\; C\,N^{\vartheta-\alpha}$$

and

$$\begin{aligned}\widetilde{\mathcal{C}} \;\coloneqq\; & \sup_{\boldsymbol{\theta}\,\in\,\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\,\epsilon^\star)} \kappa_\infty(\boldsymbol{\theta}) \;\leq\; 5\,(1 + \exp(2\,(2+3\,D)\,\|\boldsymbol{\theta}^\star\|_\infty)))^3 \\ \leq\; & 40\,\exp(2\,(3+D)\,\|\boldsymbol{\theta}^\star\|_\infty)))^3 \;\leq\; 40\,\exp\left(\frac{L + (1-\vartheta)\,\log N)}{2}\right)^3 \\ =\; & 40\,\exp\left(\frac{3\,L + 3\,(1-\vartheta)\,\log N)}{2}\right) \;=\; U\,N^{3\,(1-\vartheta)/2},\end{aligned}$$

provided that (D.5) is satisfied, where $U \coloneqq 40 \exp(3\, L\,/\,2) \in [40, \infty)$ is independent of N and p, because $L \geq 0$ is independent of N and p.

Lemma 6. *Consider Models 2–4 with a known level of sparsity $\alpha \in [0, 1/2)$. Assume that the data-generating parameter vector $\boldsymbol{\theta}^\star \in \mathbb{R}^p$ satisfies*

$$\|\boldsymbol{\theta}^\star\|_\infty \;\;\leq\;\; \frac{L + (1 - \vartheta)\, \log N}{16\,(3 + D)}, \tag{D.6}$$

where $L \geq 0$ and $\vartheta \in (4/5,\, 1]$ are finite constants, independent of N and p, and $D \coloneqq \max_{1 \leq i < j \leq N} |\mathfrak{N}_{i,j}| \geq 1$. Then there exists a finite constant $C > 0$, independent of N and p, such that, for all $\epsilon^\star \in (0,\, \|\boldsymbol{\theta}^\star\|_\infty]$,

$$\widetilde{\Lambda} \;\;\coloneqq\;\; \inf_{\boldsymbol{\theta} \,\in\, \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)} \sqrt[p]{\det(-\mathbb{E}\, \nabla_{\boldsymbol{\theta}}^2\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}))} \;\;\geq\;\; C\, N^{\vartheta - \alpha},$$

provided $\min_{1 \leq k \leq K} |\mathcal{A}_k| \geq 3$ and N is large enough.

Proof of Lemma 6. Using (D.2), the determinant of the expected Hessian $-\mathbb{E}\, \nabla_{\boldsymbol{\theta}}^2\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})$ can be written as

$$\det(-\mathbb{E}\, \nabla_{\boldsymbol{\theta}}^2\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})) \;=\; \det(\boldsymbol{A}(\boldsymbol{\theta}))\; (v(\boldsymbol{\theta}) - \boldsymbol{c}(\boldsymbol{\theta})^\top\, \boldsymbol{A}(\boldsymbol{\theta})^{-1}\, \boldsymbol{c}(\boldsymbol{\theta})).$$

Observe that the assumptions of Lemma 5 are satisfied, so we may conclude that there exists, for all $\epsilon^\star \in (0,\, \|\boldsymbol{\theta}^\star\|_\infty]$, a finite constant $C_1 > 0$, independent of N and p, such that

$$\det(\boldsymbol{A}(\boldsymbol{\theta})) \;\;\geq\;\; \lambda_{\min}(\boldsymbol{\theta})^N \;\;\geq\;\; (C_1\, N^{\vartheta - \alpha})^N,$$

provided (D.6) is satisfied. Thus,

$$\begin{aligned}
\det(-\mathbb{E}\, \nabla_{\boldsymbol{\theta}}^2\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X})) \;\;&=\;\; \det(\boldsymbol{A}(\boldsymbol{\theta}))\;\, (v(\boldsymbol{\theta}) - \boldsymbol{c}(\boldsymbol{\theta})^\top\, \boldsymbol{A}(\boldsymbol{\theta})^{-1}\, \boldsymbol{c}(\boldsymbol{\theta}))\\
&\geq\;\; (C_1\, N^{\vartheta - \alpha})^N\; (v(\boldsymbol{\theta}) - \boldsymbol{c}(\boldsymbol{\theta})^\top\, \boldsymbol{A}(\boldsymbol{\theta})^{-1}\, \boldsymbol{c}(\boldsymbol{\theta}))\\
&=\;\; (C_1\, N^{\vartheta - \alpha})^N\; (\boldsymbol{c}(\boldsymbol{\theta})^\top\, \boldsymbol{c}(\boldsymbol{\theta}))\; \left(\frac{v(\boldsymbol{\theta})}{\boldsymbol{c}(\boldsymbol{\theta})^\top\, \boldsymbol{c}(\boldsymbol{\theta})} - R(\boldsymbol{A}(\boldsymbol{\theta})^{-1},\, \boldsymbol{c}(\boldsymbol{\theta}))\right),
\end{aligned}$$

where

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

is the Rayleigh quotient of $\boldsymbol{A}(\boldsymbol{\theta})^{-1} \in \mathbb{R}^{N \times N}$, assuming $\boldsymbol{c}(\boldsymbol{\theta}) \in \mathbb{R}^N \setminus \mathbf{0}$. Here, $\mathbf{0} \in \mathbb{R}^N$ denotes the N-dimensional zero vector. We bound the terms $\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta})$, $v(\boldsymbol{\theta}) \,/\, (\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta}))$, and $R(\boldsymbol{A}(\boldsymbol{\theta})^{-1}, \boldsymbol{c}(\boldsymbol{\theta}))$ one by one, assuming

$$\|\boldsymbol{\theta}^\star\|_\infty \;\; \leq \;\; \frac{L + (1 - \vartheta)\, \log N}{16\, (3 + D)},$$

so that the conditions of all lemmas are satisfied.

First, by Lemma 7, there exists a constant $C_2 > 0$ such that

$$\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta}) \;\; \geq \;\; C_2\, N^\vartheta \;\; > \;\; 0 \quad \text{for all} \quad \boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star).$$

Second, by Lemma 9, there exists a constant $C_3 > 0$ such that

$$\frac{v(\boldsymbol{\theta})}{\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta})} \;\; \geq \;\; \frac{C_3}{D^8\, N^{1-\vartheta}} \;\; > \;\; 0 \quad \text{for all} \quad \boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star).$$

Last, but not least, to bound the Rayleigh quotient $R(\boldsymbol{A}(\boldsymbol{\theta})^{-1}, \boldsymbol{c}(\boldsymbol{\theta}))$, note that if $\lambda_1(\boldsymbol{\theta}), \ldots, \lambda_N(\boldsymbol{\theta})$ are the eigenvalues of $\boldsymbol{A}(\boldsymbol{\theta})$, then $1/\lambda_1(\boldsymbol{\theta})$, $\ldots$, $1/\lambda_N(\boldsymbol{\theta})$ are the eigenvalues of $\boldsymbol{A}(\boldsymbol{\theta})^{-1}$. Let $\lambda_{\min}(\boldsymbol{\theta})$ be the smallest eigenvalue of $\boldsymbol{A}(\boldsymbol{\theta})$, so $1/\lambda_{\min}(\boldsymbol{\theta})$ is the largest eigenvalue of $\boldsymbol{A}(\boldsymbol{\theta})^{-1}$. Since the Rayleigh quotient is bounded above by the largest eigenvalue of $\boldsymbol{A}(\boldsymbol{\theta})^{-1}$, we obtain, using Lemma 5,

$$R(\boldsymbol{A}(\boldsymbol{\theta})^{-1},\, \boldsymbol{c}(\boldsymbol{\theta})) \;\; \leq \;\; \frac{1}{\lambda_{\min}(\boldsymbol{\theta})} \;\; \leq \;\; \frac{1}{C_1\, N^{\vartheta-\alpha}}.$$

As a result,

$$\begin{aligned}
& (C_1\, N^{\vartheta-\alpha})^N\, (\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta})) \left(\frac{v(\boldsymbol{\theta})}{\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta})} - R(\boldsymbol{A}(\boldsymbol{\theta})^{-1},\, \boldsymbol{c}(\boldsymbol{\theta}))\right) \\
\geq \;\; & (C_1\, N^{\vartheta-\alpha})^N\, C_2\, N^\vartheta \left(\frac{C_3}{D^8\, N^{1-\vartheta}} - \frac{1}{C_1\, N^{\vartheta-\alpha}}\right) \\
= \;\; & C_1^N\, C_2\, N^{(\vartheta-\alpha)\, N}\, \frac{N^\vartheta}{N^{1-\vartheta}} \left(\frac{C_3}{D^8} - \frac{1}{C_1\, N^{2\,\vartheta-\alpha-1}}\right) \\
= \;\; & C_1^N\, C_2\, N^{(\vartheta-\alpha)\,(N+1)}\, \frac{N^\alpha}{N^{1-\vartheta}} \left(\frac{C_3}{D^8} - \frac{1}{C_1\, N^{2\,\vartheta-\alpha-1}}\right).
\end{aligned}$$

By assumption, $2\,\vartheta - \alpha - 1 > 0$, because $\alpha = 0$ and $\vartheta \in (4/5,\, 1]$ under Models 2 and 4, whereas $\alpha \in (0,\, 1/2)$ and $\vartheta \in (4/5,\, 1]$ under Model 3.

Thus, $N^{2\,\vartheta-\alpha-1} \to \infty$ as $N \to \infty$ and there exist finite constants $C_4 > 0$ and $C_5 > 0$, independent of N and p, such that

$$\begin{array}{lll}\widetilde{\Lambda} \;\coloneqq\; \inf\limits_{\boldsymbol{\theta} \,\in\, \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\,\epsilon^\star)} \sqrt[N+1]{\det(-\mathbb{E}\,\nabla_{\boldsymbol{\theta}}^2\, \widetilde{\ell}(\boldsymbol{\theta}; \boldsymbol{X}))} & \geq & C_4\, N^{\vartheta-\alpha} \left(\dfrac{N^{\alpha}}{D^8\, N^{1-\vartheta}}\right)^{\frac{1}{N+1}} \\[2ex] & \geq & C_5\, N^{\vartheta-\alpha},\end{array}$$

provided N is large enough.

Lemma 7. *Consider Models 2–4. Assume that the data-generating parameter vector $\boldsymbol{\theta}^\star \in \mathbb{R}^{N+1}$ satisfies*

$$\|\boldsymbol{\theta}^\star\|_\infty \;\;\leq\;\; \frac{L + (1-\vartheta)\, \log N}{16\,(3+D)},$$

where $L \geq 0$ and $\vartheta \in (4/5,\, 1]$ are finite constants, independent of N and p, and $D \coloneqq \max_{1 \leq i < j \leq N} |\mathfrak{N}_{i,j}| \geq 1$. Then, for all $\epsilon^\star \in (0,\, \|\boldsymbol{\theta}^\star\|_\infty]$ and all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$,

$$\boldsymbol{c}(\boldsymbol{\theta})^\top\, \boldsymbol{c}(\boldsymbol{\theta}) \;\;\geq\;\; C\, N^{\vartheta},$$

where $C \coloneqq \exp(-L)\,/\,256 \in (0,\, 1/256]$ is a finite constant, independent of N and p.

Proof of Lemma 7. By Lemma 8, for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$,

$$\begin{array}{lll}\boldsymbol{c}(\boldsymbol{\theta})^\top\, \boldsymbol{c}(\boldsymbol{\theta}) & \geq & \dfrac{N}{(1 + \exp((3+D)\,(\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)))^8} \\[2ex] & = & \dfrac{N}{256\, \exp((3+D)\,(\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)))^8} \\[2ex] & \geq & \dfrac{N}{256\, \exp(2\,(3+D)\, \|\boldsymbol{\theta}^\star\|_\infty)^8},\end{array}$$

using $\epsilon^\star \in (0,\, \|\boldsymbol{\theta}^\star\|_\infty]$. Observe that the lower bound on $\boldsymbol{c}(\boldsymbol{\theta})^\top\, \boldsymbol{c}(\boldsymbol{\theta})$ does not involve α, because the elements of the vector $\boldsymbol{c}(\boldsymbol{\theta}) \in \mathbb{R}^N$ correspond to the covariances of the degrees of nodes and the number of brokered edges of nodes with intersecting neighborhoods, and brokered edges among nodes in intersecting neighborhoods are not penalized by α.

We demonstrate that there exists a finite constant $C > 0$, independent of N and p, such that

$$\frac{N}{256\, \exp(2\,(3+D)\, \|\boldsymbol{\theta}^\star\|_\infty)^8} \;\;\geq\;\; C\, N^{\vartheta}, \tag{D.7}$$

which in turn implies that $\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta}) \geq C\, N^{\vartheta}$. Upon rearranging (D.7), we obtain

$$\left(\frac{N^{1-\vartheta}}{256\, C}\right)^{1/8} \quad \geq \quad \exp(2\,(3+D)\,\|\boldsymbol{\theta}^\star\|_\infty).$$

Upon taking the natural logarithm on both sides, we obtain

$$\frac{(1-\vartheta)\, \log N - \log 256\, C}{16\,(3+D)} \quad \geq \quad \|\boldsymbol{\theta}^\star\|_\infty.$$

Let $C \coloneqq \exp(-L)\,/\,256 \in (0,\, 1/256]$ and note that C is independent of N and p, because $L \geq 0$ is independent of N and p. Then, for all $\epsilon^\star \in (0,\, \|\boldsymbol{\theta}^\star\|_\infty]$ and all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$,

$$\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta}) \quad \geq \quad C\, N^{\vartheta},$$

provided that

$$\|\boldsymbol{\theta}^\star\|_\infty \quad \leq \quad \frac{L + (1-\vartheta)\, \log N}{16\,(3+D)}.$$

Lemma 8. *Consider Models 2–4. Then, for all $\epsilon^\star \in (0,\, \|\boldsymbol{\theta}^\star\|_\infty]$, all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$, and all $N \geq 3$,*

$$\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta}) \quad \geq \quad \frac{N}{(1 + \exp((3+D)\,(\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)))^8},$$

where $D \coloneqq \max_{1 \leq i < j \leq N} |\mathfrak{N}_{i,j}| \geq 1$.

Proof of Lemma 8. The coordinates of $\boldsymbol{c}(\boldsymbol{\theta}) \in \mathbb{R}^N$ are given by

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

noting that edges $X_{t,a}$ ($\{t,a\} \neq \{i,j\}$) are almost surely constant conditional on $\boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}$ and that $s_{N+1}(\boldsymbol{X})$ is not a function of edges $X_{t,a}$ satisfying $\mathcal{N}_t \cap \mathcal{N}_a = \emptyset$, and would thus is almost surely constant under the conditional distribution of $X_{t,a}$ given $\boldsymbol{X}_{-\{t,a\}} = \boldsymbol{x}_{-\{t,a\}}$. We obtain

$$\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta}) \;=\; \sum_{t=1}^{N} \left(\sum_{a \in \mathcal{N} \setminus \{t\}:\, \mathcal{N}_a \cap \mathcal{N}_t \neq \emptyset} \mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, s_{N+1}(\boldsymbol{X})) \right)^2.$$

It is therefore enough to demonstrate that

$$\begin{aligned} &\left(\sum_{a \in \mathcal{N} \setminus \{t\}:\, \mathcal{N}_a \cap \mathcal{N}_t \neq \emptyset} \mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, s_{N+1}(\boldsymbol{X})) \right)^2 \\ \geq\;\; &\frac{1}{(1 + \exp((3 + D)\,(\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)))^8}, \end{aligned}$$

where $D \geq 1$. Let

$$I_{t,a}(\boldsymbol{X}) \;=\; \mathbb{1}\left(\sum_{h \in \mathcal{N}_t \cap \mathcal{N}_a} X_{t,h}\, X_{a,h} \,>\, 0 \right), \qquad \{t,a\} \subset \mathcal{N},$$

and note that

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

where $I_{i,j}(\boldsymbol{X}) = 0$ almost surely for all $\{i,j\} \subset \mathcal{N}$ satisfying $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset$. Using properties of covariances,

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

The FKG inequality implies that

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

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

$$\mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, s_{N+1}(\boldsymbol{X})) \;\;\geq\;\; \mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, X_{t,a}\, I_{t,a}(\boldsymbol{X})).$$

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 $b \in \mathcal{N}_i \cap \mathcal{N}_j$ such that

$$\begin{aligned} & \mathbb{P}_{\boldsymbol{\theta}}(X_{t,a}\, I_{t,a}(\boldsymbol{X}) = 1 \mid \boldsymbol{X}_{-\{t,a\}} = \boldsymbol{x}_{-\{t,a\}}) \\ \geq \;\; & \mathbb{P}_{\boldsymbol{\theta}}(X_{t,a}\, X_{t,b}\, X_{a,b} = 1 \mid \boldsymbol{X}_{-\{t,a\}} = \boldsymbol{x}_{-\{t,a\}}), \end{aligned}$$

because the event $X_{t,a}\, X_{t,b}\, X_{a,b} = 1$ implies the event $X_{t,a}\, I_{t,a}(\boldsymbol{X}) = 1$. By Lemma 12,

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

and

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

for all pairs of nodes $\{i, j\} \subset \mathcal{N}$ satisfying $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$. 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

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

Thus, using the law of total expectation,

$$\begin{aligned} & \mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, X_{t,a}\, I_{t,a}(\boldsymbol{X})) \\ = \;\; & \mathbb{P}(I_{t,a}(\boldsymbol{X}) = 1)\, \mathbb{V}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}\, X_{t,a} \\ \geq \;\; & \mathbb{P}(X_{t,b}\, X_{a,b} = 1) \left(\frac{1}{1 + \exp((3 + D)\, \|\boldsymbol{\theta}\|_\infty)}\right)^2 \\ \geq \;\; & \left(\frac{1}{1 + \exp((3 + D)\, \|\boldsymbol{\theta}^\star\|_\infty)}\right)^2 \left(\frac{1}{1 + \exp((3 + D)\, \|\boldsymbol{\theta}\|_\infty)}\right)^2 \\ \geq \;\; & \frac{1}{(1 + \exp((3 + D)\, \max\{\|\boldsymbol{\theta}^\star\|_\infty,\, \|\boldsymbol{\theta}\|_\infty\}))^4}. \end{aligned}$$

Thus, we have shown that, for all $t \in \{1, \dots, N\}$,

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

which in turn implies that

$$\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta}) \quad \geq \quad \frac{N}{(1 + \exp((3 + D)\, \max\{\|\boldsymbol{\theta}^\star\|_\infty,\, \|\boldsymbol{\theta}\|_\infty\}))^8}.$$

Since $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$, we know that $\|\boldsymbol{\theta}\|_\infty \leq \|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star$ by the reverse triangle inequality, which implies that, for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$,

$$\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta}) \quad \geq \quad \frac{N}{(1 + \exp((3 + D)\, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)))^8}.$$

Lemma 9. *Consider Models 2–4. Assume that the data-generating parameter vector $\boldsymbol{\theta}^\star \in \mathbb{R}^{N+1}$ satisfies*

$$\|\boldsymbol{\theta}^\star\|_\infty \quad \leq \quad \frac{L + (1 - \vartheta)\, \log N}{16\, (3 + D)},$$

where $L \geq 0$ and $\vartheta \in (4/5,\, 1]$ are finite constants, independent of N and p, and $D \coloneqq \max_{1 \leq i < j \leq N} |\mathfrak{N}_{i,j}| \geq 1$. Then, for all $\epsilon^\star \in (0,\, \|\boldsymbol{\theta}^\star\|_\infty]$ and all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$,

$$\frac{v(\boldsymbol{\theta})}{\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta})} \quad \geq \quad \frac{C}{D^8\, N^{1-\vartheta}},$$

where $C \coloneqq \exp(-L) \,/\, 16 \in (0,\, 1/16]$ is a finite constant, independent of N and p.

Proof of Lemma 9. By Lemma 10, for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$,

$$\begin{aligned} \frac{v(\boldsymbol{\theta})}{\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta})} \quad &\geq \quad \frac{1}{D^8\, (1 + \exp((3 + D)\, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)))^4} \\ &\geq \quad \frac{1}{D^8\, (2\, \exp(2\, (3 + D)\, \|\boldsymbol{\theta}^\star\|_\infty))^4}, \end{aligned} \tag{D.8}$$

using $\epsilon^\star \in (0,\, \|\boldsymbol{\theta}^\star\|_\infty]$. We show that there exists a finite constant $C > 0$, independent of N and p, such that

$$\frac{1}{D^8\, (2\, \exp(2\, (3 + D)\, \|\boldsymbol{\theta}^\star\|_\infty))^4} \quad \geq \quad \frac{C}{D^8\, N^{1-\vartheta}}.$$

Upon rearranging terms, we obtain

$$\left(\frac{N^{1-\vartheta}}{16\, C}\right)^{1/4} \quad \geq \quad \exp(2\,(3+D)\,\|\boldsymbol{\theta}^\star\|_\infty).$$

Upon taking the natural logarithm on both sides, we obtain

$$\frac{(1-\vartheta)\,\log\, N - \log 16\, C}{8\,(3+D)} \quad \geq \quad \|\boldsymbol{\theta}^\star\|_\infty.$$

Let $C \coloneqq \exp(-L)\,/\,16 \in (0,\, 1/16]$ and note that C is independent of N and p, because $L \geq 0$ is independent of N and p. Then, for all $\epsilon^\star \in (0,\, \|\boldsymbol{\theta}^\star\|_\infty]$ and all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$,

$$\frac{v(\boldsymbol{\theta})}{\boldsymbol{c}(\boldsymbol{\theta})^\top\, \boldsymbol{c}(\boldsymbol{\theta})} \quad \geq \quad \frac{C}{D^8\, N^{1-\vartheta}},$$

provided that

$$\|\boldsymbol{\theta}^\star\|_\infty \quad \leq \quad \frac{L + (1-\vartheta)\,\log N}{16\,(3+D)} \quad \leq \quad \frac{L + (1-\vartheta)\,\log N}{8\,(3+D)}.$$

Lemma 10. *Consider Models 2–4. Then, for all* $\epsilon^\star \in (0,\, \|\boldsymbol{\theta}^\star\|_\infty]$ *and all* $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$,

$$\frac{v(\boldsymbol{\theta})}{\boldsymbol{c}(\boldsymbol{\theta})^\top\, \boldsymbol{c}(\boldsymbol{\theta})} \quad \geq \quad \frac{1}{D^8\,(1+\exp((3+D)\,(\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)))^4},$$

where $D \coloneqq \max_{1 \leq i < j \leq N} |\mathfrak{N}_{i,j}| \geq 1$.

Proof of Lemma 10. We have

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

where

$$I_{a,b}(\boldsymbol{X}) \;=\; \mathbb{1}\left(\sum_{h \,\in\, \mathcal{N}_a \,\cap\, \mathcal{N}_b} X_{a,h}\, X_{b,h} \;>\; 0\right), \qquad \{a, b\} \subset \mathcal{N}.$$

Given any pair of nodes $\{i,j\} \subset \mathcal{N}$, notice that each random variable $X_{a,b}\, I_{a,b}(\boldsymbol{X})$ is a monotone non-decreasing function of $X_{i,j}$ and, conditional on $\boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}$, all other edge variables are constant. As a consequence, the FKG inequality implies that

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

for all pairs of nodes $\{a,b\} \subset \mathcal{N}$ and $\{r,t\} \subset \mathcal{N}$. Thus,

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

Using the law of total expectation,

$$\begin{aligned} &\mathbb{E}\, \mathbb{V}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{i,j\}}} \left(X_{i,j}\, I_{a,b}(\boldsymbol{X})\right) \\ = \quad &\mathbb{E}\,(\mathbb{V}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{i,j\}}}\, X_{a,b}\, I_{a,b}(\boldsymbol{X}) \mid I_{a,b}(\boldsymbol{X}) = 1)\, \mathbb{P}(I_{a,b}(\boldsymbol{X}) = 1) \\ + \quad &\mathbb{E}\,(\mathbb{V}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{i,j\}}}\, X_{a,b}\, I_{a,b}(\boldsymbol{X}) \mid I_{a,b}(\boldsymbol{X}) = 0)\, \mathbb{P}(I_{a,b}(\boldsymbol{X}) = 0), \end{aligned}$$

which shows

$$\mathbb{E}\, \mathbb{V}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{i,j\}}} \left(X_{i,j}\, I_{a,b}(\boldsymbol{X})\right) \quad = \quad \mathbb{E}\,(\mathbb{V}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{i,j\}}}\, X_{a,b} \mid I_{a,b}(\boldsymbol{X}) = 1)\, \mathbb{P}(I_{a,b}(\boldsymbol{X}) = 1),$$

owing to the fact that $X_{a,b}\, I_{a,b}(\boldsymbol{X}) = 0$ almost surely conditional on the event $I_{a,b}(\boldsymbol{X}) = 0$. Hence

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

noting that $\mathbb{V}_{\boldsymbol{\theta},\boldsymbol{X}_{-\{i,j\}}}\, X_{a,b} = 0$ for all $\{a,b\} \neq \{i,j\}$ and $\mathbb{P}(I_{i,j}(\boldsymbol{X}) = 1) = 0$ for all $\{i,j\} \subset \mathcal{N}$ satisfying $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset$. We bound

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

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

$$\mathbb{V}_{\boldsymbol{\theta},\boldsymbol{x}_{-\{i,j\}}} X_{i,j} \;\; \geq \;\; \frac{1}{(1+\exp((3+2\,D)\,\|\boldsymbol{\theta}\|_\infty))^2}. \tag{D.9}$$

The lower bound in (D.9) follows from Lemma 12, which shows that, for all $\boldsymbol{x}_{-\{i,j\}} \in \{0,1\}^{\binom{N}{2}-1}$,

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

and

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

where $D \geq 1$. Hence

$$\begin{aligned} v(\boldsymbol{\theta}) \;\; &\geq \;\; \sum_{i<j:\, \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset}^{N} \frac{1}{(1+\exp((3+D)\,\|\boldsymbol{\theta}\|_\infty))^2}\;\, \mathbb{P}(I_{i,j}(\boldsymbol{X}) = 1) \\ &\geq \;\; \sum_{i<j:\, \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset}^{N} \frac{1}{(1+\exp((3+D)\, \max\{\|\boldsymbol{\theta}\|_\infty, \|\boldsymbol{\theta}^\star\|_\infty\}))^4}. \end{aligned}$$

The lower bound on the probability of event $I_{i,j}(\boldsymbol{X}) = 1$ follows from the observation that, for any given node $h \in \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$, the event $X_{i,h}\, X_{j,h} = 1$ implies the event $I_{i,j}(\boldsymbol{X}) = 1$, so

$$\mathbb{P}(I_{i,j}(\boldsymbol{X}) = 1) \;\geq\; \mathbb{P}(X_{i,h}\, X_{j,h} = 1) \;=\; \mathbb{P}(X_{i,h} = 1 \mid X_{j,h} = 1)\;\, \mathbb{P}(X_{j,h} = 1).$$

By using the law of total probability along with the observation that, for all $\boldsymbol{x}_{-\{i,j\}} \in \{0,1\}^{\binom{N}{2}-1}$,

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

one can show that both the conditional and unconditional probabilities are bounded below by $1\,/\,(1+\exp((3+D)\,\|\boldsymbol{\theta}^\star\|_\infty))$. Last, but not least, we consider the number of pairs of nodes $\{i,j\} \subset \{1,\dots,N\}$ for which $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$. There exists, for each node $i \in \mathcal{N}$, a node $h \in \mathcal{N} \setminus \{i\}$ such that $\mathcal{N}_i \cap \mathcal{N}_h \neq \emptyset$, so

$$\begin{aligned} &\sum_{i<j:\, \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset}^{N} \frac{1}{(1+\exp((3+D)\, \max\{\|\boldsymbol{\theta}\|_\infty, \|\boldsymbol{\theta}^\star\|_\infty\}))^2} \\ \geq \;\; &\frac{N}{(1+\exp((3+D)\, \max\{\|\boldsymbol{\theta}\|_\infty, \|\boldsymbol{\theta}^\star\|_\infty\}))^4}, \end{aligned}$$

which implies that

$$v(\boldsymbol{\theta}) \quad \geq \quad \frac{N}{(1 + \exp((3 + D)\, \max\{\|\boldsymbol{\theta}\|_\infty, \|\boldsymbol{\theta}^\star\|_\infty\}))^4}.$$

Next, we proceed to demonstrate that there exists a finite constant $C_1 > 0$ such that

$$\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta}) \quad \leq \quad C_1\, N.$$

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

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

which may be written as

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

noting $s_{N+1}(\boldsymbol{X})$ is not a function of edge variables $X_{t,a}$ when $\mathcal{N}_t \cap \mathcal{N}_a = \emptyset$, and is hence almost surely constant under the conditional distribution of $X_{t,a}$ given $\boldsymbol{X}_{-\{t,a\}} = \boldsymbol{x}_{-\{t,a\}}$. Hence, we have

$$\begin{aligned} & \sum_{a \,\in\, \mathcal{N} \setminus \{t\}:\, \mathcal{N}_a \,\cap\, \mathcal{N}_t \,\neq\, \emptyset} \mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, s_{N+1}(\boldsymbol{X})) \\ \leq \quad & D^2 \max_{a \,\in\, \mathcal{N} \setminus \{t\}:\, \mathcal{N}_a \,\cap\, \mathcal{N}_t \,\neq\, \emptyset} \mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, s_{N+1}(\boldsymbol{X})), \end{aligned}$$

because $\mathcal{N}_a \cap \mathcal{N}_t \neq \emptyset$ if one of the two following conditions is satisfied:

- $a \in \mathcal{N}_t$, the number of which is bounded above by D;
- $a \notin \mathcal{N}_t$, but there exists $h \in \mathcal{N}_t$ such that $h \in \mathcal{N}_a$; the number of such a for each $h \in \mathcal{N}_t$ is bounded above by D. Noting that $|\mathcal{N}_t| \leq D$, the total number of such $a \in \mathcal{N} \setminus \{t\}$ satisfying $\mathcal{N}_a \cap \mathcal{N}_t \neq \emptyset$ is bounded above by D^2.

To bound $\max_{a \in \mathcal{N} \setminus \{t\}:\, \mathcal{N}_a \cap \mathcal{N}_t \neq \emptyset}\ \mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, s_{N+1}(\boldsymbol{X}))$, recall that $s_{N+1}(\boldsymbol{X}) = \sum_{i<j}^{N} X_{i,j}\, I_{i,j}(\boldsymbol{X})$. The number of indicator functions $I_{i,j}(\boldsymbol{X})$ that are functions of $X_{t,a}$ and are not constant conditional on $X_{-\{t,a\}}$ is bounded above by D^2, applying the same argument as above. Hence,

$$\mathbb{E}\, \mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, s_{N+1}(\boldsymbol{X})) \;\;\leq\;\; D^2,$$

because

$$\mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, X_{i,j}\, I_{i,j}(\boldsymbol{X})) \;\;\leq\;\; 1$$

for all $\{i, j\} \subset \{1, \ldots, N\}$, implying

$$\mathbb{C}_{\boldsymbol{\theta}, \boldsymbol{X}_{-\{t,a\}}}(X_{t,a},\, s_{N+1}(\boldsymbol{X})) \;\;\leq\;\; D^4.$$

Thus, $c_t(\boldsymbol{\theta}) \leq D^4$, so $\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta}) \leq D^8\, N$.

Collecting terms shows that

$$\begin{aligned} \frac{v(\boldsymbol{\theta})}{\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta})} \;\;&\geq\;\; \frac{N}{(1 + \exp((3 + D)\, \max\{\|\boldsymbol{\theta}\|_\infty,\, \|\boldsymbol{\theta}^\star\|_\infty\}))^4} \left(\frac{1}{D^8\, N}\right) \\ &=\;\; \frac{1}{D^8\, (1 + \exp((3 + D)\, \max\{\|\boldsymbol{\theta}\|_\infty,\, \|\boldsymbol{\theta}^\star\|_\infty\}))^4}. \end{aligned}$$

Since $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star) \subset \mathbb{R}^{N+1}$, the reverse implies that $\|\boldsymbol{\theta}\|_\infty \leq \|\boldsymbol{\theta}^\star\|_\infty + \epsilon$, which in turn implies that, for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$,

$$\frac{v(\boldsymbol{\theta})}{\boldsymbol{c}(\boldsymbol{\theta})^\top \boldsymbol{c}(\boldsymbol{\theta})} \;\;\geq\;\; \frac{1}{D^8\, (1 + \exp((3 + D)\, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)))^4}.$$

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

D.2.1. *Undirected graphical models of random graphs.* Let $\mathcal{G}(\mathcal{V}, \mathcal{E})$ be an undirected graph with a set of vertices $\mathcal{V}$ and a set of edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. An undirected graphical model of a random graph [32] is a family of probability measures $\{\mathbb{P}_{\boldsymbol{\theta}},\, \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ dominated by a σ-finite measure ν, with densities of the form

$$f_{\boldsymbol{\theta}}\,(\boldsymbol{x}) \;\;\propto\;\; \prod_{\mathcal{C} \in \mathfrak{C}} g_{\mathcal{C}}(\boldsymbol{x}_{\mathcal{C}};\, \boldsymbol{\theta}), \qquad \boldsymbol{x} \in \mathbb{X}, \tag{D.10}$$

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 $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. 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 $\mathcal{G}$. Here, as elsewhere [e.g., 32], 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. It is well-known that the factorization property of probability density function (D.10) implies conditional independence properties, that is, Markov properties.

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

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) \;\propto\; \prod_{i<j}^{N} \varphi_{i,j}(\boldsymbol{x}_{\mathcal{S}_{i,j}}; \boldsymbol{\theta}), \qquad \boldsymbol{x} \in \mathbb{X}, \tag{D.11}$$

where $\mathcal{S}_{i,j} \subset \mathcal{N} \setminus \{i, j\}$ $(i < j = 1, \ldots, N)$. Probability density functions of the form (D.11) can be represented as probability density functions of the form (D.10) 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 graphs of Models 2–4 have edges, representing conditional dependencies induced by brokerage in networks. A graphical representation of the conditional independence graphs of Models 2–4 is shown in Figure 2—note that all three models have the same conditional independence graph.

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

D.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.

Generalized β-models with dependent edges constrain the dependence among edges to the intersections $\mathcal{N}_i \cap \mathcal{N}_j$ of neighborhoods $\mathcal{N}_i$ and $\mathcal{N}_j$ of nodes $i \in \mathcal{N}$ and $j \in \mathcal{N}$, and hence possess the following property:

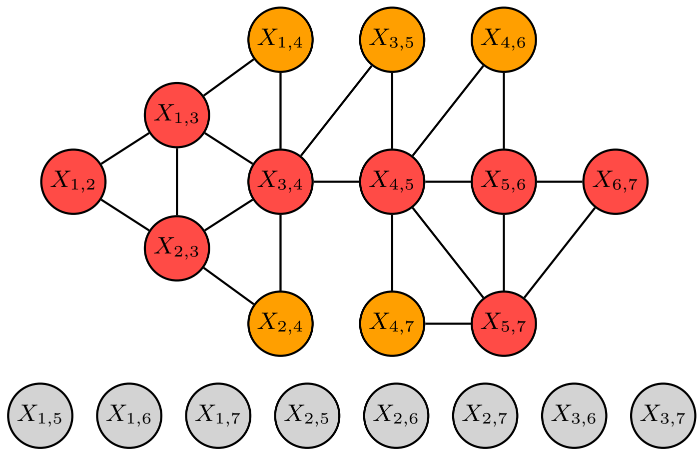

FIG 2. *The conditional independence graph of Models 2–4 with population of nodes* $\mathcal{N} = \{1, \ldots, 7\}$ *consisting of overlapping subpopulations* $\mathcal{A}_1 = \{1, 2, 3\}$, $\mathcal{A}_2 = \{3, 4\}$, $\mathcal{A}_3 = \{4, 5\}$, *and* $\mathcal{A}_4 = \{5, 6, 7\}$. *If nodes i and j belong to the same subpopulation, edge variable* $X_{i,j}$ *is colored red. If nodes i and j do not belong to the same subpopulation, edge variable* $X_{i,j}$ *is colored orange if the subpopulations of i and j overlap and is colored gray otherwise.*

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

The neighborhood intersection property implies conditional independence properties, including—but not limited to—the following.

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* $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset$:

$$X_{i,j} \perp\!\!\!\perp \boldsymbol{X} \setminus X_{i,j}.$$

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* $\{i, j\} \subset \mathcal{A}_k$:

$$X_{i,j} \quad \perp\!\!\!\perp \quad \boldsymbol{X} \setminus (X_{i,j}, \boldsymbol{X}_{\{i,j\}, \mathcal{N}_i \cup \mathcal{N}_j}) \mid \boldsymbol{X}_{\{i,j\}, \mathcal{N}_i \cup \mathcal{N}_j} \setminus X_{i,j}$$

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* $\{i, j\} \subset \mathcal{A}_k$:

$$X_{i,j} \quad \perp\!\!\!\perp \quad \boldsymbol{X} \setminus (X_{i,j}, \boldsymbol{X}_{\{i,j\}, \mathcal{N}_i \cap \mathcal{N}_j}) \mid \boldsymbol{X}_{\{i,j\}, \mathcal{N}_i \cap \mathcal{N}_j}$$

PROOF OF PROPOSITION 2. In the following, we use the characterizations of conditional independence due to Dawid [18], which relate factorization properties of probability density functions to conditional independence properties. Using these characterizations of conditional independence, we establish the conditional independence properties stated in Proposition 2 by showing that, for each pair of nodes $\{i, j\} \subset \mathcal{N}$, there exists a subset of edge indices $\mathfrak{S} \subset \{\{i, j\} : i \in \mathcal{N}, j \in \mathcal{N}\} \setminus \{i, j\}$ and non-negative functions g and h such that the probability density function can be written as

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

implying that

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

understanding $\boldsymbol{x} \setminus (x_{i,j},\, \boldsymbol{x}_{\mathfrak{S}})$ to be all edge variables in $\boldsymbol{x}$ except those edge variables contained in $(x_{i,j},\, \boldsymbol{x}_{\mathfrak{S}})$. Proposition 2 assumes that the neighborhood intersection property is satisfied, allowing us to write

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) \quad \propto \quad \prod_{a\,<\,b}^{N} \varphi_{a,b}(x_{a,b},\, \boldsymbol{x}_{\{a,b\},\, \mathcal{N}_a \cap \mathcal{N}_b}).$$

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 $\{i, j\} \times \{\mathcal{N}_i \cap \mathcal{N}_j\} = \emptyset$,

$$\varphi_{i,j}(x_{i,j},\, \boldsymbol{x}_{\{i,j\},\, \mathcal{N}_i \cap \mathcal{N}_j}) \quad = \quad \varphi_{i,j}(x_{i,j}).$$

It remains to check whether any $\varphi_{a,b}$ ($\{a, b\} \neq \{i, j\}$) can be a function of $x_{i,j}$, which would require that

$$(i, j) \in \{a, b\} \times \{\mathcal{N}_a \cap \mathcal{N}_b\} \quad \text{or} \quad (j, i) \in \{a, b\} \times \{\mathcal{N}_a \cap \mathcal{N}_b\}.$$

We prove by contradiction that $\varphi_{a,b}$ ($\{a, b\} \neq \{i, j\}$) cannot be a function of $x_{i,j}$. Consider any $a \in \mathcal{N} \setminus \{i, j\}$. For $\varphi_{a,i}$ to be a function of $x_{i,j}$, it must be that $(i, j) \in \{a, i\} \times \{\mathcal{N}_a \cap \mathcal{N}_i\}$, as $(j, i) \in \{a, i\} \times \{\mathcal{N}_a \cap \mathcal{N}_i\}$ is not possible because $\{a, b\} \neq \{i, j\}$ so that $a \neq j$, which would require that $j \in \mathcal{N}_a \cap \mathcal{N}_i$, implying $j \in \mathcal{N}_i$. Recall the definition of $\mathcal{N}_i$:

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

Hence, $j \in \mathcal{N}_i$ is possible if and only if there exists $k \in \{1, \ldots, K\}$ such that $\{i, j\} \subset \mathcal{A}_k$. As it is assumed that $|\mathcal{A}_k| \geq 3$ for all $k \in \{1, \ldots, K\}$, there would exist $h \in \mathcal{A}_k \setminus \{i,j\}$ such that $h \in \mathcal{N}_i \cap \mathcal{N}_j$, violating the assumption that $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset$. 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

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

and

$$h(\boldsymbol{x} \setminus x_{i,j}) \quad = \prod_{a<b\colon \{a,b\} \neq \{i,j\}}^{N} \varphi_{a,b}(x_{a,b},\, \boldsymbol{x}_{\{a,b\},\, \mathcal{N}_a \cap \mathcal{N}_b})$$

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

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) \quad \propto \quad g(x_{i,j})\, h(\boldsymbol{x} \setminus x_{i,j}),$$

which implies

$$X_{i,j} \perp\!\!\!\perp \boldsymbol{X} \setminus X_{i,j}.$$

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 $k \in \{1,\ldots,K\}$ such that $\{i,j\} \subset \mathcal{A}_k$. By definition, $\varphi_{i,j}$ is a function of $x_{i,j}$. For any $\varphi_{a,b}$ ($\{a,b\} \neq \{i,j\}$) to be a function of $x_{i,j}$, we must have either

$$(i,j) \in \{a,b\} \times \{\mathcal{N}_a \cap \mathcal{N}_b\} \quad \text{or} \quad (j,i) \in \{a,b\} \times \{\mathcal{N}_a \cap \mathcal{N}_b\},$$

which requires that two conditions hold:

1. $\{a,b\} \cap \{i,j\} \neq \emptyset$;
2. Either $i \in \mathcal{N}_a \cap \mathcal{N}_b$ or $j \in \mathcal{N}_a \cap \mathcal{N}_b$.

From the first condition, either $\{a,b\} = \{i,b\}$ or $\{a,b\} = \{a,j\}$. Without loss, we take $a = i$ so that we take $\{a,b\} = \{i,b\}$. From the second condition, this implies $j \in \mathcal{N}_i \cap \mathcal{N}_b$. By assumption, there exists $k \in \{1,\ldots,K\}$ such that $\{i,j\} \subset \mathcal{A}_k$, implying $j \in \mathcal{N}_i$. For $j \in \mathcal{N}_b$ we would have $b \in \mathcal{N}_j$. Let $c = \{a,b\} \setminus \{i,j\}$ and $d = \{a,b\} \cap \{i,j\}$ and assume, without loss, that $c < d$. Then the subset of pairs of nodes $\{c,d\}$ such that $\varphi_{c,d}$ is a function of $x_{i,j}$ is the subset of all nodes c such that either $c \in \mathcal{N}_i$ or $c \in \mathcal{N}_j$ or both, so c must satisfy $c \in \mathcal{N}_i \cup \mathcal{N}_j \setminus \{i,j\}$. Therefore, there exist non-negative functions g and h such that the probability density function can be written as follows:

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) \quad \propto \quad g(\boldsymbol{x}_{\{i,j\},\, \mathcal{N}_i \cup \mathcal{N}_j})\, h(\boldsymbol{x} \setminus \boldsymbol{x}_{\{i,j\},\, \mathcal{N}_i \cup \mathcal{N}_j}),$$

which implies

$$X_{i,j} \quad \perp\!\!\!\perp \quad \boldsymbol{X} \setminus (X_{i,j}, \boldsymbol{X}_{\{i,j\},\, \mathcal{N}_i \cup \mathcal{N}_j}) \mid \boldsymbol{X}_{\{i,j\},\, \mathcal{N}_i \cup \mathcal{N}_j} \setminus X_{i,j}$$

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 $\{i,j\} \subset \mathcal{A}_k$. Since i and j do not share any subpopulation, $\varphi_{a,b}$ can be a function of $x_{i,j}$ if and only if $\{a,b\} \cap \{i,j\} \neq \emptyset$ and $c = \{a,b\} \setminus \{i,j\}$ is contained in $\mathcal{N}_i \cap \mathcal{N}_j$. This can be seen from the proof of Condition 2, as $i \notin \mathcal{N}_j$ and $j \notin \mathcal{N}_i$, because by assumption, $\{i,j\} \not\subset \mathcal{A}_k$ for all $k \in \{1,\ldots,K\}$. As a result, there exists no $\{a,b\} \neq \{i,j\}$ such that

$$(i,j) \in \{a,b\} \times \{\mathcal{N}_a \cap \mathcal{N}_b\} \quad \text{or} \quad (j,i) \in \{a,b\} \times \{\mathcal{N}_a \cap \mathcal{N}_b\}.$$

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

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) \quad \propto \quad g(x_{i,j},\, \boldsymbol{x}_{\{i,j\},\,\mathcal{N}_i \cap \mathcal{N}_j})\; h(\boldsymbol{x} \setminus\, (x_{i,j},\, \boldsymbol{x}_{\{i,j\},\,\mathcal{N}_i \cap \mathcal{N}_j})),$$

which implies

$$X_{i,j} \quad \perp\!\!\!\perp \quad \boldsymbol{X} \setminus (X_{i,j},\, \boldsymbol{X}_{\{i,j\},\,\mathcal{N}_i \cap \mathcal{N}_j}) \mid \boldsymbol{X}_{\{i,j\},\,\mathcal{N}_i \cap \mathcal{N}_j}.$$

D.2.3. *Bounding the spectral norm of the coupling matrix.* We bound the spectral norm $|||\mathcal{D}|||_2$ of the coupling matrix $\mathcal{D}$. Throughout, we adopt the notation used in Section 3 of the manuscript for indexing edge variables in $\boldsymbol{X}$, that is, we denote the number of edge variables by $M = \binom{N}{2}$ and edge variables by $X_1, \ldots, X_M$, where N is the number of nodes. The edge variables $X_1, \ldots, X_M$ form the vertices of the conditional independence graph. Recall that the edge variables $X_1, \ldots, X_m$ form the vertex set of the conditional independence graph $\mathcal{G}$ for $\boldsymbol{X}$. We proceed to bound $|||\mathcal{D}|||_2$ under Models 2–4, using Assumption A from Section 3.4.

Lemma 11. *Consider Models 2–4. Assume that Assumption A is satisfied.*

1. *If the subpopulations do not intersect ($\omega_1 = \omega_2 = 0$) and condition (3.7) is satisfied with $\vartheta \in (4/5,\, 1]$, then*

$$|||\mathcal{D}|||_2 \quad \leq \quad 1 + D.$$

2. *If the subpopulations do intersect and condition (3.7) is satisfied with $\vartheta = 1$, then there exist finite constants $C_1 > 0$ and $C_2 > 0$, independent of N and p, such that*

$$|||\mathcal{D}|||_2 \quad \leq \quad C_1\, D^2\, \exp(C_2\, D^2).$$

PROOF OF LEMMA 11. We adapt the coupling approach of van den Berg and Maes [48] 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 $\mathcal{G}$ and consider any $\boldsymbol{x}_{1:i-1} \in \{0,1\}^{i-1}$. Define

$$\mathbb{P}_{\boldsymbol{x}_{1:i-1},0}(\boldsymbol{X}_{i+1:M} = \boldsymbol{a}) \;\; := \;\; \mathbb{P}(\boldsymbol{X}_{i+1:M} = \boldsymbol{a} \mid \boldsymbol{X}_{1:i-1} = \boldsymbol{x}_{1:i-1}, X_i = 0)$$

and

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

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 $\mathcal{D}$.

III. Bounding the spectral norm $|||\mathcal{D}|||_2$ of the coupling matrix $\mathcal{D}$.

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 $(\boldsymbol{X}^\star, \boldsymbol{X}^{\star\star}) \in \{0,1\}^{M-i} \times \{0,1\}^{M-i}$ of the conditional 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 [35].

To simplify the notation, we assume that the couple $(\boldsymbol{X}^\star, \boldsymbol{X}^{\star\star})$ takes on values in $\{0,1\}^M \times \{0,1\}^M$ rather than $\{0,1\}^{M-i} \times \{0,1\}^{M-i}$ and 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 one. 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 then construct a coupling of the conditional distributions $\mathbb{P}_{i,\boldsymbol{x}_{1:i-1},0}$ and $\mathbb{P}_{i,\boldsymbol{x}_{1:i-1},1}$ as follows:

1. Initialize $\mathfrak{V} = \{1, \ldots, i\}$.

2. Check whether there exists a vertex $j \in \mathcal{V} \setminus \mathfrak{V}$ that is connected to a vertex $v \in \mathfrak{V}$ in $\mathcal{G}$ and satisfies $X^\star_v \neq X^{\star\star}_v$:

   (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 $(X^\star_j,\, X^{\star\star}_j)$ be distributed according to an optimal coupling of the distributions $\mathbb{P}(X_j = \cdot \mid \boldsymbol{X}_{\mathfrak{V}} = \boldsymbol{x}_{\mathfrak{V}}^\star)$ and $\mathbb{P}(X_j = \cdot \mid \boldsymbol{X}_{\mathfrak{V}} = \boldsymbol{x}_{\mathfrak{V}}^{\star\star})$. The optimal coupling ensures that $X_j^\star = X_j^{\star\star}$ with probability 1, so the total variation distance is 0.

In both steps, an optimal coupling exists [35, Theorem 5.2, p. 19], but may not be unique. However, 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 14 verifies that the algorithm above 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 $\mathbb{P}_{i,\boldsymbol{x}_{1:i-1},1}$ and $\mathbb{P}_{i,\boldsymbol{x}_{1:i-1},0}$, respectively.

By construction, the coupling has useful properties. 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 $i \not\leftrightarrow j$ to be the event that there exists a path from i to j in $\mathcal{G}$ such that $X_v^\star \neq X_v^{\star\star}$ for all vertices v along the path. Such paths are known as *paths of disagreement* in the probability literature on Gibbs measures and Markov random fields. Theorem 1 of van den Berg and Maes [48, p. 753] shows that

$$\text{(D.12)} \qquad \mathbb{Q}_{i,\boldsymbol{x}_{1:i-1}}(X_j^\star \neq X_j^{\star\star}) \;=\; \mathbb{Q}_{i,\boldsymbol{x}_{1:i-1}}(i \not\leftrightarrow j) \;\leq\; \mathbb{B}_{\boldsymbol{\pi}}(i \not\leftrightarrow j),$$

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

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

where

$$\pi_{v,\, \boldsymbol{x}_{-v},\, \boldsymbol{x}'_{-v}} \;=\; \|\mathbb{P}(\,\cdot \mid \boldsymbol{X}_{-v} = \boldsymbol{x}_{-v}) - \mathbb{P}(\,\cdot \mid \boldsymbol{X}'_{-v} = \boldsymbol{x}'_{-v})\|_{\text{TV}}.$$

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,* corresponding to the event that $X_v^\star \neq X_v^{\star\star}$ 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 one, and vertex i is open with probability one.

The coupling argument of van den Berg and Maes [48] 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 $\mathcal{D}_{i,j}$ of the coupling matrix $\mathcal{D}$ by

$$\mathcal{D}_{i,j} \;\; = \;\; \sup_{\boldsymbol{x}_{1:i-1} \in \{0,1\}^{i-1}} \mathbb{Q}_{i,\boldsymbol{x}_{1:i-1}}(X_j^\star \neq X_j^{\star\star}) \;\; \leq \;\; \mathbb{B}_{\boldsymbol{\pi}}(i \not\leftrightarrow j). \tag{D.13}$$

By construction of $\mathcal{D}$, the below-diagonal and diagonal elements of $\mathcal{D}$ are known to be 0 and 1, respectively. We define $\pi^\star \in (0,1)$ by

$$\pi^\star \;\; = \;\; \max_{1 \leq v \leq M} \; \max_{\boldsymbol{x}_{-v} \in \{0,1\}^{M-1}} \mathbb{P}(X_v = 1 \mid \boldsymbol{X}_{-v} = \boldsymbol{x}_{-v}),$$

and note that Lemma 15, together with the assumption that $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta}_N = \mathbb{R}^p$ satisfies (3.7) with $\vartheta = 1$, implies that

$$\pi^\star \;\; \leq \;\; \frac{1}{1 + \exp(-L \,/\, 16)} \;\; < \;\; 1,$$

where $L > 0$ is the same constant as in (3.7), independent of N and p. In the following, we assume that $\vartheta = 1$. We consider the case when $\vartheta < 1$, relevant to the first result stated in Lemma 11, at the end of the proof. Define

$$\xi \;\; \coloneqq \;\; \frac{1}{1 + \exp(-L \,/\, 16)},$$

noting that pp is independent of N and p, because $L > 0$ is independent of N and p, and define the vector $\boldsymbol{\xi} \in [0,1]^M$ by

$$(\boldsymbol{\xi})_i \;\; \coloneqq \;\; \begin{cases} 0 & \text{if } v \in \{1, \ldots, i-1\} \\ 1 & \text{if } v = i \\ \xi & \text{if } v \in \{i+1, \ldots, M\} \end{cases}.$$

Observe that the probability $\mathbb{B}_{\boldsymbol{\pi}}(i \not\leftrightarrow j)$ is non-decreasing in the components of $\boldsymbol{\pi}$. Hence,

$$\mathbb{B}_{\boldsymbol{\pi}}(i \not\leftrightarrow j) \;\; \leq \;\; \mathbb{B}_{\boldsymbol{\xi}}(i \not\leftrightarrow j).$$

We bound the elements of $\mathcal{D}$ by bounding the probability $\mathbb{B}_{\boldsymbol{\xi}}(i \not\leftrightarrow j)$.

II. Bounding the elements of the coupling matrix $\mathcal{D}$. To bound the elements $\mathcal{D}_{i,j}$ of the coupling matrix $\mathcal{D}$, we bound terms $\mathbb{B}_{\boldsymbol{\xi}}(i \not\leftrightarrow j)$ on the right-hand side of (D.13) using Assumption A. To do so, define

$$\mathcal{S}_{i,k} \;\; = \;\; \{v \in \mathcal{V} \setminus \{i\} \,:\, d_{\mathcal{G}}(i, v) = k\}\,, \qquad k = 1, \ldots, M-1,$$

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 $\mathcal{G}$. Thus, the set $\mathcal{S}_{i,k} \subseteq \mathcal{V}$ is the subset of vertices in the conditional independence graph $\mathcal{G}$ at graph distance k from i in $\mathcal{G}$. We bound probabilities $\mathbb{B}_{\boldsymbol{\xi}}(i \not\leftrightarrow j)$ by placing restrictions on the subpopulation structure, which determines which edges are present in $\mathcal{G}$, influencing probabilities $\mathbb{B}_{\boldsymbol{\xi}}(i \not\leftrightarrow j)$. To do so, we define the subpopulation graph $\mathcal{G}_{\mathcal{A}}$ to be a graph with set of vertices $\{\mathcal{A}_1, \ldots, \mathcal{A}_K\}$ and edges between vertices $\mathcal{A}_r$ and $\mathcal{A}_l$ if $\mathcal{A}_r \cap \mathcal{A}_l \neq \emptyset$. To measure distances between subpopulations, we define $d_{\mathcal{G}_{\mathcal{A}}} : \{\mathcal{A}_1, \ldots, \mathcal{A}_K\}^2 \mapsto \{0, 1, 2, \ldots\} \cup \{\infty\}$ to be the length of the shortest path between two vertices $\mathcal{A}_r$ and $\mathcal{A}_l$ in $\mathcal{G}_{\mathcal{A}}$, adopting that usual convention that $d_{\mathcal{G}_{\mathcal{A}}}(\mathcal{A}_r, \mathcal{A}_l) = \infty$ if there exists no path connecting $\mathcal{A}_r$ to $\mathcal{A}_l$ in $\mathcal{G}_{\mathcal{A}}$. Define

$$\mathcal{V}_{\mathcal{A}_r,k} \;\; = \;\; \{\mathcal{A}_l \in \{\mathcal{A}_1, \ldots, \mathcal{A}_K\} \,:\, d_{\mathcal{G}_{\mathcal{A}}}(\mathcal{A}_r, \mathcal{A}_l) = k\}\,,$$

for all $\mathcal{A}_r \in \{\mathcal{A}_1, \ldots, \mathcal{A}_K\}$ and $k \geq 1$. Let $g : \{1, 2, \ldots\} \mapsto \mathbb{R}^+$ be such that

$$|\mathcal{V}_{\mathcal{A}_r,k}| \;\; \leq \;\; g(k), \qquad k \in \{1, 2, \ldots\}.$$

In words, $g(k)$ bounds the number of subpopulations at graph distance k from any given subpopulation in $\mathcal{G}_{\mathcal{A}}$. It is straightforward to verify that Models 2–4 satisfy Definition 1 and thus posses the neighborhood intersection property, satisfying the assumptions of Proposition 2. By Proposition 2, the Markov blanket of any edge X_i between nodes $\{a, b\} \subset \mathcal{N}$ is not larger than the subset of edges contained in the subgraph $\boldsymbol{X}_{\{a,b\},\, \mathcal{N}_a \cup \mathcal{N}_b}$, which contains edges between nodes in the set $\{a, b\}$ and nodes in the set $\mathcal{N}_a \cup \mathcal{N}_b$. 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 as $\mathcal{G}$ and the same set of edges.

2. For each vertex X_i in $\mathcal{G}$ corresponding to an edge between a pair of nodes $\{a, b\} \subset \mathcal{N}$ with degree greater than 0 in $\mathcal{G}$, add edges between X_i and any other edge X_j contained in the subgraph $\boldsymbol{X}_{\{a,b\},\, \mathcal{N}_a \cup \mathcal{N}_b}$ which are not already in $\mathcal{G}$.

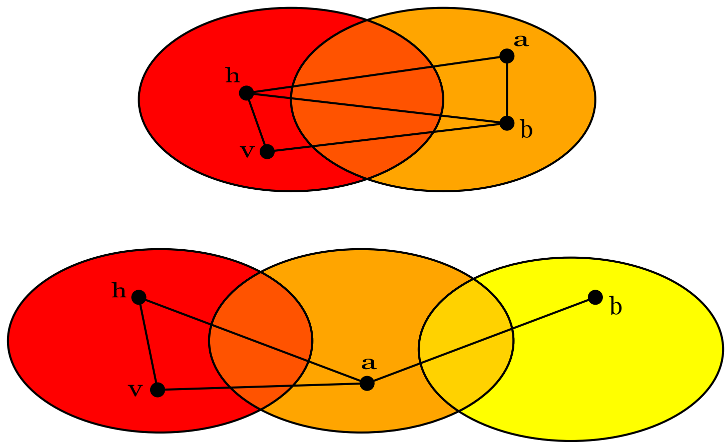

FIG 3. *Demonstrating how vertices X_i and X_j of the cover $\mathcal{G}^\star$ of conditional independence graph $\mathcal{G}$ can have graph distance $d_{\mathcal{G}^\star}(i,\, j) = 2$ in $\mathcal{G}^\star$. Suppose that the indicator X_j of an edge between nodes v and h depends on the indicator X_i of an edge between nodes a and b through the indicator X_l of an edge between nodes a and h. By construction, v either belongs to a subpopulation that intersects with one of the subpopulations of b, or a act as a broker for v and b.*

The construction of $\mathcal{G}^\star$ ensures that the neighborhood of any given vertex X_i in $\mathcal{G}^\star$ corresponding to the edge between pair of nodes $\{a, b\} \subset \mathcal{N}$ is either empty or is equal to the set of vertices corresponding to edges contained in the subgraph $\boldsymbol{X}_{\{a,b\},\, \mathcal{N}_a \cup \mathcal{N}_b}$. Moreover, the fact that $\mathcal{G} \subseteq \mathcal{G}^\star$ implies

$$\mathbb{B}_{\boldsymbol{\xi}}(i \not\leftrightarrow j \text{ in } \mathcal{G}) \quad \leq \quad \mathbb{B}_{\boldsymbol{\xi}}(i \not\leftrightarrow j \text{ in } \mathcal{G}^\star).$$

In words, the probability of the existence of a path of disagreement between any two vertices in a graph does not decrease through the addition of edges in the graph. Henceforth, the event $i \not\leftrightarrow j$ represents a path of disagreement in the graph covering $\mathcal{G}^\star$ of $\mathcal{G}$.

We bound each $|\mathcal{S}_{i,k}|$ ($k \geq 1$) for arbitrary $i \in \mathcal{V}$ in turn:

- Consider any $j \in \mathcal{S}_{i,1}$. The neighborhood of edge X_i between nodes $\{a, b\} \subset \mathcal{N}$ in $\mathcal{G}^\star$ is equal to the set of edges contained in the subgraph $\boldsymbol{X}_{\{a,b\},\, \mathcal{N}_a \cup \mathcal{N}_b}$, which contains edges between nodes in the set $\{a, b\}$ and nodes in the set $\mathcal{N}_a \cup \mathcal{N}_b$, the number of which is bounded above by $D = 2 \max_{\{v,w\} \subset \mathcal{N}} |\mathcal{N}_v \cup \mathcal{N}_w|$. Hence, $|\mathcal{S}_{i,1}| \leq D$.
- Consider any $j \in \mathcal{S}_{i,2}$. Then, the shortest path between edge X_i and X_j in $\mathcal{G}^\star$ is of length 2, implying the following facts:

  (F.1) X_j is not in the neighborhood of X_i in $\mathcal{G}^\star$.

(F.2) In $\mathcal{G}^\star$, there is at least one edge X_l in the neighborhood of X_i such that X_j is in the neighborhood of X_l.

Let X_i be the edge between pair of nodes $\{a, b\} \subset \mathcal{N}$ and consider the graph $\mathcal{G}^\star$. Then, (F.2) implies there exists $h \in (\mathcal{N}_a \cup \mathcal{N}_b) \setminus \{a, b\}$ such that X_l is an edge between either pair of nodes $\{a, h\}$ or $\{b, h\}$, which without loss, we take to be $\{a, h\}$. Next, (F.2) states X_j lies in the neighborhood of X_l, implying there exists $v \in \mathcal{N}_a \cup \mathcal{N}_h \setminus \{a, b, h\}$ such that X_j is an edge between pair of nodes $\{v, h\}$ or $\{v, a\}$. It must be $\{v, h\}$, because (F.1) implies that $v \notin \mathcal{N}_a \cup \mathcal{N}_b$, implying $v \notin \mathcal{N}_a$, because if $v \in \mathcal{N}_a \cup \mathcal{N}_b$, either X_j would be in the neighborhood of X_i, by the construction of $\mathcal{G}^\star$, implying $d_{\mathcal{G}^\star}(i, j) = 1$, or X_j would be independent of all other edges and thus would have infinite graph distance to all other vertices in $\mathcal{G}$ and thus $\mathcal{G}^\star$, in which case $j \notin \mathcal{S}_{i,2}$. Using these facts, we bound the number of such $v \in \mathcal{N}_h \setminus (\mathcal{N}_a \cup \mathcal{N}_b)$:

- The total number of subpopulations $\mathcal{A}_r$ to which a and b can belong is bounded above by $2\,(g(1) + 1) \leq 4\,g(1)$, as $g(1)$ bounds the number of subpopulations to which any subpopulation can overlap.
- As $v \in \mathcal{N}_h \setminus (\mathcal{N}_a \cup \mathcal{N}_b)$, there exists $\mathcal{A}_r \in \{\mathcal{A}_1, \ldots, \mathcal{A}_K\}$ such that $\{v, h\} \subset \mathcal{A}_r$, and $a \notin \mathcal{A}_r$ and $b \notin \mathcal{A}_r$, because $v \notin \mathcal{N}_a \cup \mathcal{N}_b$. Moreover, there exists a subpopulation $\mathcal{A}_l \in \{\mathcal{A}_1, \ldots, \mathcal{A}_K\}$ such that $\{h, a\} \subset \mathcal{A}_l$ and/or $\{h, b\} \subset \mathcal{A}_l$, implying that v lies in a subpopulation $\mathcal{A}_r$ such that either $d_{\mathcal{G}_\mathcal{A}}(\mathcal{A}_r, \mathcal{A}_l) = 1$ or $d_{\mathcal{G}_\mathcal{A}}(\mathcal{A}_r, \mathcal{A}_l) = 2$. An illustration of this is shown in Figure 3. Thus, the number of such subpopulations $\mathcal{A}_r$ is not greater than $4\,g(1)\,(g(1) + g(2))$.
- Lastly, each of the no more than $g(1) + g(2)$ subpopulations contain no more than $D = \max_{\{v,w\} \subset \mathcal{N}} |\mathcal{N}_v \cup \mathcal{N}_w|$ nodes.

Taken together,

$$|\mathcal{S}_{i,2}| \quad \leq \quad 4\,D\,g(1)\,(g(1) + g(2))$$

- Consider any $k \geq 3$ and $j \in \mathcal{S}_{i,k}$. We can repeat the previous argument for any $k \geq 3$ to show

$$|\mathcal{S}_{i,k}| \quad \leq \quad 4\,D\,g(1)\,(g(k-1) + g(k)),$$

proving the result by induction.

As $g(1) \leq D$, we may take the bound

$$|\mathcal{S}_{i,k}| \quad \leq \quad 4\,D^2\,(g(k-1) + g(k)), \qquad k \geq 2.$$

We proceed with bounding $\mathcal{D}_{i,j}$ under Assumption A. Define the function $g : \{1, 2, \dots\} \mapsto \mathbb{R}^+$ by

$$g(k) \quad = \quad \omega_1 + \frac{\omega_2}{8D^2} \log k, \qquad k \in \{1, 2, \dots\},$$

where $\omega_1 \geq 0$ and $\omega_2 \geq 0$, and $\omega_2 \geq 0$ additionally satisfies

$$\omega_2 \quad < \quad \frac{1}{|\log(1 - \xi)|},$$

as $\omega_2 \in [0, |\log(1 - \xi)|^{-1})$ is independent of N and p, by Assumption A. Hence,

$$\mathbb{B}_{\boldsymbol{\xi}}(v \text{ is open in } \mathcal{G}^\star) \quad \leq \quad \xi \quad < \quad 1, \qquad v \in \{i + 1, \dots, M\}.$$

Assumption A assumes that, for all subpopulations $\mathcal{A}_r$,

$$|\mathcal{V}_{\mathcal{A}_r,k}| \quad \leq \quad g(k), \qquad k \in \{1, 2, \dots\},$$

implying that, for all $i \in \mathcal{V}$ and all $k \in \{2, 3, \dots\}$,

$$|\mathcal{S}_{i,k}| \quad \leq \quad 4\, D^2\, (g(k-1) + g(k)) \quad \leq \quad 8\, D^2\, g(k) \quad \leq \quad 8\, D^2\, \omega_1 + \omega_2 \log k,$$

using the above definition of $g(k)$ and the bound $g(k - 1) + g(k) \leq 2\, g(k)$, where the bound follows from the monotonicity of $g(k)$. For there to be a path of disagreement $i \not\leftrightarrow j$ in $\mathcal{G}^\star$, 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; note that i is open with probability 1. The probability that there exists at least one open vertex $v \in \mathcal{S}_{i,k}$ is bounded above by

$$1 - (1 - \xi)^{|\mathcal{S}_{i,k}|} \quad \leq \quad 1 - (1 - \xi)^{8D^2\omega_1 + \omega_2 \log k}, \qquad k \in \{1, 2, \dots\}.$$

Since the events that vertices are open are independent under the Bernoulli product measure $\mathbb{B}_{\boldsymbol{\xi}}$, we obtain

$$\begin{aligned}
\mathbb{B}_{\boldsymbol{\xi}}(i \not\leftrightarrow j) \quad &\leq \quad \xi \prod_{l=1}^{k-1} \left[1 - (1 - \xi)^{8D^2\omega_1 + \omega_2 \log l}\right] \\
&\leq \quad \left[1 - (1 - \xi)^{8D^2\omega_1 + \omega_2 \log k}\right]^k,
\end{aligned}$$

using the inequality $\log l \leq \log k$ provided $l \leq k$, along with the fact that $\xi \leq 1-(1-\xi)^{8D^2\omega_1+\omega_2 \log k}$. We then write

$$\begin{aligned} 1-(1-\xi)^{8D^2\omega_1+\omega_2 \log k} &\leq \exp(-(1-\xi)^{8D^2\omega_1+\omega_2 \log k}) \\ &= \exp(-\exp((8D^2\omega_1+\omega_2 \log(k)) \log(1-\xi))) \\ &= \exp(-\exp(-(8D^2\omega_1+\omega_2 \log(k))\,|\log(1-\xi)|)) \\ &= \exp(-\exp(-8D^2\omega_1\,|\log(1-\xi)|)\,k^{-\omega_2\,|\log(1-\xi)|}), \end{aligned}$$

using the inequality $1-z \leq \exp(-z)$ $(z \in (0,1))$. Let

$$A = \exp(-8\,D^2\,\omega_1\,|\log(1-\xi)|) \in (0,1).$$

Then

$$\begin{aligned} \left[1-(1-\xi)^{8D^2\omega_1+\omega_2 \log k}\right]^k &\leq \left[\exp(-A\,k^{-\omega_2\,|\log(1-\xi)|})\right]^k \\ &= \exp(-A\,k^{1-\omega_2\,|\log(1-\xi)|}). \end{aligned}$$

The probability of the event that vertex j is open is bounded above by $\xi \in (0,1)$, so

$$\mathbb{B}_{\boldsymbol{\pi}}(i \not\leftrightarrow j) \leq \mathbb{B}_{\boldsymbol{\xi}}(i \not\leftrightarrow j) \leq \xi \exp(-A\,k^{1-\omega_2\,|\log(1-\xi)|}).$$

We hence obtain

$$\mathcal{D}_{i,j} \leq \exp(-A\;k^{1-\omega_2\,|\log(1-\xi)|}). \tag{D.14}$$

III. Bounding the spectral norm $|||\mathcal{D}|||_2$ of the coupling matrix $\mathcal{D}$. To bound the spectral norm $|||\mathcal{D}|||_2$ of the coupling matrix $\mathcal{D}$, we first use Hölder's inequality to obtain

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

and then bound the elements $\mathcal{D}_{i,j}$ of $\mathcal{D}$. To facilitate bounds on $|||\mathcal{D}|||_2$, we form a symmetric $M \times M$ matrix $\mathcal{T}$ by defining

$$\mathcal{T} = \mathcal{D} + \mathcal{D}^\top - \operatorname{diag}(\mathcal{D}),$$

where $\mathcal{D}^\top$ is the $M \times M$ transpose of $\mathcal{D}$ and $\text{diag}(\mathcal{D})$ is the $M \times M$ diagonal matrix with elements $\mathcal{D}_{1,1}, \dots, \mathcal{D}_{M,M}$ on the main diagonal. By construction of $\mathcal{T}$, the elements $\mathcal{T}_{i,j}$ of $\mathcal{T}$ are given by

$$\mathcal{T}_{i,j} \quad = \quad \begin{cases} \mathcal{D}_{i,j} & \text{if } i < j \\ \mathcal{D}_{i,i} & \text{if } i = j \\ \mathcal{D}_{j,i} & \text{if } j < i \end{cases},$$

where $\mathcal{D}_{i,i} = 1$ by definition of $\mathcal{D}_{i,i}$ $(i = 1, \dots, M)$. Using the fact that $\mathcal{T}_{i,j} = \max(\mathcal{D}_{i,j},\, \mathcal{D}_{j,i})$ $(i, j = 1, \dots, M)$, we obtain

$$|||\mathcal{D}|||_1 \quad = \quad \max_{1 \leq j \leq M} \sum_{i=1}^{M} |\mathcal{D}_{i,j}| \quad \leq \quad \max_{1 \leq j \leq M} \sum_{i=1}^{M} |\mathcal{T}_{i,j}| \quad = \quad |||\mathcal{T}|||_1$$

and

$$|||\mathcal{D}|||_\infty \quad = \quad \max_{1 \leq i \leq M} \sum_{j=1}^{M} |\mathcal{D}_{i,j}| \quad \leq \quad \max_{1 \leq i \leq M} \sum_{j=1}^{M} |\mathcal{T}_{i,j}| \quad = \quad |||\mathcal{T}|||_\infty.$$

In addition, we know that $\mathcal{T}_{i,j} = \mathcal{T}_{j,i}$ $(i, j = 1, \dots, M)$, which implies that

$$|||\mathcal{T}|||_1 \quad = \quad |||\mathcal{T}^\top|||_\infty \quad = \quad |||\mathcal{T}|||_\infty.$$

As a consequence, we obtain

$$|||\mathcal{D}|||_2 \quad \leq \quad \sqrt{|||\mathcal{D}|||_1\, |||\mathcal{D}|||_\infty} \quad \leq \quad \sqrt{|||\mathcal{T}|||_1\, |||\mathcal{T}|||_\infty} \quad = \quad |||\mathcal{T}|||_\infty,$$

where $|||\mathcal{T}|||_\infty$ can be bounded above by using (D.13):

$$|||\mathcal{T}|||_\infty \quad = \quad \max_{1 \leq i \leq M} \sum_{j=1}^{M} |\mathcal{T}_{i,j}| \quad \leq \quad 1 + \max_{1 \leq i \leq M} \sum_{j=1:\, j \neq i}^{M} \mathbb{B}_{\boldsymbol{\xi}}(i \not\leftrightarrow j).$$

Using (D.14) along with Assumption A,

$$\begin{aligned} |||\mathcal{D}|||_2 \quad &\leq \quad 1 + \max_{1 \leq i \leq M} \sum_{j=1:\, j \neq i}^{M} \mathbb{B}_{\boldsymbol{\xi}}(i \not\leftrightarrow j) \\ &\leq \quad 1 + D + \sum_{k=2}^{\infty} 8\, D^2\, g(k)\, \exp(-A\, k^{1-\omega_2\, |\log(1-\xi)|}), \end{aligned}$$

noting that the number of vertices $j \in \mathcal{V}$ at distance k in $\mathcal{G}^\star$ to any given vertex i is bounded above by D when $k = 1$ and is bounded above by $8\, D^2 g(k) = 8\, D^2 \omega_1 + \omega_2 \log k$ when $k \geq 2$, as derived above. We then have the upper bound

$$\begin{aligned} & \sum_{k=2}^{\infty} 8\, D^2\, g(k)\, \exp(-A\, k^{1-\omega_2\, |\log(1-\xi)|}) \\ \leq \quad & 8\, D^2 \sum_{k=2}^{\infty} (\omega_1 + \omega_2\, \log\, k)\, \exp(-A\, k^{1-\omega_2\, |\log(1-\xi)|}), \end{aligned}$$

and focus on bounding the infinite series

$$\sum_{k=2}^{\infty} (\omega_1 + \omega_2\, \log(k))\, \exp(-A\, k^{1-\omega_2\, |\log(1-\xi)|}). \tag{D.15}$$

Note that the infinite series is equal to 0 when $\omega_1 = 0$ and $\omega_2 = 0$, a fact which we will make use of at the end of the proof. The inequality $\exp(z) \geq z^u / u! + 1 > z^u / u!$, which holds for any $z > 0$ and any $u > 0$, implies the inequality $\exp(-z) < u! / z^u$. Hence, for $u > 0$,

$$\begin{aligned} & \sum_{k=1}^{\infty} (\omega_1 + \omega_2\, \log(k))\, \exp\left(-A\, k^{1-\omega_2\, |\log(1-\xi)|}\right) \\ < \quad & \sum_{k=1}^{\infty} (\omega_1 + \omega_2\, \log(k))\, \frac{u!}{A^u\, k^{u\,(1-\omega_2\, |\log(1-\xi)|)}} \\ \leq \quad & \sum_{k=1}^{\infty} (\omega_1 + \omega_2)\, k\, \frac{u!}{A^u\, k^{u\,(1-\omega_2\, |\log(1-\xi)|)}}, \end{aligned}$$

where the last inequality follows from the assumption $\omega_1 \geq 0$ and $\omega_2 \geq 0$, and the fact that $\log k \leq k$ provided $k \geq 1$. Recall Assumption A assumes

$$\omega_2 \quad < \quad \frac{1}{|\log(1-\xi)|}.$$

Since we are free to choose u as long as $u > 0$, we choose

$$u \quad = \quad \frac{3}{1 - \omega_2\, |\log(1-\xi)|},$$

where u is independent of N and p, because both ξ and ω_2 are independent of N and p. Note $\omega_2\, |\log(1-\xi)| < 1$, by the choice of ω_2, which ensures

$u \geq 3$, because $u\,(1 - \omega_2\,|\log(1-\xi)|) = 3$ and $(1 - \omega_2\,|\log(1-\xi)|) \in (0, 1)$. Thus,

$$\begin{aligned}
\sum_{k=1}^{\infty} (\omega_1 + \omega_2)\; k\; \frac{u!}{A^u\; k^{u\,(1-\omega_2\,|\log(1-\xi)|)}} &\leq \sum_{k=1}^{\infty} (\omega_1 + \omega_2)\; k\; \frac{u!}{A^u\; k^3} \\
&= \frac{(\omega_1 + \omega_2)\, u!}{A^u} \sum_{k=1}^{\infty} \frac{1}{k^2} \\
&= \frac{\pi^2\,(\omega_1 + \omega_2)\, u!}{6\,\exp(-u\, 8\, D^2\, \omega_1\, |\log(1-\xi)|)},
\end{aligned}$$

using the fact that $\sum_{k=1}^{\infty} k^{-2} = \pi^2 \,/\, 6$ (here, $\pi \leq 3.142$ is the mathematical constant, which should not be confused with the probabilities $\pi^\star$ and π_v). Let $C_1 = u\, 8\, \omega_1\, |\log(1-\xi)| > 0$ be a constant, independent of N and p since it is defined in terms of constants independent of N and p. We then obtain the upper bound

$$\frac{\pi^2\,(\omega_1 + \omega_2)\, u!}{6\,\exp(-u\, 8\, D^2\, \omega_1\, |\log(1-\xi)|)} \;\leq\; \frac{3.142^2\,(\omega_1 + \omega_2)\, u!}{6\,\exp(-C_1\, D^2)} \;\leq\; C_2\, \exp(-C_1\, D^2),$$

taking $C_2 = 3.142^2\, u! \,/\, 6 > 0$, also independent of N and p, because u is independent of N and p. Thus, we have shown there exist constants $C_1 > 0$ and $C_2 > 0$, independent of N and p, such that

$$\sum_{k=2}^{\infty} g(k)\, \exp(-A\, k^{1-\omega_2\,|\log(1-\pi^\star)|}) \;\leq\; (\omega_1 + \omega_2)\, C_2\, \exp(C_1\, D^2)$$

and

$$\begin{aligned}
|||\mathcal{D}|||_2 &\leq 1 + D + 8\, D^2\,(\omega_1 + \omega_2)\, C_2\, \exp(C_1\, D^2) \\
&\leq C_3\, D^2\, \exp(C_1\, D^2),
\end{aligned}$$

taking $C_3 = 8\,(\omega_1 + \omega_2) > 0$; note that $C_3 > 0$ is independent of N and p because ω_1 and ω_2 are independent of N and p.

Last, but not least, in the scenario in which $\omega_1 = \omega_2 = 0$ and condition (3.7) is satisfied with $\vartheta \in (4/5,\, 1]$, the infinite series in (D.15) is equal to 0, and we obtain the upper bound

$$|||\mathcal{D}|||_2 \;\leq\; 1 + D.$$

Lemma 12. *Consider Models 1–4 with $\boldsymbol{\theta} \in \mathbb{R}^p$, where*

- $p = N$ *under Model 1 and* $p = N + 1$ *under Models 2–4;*
- $\alpha = 0$ *under Models 1, 2, and 4 and* $\alpha \in (0, 1/2)$ *under Model 3.*

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* $\boldsymbol{x}_{-\{i,j\}} \in \{0,1\}^{M-1}$,

$$0 \;<\; L_k(\boldsymbol{\theta}) \;\leq\; \mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = k \mid \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}) \;\leq\; U_k(\boldsymbol{\theta}) \;<\; 1.$$

Under the sparse β*-model (which includes Model 1 as a special case with* $\alpha = 0$*) and under Model 3, for all pairs of nodes* $\{i, j\}$ *satisfying* $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset$,

$$L_0(\boldsymbol{\theta}) \;=\; \frac{1}{1 + \exp(2\,\|\boldsymbol{\theta}\|_\infty)} \quad \text{and} \quad U_0(\boldsymbol{\theta}) \;=\; \frac{1}{1 + \exp(-2\,\|\boldsymbol{\theta}\|_\infty)\,N^{-\alpha}}$$

$$L_1(\boldsymbol{\theta}) \;=\; \frac{N^{-\alpha}}{1 + \exp(2\,\|\boldsymbol{\theta}\|_\infty)} \quad \text{and} \quad U_1(\boldsymbol{\theta}) \;=\; \frac{1}{1 + \exp(-2\,\|\boldsymbol{\theta}\|_\infty)}.$$

Under Model 3, for all pairs of nodes $\{i, j\}$ *satisfying* $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$*, and under Models 2 and 4,*

$$\begin{aligned} L_0(\boldsymbol{\theta}) \;&=\; L_1(\boldsymbol{\theta}) \;=\; \frac{1}{1 + \exp((3 + D)\,\|\boldsymbol{\theta}\|_\infty)} \\ U_0(\boldsymbol{\theta}) \;&=\; U_1(\boldsymbol{\theta}) \;=\; \frac{1}{1 + \exp(-(3 + D)\,\|\boldsymbol{\theta}\|_\infty)}, \end{aligned}$$

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

Proof of Lemma 12. To prove Lemma 12, we return to the notation used in Section 2 of the manuscript, denoting edge variables by $X_{i,j}$ ($\{i, j\} \subset \mathcal{N}$). 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 probabilities

$$\mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = x_{i,j} \mid \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}})$$

as

$$\begin{aligned} &\frac{\exp(\langle \boldsymbol{\theta},\, s(\boldsymbol{x}_{-\{i,j\}},\, x_{i,j})\rangle)\, N^{-\alpha\, x_{i,j}}}{\exp(\langle \boldsymbol{\theta},\, s(\boldsymbol{x}_{-\{i,j\}},\, x_{i,j} = 0)\rangle) + \exp(\langle \boldsymbol{\theta},\, s(\boldsymbol{x}_{-\{i,j\}},\, x_{i,j} = 1)\rangle)\, N^{-\alpha}} \\ =\;\; &\frac{1}{g(0;\, \boldsymbol{x}_{-\{i,j\}},\, x_{i,j},\, \boldsymbol{\theta})\, N^{\alpha\, x_{i,j}} + g(1;\, \boldsymbol{x}_{-\{i,j\}},\, x_{i,j},\, \boldsymbol{\theta})\, N^{\alpha\,(x_{i,j}-1)}} \end{aligned}$$

provided $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset$, and

$$\frac{\exp(\langle \boldsymbol{\theta},\, s(\boldsymbol{x}_{-\{i,j\}},\, x_{i,j})\rangle)}{\exp(\langle \boldsymbol{\theta},\, s(\boldsymbol{x}_{-\{i,j\}},\, x_{i,j}=0)\rangle) + \exp(\langle \boldsymbol{\theta},\, s(\boldsymbol{x}_{-\{i,j\}},\, x_{i,j}=1)\rangle)}$$

$$= \quad \frac{1}{g(0;\, \boldsymbol{x}_{-\{i,j\}},\, x_{i,j},\, \boldsymbol{\theta}) + g(1;\, \boldsymbol{x}_{-\{i,j\}},\, x_{i,j},\, \boldsymbol{\theta})}$$

provided $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$. Here, $\alpha = 0$ under Models 1, 2, and 4 and $\alpha \in (0,\, 1/2)$ under Model 3, and

$$g(y;\, \boldsymbol{x}_{-\{i,j\}},\, x_{i,j},\, \boldsymbol{\theta}) \quad = \quad \exp(\langle \boldsymbol{\theta},\, s(\boldsymbol{x}_{-\{i,j\}},\, y) - s(\boldsymbol{x}_{-\{i,j\}},\, x_{i,j})\rangle),$$

where $y \in \{0, 1\}$. We have

$$\max_{\boldsymbol{x}_{-\{i,j\}} \in \{0,1\}^{M-1}} \left| s_l(\boldsymbol{x}_{-\{i,j\}},\, x_{i,j}=0) - s_l(\boldsymbol{x}_{-\{i,j\}},\, x_{i,j}=1) \right|$$

$$= \begin{cases} 0 & \text{if } l \notin \{1, \dots, N\} \setminus \{i, j\} \\ 1 & \text{if } l \in \{1, j\} \\ 1 + D & \text{if } l = N + 1. \end{cases}$$

The bound on s_{N+1} follows from Lemma 13, and 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 and the above bounds reveal

$$\left| \left\langle \boldsymbol{\theta},\, s(\boldsymbol{x}_{-\{i,j\}},\, x_{i,j}=0) \right\rangle - \left\langle \boldsymbol{\theta},\, s(\boldsymbol{x}_{-\{i,j\}},\, x_{i,j}=0) \right\rangle \right| \quad \leq \quad (3 + D)\, \|\boldsymbol{\theta}\|_\infty.$$

We thus obtain the bounds stated in Lemma 12.

Lemma 13. *Consider Models 2–4 with statistic*

$$s_{N+1}(\boldsymbol{x}) \quad = \quad \sum_{i<j}^{N} x_{i,j}\, I_{i,j}(\boldsymbol{x}),$$

where

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

Then, for all $\{i,j\} \subset \mathcal{N}$,

$$\max_{(\boldsymbol{x},\, \boldsymbol{x}') \,\in\, \mathbb{X} \times \mathbb{X}:\; x_{v,w} = x'_{v,w},\; \{v,\, w\} \neq \{i,\, j\}} |s_{N+1}(\boldsymbol{x}) - s_{N+1}(\boldsymbol{x}')| \;\;\leq\;\; 1 + D,$$

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

Proof of Lemma 13. Consider any $\{i,j\} \subset \mathcal{N}$. The proof of Proposition 2 shows that the number of summands $x_{a,b}\, I_{a,b}(\boldsymbol{x})$ ($\{a,b\} \neq \{i,j\}$) that are a function of $x_{i,j}$ is bounded above by $2\,|\mathcal{N}_i \cup \mathcal{N}_j|$. In addition, Proposition 2 shows that $\mathfrak{N}_{i,j} \subseteq \{i,j\} \times (\mathcal{N}_i \cup \mathcal{N}_j)$ for all $\{i,j\} \subset \mathcal{N}$. We therefore obtain the bound $|\mathfrak{N}_{i,j}| \leq 2\,|\mathcal{N}_i \cup \mathcal{N}_j| \leq D$, where $D \coloneqq \max_{\{i,j\} \subset \mathcal{N}} |\mathfrak{N}_{i,j}| = \max_{\{i,j\} \subset \mathcal{N}} 2\,|\mathcal{N}_i \cup \mathcal{N}_j|$. Thus, the number of summands that are a function of $x_{i,j}$ is bounded above by $1 + D$. 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,

$$|s_{N+1}(\boldsymbol{x}) - s_{N+1}(\boldsymbol{x}')| \;\;\leq\;\; \sum_{\{a,b\} \subset \mathcal{N}} |x_{a,b}\, I_{a,b}(\boldsymbol{x}) - x'_{a,b}\, I_{a,b}(\boldsymbol{x}')| \;\;\leq\;\; 1 + D,$$

using $x_{a,b}\, I_{a,b}(\boldsymbol{x}) \leq 1$ for all $\{a,b\} \subset \mathcal{N}$ and all $\boldsymbol{x} \in \mathbb{X}$.

D.2.4. *Auxiliary results.* We prove Lemmas 14 and 15.

Lemma 14. *Choose any $i \in \{1, \ldots, M\}$ and any $\boldsymbol{x}_{1:i-1} \in \{0,1\}^{i-1}$. Then the coupling of the conditional distributions*

$$\mathbb{P}(\,\cdot \mid \boldsymbol{X}_{1:i-1} = \boldsymbol{x}_{1:i-1},\; X_i = 0)$$

and

$$\mathbb{P}(\,\cdot \mid \boldsymbol{X}_{1:i-1} = \boldsymbol{x}_{1:i-1},\; X_i = 1)$$

constructed in Lemma 11 is a valid coupling.

Proof of Lemma 14. Denote the coupling distribution generated by the algorithm in Lemma 11 by $\mathbb{Q}_{\boldsymbol{x}_{1:i-1}}$ and let $v_1, \ldots, v_{M-i}$ be the vertices added to the set $\mathfrak{V}$ at iteration $1, \ldots, M-i$ of the algorithm. To reduce the notational burden, define

$$\begin{aligned} & q(\boldsymbol{x}^\star_{a:b},\; \boldsymbol{x}^{\star\star}_{a:b} \mid \boldsymbol{x}^\star_{c:d},\; \boldsymbol{x}^{\star\star}_{c,d}) \\ =\;\; & \mathbb{Q}_{\boldsymbol{x}_{1:i-1}}(\boldsymbol{X}^\star_{a:b} = \boldsymbol{x}^\star_{a:b},\; \boldsymbol{X}^{\star\star}_{a:b} = \boldsymbol{x}^{\star\star}_{a:b} \mid \boldsymbol{X}^\star_{c:d} = \boldsymbol{x}^\star_{c:d},\; \boldsymbol{X}^{\star\star}_{c:d} = \boldsymbol{x}^{\star\star}_{c:d}), \end{aligned}$$

where $a, b, c, d \in \{1, \ldots, M\}$ are distinct integers and $\{a, \ldots, b\} \cap \{c, \ldots, d\} = \emptyset$. By construction,

$$q(\boldsymbol{x}^\star_{i+1:M},\; \boldsymbol{x}^{\star\star}_{i+1:M}) \;\;=\;\; q(x^\star_{v_1},\; x^{\star\star}_{v_1}) \prod_{l=2}^{M-i} q(x^\star_{v_l},\; x^{\star\star}_{v_l} \mid \boldsymbol{x}^\star_{v_1,\ldots,v_{l-1}},\; \boldsymbol{x}^{\star\star}_{v_1,\ldots,v_{l-1}}).$$

Observe that

$$\sum_{x^{\star}_{v_{M-i}} \in \{0,1\}} q(x^{\star}_{v_{M-i}},\ x^{\star\star}_{v_{M-i}} \mid \boldsymbol{x}^{\star}_{v_1,\dots,v_{M-i-1}},\ \boldsymbol{x}^{\star\star}_{v_1,\dots,v_{M-i-1}})$$

$$= \mathbb{P}(X_{v_{M-i}} = x^{\star\star}_{v_{M-i}} \mid \boldsymbol{X}_{1:i-1} = \boldsymbol{x}_{1:i-1},\ X_i = 1,\ \boldsymbol{X}_{v_1,\dots,v_{M-i-1}} = \boldsymbol{x}^{\star\star}_{v_1,\dots,v_{M-i-1}})$$

and

$$\sum_{x^{\star\star}_{v_{M-i}} \in \{0,1\}} q(x^{\star}_{v_{M-i}},\ x^{\star\star}_{v_{M-i}} \mid \boldsymbol{x}^{\star}_{v_1,\dots,v_{M-i-1}},\ \boldsymbol{x}^{\star\star}_{v_1,\dots,v_{M-i-1}})$$

$$= \mathbb{P}(X_{v_{M-i}} = x^{\star}_{v_{M-i}} \mid \boldsymbol{X}_{1:i-1} = \boldsymbol{x}_{1:i-1},\ X_i = 1,\ \boldsymbol{X}_{v_1,\dots,v_{M-i-1}} = \boldsymbol{x}^{\star}_{v_1,\dots,v_{M-i-1}}),$$

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

$$\mathbb{P}(X_{v_{M-i}} = \,\cdot \mid \boldsymbol{X}_{1:i-1} = \boldsymbol{x}_{1:i-1},\ X_i = 1,\ \boldsymbol{X}_{v_1,\dots,v_{M-i-1}} = \boldsymbol{x}^{\star\star}_{v_1,\dots,v_{M-i-1}}).$$

By induction,

$$\sum_{x^{\star}_{v_1} \in \{0,1\}} \cdots \sum_{x^{\star}_{v_{M-i}} \in \{0,1\}} q(\boldsymbol{x}^{\star}_{v_1,\dots,v_{M-i}},\ \boldsymbol{x}^{\star\star}_{i+1:M})$$

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

and

$$\sum_{x^{\star\star}_{v_1} \in \{0,1\}} \cdots \sum_{x_{v^{\star\star}_{M-i}} \in \{0,1\}} q(\boldsymbol{x}^{\star}_{i+1:M},\ \boldsymbol{x}^{\star\star}_{v_1,\dots,v_{M-i}})$$

$$= \mathbb{P}(\boldsymbol{X}_{i+1:M} = \boldsymbol{x}^{\star}_{1+i:M} \mid \boldsymbol{X}_{1:i-1} = \boldsymbol{x}_{1:i-1},\ X_i = 0),$$

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 15. *Consider Models 1–4, any $v \in \{1, \ldots, M\}$, 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

$$\pi^\star \;\;=\;\; \max_{1 \leq v \leq M} \;\; \max_{(\boldsymbol{x}_{-v},\, \boldsymbol{x}'_{-v}) \,\in\, \{0,1\}^{M-1} \times \{0,1\}^{M-1}} \; \pi_{v,\, \boldsymbol{x}_{-v},\, \boldsymbol{x}'_{-v}}.$$

Then

$$\pi^\star \;\;\leq\;\; \begin{cases} 0 & \text{under Model 1} \\[2ex] \dfrac{1}{1 + \exp(-(2 + 3\, \max_{1 \leq i \leq N} |\mathcal{N}_i|)\, \|\boldsymbol{\theta}^\star\|_\infty)} & \text{under Models 2–4.} \end{cases}$$

Proof of Lemma 15. 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 2–4, we distinguish two cases:

(a) If edge variable X_v corresponds to a pair of nodes with non-intersecting neighborhoods, 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 $(\boldsymbol{x}_{-v},\, \boldsymbol{x}'_{-v}) \in \{0,1\}^{M-1} \times \{0,1\}^{M-1}$, so $\pi^\star = 0$.

(b) If edge variable X_v corresponds to a pair of nodes with intersecting neighborhoods, 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 12 shows that, under Models 2–4,

$$\mathbb{P}(X_v = 0 \mid \boldsymbol{X}_{-v} = \boldsymbol{x}_{-v}) \;\;\leq\;\; \frac{1}{1 + \exp(-(2 + 3\, \max_{1 \leq i \leq N} |\mathcal{N}_i|)\, \|\boldsymbol{\theta}^\star\|_\infty)}$$

and

$$\mathbb{P}(X_v = 1 \mid \boldsymbol{X}_{-v} = \boldsymbol{x}_{-v}) \;\;\leq\;\; \frac{1}{1 + \exp(-(2 + 3\, \max_{1 \leq i \leq N} |\mathcal{N}_i|)\, \|\boldsymbol{\theta}^\star\|_\infty)}.$$

We may therefore conclude that, under Models 2–4,

$$\begin{aligned} \pi^\star \;\;&\leq\;\; \min\left\{\max\{a_0,\, b_0\},\; \max\{a_1,\, b_1\}\right\} \\ &\leq\;\; \frac{1}{1 + \exp(-(2 + 3\, \max_{1 \leq i \leq N} |\mathcal{N}_i|)\, \|\boldsymbol{\theta}^\star\|_\infty)}. \end{aligned}$$

APPENDIX E: 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

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\theta}} \left(\sum_{j \neq i}^{N} X_{i,j}\right) \;\;&=\;\; \sum_{j\,\in\,\mathfrak{A}_{i,1}} \mathbb{E}_{\boldsymbol{\theta}}\, X_{i,j} + \sum_{j\,\in\,\mathfrak{A}_{i,2}} \mathbb{E}_{\boldsymbol{\theta}}\, X_{i,j} \\ &\leq\;\; |\mathfrak{A}_{i,1}| \max_{j\,\in\,\mathfrak{A}_{i,1}} \mathbb{E}_{\boldsymbol{\theta}}\, X_{i,j} + |\mathfrak{A}_{i,2}| \max_{j\,\in\,\mathfrak{A}_{i,2}} \mathbb{E}_{\boldsymbol{\theta}}\, X_{i,j}, \end{aligned}$$

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

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\theta}}\, X_{i,j} \;\;&=\;\; \mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = 1) \\ &\leq\;\; \max_{\boldsymbol{x}_{-\{i,\,j\}}\,\in\,\{0,1\}^{N(N-1)/2-1}} \mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = 1 \mid \boldsymbol{X}_{-\{i,\,j\}} = \boldsymbol{x}_{-\{i,\,j\}}). \end{aligned}$$

For any $j \in \mathfrak{A}_{i,1}$, $\mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = 1 \mid \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}) \leq 1 \leq \exp(2\,\|\boldsymbol{\theta}\|_\infty)$ for all $\boldsymbol{x}_{-\{i,j\}} \in \{0,1\}^{N(N-1)/2-1}$. In addition, for any $j \in \mathfrak{A}_{i,2}$, Lemma 12 in Appendix D.2.3 shows that

$$\mathbb{P}_{\boldsymbol{\theta}}(X_{i,j} = 1 \mid \boldsymbol{X}_{-\{i,j\}} = \boldsymbol{x}_{-\{i,j\}}) \leq \frac{1}{1 + N^\alpha \exp(-2\,||\boldsymbol{\theta}||_\infty)} < \frac{\exp(2\,\|\boldsymbol{\theta}\|_\infty)}{N^\alpha}$$

for all $\boldsymbol{x}_{-\{i,j\}} \in \{0,1\}^{N(N-1)/2-1}$. Hence,

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

Bounding $|\mathfrak{A}_{i,1}|$. To bound $|\mathfrak{A}_{i,1}|$, we distinguish two cases:

- $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$ where $j \in \mathcal{N}_i$, which implies that there exists $k \in \{1,\dots,K\}$ such that $\{i, j\} \subset \mathcal{A}_k$ and $j \in \mathcal{N}_i$.
- $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$ where $j \notin \mathcal{N}_i$, in which case there exists $h \in \mathcal{N}_i \cap \mathcal{N}_j$ such that $j \in \mathcal{N}_h$, which in turn implies that there exists $k \in \{1,\dots,K\}$ such that $\{j, h\} \subset \mathcal{A}_k$ and $h \in \mathcal{N}_j$.

The number of $j \in \mathcal{N}$ satisfying the first case is bounded above by $|\mathcal{N}_i| \leq \max_{1\leq h\leq N} |\mathcal{N}_h|$, and the number of $j \in \mathcal{N}$ satisfying the second case is bounded above by

$$|\mathcal{N}_i| \max_{1\leq h\leq N} |\mathcal{N}_h| \;\leq\; \left(\max_{1\leq h\leq N} |\mathcal{N}_h|\right)^2.$$

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 $i \in \mathcal{N}$,

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

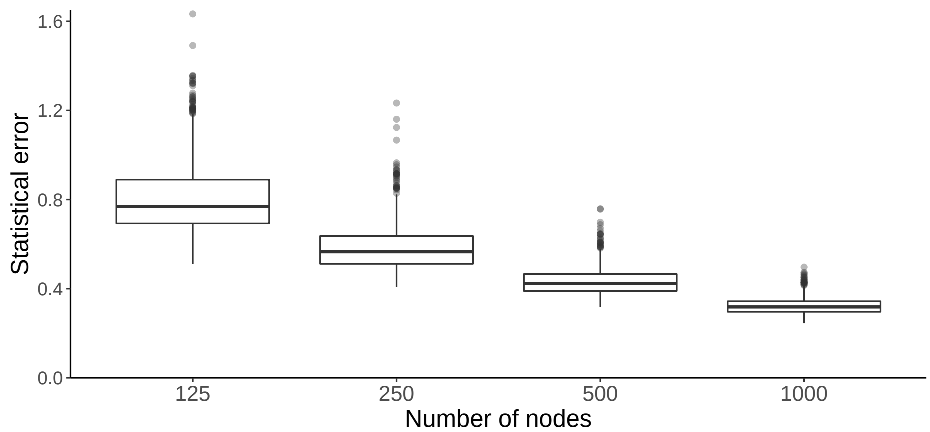

FIG 4. *The statistical error* $\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star\|_\infty$ *of maximum pseudo-likelihood estimator* $\widetilde{\boldsymbol{\theta}}$ *as an estimator of* $\boldsymbol{\theta}^\star \in \mathbb{R}^{N+1}$ *plotted against the number of nodes* N.

APPENDIX F: 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:

- The number of subpopulations K is $N \,/\, 25$.
- Each node $i \in \mathcal{N}$ belongs to $1 + Y_i$ subpopulations, where $Y_i \overset{\text{iid}}{\sim} \text{Binomial}(K-1,\; 1\,/\,K)$ $(i = 1, \dots, N)$.
- For node $i = 1, \dots, N$, the $1 + Y_i$ subpopulation memberships are sampled from the Multinomial$(p_1^{(i)}, \dots, p_K^{(i)})$ distribution with

$$p_k^{(i)} \;=\; \begin{cases} \dfrac{1}{K} & \text{if } i = 1 \\[2ex] \dfrac{1}{K-1}\,\left(1 - \dfrac{N_k^{(i-1)}}{N_1^{(i-1)} + \ldots + N_K^{(i-1)}}\right) & \text{if } i \in \{2, \dots, N\}, \end{cases}$$

  where $N_k^{(i-1)}$ is the number of nodes in $\{1, \dots, i-1\}$ that belong to subpopulation $\mathcal{A}_k$ $(k = 1, \dots, K)$.

We consider Model 2 with degree parameters $\theta_1^\star, \dots, \theta_N^\star$ drawn from Uniform$(-1.25,\, -.75)$ and brokerage parameter $\theta_{N+1}^\star = .25$. For each value

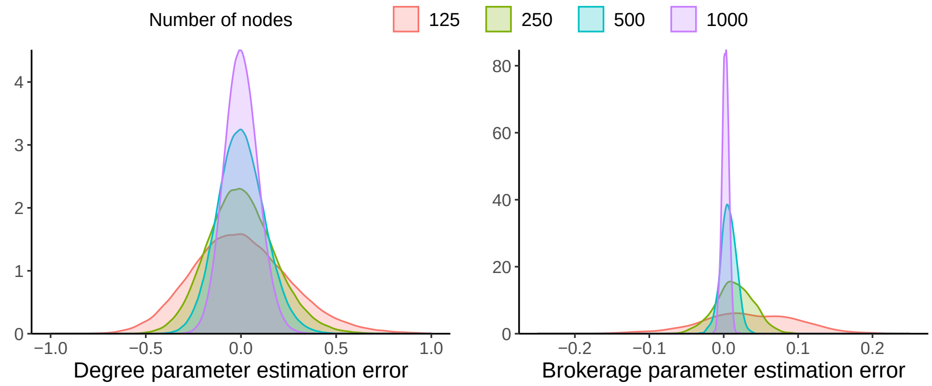

Fig 5. *The maximum deviation* $\max_{1 \leq i \leq N} |\widetilde{\theta}_i - \theta_i^\star|$ *of the maximum pseudo-likelihood estimators* $\widetilde{\theta}_i$ *from the data-generating degree parameters* $\theta_i^\star$ $(i = 1, \ldots, N)$ *(left) and the deviation* $|\widetilde{\theta}_{N+1} - \theta_{N+1}^\star|$ *of the maximum pseudo-likelihood estimator* $\widetilde{\theta}_{N+1}$ *from the data-generating brokerage parameter* $\theta_{N+1}^\star$ *(right).*

of N and each population of size N, 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 4 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 the number of nodes N increases. Figure 5 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 5 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.

REFERENCES

[1] Airoldi, E., Blei, D., Fienberg, S., and Xing, E. (2008), "Mixed membership stochastic blockmodels," *Journal of Machine Learning Research*, 9, 1981–2014.

[2] 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.

[3] Bhamidi, S., Bresler, G., and Sly, A. (2011), "Mixing time of exponential random graphs," *The Annals of Applied Probability*, 21, 2146–2170.

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

[5] 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.

[6] Bickel, P. J., Chen, A., and Levina, E. (2011), "The method of moments and degree distributions for network models," *The Annals of Statistics*, 39, 2280–2301.

[7] 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.

[8] Brown, L. (1986), *Fundamentals of Statistical Exponential Families*, Hayworth, CA, USA: Institute of Mathematical Statistics.

[9] Butts, C. T. (2011), "Bernoulli graph bounds for general random graph models," *Sociological Methodology*, 41, 299–345.

[10] 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.

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

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

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

[14] 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.

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

[16] 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.

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

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

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

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

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

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

[23] 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.

[24] Hillar, C. J., Lin, S., and Wibisono, A. (2012), "Inverses of symmetric, diagonally dominant positive matrices and applications," *arXiv preprint arXiv:1203.6812*.

[25] 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.

[26] 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.

[27] 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.

[28] Hunter, D. R., Krivitsky, P. N., and Schweinberger, M. (2012), "Computational statistical methods for social network models," *Journal of Computational and Graphical Statistics*, 21, 856–882.

[29] Jensen, J. L., and Møller, J. (1991), "Pseudolikelihood for exponential family models of spatial point processes," *The Annals of Applied Probability*, 1, 445–461.

[30] 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.

[31] Krivitsky, P. N., and Kolaczyk, E. D. (2015), "On the question of effective sample size in network modeling: An asymptotic inquiry," *Statistical Science*, 30, 184–198.

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

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

[34] Lei, J., and Rinaldo, A. (2015), "Consistency of spectral clustering in stochastic block models," *The Annals of Statistics*, 43, 215–237.

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

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

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

[38] 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.

[39] 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.

[40] 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.

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

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

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

[44] 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.

[45] 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.

[46] Stewart, J. R., and Schweinberger, M. (2022), "Pseudo-likelihood-based M-estimators for random graphs with dependent edges and parameter vectors of increasing dimension," Tech. rep., Florida State University, arXiv:2012.07167.

[47] Tang, M., Sussman, D. L., and Priebe, C. E. (2013), "Universally consistent vertex classification for latent positions graphs," *The Annals of Statistics*, 41, 1406–1430.

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

[49] Wainwright, M. J., and Jordan, M. I. (2008), "Graphical models, exponential families, and variational inference," *Foundations and Trends in Machine Learning*, 1, 1–305.

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

[51] 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.

[52] 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