Bernoulli **26**(2), 2020, 1205–1233
https://doi.org/10.3150/19-BEJ1153

Consistent structure estimation of exponential-family random graph models with block structure

MICHAEL SCHWEINBERGER[1]

[1]*Department of Statistics, Rice University, 6100 Main St, MS-138, Houston, TX 77005-1827, USA. E-mail: m.s@rice.edu*

We consider the challenging problem of statistical inference for exponential-family random graph models based on a single observation of a random graph with complex dependence. To facilitate statistical inference, we consider random graphs with additional structure in the form of block structure. We have shown elsewhere that when the block structure is known, it facilitates consistency results for M-estimators of canonical and curved exponential-family random graph models with complex dependence, such as transitivity. In practice, the block structure is known in some applications (e.g., multilevel networks), but is unknown in others. When the block structure is unknown, the first and foremost question is whether it can be recovered with high probability based on a single observation of a random graph with complex dependence. The main consistency results of the paper show that it is possible to do so under weak dependence and smoothness conditions. These results confirm that exponential-family random graph models with block structure constitute a promising direction of statistical network analysis.

Keywords: exponential families; exponential random graph models; network data; random graphs; stochastic block models

1. Introduction

Exponential-family random graph models [19,30,35,63,69] are models of network data, such as disease transmission networks, insurgent and terrorist networks, social networks, and the World Wide Web [42]. Such models can be viewed as generalizations of Bernoulli random graphs with independent edges [18,22] to random graphs with dependent edges. Exponential-family random graph models are popular among network scientists [42], because network data are dependent data and exponential-family random graph models enable network scientists to model a wide range of dependencies found in network data.

Exponential-family random graph models of dependent network data were pioneered by [19]. The models of [19] and more general models [30,35,63,69] are discrete exponential families of densities with countable support $\mathbb{X}$ – the set of possible graphs with n nodes and binary or non-binary, count-valued edges – of the form

$$p_{\boldsymbol{\eta}}(\boldsymbol{x}) = \exp\big(\langle \boldsymbol{\eta}, s(\boldsymbol{x})\rangle - \psi(\boldsymbol{\eta})\big), \quad \boldsymbol{x} \in \mathbb{X}, \tag{1.1}$$

1350-7265

where $\langle \boldsymbol{\eta}, s(\boldsymbol{x})\rangle$ denotes the inner product of a vector of natural parameters $\boldsymbol{\eta} \in \{\boldsymbol{\eta} \in \mathbb{R}^{\dim(\boldsymbol{\eta})} : \psi(\boldsymbol{\eta}) < \infty\}$ and a vector of sufficient statistics $s : \mathbb{X} \mapsto \mathbb{R}^{\dim(\boldsymbol{\eta})}$ and $\psi(\boldsymbol{\eta})$ ensures that $\sum_{\boldsymbol{x}' \in \mathbb{X}} p_{\boldsymbol{\eta}}(\boldsymbol{x}') = 1$.

In general, statistical inference for exponential-family random graph models is challenging [6,13,23,54,60], because exponential-family random graph models induce complex dependence (e.g., transitivity, [42]) and many network data sets either consist of a single observation of a population graph or subgraphs sampled from a population graph. For example, epidemiologists studying the spread of infectious diseases (e.g., HIV, Ebola) may be able to observe whether population members were in contact during an epidemic, but may not be able to obtain independent or repeated observations of contacts over time. As a result, the epidemiologists may have to be content with a single observation of the population contact network of interest or subgraphs sampled from the population contact network. The fact that many network data sets consist of a single observation of a population graph or sampled subgraphs means that concentration and consistency results cannot be obtained along the lines of classical and high-dimensional statistics, which rely on independent observations from the same source (in a well-defined sense). In addition, the complex dependence induced by these models implies that establishing concentration, consistency, and weak convergence results for estimators requires concentration-of-measure results for dependent random variables, which are more challenging than concentration-of-measure results for independent random variables (e.g., [34]).

1.1. Advantages of block structure

While statistical inference for exponential-family random graph models is challenging, statistical inference for models with additional structure has advantages.

To demonstrate the advantages of additional structure, we consider a natural form of additional structure known as block structure. Block structure is popular in the large and growing body of literature on stochastic block models (e.g., [1,3,7–9,12,15,21,32,39,44,45,52,53,74,75]). We focus here on exponential-family random graph models with block structure, which allow edges within blocks to be dependent [56]. Such models are less restrictive than stochastic block models [45], which assume that edges within blocks are independent Bernoulli random variables. Indeed, sensible specifications of exponential-family random graph models can capture excesses in transitivity and many other interesting features of random graphs that induce complex dependence among edges within blocks [56]. We have shown elsewhere that when the block structure is known, exponential-family random graph models with block structure have important advantages:

- If edges depend on other edges within the same block but do not depend on edges outside of the block, models induce local dependence within blocks. Local dependence makes sense in applications, because network data are dependent data but network dependence is more local than global [46,56].
- Models with block structure are weakly projective in the sense that the probability mass function of a random graph with block structure is consistent with the probability mass function of a larger random graph with more blocks [56,59], whereas many models without block structure are not projective [16,37,60,62].

- Local dependence induces weak dependence as long as the blocks are not too large. Weak dependence facilitates concentration and consistency results for M-estimators, including maximum likelihood estimators [59]. These results are of fundamental importance, because they are the first consistency results for models with transitivity and other interesting features of random graphs that induce complex dependence. Transitivity is interesting in practice [68], but is challenging from a theoretical point of view (e.g., [13,60]), and indeed no other consistency results exist for transitivity.

In other words, block structure is not only useful for community detection in social networks, for which stochastic block models can be used, but also facilitates statistical inference for random graphs with complex dependence induced by transitivity and many other interesting features of random graphs.

1.2. Recovery of unknown block structure

In some applications, the block structure is known. An example is multilevel networks, which are popular in network science (e.g., [26,27,41,61,66,73]): for example, the blocks may correspond to school classes in schools, units of armed forces, and departments in universities.

While the block structure is known in some applications, it is unknown in others. When the block structure is unknown, the first and foremost question is whether it can be recovered with high probability. A large and growing body of consistency results for stochastic block models shows that it is possible to recover the block structure of stochastic block models with high probability (e.g., [1,3,7–9,12,15,21,32,39,44,45,52,53,74,75]). While it is encouraging that the block structure of stochastic block models can be recovered with high probability, these results are restricted to models with independent edges within and between blocks. It is not at all obvious whether the block structure of the more complex exponential-family random graph models can be recovered with high probability.

We show that consistent recovery of block structure is not limited to stochastic block models, but is possible for the more complex exponential-family random graph models. The main consistency results of the paper show that it is possible to recover the block structure with high probability under weak dependence and smoothness conditions. Among other things, these consistency results demonstrate that the conditional independence assumptions underlying stochastic block models are not necessary for consistent recovery of block structure. In other words, these results suggest that it is possible to obtain consistency results for many interesting models with block structure, both stochastic block models with independent edges within blocks and richer models with dependent edges within blocks, such as the models and methods proposed by [56] and [67]. An indepth investigation of all of these models and methods is beyond the scope of a single paper: each of them is challenging, owing to the complex dependence within blocks and the wide range of model terms and canonical and curved exponential-family parameterizations. However, the main consistency results reported here suggest that statistical inference for these models and methods is possible and worth exploring in more depth.

The paper is structured as follows. Section 2 introduces exponential-family random graph models with additional structure in the form of block structure. Section 3 discusses the main consistency results. Section 4 presents simulation results. Section 5 proves the main consistency results.

1.3. Other, related literature

It is worth noting that two broad classes of exponential-family random graph models can be distinguished based on the underlying dependence assumptions: one class of models assumes that edges or pairs of directed edges are independent (e.g., the β-model and the p_1-model, [14,25,51,70–72]), while the other class of models allows edges or pairs of directed edges to be dependent [19,30,63]. The independence assumptions of the first class of models are restrictive, because it is known that edges in real-world networks tend to depend on other edges [24]. The dependence assumptions of the second class of models are problematic, because some of these models allow edges to depend on many other edges: e.g., the conditional independence assumptions of [19] allow the conditional distribution of each edge variable to depend on $2(n-2)$ other edge variables. Some – but not all – of these models induce strong dependence in large random graphs and therefore have undesirable properties, such as model near-degeneracy [6,13,16,23,54,60]. Exponential-family random graph models with block structure strike a middle ground between models with independence assumptions and models with strong dependence assumptions, because sensible specifications of these models induce weak dependence. As a consequence, sensible specifications of these models have desirable properties, as explained above.

2. Exponential-family random graph models with additional structure

In general, statistical inference for exponential-family random graph models is challenging, as discussed in Section 1. We facilitate statistical inference by endowing exponential-family random graph models with additional structure that induces weak dependence and hence facilitates consistency results.

Throughout, we consider random graphs with a set of nodes $\mathcal{A} = \{1, \ldots, n\}$ and a set of edges $\mathcal{E} \subseteq \mathcal{A} \times \mathcal{A}$, where edges between pairs of nodes $(i, j) \in \mathcal{A} \times \mathcal{A}$ are regarded as random variables $X_{i,j}$ with countable sample spaces $\mathbb{X}_{i,j}$. We focus on undirected graphs without self-edges – that is, $X_{i,i} = 0$ and $X_{i,j} = X_{j,i}$ with probability 1 – but extensions to directed random graphs are straightforward. We write $\boldsymbol{X} = (X_{i,j})_{i<j}^{n}$ and $\mathbb{X} = \times_{i<j}^{n} \mathbb{X}_{i,j}$.

To facilitate statistical inference, we assume that the random graph is endowed with additional structure in the form of a partition of the set of nodes $\mathcal{A}$ into $K \geq 2$ subsets of nodes $\mathcal{A}_1, \ldots, \mathcal{A}_K$, called blocks. To obtain concentration and consistency results, it is important that the additional structure induces weak dependence, because strong dependence can make concentration results impossible (e.g., [34]). We induce weak dependence by restricting dependence to within-block subgraphs $\boldsymbol{X}_{k,k} = (X_{i,j})_{i<j:\, i\in\mathcal{A}_k, j\in\mathcal{A}_k}$ $(k = 1, \ldots, K)$. The resulting exponential families induce a form of local dependence defined as follows [56].

Definition (Exponential families with local dependence). An exponential family of densities of the form (1.1) with countable support $\mathbb{X}$ satisfies local dependence as long as its densities satisfy

$$p_{\boldsymbol{\eta}}(\boldsymbol{x}) = \prod_{k=1}^{K} p_{\boldsymbol{\eta}}(\boldsymbol{x}_{k,k}) \prod_{l=1}^{k-1} \prod_{i\in\mathcal{A}_k, j\in\mathcal{A}_l} p_{\boldsymbol{\eta}}(x_{i,j}) \quad \text{for all } \boldsymbol{x} \in \mathbb{X}. \tag{2.1}$$

We give examples of canonical and curved exponential families with local dependence in Sections 2.1 and 2.2, respectively. We discuss the well known, but restrictive special case of stochastic block models in Section 2.3 and demonstrate the added value of exponential families with local dependence relative to stochastic block models in Section 2.4.

2.1. Example: Canonical exponential families with local dependence

An example of canonical exponential families with local dependence and support $\mathbb{X} = \{0, 1\}^{\binom{n}{2}}$ is given by exponential families with block-dependent edge and transitive edge terms of the form

$$p_{\boldsymbol{\eta}}(\boldsymbol{x}) \propto \exp\left(\sum_{k \le l}^{K} \eta_{1,k,l} \sum_{i \in \mathcal{A}_k, j \in \mathcal{A}_l} x_{i,j} + \sum_{k=1}^{K} \eta_{2,k,k}\, s_{k,k}(\boldsymbol{x})\right), \tag{2.2}$$

where

$$s_{k,k}(\boldsymbol{x}) = \sum_{i<j: i \in \mathcal{A}_k, j \in \mathcal{A}_k} x_{i,j}\, \mathbb{1}_{i,j}(\boldsymbol{x}). \tag{2.3}$$

Here, $\mathbb{1}_{i,j}(\boldsymbol{x}) = 1$ if the number of shared partners of nodes $i \in \mathcal{A}_k$ and $j \in \mathcal{A}_k$ in block $\mathcal{A}_k$ satisfies $\sum_{h \in \mathcal{A}_k, h \neq i,j} x_{h,i}\, x_{h,j} > 0$ and $\mathbb{1}_{i,j}(\boldsymbol{x}) = 0$ otherwise. If $x_{i,j}\, \mathbb{1}_{i,j}(\boldsymbol{x}) = 1$, the edge between nodes i and j is called transitive. We note that in recent work [31,35,36,59] transitive edge terms have turned out to be attractive alternatives to the triangle terms, which have been used since the classic work of [19] but which possess undesirable properties [13,23,54].

2.2. Example: Curved exponential families with local dependence

An example of curved exponential families with local dependence and support $\mathbb{X} = \{0, 1\}^{\binom{n}{2}}$ is given by exponential families with block-dependent edge and geometrically weighted edgewise shared partner terms of the form

$$p_{\boldsymbol{\eta}}(\boldsymbol{x}) \propto \exp\left(\sum_{k \le l}^{K} \eta_{1,k,l} \sum_{i \in \mathcal{A}_k, j \in \mathcal{A}_l} x_{i,j} + \sum_{k=1}^{K} \sum_{t=1}^{|\mathcal{A}_k|-2} \eta_{2,k,k,t}\, s_{k,k,t}(\boldsymbol{x})\right), \tag{2.4}$$

where

$$s_{k,k,t}(\boldsymbol{x}) = \sum_{i<j: i \in \mathcal{A}_k, j \in \mathcal{A}_k} x_{i,j}\, \mathbb{1}_{i,j,t}(\boldsymbol{x}). \tag{2.5}$$

Here, $\mathbb{1}_{i,j,t}(\boldsymbol{x}) = 1$ if the number of shared partners of nodes $i \in \mathcal{A}_k$ and $j \in \mathcal{A}_k$ in block $\mathcal{A}_k$ satisfies $\sum_{h \in \mathcal{A}_k, h \neq i,j} x_{h,i}\, x_{h,j} = t$ and $\mathbb{1}_{i,j,t}(\boldsymbol{x}) = 0$ otherwise. A curved exponential-family pa-

rameterization is given by

$$\begin{aligned}\eta_{1,k,l}(\boldsymbol{\theta}) &= \theta_{1,k,l},\\ \eta_{2,k,k,t}(\boldsymbol{\theta}) &= \theta_{2,k}\left\{\theta_{3,k}\left[1-\left(1-\frac{1}{\theta_{3,k}}\right)^t\right]\right\}, \quad \theta_{3,k} > \frac{1}{2}.\end{aligned} \tag{2.6}$$

Such terms are called geometrically weighted edgewise shared partner terms [29,30], because the natural parameters $\eta_{2,k,k,t}(\boldsymbol{\theta})$ are based on the geometric sequence $(1-1/\theta_{3,k})^t$, $t = 1, 2, \ldots$. It is worth noting that the corresponding geometric series converges as long as $\theta_{3,k} > 1/2$ and that $\theta_{3,k} \leq 1/2$ is problematic on probabilistic and statistical grounds [54,59]. The parameterization is called a curved exponential-family parameterization, because the natural parameter vector $\boldsymbol{\eta}(\boldsymbol{\theta})$ is a non-affine function of a lower-dimensional parameter vector $\boldsymbol{\theta}$; see Remark 5 in Section 3.2. Last, but not least, note that in the special case $\theta_{3,k} = 1$ $(k = 1, \ldots, K)$ the curved exponential family reduces to the canonical exponential family described in Section 2.1.

2.3. Example: Stochastic block models

A well-known, but restrictive special case of exponential families with local dependence and support $\mathbb{X} = \{0,1\}^{\binom{n}{2}}$ are stochastic block models [45]. Stochastic block models assume that all edge variables $X_{i,j}$ are independent given the block structure, which implies that $p_{\boldsymbol{\eta}}(\boldsymbol{x})$ can be written as

$$p_{\boldsymbol{\eta}}(\boldsymbol{x}) \propto \exp\left(\sum_{k\leq l}^{K} \eta_{1,k,l} \sum_{i\in\mathcal{A}_k,\, j\in\mathcal{A}_l} x_{i,j}\right), \tag{2.7}$$

where $\eta_{1,k,l}$ is the log odds of the probability of an edge between nodes in blocks $\mathcal{A}_k$ and $\mathcal{A}_l$.

2.4. Added value of exponential families with local dependence

Exponential families with local dependence can capture many features of random graphs within blocks, in contrast to stochastic block models, and can therefore be worth the additional costs in terms of model complexity.

To demonstrate the added value of exponential families with local dependence compared with stochastic block models, first note that many network data sets show evidence of systematic deviations from models which assume that edges are independent, as has been well-documented since the 1970s (see, e.g., [24,48,49]). Stochastic block models assume that edges are independent within and between blocks and cannot capture such systematic deviations from independence. For example, suppose that $\boldsymbol{x} \in \mathbb{X}$ is observed and the block structure is known, and let $s_{1,k,k}(\boldsymbol{x})$ be the observed number of edges and $s_{2,k,k}(\boldsymbol{x})$ be the observed number of transitive edges in block $\mathcal{A}_k$ $(k = 1, \ldots, K)$. A helpful observation for comparing exponential families with local dependence and stochastic block models is that stochastic block models are special cases of exponential families with local dependence and natural parameter vectors $\boldsymbol{\eta}_{k,k} = (\eta_{1,k,k}, \eta_{2,k,k}) = (\eta_{1,k,k}, 0)$

as described in Section 2.1, where $\eta_{1,k,k}$ and $\eta_{2,k,k}$ are the natural edge and transitive edge parameter of block $\mathcal{A}_k$, respectively. If the natural parameter vector $\boldsymbol{\eta}_{k,k} = (\eta_{1,k,k}, 0)$ of block $\mathcal{A}_k$ is estimated by the maximum likelihood estimator $\widehat{\boldsymbol{\eta}}_{k,k} = (\widehat{\eta}_{1,k,k}, 0)$ under stochastic block models with known block structure, then the maximum likelihood estimator solves

$$\mathbb{E}_{\widehat{\eta}_{1,k,k},\eta_{2,k,k}=0}\, s_{1,k,k}(\boldsymbol{X}) = s_{1,k,k}(\boldsymbol{x}), \quad k = 1, \ldots, K, \tag{2.8}$$

provided the maximum likelihood estimator exists [23,50]. However, network data sets may have many more transitive edges within blocks than expected under stochastic block models. In other words, we may observe that

$$s_{2,k,k}(\boldsymbol{x}) \gg \mathbb{E}_{\widehat{\eta}_{2,k,k},\eta_{2,k,k}=0}\, s_{2,k,k}(\boldsymbol{X}) \quad \text{for some or all } k \in \{1, \ldots, K\}. \tag{2.9}$$

To capture such systematic deviations from stochastic block models, exponential families with local dependence can be useful. To see that, note that classic exponential-family theory ([11], Corollary 2.5, page 37) implies that, for any $\eta_{2,k,k} > 0$,

$$\mathbb{E}_{\eta_{1,k,k},\eta_{2,k,k}>0}\, s_{2,k,k}(\boldsymbol{X}) > \mathbb{E}_{\eta_{1,k,k},\eta_{2,k,k}=0}\, s_{2,k,k}(\boldsymbol{X}), \quad k = 1, \ldots, K. \tag{2.10}$$

In other words, the expected number of transitive edges in block $\mathcal{A}_k$ is greater under exponential families with local dependence with $\eta_{2,k,k} > 0$ than under stochastic block models with $\eta_{2,k,k} = 0$, assuming that both have the same edge parameters $\eta_{1,k,k}$ ($k = 1, \ldots, K$). As a consequence, exponential families with local dependence can capture an excess in the expected number of transitive edges within blocks, relative to stochastic block models. In fact, the maximum likelihood estimator $\widehat{\boldsymbol{\eta}}_{k,k} = (\widehat{\eta}_{1,k,k}, \widehat{\eta}_{2,k,k})$ of block $\mathcal{A}_k$ under exponential families with local dependence and known block structure solves

$$\begin{aligned} \mathbb{E}_{\widehat{\eta}_{1,k,k},\widehat{\eta}_{2,k,k}}\, s_{1,k,k}(\boldsymbol{X}) &= s_{1,k,k}(\boldsymbol{x}), \quad k = 1, \ldots, K, \\ \mathbb{E}_{\widehat{\eta}_{1,k,k},\widehat{\eta}_{2,k,k}}\, s_{2,k,k}(\boldsymbol{X}) &= s_{2,k,k}(\boldsymbol{x}), \quad k = 1, \ldots, K, \end{aligned} \tag{2.11}$$

provided the maximum likelihood estimator exists [23,50]. Thus, exponential families with local dependence can match both the observed number of edges and transitive edges within blocks, in contrast to stochastic block models. As a consequence, exponential families with local dependence can outperform stochastic block models in terms of transitivity (see, e.g., the empirical results of [65], where the blocks are known and correspond to school classes in schools).

More generally, exponential families with local dependence can capture many features of random graphs that induce dependence among edges within blocks, including – but not limited to – transitivity. The flexibility of the exponential-family framework and its ability to capture many features of random graphs within blocks is one of its greatest advantages. However, it is worth noting that not all specifications of exponential-family models with local dependence are equally useful: for example, it is well known that exponential-family models with k-star and triangle terms can induce undesirable behavior in large random graphs, such as model near-degeneracy [13,23,33,54]. Thus, within-block k-star and triangle terms can be used as long as the blocks are not too large, but should not be used when the blocks are large. Other specifications of exponential-family models are more appropriate for large blocks, for example, the specifications

described in Sections 2.1 and 2.2: each of them implies that the value added by additional triangles to the log odds of the conditional probability of an edge, given all other edges, decays (see, e.g., [28,30,57,63]). By contrast, models with triangle terms make the implicit assumption that the added value of additional triangles does not decay, which can lead to undesirable behavior in large random graphs and hence large within-block subgraphs [13,23,33,54]. However, note that the restriction that blocks cannot be too large – which we discuss in Remark 3 in Section 3 – ensures that the effect of less appropriate within-block specifications (such as within-block triangle or k-star terms) on the random graph remains limited.

2.5. Notation

Throughout, $\mathbb{E} f(\boldsymbol{X})$ denotes the expectation of a function $f : \mathbb{X} \mapsto \mathbb{R}$ of a random graph with respect to exponential-family distributions $\mathbb{P}$ admitting densities of the form (2.1). We write $\mathbb{P} \equiv \mathbb{P}_{\boldsymbol{\eta}^\star}$ and $\mathbb{E} \equiv \mathbb{E}_{\boldsymbol{\eta}^\star}$, where $\boldsymbol{\eta}^\star \in \boldsymbol{\Xi} \subseteq \text{int}(\mathbb{N})$ denotes the data-generating natural parameter vector and $\boldsymbol{\Xi} \subseteq \text{int}(\mathbb{N})$ denotes a subset of the interior $\text{int}(\mathbb{N})$ of the natural parameter space $\mathbb{N} = \{\boldsymbol{\eta} \in \mathbb{R}^{\dim(\boldsymbol{\eta})} : \psi(\boldsymbol{\eta}) < \infty\}$. We assume that $\boldsymbol{\eta} : \boldsymbol{\Theta} \times \mathbb{Z} \mapsto \boldsymbol{\Xi}$ is a function of $(\boldsymbol{\theta}, z) \in \boldsymbol{\Theta} \times \mathbb{Z}$, where

$$\boldsymbol{\Theta} \times \mathbb{Z} = \big\{(\boldsymbol{\theta}, z) \in \mathbb{R}^{\dim(\boldsymbol{\theta})} \times \{1, \ldots, K\}^n : \psi\big(\boldsymbol{\eta}(\boldsymbol{\theta}, z)\big) < \infty\big\}. \tag{2.12}$$

Here, $\boldsymbol{\theta}$ is a vector of block-dependent parameters of dimension $\dim(\boldsymbol{\theta}) \leq \dim(\boldsymbol{\eta})$ while z is a vector of block memberships of nodes. The natural parameter vectors of the canonical and curved exponential families described in Sections 2.1 and 2.2 can be represented in this form. The data-generating value of $(\boldsymbol{\theta}, z) \in \boldsymbol{\Theta} \times \mathbb{Z}$ is denoted by $(\boldsymbol{\theta}^\star, z^\star)$. The ℓ_1-, ℓ_2-, and ℓ_∞-norm of vectors are denoted by $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$, respectively. Uppercase letters $A, B, C > 0$ denote unspecified constants, which may be recycled from line to line.

3. Consistent estimation of block structure

We present here the first consistency results which show that it is possible to recover the block structure with high probability under weak dependence and smoothness conditions. These consistency results are non-trivial, because we cover exponential families with (a) countable support; (b) a wide range of dependencies within blocks; and (c) a wide range of canonical and curved exponential-family parameterizations.

To recover the block structure along with the parameters given an observation $\boldsymbol{x}$ of $\boldsymbol{X}$, we consider the following restricted maximum likelihood estimator:

$$(\widehat{\boldsymbol{\theta}}, \widehat{z}) \in \underset{(\boldsymbol{\theta}, z) \in \boldsymbol{\Theta}_0 \times \mathbb{Z}_0}{\arg\max}\ \ell\big(\boldsymbol{\theta}, z; s(\boldsymbol{x})\big), \tag{3.1}$$

where

$$\ell\big(\boldsymbol{\theta}, z; s(\boldsymbol{x})\big) = \big\langle \boldsymbol{\eta}(\boldsymbol{\theta}, z), s(\boldsymbol{x})\big\rangle - \psi\big(\boldsymbol{\eta}(\boldsymbol{\theta}, z)\big) \tag{3.2}$$

denotes the loglikelihood function of $(\boldsymbol{\theta}, z) \in \boldsymbol{\Theta}_0 \times \mathbb{Z}_0$ and $\boldsymbol{\Theta}_0 \times \mathbb{Z}_0$ is a subset of $\boldsymbol{\Theta} \times \mathbb{Z}$ to be specified. Computational implications are discussed in Section 6. We assume that the number of

blocks K is known and that both $\boldsymbol{\theta}$ and $\boldsymbol{z}$ are parameters, which is commonplace in the special case of stochastic block models (e.g., [3,7,15]). It is worth noting that the maximum likelihood estimator $(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{z}})$ is not unique, because the likelihood function is invariant to the labeling of blocks. All following statements are therefore understood as statements about equivalence classes of block structures.

We call the maximum likelihood estimator $(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{z}})$ restricted, because we restrict maximum likelihood estimation to a subset $\boldsymbol{\Theta}_0 \times \mathbb{Z}_0$ of $\boldsymbol{\Theta} \times \mathbb{Z}$. We need to do so, because without additional restrictions exponential families with local dependence can induce strong dependence and smoothness problems. To motivate the restrictions on $\boldsymbol{\Theta} \times \mathbb{Z}$, it is instructive to discuss the following concentration result, which is instrumental to deriving the main consistency results of the paper.

Lemma 1. *Suppose that a random graph is governed by an exponential family with local dependence and countable support* $\mathbb{X}$. *Let* $f : \mathbb{X} \mapsto \mathbb{R}$ *be Lipschitz with respect to the Hamming metric* $d : \mathbb{X} \times \mathbb{X} \mapsto \{0, \dots, \binom{n}{2}\}$ *defined by*

$$d(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sum_{i<j}^{n} \mathbb{1}_{x_{1,i,j} \neq x_{2,i,j}}, \quad (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{X} \times \mathbb{X}, \tag{3.3}$$

with Lipschitz coefficient $\|f\|_{\text{Lip}} > 0$ *and expectation* $\mathbb{E}|f(\boldsymbol{X})| < \infty$. *Then there exists* $C > 0$ *such that, for all* $n > 0$ *and all* $t > 0$,

$$\mathbb{P}\big(\big|f(\boldsymbol{X}) - \mathbb{E}f(\boldsymbol{X})\big| \geq t\big) \leq 2\exp\bigg(-\frac{t^2}{Cn^2\|\mathcal{A}\|_\infty^4 \|f\|_{\text{Lip}}^2}\bigg), \tag{3.4}$$

where $\|\mathcal{A}\|_\infty = \max_{1 \leq k \leq K} |\mathcal{A}_k| > 0$ *denotes the size of the largest data-generating block.*

The proof of Lemma 1 can be found in the supplementary materials. The proof relies on concentration of measure inequalities for dependent random variables [34] and bounds mixing coefficients – which quantify the strength of dependence induced by exponential families with local dependence – in terms of $\|\mathcal{A}\|_\infty$.

Lemma 1 demonstrates that the probability mass of a function $f(\boldsymbol{X})$ of a random graph concentrates around the corresponding expectation $\mathbb{E}f(\boldsymbol{X})$ as long as the data-generating exponential family induces weak dependence and the function $f(\boldsymbol{X})$ satisfies smoothness conditions. We are interested in applying Lemma 1 to concentrate exponential-family loglikelihood functions of the form $\ell(\boldsymbol{\theta}, \boldsymbol{z}; s(\boldsymbol{X})) = \log p_{\boldsymbol{\eta}(\boldsymbol{\theta},\boldsymbol{z})}(\boldsymbol{X})$. To make sure that the probability mass of $\log p_{\boldsymbol{\eta}(\boldsymbol{\theta},\boldsymbol{z})}(\boldsymbol{X})$ concentrates around the expectation $\mathbb{E}\log p_{\boldsymbol{\eta}(\boldsymbol{\theta},\boldsymbol{z})}(\boldsymbol{X})$, we need to impose additional restrictions on $\mathbb{Z}$ for at least two reasons. First of all, large blocks can induce strong dependence, which weakens concentration results – as can be seen from the term $\|\mathcal{A}\|_\infty$ in Lemma 1. Second, changes of edges in large blocks can give rise to large changes of $\log p_{\boldsymbol{\eta}(\boldsymbol{\theta},\boldsymbol{z})}(\boldsymbol{x})$, which weakens concentration results as well – as can be seen from the Lipschitz coefficient $\|f\|_{\text{Lip}}$ in Lemma 1. Thus, to deal with strong dependence and smoothness problems, restrictions need to be imposed on the sizes of blocks in $\mathbb{Z}$. An additional issue is that the unrestricted maximum likelihood estimator fails to exist with non-negligible probability [23,50]. These observations motivate the following assumptions.

3.1. Assumptions

We assume that the data-generating natural parameter vector $\boldsymbol{\eta}^\star \in \boldsymbol{\Xi} \subseteq \text{int}(\mathbb{N})$ is in the interior $\text{int}(\mathbb{N})$ of the natural parameter space $\mathbb{N}$, which implies that the expectation $\mathbb{E}\, s(\boldsymbol{X})$ exists and is finite ([11], Theorem 2.2, pages 34–35) and so does the expectation $\mathbb{E}\, \ell(\boldsymbol{\theta}, z; s(\boldsymbol{X}))$, because

$$\mathbb{E}\, \ell\big(\boldsymbol{\theta}, z; s(\boldsymbol{X})\big) = \big\langle \boldsymbol{\eta}(\boldsymbol{\theta}, z), \mathbb{E}\, s(\boldsymbol{X})\big\rangle - \psi\big(\boldsymbol{\eta}(\boldsymbol{\theta}, z)\big) = \ell\big(\boldsymbol{\theta}, z; \mathbb{E}\, s(\boldsymbol{X})\big). \tag{3.5}$$

Let $\boldsymbol{\mu}(\boldsymbol{\eta}) = \mathbb{E}_{\boldsymbol{\eta}}\, s(\boldsymbol{X})$ be the mean-value parameter vector of an exponential family with natural parameter vector $\boldsymbol{\eta} \equiv \boldsymbol{\eta}(\boldsymbol{\theta}, z)$ and $\mathbb{M} = \text{rint}(\mathbb{C})$ be the mean-value parameter space, where $\text{rint}(\mathbb{C})$ is the relative interior of the convex hull $\mathbb{C} = \text{conv}\{s(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{X}\}$ of the set $\{s(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{X}\}$. It is well known that in minimal exponential families the mapping between the relative interior of the mean-value and natural parameter space is one-to-one ([11], Theorem 3.6, page 74), and that all non-minimal exponential families can be reduced to minimal exponential families ([11], Theorem 1.9, page 13). Denote by $\boldsymbol{\mu}^\star \equiv \boldsymbol{\mu}(\boldsymbol{\eta}^\star)$ the data-generating mean-value parameter vector. For any $\alpha > 0$, let

$$\mathbb{M}(\alpha) = \big\{\boldsymbol{\mu} \in \mathbb{M} : \big|\ell\big(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}\big) - \ell\big(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star\big)\big| < \alpha\big|\ell\big(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star\big)\big|\big\} \tag{3.6}$$

be the subset of mean-value parameter vectors $\boldsymbol{\mu} \in \mathbb{M}$ that are close to the data-generating mean-value parameter vector $\boldsymbol{\mu}^\star \in \mathbb{M}$ in the sense that $|\ell(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}) - \ell(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star)| < \alpha|\ell(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star)|$. The advantage of introducing the subset $\mathbb{M}(\alpha)$ of $\mathbb{M}$ is that the main assumptions stated below can be weakened, because some of them need to hold on $\mathbb{M}(\alpha)$, but need not hold on $\mathbb{M} \setminus \mathbb{M}(\alpha)$.

The main assumptions can be stated as follows; note that conditions [C.2] and [C.3] are assumed to hold on $\mathbb{M}(\alpha)$, but need not hold on $\mathbb{M} \setminus \mathbb{M}(\alpha)$.

[C.1] For any fixed $z \in \mathbb{Z}$, the map $\boldsymbol{\eta} : \boldsymbol{\Theta} \times \mathbb{Z} \mapsto \boldsymbol{\Xi}$ is one-to-one and continuous on $\boldsymbol{\Theta}$.

[C.2] For any fixed $z \in \mathbb{Z}$ and any fixed $\boldsymbol{\mu} \in \mathbb{M}(\alpha)$, the loglikelihood function $\ell(\boldsymbol{\theta}, z; \boldsymbol{\mu})$ is upper semicontinuous on $\boldsymbol{\Theta}$.

[C.3] There exist $A_1 > 0$ and $n_1 > 0$ such that, for all $n > n_1$, all $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \boldsymbol{\Theta} \times \boldsymbol{\Theta}$, all $z \in \mathbb{Z}$, and all $\boldsymbol{\mu} \in \mathbb{M}(\alpha)$,

$$\big|\big\langle \boldsymbol{\eta}(\boldsymbol{\theta}_1, z) - \boldsymbol{\eta}(\boldsymbol{\theta}_2, z), \boldsymbol{\mu}\big\rangle\big| \leq A_1 \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|_2 \,\big|\ell\big(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star\big)\big|. \tag{3.7}$$

[C.4] There exist $A_2 > 0$ and $n_2 > 0$ such that, for all $n > n_2$, all $(\boldsymbol{\theta}, z) \in \boldsymbol{\Theta} \times \mathbb{Z}$, and all $(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{X} \times \mathbb{X}$,

$$\big|\big\langle \boldsymbol{\eta}(\boldsymbol{\theta}, z), s(\boldsymbol{x}_1) - s(\boldsymbol{x}_2)\big\rangle\big| \leq A_2\, d(\boldsymbol{x}_1, \boldsymbol{x}_2) L(z), \tag{3.8}$$

where $L(z)$ is the size of the largest block under z.

[C.5] The data-generating parameter $(\boldsymbol{\theta}^\star, z^\star)$ is contained in $\boldsymbol{\Theta}_0 \times \mathbb{Z}_0 \subseteq \boldsymbol{\Theta} \times \mathbb{Z}$, where

(a) $\boldsymbol{\Theta}_0$ has dimension $\dim(\boldsymbol{\theta}) \leq An$ and can be covered by $\exp(Cn)$ closed balls $\mathcal{B}(\boldsymbol{\theta}_q, B)$ with centers $\boldsymbol{\theta}_q \in \boldsymbol{\Theta}$ and radius $B > 0$, i.e., $\boldsymbol{\Theta}_0 \subseteq \bigcup_{1 \leq q \leq \exp(Cn)} \mathcal{B}(\boldsymbol{\theta}_q, B)$, where $A, B, C > 0$.

(b) $\mathbb{Z}_0$ consists of all block structures for which the size of each of the K blocks is bounded above by L, where K and L can increase as a function of the number of nodes n.

Corollaries 1 and 2 in Section 3.2 show that conditions [C.1]–[C.4] are satisfied by a wide range of canonical and curved exponential families with local dependence. Condition [C.1] along with the assumption that the exponential family is minimal ensures that $\mathbb{P}_{\boldsymbol{\eta}(\boldsymbol{\theta}_1, z)} \neq \mathbb{P}_{\boldsymbol{\eta}(\boldsymbol{\theta}_2, z)}$ for all $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, for any fixed $z \in \mathbb{Z}$. Conditions [C.2]–[C.4] are smoothness conditions. Condition [C.2] is a weak assumption: it is well known that canonical exponential-family loglikelihood functions are upper semicontinuous ([11], Lemma 5.3, page 146), and it turns out that the most interesting curved exponential-family loglikelihood functions are upper semicontinuous as well, which is verified by Corollaries 1 and 2 in Section 3.2. Condition [C.3] imposes restrictions on how much $\log p_{\boldsymbol{\eta}(\boldsymbol{\theta}, z)}(\boldsymbol{x})$ can change as a function of $\boldsymbol{\eta}(\boldsymbol{\theta}, z)$, whereas condition [C.4] imposes restrictions on how much $\log p_{\boldsymbol{\eta}(\boldsymbol{\theta}, z)}(\boldsymbol{x})$ can change as a function of $\boldsymbol{x}$. Condition [C.3] is stated in terms of $|\ell(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star)|$ to accommodate both sparse and dense random graphs; we discuss the notion of sparse and dense random graphs in Remark 2 in Section 3.2. Condition [C.5](a) allows the dimension $\dim(\boldsymbol{\theta})$ of the parameter space $\boldsymbol{\Theta}_0$ to increase as a function of the number of nodes n and hence allows the model to be flexible while ensuring that $\boldsymbol{\Theta}_0$ cannot be too large. We need these conditions, because we have a single observation of a random graph and therefore cannot use conventional arguments to prove that estimators fall with high probability into compact subsets of the parameter space when the number of observations N is large (e.g., [5]). Condition [C.5](b) complements condition [C.4] and helps ensure that $\log p_{\boldsymbol{\eta}(\boldsymbol{\theta}, z)}(\boldsymbol{x})$ is not too sensitive to changes of $\boldsymbol{x}$ by restricting the set of block structures to blocks whose size is bounded above by L. The main consistency results of the paper, Proposition 1 and Theorem 1 in Section 3.2, impose restrictions on L.

3.2. Main consistency results

We discuss the main consistency results concerning the recovery of block structure given an observation of a random graph with complex dependence.

The recovery of block structure is made possible by the following fundamental concentration result. The concentration result shows that with high probability the distribution parameterized by the restricted maximum likelihood estimator $(\widehat{\boldsymbol{\theta}}, \widehat{z})$ is close to the distribution parameterized by the data-generating parameter $(\boldsymbol{\theta}^\star, z^\star)$ in terms of Kullback–Leibler divergence $\mathrm{KL}(\boldsymbol{\theta}^\star, z^\star; \widehat{\boldsymbol{\theta}}, \widehat{z}) = \ell(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star) - \ell(\widehat{\boldsymbol{\theta}}, \widehat{z}; \boldsymbol{\mu}^\star)$ provided that the number of nodes n is sufficiently large. The result covers a wide range of canonical and curved exponential families with local dependence.

Proposition 1. *Suppose that an observation of a random graph is generated by an exponential family with local dependence and countable support* $\mathbb{X}$ *satisfying conditions* [C.1]–[C.5]. *Assume that, for all* $C_1 > 0$, *however large, there exists* $n_1 > 0$ *such that, for all* $n > n_1$,

$$\left|\ell\left(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star\right)\right| \geq C_1\, n^{3/2}\, \|\mathcal{A}\|_\infty^2\, L\, \sqrt{\log n}, \tag{3.9}$$

where $L = \max_{z \in \mathbb{Z}_0} L(z)$. *Then there exist* $C > 0$, $C_2 > 0$, *and* $n_2 > 0$ *such that, for all* $n > n_2$, *with at least probability* $1 - 2\exp(-\alpha^2 C_2 n \log n)$, *the restricted maximum likelihood estimator* $(\widehat{\boldsymbol{\theta}}, \widehat{z}) \in \boldsymbol{\Theta}_0 \times \mathbb{Z}_0$ *exists and, for all* $\epsilon > 0$,

$$\mathbb{P}\left(\mathrm{KL}\left(\boldsymbol{\theta}^\star, z^\star; \widehat{\boldsymbol{\theta}}, \widehat{z}\right) < \epsilon \left|\ell\left(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star\right)\right|\right) \geq 1 - 4\exp\left(-\min\left(\alpha^2, \epsilon^2\right) C\, n\, \log n\right), \tag{3.10}$$

where $\alpha > 0$ is identical to the constant α used in the construction of the subset $\mathbb{M}(\alpha)$ of the mean-value parameter space $\mathbb{M}$.

The concentration result in Proposition 1 paves the ground for the main consistency result. The consistency result is generic and covers a wide range of canonical and curved exponential families with local dependence. It states that the discrepancy between the estimated and data-generating block structure is small with high probability provided that the number of nodes n is sufficiently large. To define the discrepancy between the estimated and data-generating block structure, let $\delta : \mathbb{Z} \times \mathbb{Z} \mapsto [0, n]$ be a discrepancy measure that is invariant to the labeling of blocks. An example is given by $\delta(z^\star, \widehat{z}) = \min_\pi \sum_{i=1}^n \mathbb{1}_{z_i^\star \neq \pi(\widehat{z}_i)}$, the minimum Hamming distance between $z^\star$ and $\widehat{z}$, where the minimum is taken with respect to all possible permutations π of $\widehat{z}$. The following consistency result holds for all discrepancy measures $\delta : \mathbb{Z} \times \mathbb{Z} \mapsto [0, n]$ satisfying assumption (3.11) of the following result.

Theorem 1. *Suppose that an observation of a random graph is generated by an exponential family with local dependence and countable support $\mathbb{X}$ satisfying conditions* [C.1]–[C.5]. *If the random graph satisfies assumption* (3.9) *of Proposition* 1 *and there exist $C_1 > 0$ and $n_1 > 0$ such that, for all $n > n_1$ and all $(\boldsymbol{\theta}, z) \in \boldsymbol{\Theta}_0 \times \mathbb{Z}_0$,*

$$\mathrm{KL}\big(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\theta}, z\big) \geq \frac{\delta(z^\star, z)\, C_1\, |\ell(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star)|}{n}, \tag{3.11}$$

then there exist $C > 0$, $C_2 > 0$, and $n_2 > 0$ such that, for all $n > n_2$, with at least probability $1 - 2\exp(-\alpha^2 C_2 n \log n)$, the restricted maximum likelihood estimator $(\widehat{\boldsymbol{\theta}}, \widehat{z}) \in \boldsymbol{\Theta}_0 \times \mathbb{Z}_0$ exists and, for all $\epsilon > 0$,

$$\mathbb{P}\left(\frac{\delta(z^\star, \widehat{z})}{n} < \epsilon\right) \geq 1 - 4\exp\big(-\min\big(\alpha^2, \epsilon^2\big)\, C\, n\, \log n\big), \tag{3.12}$$

where $\alpha > 0$ is identical to the constant α used in the construction of the subset $\mathbb{M}(\alpha)$ of the mean-value parameter space $\mathbb{M}$.

We discuss implications of Proposition 1 and Theorem 1, starting with a short comparison with stochastic block models (Remark 1) and then discussing assumption (3.9) (Remark 2) and its implications in terms of the sizes of blocks (Remark 3) and the number of blocks (Remark 4). We then proceed with a discussion of conditions [C.1]–[C.4] (Remark 5) and assumption (3.11) (Remark 6) and conclude with some comments on parameter estimation (Remark 7). Last, but not least, we discuss the sharpness of the results (Remark 8).

Remark 1 (Comparison with stochastic block models). There is a large and growing body of consistency results on stochastic block models (e.g., [3,7,8,12,15,21,39,52,53]). In the language of stochastic block models, the consistency result in Theorem 1 is a weak consistency result in the sense that the discrepancy between the estimated and data-generating block structure is small with high probability. In contrast to stochastic block models, we cover exponential families with (a) countable support; (b) a wide range of dependencies within blocks; and (c) a wide

range of canonical and curved exponential-family parameterizations. These dependencies and parameterizations make theoretical results more challenging from a statistical point of view, but more relevant from a scientific point of view. However, these results come at a cost: in contrast to stochastic block models, we need to restrict the sizes of blocks from above to deal with strong dependence and smoothness problems, as we pointed out in the discussion of Lemma 1. The restrictions on the sizes of blocks are detailed in Remark 3.

Remark 2 (Assumption (3.9): Sparse and dense random graphs). Assumption (3.9) of Proposition 1 and Theorem 1 is stated in terms of the absolute value of the expected loglikelihood function $|\mathbb{E}\,\ell(\boldsymbol{\theta}^\star, z^\star; s(\boldsymbol{X}))| = |\ell(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star)|$ to accommodate sparse and dense random graphs. We first explain why $|\ell(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star)|$ may be interpreted as the level of sparsity of a random graph, and then return to assumption (3.9).

To demonstrate that $|\ell(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star)|$ may be interpreted as the level of sparsity of a random graph, consider the classic Bernoulli(ω) random graphs, under which edges $X_{i,j}$ are independent Bernoulli(ω) random variables [17,18]. It is natural, and conventional, to use the expected number of edges $\mathbb{E}\sum_{i<j}^n X_{i,j}$ to quantify the sparsity of Bernoulli(ω) random graphs, because $\sum_{i<j}^n X_{i,j}$ is a sufficient statistic for the natural parameter $\theta = \operatorname{logit}(\omega)$ of the canonical exponential-family representation of Bernoulli(ω) random graphs. If an exponential family contains more than one natural parameter and one sufficient statistic, it makes sense to quantify the sparsity of a random graph based on all sufficient statistics: in fact, in many applications, the sufficient statistics are of substantive interest, because researchers specify exponential-family models of random graphs by specifying sufficient statistics that capture features of random graphs considered relevant (e.g., the number of edges and transitive edges, see Section 2.4). The question, then, is how the sparsity of a random graph can be quantified based on all sufficient statistics, that is, all relevant features of the random graph. The absolute value of the expected loglikelihood function $|\ell(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star)|$ is a simple choice, because it is a function of all sufficient statistics and the key to likelihood-based inference. In the special case of Bernoulli(ω) random graphs, $|\ell(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star)|$ agrees with $\mathbb{E}\sum_{i<j}^n X_{i,j}$ on the level of sparsity (ignoring logarithmic factors). If a Bernoulli(ω) random graph is dense in the sense that ω does not depend on n, then both $\mathbb{E}\sum_{i<j}^n X_{i,j}$ and $|\ell(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star)|$ are of order n^2 and hence agree on the level of sparsity. If a Bernoulli(ω_n) random graph is sparse in the sense that $\omega_n \to 0$ as $n \to \infty$, then both quantities are smaller: for example, if $\omega_n = \log n/n$ (the threshold for connectivity of Bernoulli random graphs, [10]), then $\mathbb{E}\sum_{i<j}^n X_{i,j}$ and $|\ell(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star)|$ are of order $n\log n$ and $n(\log n)^2$, respectively, so both quantities agree on the level of sparsity up to a logarithmic factor. We therefore interpret $|\ell(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star)|$ as the level of sparsity of a random graph, but note that the mathematical results in Proposition 1 and Theorem 1 hold regardless of how $|\ell(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star)|$ is interpreted.

To return to assumption (3.9), the above considerations suggest that the random graph can be sparse, but cannot be too sparse in the sense that $|\ell(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star)|$ cannot be too small. If, for example, $\|\mathcal{A}\|_\infty$ and L grow as fast as $(\log n)^{\gamma_1}$ ($\gamma_1 > 0$) and $(\log n)^{\gamma_2}$ ($\gamma_2 > 0$), respectively, then $|\ell(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star)|$ must grow faster than $n^{3/2}(\log n)^{2\gamma_1+\gamma_2+1/2}$.

Remark 3 (Sizes of blocks). The sizes of blocks in $\mathbb{Z}_0$ cannot be too large, because changes of edges in large blocks can give rise to large changes of $\ell(\boldsymbol{\theta}, z; s(\boldsymbol{x})) = \log p_{\boldsymbol{\eta}(\boldsymbol{\theta},z)}(\boldsymbol{x})$, which weakens concentration results, as we pointed out in the discussion of Lemma 1. In fact, assumption

(3.9) implies that the size L of the largest possible block in $\mathbb{Z}_0$ must satisfy

$$L \leq \frac{|\ell(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star)|}{C_1\, n^{3/2}\, \|\mathcal{A}\|_\infty^2 \sqrt{\log n}}. \tag{3.13}$$

Thus, in the best-case scenario when $\|\mathcal{A}\|_\infty$ is small in the sense that $\|\mathcal{A}\|_\infty$ grows at most as fast as $(\log n)^\gamma$ ($\gamma > 0$), L can grow at most as fast as $n^{1/2}/(\log n)^{2\gamma+1/2}$, assuming that the random graph is dense. In the worst-case scenario when $\|\mathcal{A}\|_\infty$ grows as fast as L, L can grow at most as fast as $(n/\log n)^{1/6}$.

Remark 4 (Number of blocks). The fact that the sizes of blocks in $\mathbb{Z}_0$ are bounded above by L implies that the number of blocks K is bounded below by $K \geq n/L$. If, e.g., $L \leq n^{1/2}/(\log n)^{2\gamma+1/2}$ ($\gamma > 0$), then $K \geq n^{1/2}(\log n)^{2\gamma+1/2}$. Compared with stochastic block models, the number of blocks K needs to grow at least as fast as in the high-dimensional stochastic block models of Choi et al. [15] (ignoring polylogarithmic terms), where the rate of growth of K is $n^{1/2}$ [15], but K needs not grow as fast as in the highest-dimensional stochastic block models of Rohe et al. [53], where the rate of growth of K is as high as n (ignoring polylogarithmic terms) [53]. It is worth noting that allowing K to increase as a function of n makes sense in applications: Leskovec et al. [40] and others have observed that many real-world networks have small communities, which suggests that K should increase as a function of n, as Rohe, Chatterjee, and Yu ([52], page 1883) and others have pointed out.

Remark 5 (Conditions [C.1]–[C.4]). We show that conditions [C.1]–[C.4] are satisfied by a wide range of canonical and curved exponential families with local dependence. To ease the presentation, we consider dense random graphs, but the following results can be extended to sparse random graphs as long as the random graphs are not too sparse; see Remark 2.

We assume here that $\boldsymbol{\eta} : \boldsymbol{\Theta} \times \mathbb{Z} \mapsto \boldsymbol{\Xi}$ is separable in the sense that $\boldsymbol{\eta}(\boldsymbol{\theta}, z) = \boldsymbol{A}(z)\boldsymbol{b}(\boldsymbol{\theta})$, where $\boldsymbol{A} : \mathbb{Z} \mapsto \mathbb{R}^{\dim(\boldsymbol{\eta}) \times \dim(\boldsymbol{b})}$ and $\boldsymbol{b} : \boldsymbol{\Theta} \mapsto \mathbb{R}^{\dim(\boldsymbol{b})}$; note that, for example, the curved exponential-family parameterization described in Section 2.2 is separable, and so are many other canonical and curved exponential-family parameterizations. Since $\boldsymbol{\eta} : \boldsymbol{\Theta} \times \mathbb{Z} \mapsto \boldsymbol{\Xi}$ is separable, $\boldsymbol{A}(z)$ can be absorbed into the sufficient statistics vector, so that $\boldsymbol{\eta} : \boldsymbol{\Theta} \mapsto \boldsymbol{\Xi}$ can be considered as a function of $\boldsymbol{\theta}$ and $s : \mathbb{X} \times \mathbb{Z} \mapsto \mathbb{R}^{\dim(\boldsymbol{\eta})}$ can be considered as a function of $\boldsymbol{x}$ and z. As a result, we can write

$$\begin{aligned} \big\langle \boldsymbol{\eta}(\boldsymbol{\theta}, z), \boldsymbol{\mu} \big\rangle &= \big\langle \boldsymbol{\eta}(\boldsymbol{\theta}), \boldsymbol{\mu}(z) \big\rangle = \sum_{k \leq l}^{K} \big\langle \boldsymbol{\eta}_{k,l}(\boldsymbol{\theta}), \boldsymbol{\mu}_{k,l}(z) \big\rangle, \\ \big\langle \boldsymbol{\eta}(\boldsymbol{\theta}, z), s(\boldsymbol{x}) \big\rangle &= \big\langle \boldsymbol{\eta}(\boldsymbol{\theta}), s(\boldsymbol{x}, z) \big\rangle = \sum_{k \leq l}^{K} \big\langle \boldsymbol{\eta}_{k,l}(\boldsymbol{\theta}), s_{k,l}(\boldsymbol{x}, z) \big\rangle, \end{aligned} \tag{3.14}$$

where – in an abuse of notation – we write $\boldsymbol{\mu}(z) = \boldsymbol{A}(z)^\top \boldsymbol{\mu}$ ($\boldsymbol{\mu} \in \mathbb{M}(\alpha)$) and $s(\boldsymbol{x}, z) = \boldsymbol{A}(z)^\top s(\boldsymbol{x})$ ($s(\boldsymbol{x}) \in \mathbb{M}(\alpha)$). If, in addition, $\boldsymbol{b}(\boldsymbol{\theta})$ is an affine function of $\boldsymbol{\theta}$, then $\boldsymbol{\eta}(\boldsymbol{\theta})$ can be reduced to $\boldsymbol{\eta}(\boldsymbol{\theta}) = \boldsymbol{\theta}$ and $\boldsymbol{\eta}_{k,l}(\boldsymbol{\theta})$ can be reduced to $\boldsymbol{\eta}_{k,l}(\boldsymbol{\theta}) = \boldsymbol{\theta}_{k,l}$ ($k \leq l = 1, \ldots, K$), in which case we call the exponential family canonical, otherwise we call the exponential family curved. In the following, we denote by $L_k(z)$ the number of nodes in block k under block structure $z \in \mathbb{Z}_0$.

The following result shows that conditions [C.1]–[C.4] are satisfied by all canonical exponential families with local dependence satisfying reasonable scaling and smoothness conditions.

Corollary 1. *Consider canonical exponential families with local dependence and countable support* $\mathbb{X}$. *Assume that* $\boldsymbol{\eta} : \boldsymbol{\Theta} \times \mathbb{Z} \mapsto \boldsymbol{\Xi}$ *is separable with* $\dim(\boldsymbol{\theta}_{k,l}) < \infty$ $(k \le l = 1, \dots, K)$ *and that the random graph is dense. If there exist* $C_1 > 0$, $C_2 > 0$, *and* $n_0 \ge 1$ *such that, for all* $n > n_0$,

[C.3*] $\|\boldsymbol{\mu}_{k,l}(z)\|_\infty \le C_1 L_k(z) L_l(z)$ *for all* $z \in \mathbb{Z}_0$ *and all* $\boldsymbol{\mu} \in \mathbb{M}(\alpha)$ $(k \le l = 1, \dots, K)$;

[C.4*] $\sum_{k \le l}^{K} \|s_{k,l}(\boldsymbol{x}_1, z) - s_{k,l}(\boldsymbol{x}_2, z)\|_\infty \le C_2 d(\boldsymbol{x}_1, \boldsymbol{x}_2) L(z)$ *for all* $(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{X} \times \mathbb{X}$ *and all* $z \in \mathbb{Z}_0$;

then conditions [C.1]–[C.4] *are satisfied. If conditions* [C.5] *and* (3.11) *are satisfied as well, then the conclusions of Theorem* 1 *hold.*

Condition [C.3*] is satisfied by all between- and within-block sufficient statistics for which the absolute value of the expectation is bounded above by a constant multiple of the number of pairs of nodes between blocks and within blocks, respectively: for example, the number of edges and transitive edges within blocks satisfy condition [C.3*] and so do all other sufficient statistics that count the number of pairs of nodes within blocks having specified properties or being related to other nodes in the same block in some specified form. Condition [C.4*] is satisfied by most sufficient statistics, including the number of edges and transitive edges.

We turn to curved exponential families with local dependence. We consider curved exponential families of densities of the form

$$p_{\boldsymbol{\eta}(\boldsymbol{\theta},z)}(\boldsymbol{x}) \propto \exp\bigl(\bigl\langle \boldsymbol{\eta}(\boldsymbol{\theta}), s(\boldsymbol{x}, z)\bigr\rangle\bigr), \tag{3.15}$$

where

$$\bigl\langle \boldsymbol{\eta}(\boldsymbol{\theta}), s(\boldsymbol{x}, z)\bigr\rangle = \sum_{k \le l}^{K} \eta_{1,k,l}(\boldsymbol{\theta}) \sum_{i,j : z_i = k, z_j = l} x_{i,j} + \sum_{k=1}^{K} \sum_{t=1}^{T_k} \eta_{2,k,k,t}(\boldsymbol{\theta}) s_{k,k,t}(\boldsymbol{x}, z), \tag{3.16}$$

where $s_{k,k,t}(\boldsymbol{x}, z)$ are sufficient statistics that induce dependence within blocks (e.g., in case $\mathbb{X} = \{0,1\}^{\binom{n}{2}}$, $s_{k,k,t}(\boldsymbol{x}, z)$ may be the number of pairs of nodes with t edgewise shared partners in block k). Here, the natural parameters are given by

$$\begin{aligned} \eta_{1,k,l}(\boldsymbol{\theta}) &= \theta_{1,k,l}, \\ \eta_{2,k,k,t}(\boldsymbol{\theta}) &= \theta_{2,k}\left\{\theta_{3,k}\left[1 - \left(1 - \frac{1}{\theta_{3,k}}\right)^t\right]\right\}, \quad \theta_{3,k} > \frac{1}{2}, T_k \ge 2. \end{aligned} \tag{3.17}$$

The following result shows that as long as the underlying geometric series converges, that is, as long as $\theta_{3,k} > 1/2$ $(k = 1, \dots, K)$, conditions [C.1]–[C.4] are satisfied. The result can be extended to other model terms, for example, covariate terms.

Corollary 2. *Consider curved exponential families of the form* (3.15) *with local dependence and countable support* $\mathbb{X}$. *Assume that* $\boldsymbol{\eta} : \boldsymbol{\Theta} \times \mathbb{Z} \mapsto \boldsymbol{\Xi}$ *is separable and that there exists* $B > 1/2$ *such that* $1/2 < \theta_{3,k} < B$ $(k = 1, \ldots, K)$ *and that the random graph is dense. If there exist* $C_1 > 0$, $C_2 > 0$, *and* $n_0 \geq 1$ *such that, for all* $n > n_0$,

[C.3**] $\sum_{t=1}^{T_k} |\mu_{k,k,t}(\boldsymbol{z})| \leq C_1 \binom{L_k(\boldsymbol{z})}{2}$ *for all* $\boldsymbol{z} \in \mathbb{Z}_0$, *where* $\mu_{k,k,t}(\boldsymbol{z}) = \mathbb{E}\, s_{k,k,t}(\boldsymbol{X}, \boldsymbol{z})$;

[C.4**] $|\sum_{t=1}^{T_k} s_{k,k,t}(\boldsymbol{x}_1, \boldsymbol{z}) - \sum_{t=1}^{T_k} s_{k,k,t}(\boldsymbol{x}_2, \boldsymbol{z})| \leq C_2 d(\boldsymbol{x}_{1,k,k}, \boldsymbol{x}_{2,k,k}) L(\boldsymbol{z})$ *for all* $(\boldsymbol{x}_{1,k,k}, \boldsymbol{x}_{2,k,k}) \in \mathbb{X}_{k,k}(\boldsymbol{z}) \times \mathbb{X}_{k,k}(\boldsymbol{z})$ *and all* $\boldsymbol{z} \in \mathbb{Z}_0$, *where* $\mathbb{X}_{k,k}(\boldsymbol{z})$ *denotes the set of all possible within-block subgraphs of block* k *under* $\boldsymbol{z} \in \mathbb{Z}_0$ $(k = 1, \ldots, K)$;

then conditions [C.1]–[C.4] *are satisfied. If conditions* [C.5] *and* (3.11) *are satisfied as well, then the conclusions of Theorem* 1 *hold.*

The most popular curved exponential families with geometrically weighted terms [29,30, 63] satisfy conditions [C.3**] and [C.4**] of Corollary 2. Consider, for example, geometrically weighted edgewise shared partner terms. In the case of geometrically weighted edgewise shared partner terms, $T_k = L_k(\boldsymbol{z}) - 2$ and $\sum_{t=1}^{T_k} s_{k,k,t}(\boldsymbol{x}, \boldsymbol{z})$ is the number of transitive edges in block k, hence conditions [C.3**] and [C.4**] are satisfied.

Remark 6 (Assumption (3.11)). Assumption (3.11) of Theorem 1 states that the Kullback–Leibler divergence of the distribution parameterized by $(\boldsymbol{\theta}, \boldsymbol{z})$ from the distribution parameterized by $(\boldsymbol{\theta}^\star, \boldsymbol{z}^\star)$ must increase with the discrepancy measure $\delta(\boldsymbol{z}^\star, \boldsymbol{z})$. In the special case of stochastic block models, [15] and [53] verified identifiability assumption (3.11) using the number of misclassified nodes – as defined by [15] – as a discrepancy measure, where the number of blocks can grow as fast as $n^{1/2}$ [15] and as fast as n (ignoring polylogarithmic terms) [53], respectively. In general, an application of the mean-value theorem to the expected loglikelihood function $\ell(\boldsymbol{\eta}^\star; \boldsymbol{\mu}^\star) = \langle \boldsymbol{\eta}^\star, \boldsymbol{\mu}^\star \rangle - \psi(\boldsymbol{\eta}^\star)$ shows that, for all $\boldsymbol{\eta} \in \boldsymbol{\Xi} \subseteq \text{int}(\mathbb{N})$,

$$\text{KL}\big(\boldsymbol{\eta}^\star; \boldsymbol{\eta}\big) = \ell\big(\boldsymbol{\eta}^\star; \boldsymbol{\mu}^\star\big) - \ell\big(\boldsymbol{\eta}; \boldsymbol{\mu}^\star\big) = \big\langle \boldsymbol{\eta}^\star - \boldsymbol{\eta}, \boldsymbol{\mu}\big(\boldsymbol{\eta}^\star\big) - \boldsymbol{\mu}(\dot{\boldsymbol{\eta}}) \big\rangle, \tag{3.18}$$

where $\dot{\boldsymbol{\eta}} = \lambda \boldsymbol{\eta}^\star + (1 - \lambda)\boldsymbol{\eta}$ $(0 \leq \lambda \leq 1)$; note that $\dot{\boldsymbol{\eta}} \in \text{int}(\mathbb{N})$ since $\boldsymbol{\eta}^\star \in \text{int}(\mathbb{N})$ and $\boldsymbol{\eta} \in \text{int}(\mathbb{N})$ and the natural parameter space $\mathbb{N}$ is convex. Therefore, assumption (3.11) is satisfied as long as changes of blocks give rise to large enough changes of mean-value and natural parameter vectors.

Remark 7 (Estimation of parameters). The restricted maximum likelihood estimator, as defined in (3.1), estimates the parameter vector $\boldsymbol{\theta}$ along with the block structure $\boldsymbol{z}$. We leave the study of theoretical properties of estimators of $\boldsymbol{\theta}$ to future research, but it is worth noting the following. If the blocks are known (e.g., in multilevel networks, [38]), M-estimators of canonical and curved exponential-family random graph models with local dependence are consistent under weak conditions [59]. If the blocks are unknown, M-estimators may not be consistent estimators of the data-generating parameters. Indeed, it is not too hard to see that, for any $\boldsymbol{z} \neq \boldsymbol{z}^\star$ (where $\boldsymbol{z} \in \mathbb{Z}_0$ may be an estimate of $\boldsymbol{z}^\star \in \mathbb{Z}_0$), the estimator

$$\widehat{\boldsymbol{\theta}}(\boldsymbol{z}) = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0}{\arg\max} \big[\ell\big(\boldsymbol{\theta}, \boldsymbol{z}; s(\boldsymbol{x})\big) - \ell\big(\boldsymbol{\theta}^\star, \boldsymbol{z}^\star; s(\boldsymbol{x})\big) \big] \tag{3.19}$$

estimates

$$\dot{\boldsymbol{\theta}}(z) = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0}{\arg\max}\big[\ell\big(\boldsymbol{\theta}, z; \boldsymbol{\mu}^\star\big) - \ell\big(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star\big)\big], \tag{3.20}$$

which is equivalent to minimizing the Kullback–Leibler divergence $\mathrm{KL}(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\theta}, z) = \ell(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star) - \ell(\boldsymbol{\theta}, z; \boldsymbol{\mu}^\star)$ with respect to $\boldsymbol{\theta}$, for fixed $z \in \mathbb{Z}_0$. In other words, $\widehat{\boldsymbol{\theta}}(z)$ is an estimator of the parameter vector $\dot{\boldsymbol{\theta}}(z)$ that is as close as possible to the data-generating parameter vector $\boldsymbol{\theta}^\star$ in terms of Kullback–Leibler divergence, for fixed $z \in \mathbb{Z}_0$. These considerations suggest that $\widehat{\boldsymbol{\theta}}(z)$ may be a consistent estimator of $\dot{\boldsymbol{\theta}}(z)$, but in general $\widehat{\boldsymbol{\theta}}(z)$ is not a consistent estimator of $\boldsymbol{\theta}^\star$, unless $z = z^\star$ [59].

Remark 8 (Sharpness). The results in Proposition 1 and Theorem 1 are not, and cannot be expected to be as sharp as results based on stochastic block models (e.g., [3,7–9,12,15,21,32,39,44,47,52,53,74,75]), for at least three reasons:

- *Dependence*. We are concerned with random graphs with dependent edges within blocks, and concentration results for dependent random variables tend to be weaker than concentration results for independent random variables.
- *The results cover many models with many possible forms of dependence*. One of the greatest advantages of exponential-family models of random graphs – which can be viewed as generalizations of Erdős and Rényi random graphs, GLMs, and Markov random fields for dependent network data – is the flexibility of the exponential-family framework and its ability to model many dependencies within blocks. As a consequence, we do not focus on sharp results in special cases, but on results that cover many models with many possible forms of dependence. Indeed, our concentration results are worst-case results and therefore are not, and cannot be expected to be sharp in special cases.
- *The combination of dependence and sparsity*. Many papers concerned with stochastic block models focus on sparse random graphs for which the expected number of edges grows slower than the number of possible edges $\binom{n}{2}$. While studying random graphs under sparsity assumptions makes sense and has a long tradition in classic random graph theory (e.g., [2,20,43]), it requires sharp concentration results for sparse random graphs. Such results are available for sparse random graphs with independent edges based on, for example, clever applications of Bernstein's and Talagrand's concentration inequalities [2,20,43]: e.g., Choi et al. [15] used Bernstein's concentration inequality to obtain concentration results for sparse random graphs with independent edges and the expected number of edges growing faster than $n(\log n)^{3+\beta}$ ($\beta > 0$). But Bernstein's and Talagrand's concentration inequalities are limited to random graphs with independent edges. To the best of our knowledge, no sharp concentration results have been developed for sparse random graphs with dependence among edges induced by transitivity or other network phenomena. While developing sharp concentration results for sparse random graphs with dependent edges would doubtless be an important contribution to the literature, it is beyond the scope of our paper.

In short, the sharpest results can be obtained when edges within and between blocks are independent (e.g., [3,7–9,12,15,21,32,39,44,47,52,53,74,75]), but those results come at a cost: the assumption that edges are independent within and between blocks may be violated in applications, because network data are dependent data (e.g., [24,42,68]). We remove the assumption that

edges are independent within blocks. It comes at the cost of less sharp results, but the benefit is that exponential families with local dependence can capture many features of random graphs that induce dependence among edges within blocks, including – but not limited to – transitivity, as explained in Section 2.4.

4. Simulation results

To demonstrate that the block structure can be recovered in practice, we simulate data from exponential families with block-dependent edge and transitive edge terms as described in Section 2.1. To estimate the block structure, note that (restricted) maximum likelihood estimators are intractable, because maximization over (as many as) $\exp(n \log K)$ possible partitions of a set of n nodes into K blocks is infeasible unless n is small. The same issue arises in stochastic block models, despite the simplifying assumption that edges are independent conditional on the block structure: see, e.g., Choi et al. [15] and Rohe et al. [53]. Both of these papers are concerned with theoretical results for (restricted) maximum likelihood estimators, but base simulation results on approximate methods, because (restricted) maximum likelihood estimators are intractable: Choi et al. [15] use Markov chain Monte Carlo methods, whereas Rohe et al. [53] use pseudolikelihood methods. We likewise have to resort to approximate methods, and use Bayesian auxiliary-variable methods for exponential families with local dependence [56], as implemented in `R` package `hergm` [58].

We consider networks with $n = 50$, $n = 75$, and $n = 100$ nodes and $K = 5$ blocks $\mathcal{A}_1, \ldots, \mathcal{A}_K$ of equal size. The data-generating natural parameters are given by

$$\begin{aligned} &\eta_{1,k,l} = -\log\left(\frac{n - \min(\mathcal{A}_k, \mathcal{A}_l)}{3} - 1\right), \quad k < l = 1, \ldots, K, \\ &\eta_{1,k,k} = -1, \qquad \eta_{2,k,k} = 1, \quad k = 1, \ldots, K, \end{aligned} \tag{4.1}$$

where the between-block natural parameters $\eta_{1,k,l}$ have been chosen to ensure that, for each node, the expected number of edges between blocks is 3. To deal with the so-called label-switching problem of Bayesian Markov chain Monte Carlo methods – which arises from the invariance of the likelihood function to the labeling of blocks – we follow the Bayesian decision-theoretic approach of [64] and estimate block memberships by assigning each node to its maximum-posterior-probability block [56,58].

Figure 1 shows the fraction of misclassified nodes in terms of the normalized minimum Hamming distance $\delta(\boldsymbol{z}^\star, \widehat{\boldsymbol{z}})/n = \min_\pi \sum_{i=1}^n \mathbb{1}_{z_i^\star \neq \pi(\hat{z}_i)}/n$ based on 100 simulated data sets with $n = 50$, $n = 75$, and $n = 100$ nodes and $K = 5$ blocks of equal size; note that Bayesian methods are too time-consuming to be applied to more than 100 simulated data sets. Figure 1 suggests that the fraction of misclassified nodes is small in most data sets and decreases as the number of nodes increases from $n = 50$ to $n = 100$ and hence the sizes of the blocks increase from 10 to 20.

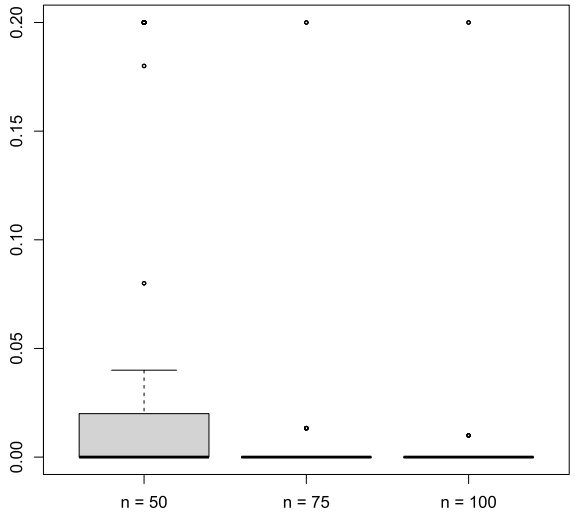

Figure 1. Fraction of misclassified nodes based on 100 simulated data sets with $n = 50$, $n = 75$, and $n = 100$ nodes and $K = 5$ blocks of equal size, where the model is estimated by Bayesian methods.

5. Proofs of main consistency results

We prove the main consistency results, Proposition 1 and Theorem 1. To prove them, we need two additional lemmas, Lemmas 2 and 3. The proofs of Lemmas 1, 2, and 3 are delegated to the supplementary materials along with the proofs of Corollaries 1 and 2.

To state Lemmas 2 and 3, note that the data-generating natural parameter vector $\boldsymbol{\eta}^\star \in \boldsymbol{\Xi} \subseteq \operatorname{int}(\mathbb{N})$ is in the interior $\operatorname{int}(\mathbb{N})$ of the natural parameter space $\mathbb{N}$. Therefore, the expectation $\mathbb{E}\, s(\boldsymbol{X})$ exists and is finite ([11], Theorem 2.2, pages 34–35) and so does the expectation $\mathbb{E}\, \ell(\boldsymbol{\theta}, \boldsymbol{z}; s(\boldsymbol{X})) = \ell(\boldsymbol{\theta}, \boldsymbol{z}; \mathbb{E}\, s(\boldsymbol{X}))$. Let

$$\mathbb{X}(\alpha) = \big\{\boldsymbol{x} \in \mathbb{X} : \big|\ell\big(\boldsymbol{\theta}^\star, \boldsymbol{z}^\star; s(\boldsymbol{x})\big) - \ell\big(\boldsymbol{\theta}^\star, \boldsymbol{z}^\star; \boldsymbol{\mu}^\star\big)\big| < \alpha\big|\ell\big(\boldsymbol{\theta}^\star, \boldsymbol{z}^\star; \boldsymbol{\mu}^\star\big)\big|\big\} \tag{5.1}$$

be the subset of $\boldsymbol{x} \in \mathbb{X}$ such that $s(\boldsymbol{x}) \in \mathbb{M}(\alpha)$, where $\alpha > 0$ is identical to the constant α used in the construction of the subset $\mathbb{M}(\alpha)$ of the mean-value parameter space $\mathbb{M}$.

Lemma 2 shows that the event $\boldsymbol{X} \in \mathbb{X}(\alpha)$ occurs with high probability provided that the number of nodes n is sufficiently large and hence all probability statements in Proposition 1 and Theorem 1 can be restricted to the high-probability subset $\mathbb{X}(\alpha)$ of $\mathbb{X}$.

Lemma 2. *Suppose that an observation of a random graph is generated by an exponential family with local dependence and countable support* $\mathbb{X}$ *satisfying condition* [C.4] *along with assumption* (3.9). *Then there exist* $C > 0$ *and* $n_0 > 0$ *such that, for all* $n > n_0$,

$$\mathbb{P}\big(\boldsymbol{X} \in \mathbb{X}(\alpha)\big) \geq 1 - 2\exp\big(-\alpha^2 C n \log n\big), \tag{5.2}$$

where $\alpha > 0$ *is identical to the constant* α *used in the construction of the subset* $\mathbb{M}(\alpha)$ *of the mean-value parameter space* $\mathbb{M}$.

Lemma 3 shows that in the event $\boldsymbol{X} \in \mathbb{X}(\alpha)$ the restricted maximum likelihood estimator $(\widehat{\boldsymbol{\theta}}, \widehat{z})$ exists, which implies that the restricted maximum likelihood estimator $(\widehat{\boldsymbol{\theta}}, \widehat{z})$ exists with high probability by Lemma 2.

Lemma 3. *Suppose that an observation of a random graph is generated by an exponential family with local dependence and countable support* $\mathbb{X}$ *satisfying conditions* [C.2] *and* [C.4] *along with assumption* (3.9). *Then the following statements hold*:

(a) *For all* $\boldsymbol{x} \in \mathbb{X}(\alpha)$, *the restricted maximum likelihood estimator* $(\widehat{\boldsymbol{\theta}}, \widehat{z})$ *exists*;
(b) *There exist* $C > 0$ *and* $n_0 > 0$ *such that, for all* $n > n_0$, *the restricted maximum likelihood estimator* $(\widehat{\boldsymbol{\theta}}, \widehat{z})$ *exists with at least probability* $1 - 2\exp(-\alpha^2 C n \log n)$;

where $\alpha > 0$ *is identical to the constant* α *used in the construction of the subset* $\mathbb{M}(\alpha)$ *of the mean-value parameter space* $\mathbb{M}$.

Armed with Lemmas 2 and 3, we can prove Proposition 1 and Theorem 1.

Proof of Proposition 1. Throughout, to ease the presentation, we use the short-hand expression

$$u(n) = \left|\ell\left(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star\right)\right|. \tag{5.3}$$

By Lemma 2, there exist $C_0 > 0$ and $n_0 > 0$ such that, for all $n > n_0$,

$$\mathbb{P}\left(\mathbb{X} \setminus \mathbb{X}(\alpha)\right) \leq 2\exp\left(-\alpha^2 C_0\, n \log n\right). \tag{5.4}$$

Thus, all following arguments can be restricted to the high-probability subset $\mathbb{X}(\alpha)$ of $\mathbb{X}$. It is therefore convenient to bound the probability of the event $\mathrm{KL}(\boldsymbol{\theta}^\star, z^\star; \widehat{\boldsymbol{\theta}}, \widehat{z}) \geq \epsilon u(n)$ by using a divide- and conquer strategy based on the inequality

$$\begin{aligned}
&\mathbb{P}\left(\mathrm{KL}\left(\boldsymbol{\theta}^\star, z^\star; \widehat{\boldsymbol{\theta}}, \widehat{z}\right) \geq \epsilon u(n)\right) \\
&\quad \leq \mathbb{P}\left(\mathrm{KL}\left(\boldsymbol{\theta}^\star, z^\star; \widehat{\boldsymbol{\theta}}, \widehat{z}\right) \geq \epsilon u(n) \cap \mathbb{X}(\alpha)\right) + \mathbb{P}\left(\mathbb{X} \setminus \mathbb{X}(\alpha)\right).
\end{aligned} \tag{5.5}$$

The advantage of doing so is that we can confine attention to observations $s(\boldsymbol{x}) \in \mathbb{M}(\alpha)$ that fall into well-behaved subsets $\mathbb{M}(\alpha)$ of the mean-value parameter space $\mathbb{M}$ satisfying conditions [C.2] and [C.3]. Observe that conditions [C.2] and [C.3] are assumed to hold on $\mathbb{M}(\alpha)$, but need not hold on $\mathbb{M} \setminus \mathbb{M}(\alpha)$.

To bound the probability of event $\mathrm{KL}(\boldsymbol{\theta}^\star, z^\star; \widehat{\boldsymbol{\theta}}, \widehat{z}) \geq \epsilon u(n) \cap \mathbb{X}(\alpha)$, note that, for any $\boldsymbol{x} \in \mathbb{X}(\alpha)$, the restricted maximum likelihood estimator $(\widehat{\boldsymbol{\theta}}, \widehat{z})$ exists by Lemma 3 and

$$\mathrm{KL}\left(\boldsymbol{\theta}^\star, z^\star; \widehat{\boldsymbol{\theta}}, \widehat{z}\right) = \ell\left(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star\right) - \ell\left(\widehat{\boldsymbol{\theta}}, \widehat{z}; \boldsymbol{\mu}^\star\right) \geq 0. \tag{5.6}$$

Since $(\widehat{\boldsymbol{\theta}}, \widehat{z}) \in \boldsymbol{\Theta}_0 \times \mathbb{Z}_0$ maximizes $\ell(\boldsymbol{\theta}, z; s(\boldsymbol{x}))$ and $(\boldsymbol{\theta}^\star, z^\star) \in \boldsymbol{\Theta}_0 \times \mathbb{Z}_0$, we have

$$\begin{aligned}
&\ell\left(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star\right) + \left[\ell\left(\boldsymbol{\theta}^\star, z^\star; s(\boldsymbol{x})\right) - \ell\left(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star\right)\right] \\
&\quad \leq \ell\left(\widehat{\boldsymbol{\theta}}, \widehat{z}; \boldsymbol{\mu}^\star\right) + \left[\ell\left(\widehat{\boldsymbol{\theta}}, \widehat{z}; s(\boldsymbol{x})\right) - \ell\left(\widehat{\boldsymbol{\theta}}, \widehat{z}; \boldsymbol{\mu}^\star\right)\right]
\end{aligned} \tag{5.7}$$

and hence $\mathrm{KL}(\boldsymbol{\theta}^\star, z^\star; \widehat{\boldsymbol{\theta}}, \widehat{z})$ can be bounded above as follows:

$$\ell\big(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star\big) - \ell\big(\widehat{\boldsymbol{\theta}}, \widehat{z}; \boldsymbol{\mu}^\star\big) \leq 2 \max_{z \in \mathbb{Z}_0} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} \big|\ell\big(\boldsymbol{\theta}, z; s(\boldsymbol{x})\big) - \ell\big(\boldsymbol{\theta}, z; \boldsymbol{\mu}^\star\big)\big|. \tag{5.8}$$

Choose any $\rho > 0$ satisfying $0 < \rho < \epsilon/(12A_1)$, where $A_1 > 0$ is equal to the constant $A_1 > 0$ in condition [C.3]. By condition [C.5], there exist $A, B, C > 0$ such that the $\dim(\boldsymbol{\theta}) \leq An$-dimensional parameter space $\boldsymbol{\Theta}_0 \subseteq \boldsymbol{\Theta}$ can be covered by $\exp(Cn)$ closed balls with centers $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and radius $B > 0$. Each of the $\exp(Cn)$ balls with radius $B > 0$ can be covered by

$$\left(\frac{4B + \rho}{\rho}\right)^{\dim(\boldsymbol{\theta})} \tag{5.9}$$

balls $\mathcal{B}(\boldsymbol{\theta}, \rho)$ with centers $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and radius $\rho > 0$. Therefore, $\boldsymbol{\Theta}_0 \subseteq \bigcup_{1 \leq q \leq Q} \mathcal{B}(\boldsymbol{\theta}_q, \rho)$ can be covered by Q balls $\mathcal{B}(\boldsymbol{\theta}_q, \rho)$ with centers $\boldsymbol{\theta}_q \in \boldsymbol{\Theta}$ and radius $\rho > 0$, where Q is bounded above by

$$Q \leq \exp\left(A \log\left(\frac{4B + \rho}{\rho}\right) n + Cn\right). \tag{5.10}$$

As a result, we can write

$$\ell\big(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star\big) - \ell\big(\widehat{\boldsymbol{\theta}}, \widehat{z}; \boldsymbol{\mu}^\star\big) \leq 2 \max_{z \in \mathbb{Z}_0} \max_{1 \leq q \leq Q} \sup_{\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}_q, \rho)} \big|\ell\big(\boldsymbol{\theta}, z; s(\boldsymbol{x})\big) - \ell\big(\boldsymbol{\theta}, z; \boldsymbol{\mu}^\star\big)\big|. \tag{5.11}$$

Collecting terms shows that

$$\begin{aligned}
&\mathbb{P}\big(\mathrm{KL}\big(\boldsymbol{\theta}^\star, z^\star; \widehat{\boldsymbol{\theta}}, \widehat{z}\big) \geq \epsilon u(n) \cap \mathbb{X}(\alpha)\big) \\
&\quad = \mathbb{P}\big(\ell\big(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star\big) - \ell\big(\widehat{\boldsymbol{\theta}}, \widehat{z}; \boldsymbol{\mu}^\star\big) \geq \epsilon u(n) \cap \mathbb{X}(\alpha)\big) \\
&\quad \leq \mathbb{P}\left(\max_{z \in \mathbb{Z}_0} \max_{1 \leq q \leq Q} \sup_{\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}_q, \rho)} \big|\ell\big(\boldsymbol{\theta}, z; s(\boldsymbol{X})\big) - \ell\big(\boldsymbol{\theta}, z; \boldsymbol{\mu}^\star\big)\big| \geq \frac{\epsilon u(n)}{2} \cap \mathbb{X}(\alpha)\right).
\end{aligned} \tag{5.12}$$

To bound the probability of the max-sup of deviations of the form $|\ell(\boldsymbol{\theta}, z; s(\boldsymbol{X})) - \ell(\boldsymbol{\theta}, z; \boldsymbol{\mu}^\star)|$, observe that, for any $\boldsymbol{x} \in \mathbb{X}(\alpha)$, the deviation reduces to

$$\big|\ell\big(\boldsymbol{\theta}, z; s(\boldsymbol{x})\big) - \ell\big(\boldsymbol{\theta}, z; \boldsymbol{\mu}^\star\big)\big| = \big|\big\langle \boldsymbol{\eta}(\boldsymbol{\theta}, z), s(\boldsymbol{x})\big\rangle - \big\langle \boldsymbol{\eta}(\boldsymbol{\theta}, z), \boldsymbol{\mu}^\star\big\rangle\big|, \tag{5.13}$$

because $\psi(\boldsymbol{\eta}(\boldsymbol{\theta}, z))$ cancels. Consider any $z \in \mathbb{Z}_0$ and any of the Q balls $\mathcal{B}(\boldsymbol{\theta}_q, \rho)$ that make up the cover $\bigcup_{1 \leq q \leq Q} \mathcal{B}(\boldsymbol{\theta}_q, \rho)$ of $\boldsymbol{\Theta}_0$. Let

$$\dot{\boldsymbol{\theta}}_q(z) = \operatorname*{arg\,max}_{\boldsymbol{\theta} \in \mathrm{cl}\, \mathcal{B}(\boldsymbol{\theta}_q, \rho)} \ell\big(\boldsymbol{\theta}, z; \boldsymbol{\mu}^\star\big), \tag{5.14}$$

where the subscript q is added to indicate the closed ball $\mathrm{cl}\, \mathcal{B}(\boldsymbol{\theta}_q, \rho)$ that contains $\dot{\boldsymbol{\theta}}_q(z)$. Observe that, for any $z \in \mathbb{Z}_0$, $\ell(\boldsymbol{\theta}, z; \boldsymbol{\mu}^\star)$ is upper semicontinuous on $\mathrm{cl}\, \mathcal{B}(\boldsymbol{\theta}_q, \rho)$ by condition [C.2] and hence assumes a maximum on $\mathrm{cl}\, \mathcal{B}(\boldsymbol{\theta}_q, \rho)$. Thus, for any $z \in \mathbb{Z}_0$, the maximizer $\dot{\boldsymbol{\theta}}_q(z)$ exists and

is unique by condition [C.1] and the assumption that the exponential family is minimal, which can be assumed without loss ([11], Theorem 1.9, page 13). The triangle inequality shows that, for any $\boldsymbol{x} \in \mathbb{X}(\alpha)$, any $z \in \mathbb{Z}_0$, any $\boldsymbol{\theta} \in \operatorname{cl}\mathcal{B}(\boldsymbol{\theta}_q, \rho)$, and any $\dot{\boldsymbol{\theta}}_q(z) \in \operatorname{cl}\mathcal{B}(\boldsymbol{\theta}_q, \rho)$,

$$\begin{aligned}
\big|\ell\big(\boldsymbol{\theta}, z; s(\boldsymbol{x})\big) - \ell\big(\boldsymbol{\theta}, z; \boldsymbol{\mu}^\star\big)\big| &= \big|\langle \boldsymbol{\eta}(\boldsymbol{\theta}, z), s(\boldsymbol{x})\rangle - \langle \boldsymbol{\eta}(\boldsymbol{\theta}, z), \boldsymbol{\mu}^\star\rangle\big| \\
&\leq \big|\langle \boldsymbol{\eta}(\boldsymbol{\theta}, z), s(\boldsymbol{x})\rangle - \langle \boldsymbol{\eta}\big(\dot{\boldsymbol{\theta}}_q(z), z\big), s(\boldsymbol{x})\rangle\big| \\
&\quad + \big|\langle \boldsymbol{\eta}\big(\dot{\boldsymbol{\theta}}_q(z), z\big), s(\boldsymbol{x})\rangle - \langle \boldsymbol{\eta}\big(\dot{\boldsymbol{\theta}}_q(z), z\big), \boldsymbol{\mu}^\star\rangle\big| \\
&\quad + \big|\langle \boldsymbol{\eta}\big(\dot{\boldsymbol{\theta}}_q(z), z\big), \boldsymbol{\mu}^\star\rangle - \langle \boldsymbol{\eta}(\boldsymbol{\theta}, z), \boldsymbol{\mu}^\star\rangle\big|. \qquad (5.15)
\end{aligned}$$

A union bound over the three terms on the right-hand side of the inequality above shows that

$$\begin{aligned}
&\mathbb{P}\bigg(\max_{z\in\mathbb{Z}_0}\max_{1\leq q\leq Q}\sup_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_q,\rho)}\big|\ell\big(\boldsymbol{\theta}, z; s(\boldsymbol{X})\big) - \ell\big(\boldsymbol{\theta}, z; \boldsymbol{\mu}^\star\big)\big| \geq \frac{\epsilon u(n)}{2} \cap \mathbb{X}(\alpha)\bigg) \\
&\quad\leq \mathbb{P}\bigg(\max_{z\in\mathbb{Z}_0}\max_{1\leq q\leq Q}\sup_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_q,\rho)}\big|\langle \boldsymbol{\eta}(\boldsymbol{\theta}, z) - \boldsymbol{\eta}\big(\dot{\boldsymbol{\theta}}_q(z), z\big), s(\boldsymbol{X})\rangle\big| \geq \frac{\epsilon u(n)}{6} \cap \mathbb{X}(\alpha)\bigg) \\
&\qquad + \mathbb{P}\bigg(\max_{z\in\mathbb{Z}_0}\max_{1\leq q\leq Q}\sup_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_q,\rho)}\big|\langle \boldsymbol{\eta}\big(\dot{\boldsymbol{\theta}}_q(z), z\big), s(\boldsymbol{X}) - \boldsymbol{\mu}^\star\rangle\big| \geq \frac{\epsilon u(n)}{6} \cap \mathbb{X}(\alpha)\bigg) \\
&\qquad + \mathbb{P}\bigg(\max_{z\in\mathbb{Z}_0}\max_{1\leq q\leq Q}\sup_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_q,\rho)}\big|\langle \boldsymbol{\eta}\big(\dot{\boldsymbol{\theta}}_q(z), z\big) - \boldsymbol{\eta}(\boldsymbol{\theta}, z), \boldsymbol{\mu}^\star\rangle\big| \geq \frac{\epsilon u(n)}{6} \cap \mathbb{X}(\alpha)\bigg). \qquad (5.16)
\end{aligned}$$

We bound the last three terms on the right-hand side of the inequality above one by one.

First term. The first term can be bounded by using condition [C.3], which implies that there exist $A_1 > 0$ and $n_1 > 0$ such that, for any $n > n_1$, any $\boldsymbol{x} \in \mathbb{X}(\alpha)$, any $z \in \mathbb{Z}_0$, any $\boldsymbol{\theta} \in \operatorname{cl}\mathcal{B}(\boldsymbol{\theta}_q, \rho)$, and any $\dot{\boldsymbol{\theta}}_q(z) \in \operatorname{cl}\mathcal{B}(\boldsymbol{\theta}_q, \rho)$,

$$\big|\langle \boldsymbol{\eta}(\boldsymbol{\theta}, z) - \boldsymbol{\eta}\big(\dot{\boldsymbol{\theta}}_q(z), z\big), s(\boldsymbol{x})\rangle\big| \leq A_1 \big\|\boldsymbol{\theta} - \dot{\boldsymbol{\theta}}_q(z)\big\|_2 u(n). \qquad (5.17)$$

Since both $\boldsymbol{\theta}$ and $\dot{\boldsymbol{\theta}}_q(z)$ are contained in the ball $\operatorname{cl}\mathcal{B}(\boldsymbol{\theta}_q, \rho)$, an application of the triangle inequality shows that

$$A_1 \big\|\boldsymbol{\theta} - \dot{\boldsymbol{\theta}}_q(z)\big\|_2 u(n) \leq A_1\, 2\,\rho\, u(n) < \frac{\epsilon u(n)}{6}, \qquad (5.18)$$

where we used the fact that $\rho > 0$ satisfies $0 < \rho < \epsilon/(12A_1)$ by construction. As a result, for all $n > n_1$, we have

$$\mathbb{P}\bigg(\max_{z\in\mathbb{Z}_0}\max_{1\leq q\leq Q}\sup_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_q,\rho)}\big|\langle \boldsymbol{\eta}(\boldsymbol{\theta}, z) - \boldsymbol{\eta}\big(\dot{\boldsymbol{\theta}}_q(z), z\big), s(\boldsymbol{X})\rangle\big| \geq \frac{\epsilon u(n)}{6} \cap \mathbb{X}(\alpha)\bigg) = 0. \qquad (5.19)$$

Second term. We are interested in bounding the probability of deviations of the form $|\langle \boldsymbol{\eta}(\dot{\boldsymbol{\theta}}_q(z), z), s(\boldsymbol{X}) - \boldsymbol{\mu}^\star\rangle|$. We make two observations. First, observe that, for any $\boldsymbol{x} \in \mathbb{X}(\alpha)$,

$$\begin{aligned}&\max_{z\in\mathbb{Z}_0}\max_{1\le q\le Q}\sup_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_q,\rho)}\big|\big\langle\boldsymbol{\eta}\big(\dot{\boldsymbol{\theta}}_q(z),z\big),s(\boldsymbol{x})-\boldsymbol{\mu}^\star\big\rangle\big|\\&\quad=\max_{z\in\mathbb{Z}_0}\max_{1\le q\le Q}\big|\big\langle\boldsymbol{\eta}\big(\dot{\boldsymbol{\theta}}_q(z),z\big),s(\boldsymbol{x})-\boldsymbol{\mu}^\star\big\rangle\big|,\end{aligned}\tag{5.20}$$

which implies that

$$\begin{aligned}&\mathbb{P}\bigg(\max_{z\in\mathbb{Z}_0}\max_{1\le q\le Q}\sup_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_q,\rho)}\big|\big\langle\boldsymbol{\eta}\big(\dot{\boldsymbol{\theta}}_q(z),z\big),s(\boldsymbol{X})-\boldsymbol{\mu}^\star\big\rangle\big|\ge\frac{\epsilon u(n)}{6}\cap\mathbb{X}(\alpha)\bigg)\\&\quad=\mathbb{P}\bigg(\max_{z\in\mathbb{Z}_0}\max_{1\le q\le Q}\big|\big\langle\boldsymbol{\eta}\big(\dot{\boldsymbol{\theta}}_q(z),z\big),s(\boldsymbol{X})-\boldsymbol{\mu}^\star\big\rangle\big|\ge\frac{\epsilon u(n)}{6}\cap\mathbb{X}(\alpha)\bigg).\end{aligned}\tag{5.21}$$

Second, bounding the probability of deviations of the form $|\langle \boldsymbol{\eta}(\dot{\boldsymbol{\theta}}_q(z), z), s(\boldsymbol{X}) - \boldsymbol{\mu}^\star\rangle|$ is equivalent to bounding the probability of deviations of the form $|f(\boldsymbol{X}) - \mathbb{E}f(\boldsymbol{X})|$, where

$$f(\boldsymbol{X})=\big\langle\boldsymbol{\eta}\big(\dot{\boldsymbol{\theta}}_q(z),z\big),s(\boldsymbol{X})\big\rangle,\qquad\mathbb{E}f(\boldsymbol{X})=\big\langle\boldsymbol{\eta}\big(\dot{\boldsymbol{\theta}}_q(z),z\big),\boldsymbol{\mu}^\star\big\rangle.\tag{5.22}$$

Here, $f : \mathbb{X} \mapsto \mathbb{R}$ is considered as a function of $\boldsymbol{X}$ for fixed $(\dot{\boldsymbol{\theta}}_q(z), z) \in \boldsymbol{\Theta}_0 \times \mathbb{Z}_0$. To bound the probability of deviations of the form $|f(\boldsymbol{X}) - \mathbb{E}f(\boldsymbol{X})|$, observe that by condition [C.4] there exist $A_2 > 0$ and $n_2 > 0$ such that, for all $n > n_2$, the Lipschitz coefficient of $f(\boldsymbol{X})$ satisfies $\|f\|_{\text{Lip}} \le A_2 L$. Thus, by applying Lemma 1 to deviations of size $t = \epsilon u(n)/6$ along with a union bound over the $|\mathbb{Z}_0|$ block structures and all Q balls that make up the cover $\bigcup_{1\le q\le Q}\mathcal{B}(\boldsymbol{\theta}_q, \rho)$ of $\boldsymbol{\Theta}_0$, there exists $C_1 > 0$ such that, for all $\epsilon > 0$ and all $n > n_2$,

$$\begin{aligned}&\mathbb{P}\bigg(\max_{z\in\mathbb{Z}_0}\max_{1\le q\le Q}\big|\big\langle\boldsymbol{\eta}\big(\dot{\boldsymbol{\theta}}_q(z),z\big),s(\boldsymbol{X})-\boldsymbol{\mu}^\star\big\rangle\big|\ge\frac{\epsilon u(n)}{6}\cap\mathbb{X}(\alpha)\bigg)\\&\quad\le\mathbb{P}\bigg(\max_{z\in\mathbb{Z}_0}\max_{1\le q\le Q}\big|\big\langle\boldsymbol{\eta}\big(\dot{\boldsymbol{\theta}}_q(z),z\big),s(\boldsymbol{X})-\boldsymbol{\mu}^\star\big\rangle\big|\ge\frac{\epsilon u(n)}{6}\bigg)\\&\quad\le 2\exp\bigg(-\frac{\epsilon^2u(n)^2}{36C_1n^2\|\mathcal{A}\|_\infty^4L^2}+\log|\mathbb{Z}_0|+\log Q\bigg).\end{aligned}\tag{5.23}$$

To bound the exponential term, observe that by assumption (3.9) of Proposition 1 there exists, for all $M > 0$, however large, $n_3 > 0$ such that, for all $n > n_3$,

$$u(n)\ge Mn^{3/2}\|\mathcal{A}\|_\infty^2L\sqrt{\log n}.\tag{5.24}$$

Therefore, for all $n > n_3$, the three terms in the exponent are bounded above by

$$\begin{aligned}&-\frac{\epsilon^2u(n)^2}{36C_1n^2\|\mathcal{A}\|_\infty^4L^2}+\log|\mathbb{Z}_0|+\log Q\\&\quad\le-\frac{\epsilon^2u(n)^2}{36C_1n^2\|\mathcal{A}\|_\infty^4L^2}+\bigg[1+A\log\bigg(\frac{4B+\rho}{\rho}\bigg)+C\bigg]n\log n,\end{aligned}\tag{5.25}$$

where we used $\log|\mathbb{Z}_0| \leq n \log K$ and $\log Q \leq (A\log(4B+\rho)/\rho + C)n$ by (5.10). Since $M > 0$ can be chosen as large as desired, we can choose

$$M > \sqrt{36C_1C_2\left[1 + A\log\left(\frac{4B+\rho}{\rho}\right) + C\right]}, \tag{5.26}$$

where $C_2 > 0$ is chosen so that $C_2\epsilon^2 > 1$. Hence there exists $C_3 > 0$ such that, for all $n > n_3$,

$$-\frac{\epsilon^2 u(n)^2}{36C_1 n^2 \|\mathcal{A}\|_\infty^4 L^2} + \left[1 + A\log\left(\frac{4B+\rho}{\rho}\right) + C\right] n\log n \leq -\epsilon^2 C_3\, n\log n. \tag{5.27}$$

Collecting terms shows that, for all $n > n_3$,

$$\begin{aligned}\mathbb{P}\Bigg(\max_{z\in\mathbb{Z}_0}\max_{1\leq q\leq Q}\sup_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_q,\rho)} \left|\left\langle \boldsymbol{\eta}\big(\dot{\boldsymbol{\theta}}_q(z), z\big), s(\boldsymbol{X}) - \boldsymbol{\mu}^\star\right\rangle\right| \geq \frac{\epsilon u(n)}{6} \cap \mathbb{X}(\alpha)\Bigg) \\ \leq 2\exp\big(-\epsilon^2 C_3\, n\log n\big).\end{aligned} \tag{5.28}$$

Third term. The third term can be bounded along the same lines as the first term, which implies that there exists $n_4 > 0$ such that, for all $n > n_4$,

$$\mathbb{P}\Bigg(\max_{z\in\mathbb{Z}_0}\max_{1\leq q\leq Q}\sup_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_q,\rho)} \left|\left\langle \boldsymbol{\eta}\big(\dot{\boldsymbol{\theta}}_q(z), z\big) - \boldsymbol{\eta}(\boldsymbol{\theta}, z), \boldsymbol{\mu}^\star\right\rangle\right| \geq \frac{\epsilon u(n)}{6} \cap \mathbb{X}(\alpha)\Bigg) = 0. \tag{5.29}$$

Conclusion. Using (5.5) and collecting terms shows that there exists $C > 0$ such that, for all $\epsilon > 0$ and all $n > \max(n_0, n_1, n_2, n_3, n_4)$,

$$\begin{aligned}\mathbb{P}\big(\mathrm{KL}\big(\boldsymbol{\theta}^\star, z^\star; \widehat{\boldsymbol{\theta}}, \widehat{z}\big) \geq \epsilon u(n)\big) &\leq 2\exp\big(-\alpha^2 C_0\, n\log n\big) + 2\exp\big(-\epsilon^2 C_3\, n\log n\big) \\ &\leq 4\exp\big(-\min\big(\alpha^2, \epsilon^2\big)\, C\, n\log n\big).\end{aligned} \tag{5.30}$$

□

Proof of Theorem 1. By assumption (3.11) of Theorem 1, there exist $C_1 > 0$ and $n_1 > 0$ such that, for all $n > n_1$,

$$\mathrm{KL}\big(\boldsymbol{\theta}^\star, z^\star; \widehat{\boldsymbol{\theta}}, \widehat{z}\big) \geq \frac{\delta(z^\star, \widehat{z})\, C_1\, |\ell(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star)|}{n} \tag{5.31}$$

provided $(\widehat{\boldsymbol{\theta}}, \widehat{z})$ exists. By Proposition 1, there exist $C_2 > 0$ and $n_2 > 0$ such that, for all $\epsilon > 0$ and all $n > n_2$, the event

$$\mathrm{KL}\big(\boldsymbol{\theta}^\star, z^\star; \widehat{\boldsymbol{\theta}}, \widehat{z}\big) < \epsilon C_1 \big|\ell\big(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star\big)\big| \tag{5.32}$$

occurs with at least probability

$$1 - 4\exp\big(-\min\big(\alpha^2, \epsilon^2\big)\, C_2\, n\, \log n\big). \tag{5.33}$$

Therefore, for all $\epsilon > 0$ and all $n > \max(n_1, n_2)$, with at least probability (5.33), we observe the event

$$\frac{\delta(z^\star, \widehat{z})\, C_1\, |\ell(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star)|}{n} \leq \mathrm{KL}\big(\boldsymbol{\theta}^\star, z^\star; \widehat{\boldsymbol{\theta}}, \widehat{z}\big) < \epsilon\, C_1 \left|\ell\big(\boldsymbol{\theta}^\star, z^\star; \boldsymbol{\mu}^\star\big)\right|,$$

that is, the event $\delta(z^\star, \widehat{z})/n < \epsilon$. □

6. Discussion

Here, and elsewhere [59], we have taken first steps to demonstrate that additional structure in the form of block structure facilitates statistical inference for exponential-family random graph models. It goes without saying that numerous open problems remain, ranging from probabilistic problems (e.g., understanding properties of probability models) and statistical problems (e.g., understanding properties of statistical methods) to computational problems (e.g., the development of computational methods for large networks).

One important problem is that the restricted maximum likelihood estimator considered here is even less tractable than (restricted) maximum likelihood estimators for stochastic block models [15,53]. The intractability stems in part from the fact that the block structure is unknown and the number of possible block structures is large and in part from the fact that the likelihood function is intractable even when the block structure is known, owing to the complex dependence within blocks. There do exist Bayesian auxiliary-variable methods for small networks [56,58] and promising directions for methods for large networks [4,67]. As pointed out in the introduction, an in depth investigation of all of these models and methods is beyond the scope of a single paper. However, the main consistency results reported here suggest that statistical inference for these models and methods is possible and worth exploring in more depth.

Acknowledgements

The author acknowledges support from the National Science Foundation (NSF awards DMS-1513644 and DMS-1812119) and is grateful to two anonymous referees, an anonymous associate editor, and Jonathan Stewart, whose careful reading and constructive suggestions led to substantial improvements of the paper.

Supplementary Material

Supplement to: Consistent structure estimation of exponential-family random graph models with block structure (DOI: 10.3150/19-BEJ1153SUPP; .pdf). The proofs of Lemmas 1, 2, and 3 and Corollaries 1 and 2 can be found in the supplementary material [55].

References

[1] Airoldi, E., Blei, D., Fienberg, S. and Xing, E. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.

[2] Alon, N. and Spencer, J.H. (2008). *The Probabilistic Method*: *With an Appendix on the Life and Work of Paul Erdős*, 3rd ed. *Wiley-Interscience Series in Discrete Mathematics and Optimization*. Hoboken, NJ: Wiley. MR2437651 https://doi.org/10.1002/9780470277331
[3] Amini, A.A., Chen, A., Bickel, P.J. and Levina, E. (2013). Pseudo-likelihood methods for community detection in large sparse networks. *Ann. Statist.* **41** 2097–2122. MR3127859 https://doi.org/10.1214/13-AOS1138
[4] Babkin, S. and Schweinberger, M. (2017). Large-scale estimation of random graph models with local dependence. Preprint. Available at arXiv:1703.09301.
[5] Berk, R.H. (1972). Consistency and asymptotic normality of MLE's for exponential models. *Ann. Math. Stat.* **43** 193–204. MR0298810
[6] Bhamidi, S., Bresler, G. and Sly, A. (2011). Mixing time of exponential random graphs. *Ann. Appl. Probab.* **21** 2146–2170. MR2895412 https://doi.org/10.1214/10-AAP740
[7] Bickel, P.J. and Chen, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA* **106** 21068–21073.
[8] Bickel, P.J., Chen, A. and Levina, E. (2011). The method of moments and degree distributions for network models. *Ann. Statist.* **39** 2280–2301. MR2906868 https://doi.org/10.1214/11-AOS904
[9] Binkiewicz, N., Vogelstein, J.T. and Rohe, K. (2017). Covariate-assisted spectral clustering. *Biometrika* **104** 361–377. MR3698259 https://doi.org/10.1093/biomet/asx008
[10] Bollobás, B. (1998). *Modern Graph Theory*. *Graduate Texts in Mathematics* **184**. New York: Springer. MR1633290 https://doi.org/10.1007/978-1-4612-0619-4
[11] Brown, L.D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. *Institute of Mathematical Statistics Lecture Notes – Monograph Series* **9**. Hayward, CA: IMS. MR0882001
[12] Celisse, A., Daudin, J.-J. and Pierre, L. (2012). Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electron. J. Stat.* **6** 1847–1899. MR2988467 https://doi.org/10.1214/12-EJS729
[13] Chatterjee, S. and Diaconis, P. (2013). Estimating and understanding exponential random graph models. *Ann. Statist.* **41** 2428–2461. MR3127871 https://doi.org/10.1214/13-AOS1155
[14] Chatterjee, S., Diaconis, P. and Sly, A. (2011). Random graphs with a given degree sequence. *Ann. Appl. Probab.* **21** 1400–1435. MR2857452 https://doi.org/10.1214/10-AAP728
[15] Choi, D.S., Wolfe, P.J. and Airoldi, E.M. (2012). Stochastic blockmodels with a growing number of classes. *Biometrika* **99** 273–284. MR2931253 https://doi.org/10.1093/biomet/asr053
[16] Crane, H. and Dempsey, W. (2019). A framework for statistical network modeling. *Statist. Sci.*. To appear.
[17] Erdős, P. and Rényi, A. (1959). On random graphs. I. *Publ. Math. Debrecen* **6** 290–297. MR0120167
[18] Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. *Magy. Tud. Akad. Mat. Kut. Intéz. Közl.* **5** 17–61. MR0125031
[19] Frank, O. and Strauss, D. (1986). Markov graphs. *J. Amer. Statist. Assoc.* **81** 832–842. MR0860518
[20] Frieze, A. and Karoński, M. (2016). *Introduction to Random Graphs*. Cambridge: Cambridge Univ. Press. MR3675279 https://doi.org/10.1017/CBO9781316339831
[21] Gao, C., Lu, Y. and Zhou, H.H. (2015). Rate-optimal graphon estimation. *Ann. Statist.* **43** 2624–2652. MR3405606 https://doi.org/10.1214/15-AOS1354
[22] Gilbert, E.N. (1959). Random graphs. *Ann. Math. Stat.* **30** 1141–1144. MR0108839 https://doi.org/10.1214/aoms/1177706098
[23] Handcock, M.S. (2003). Assessing degeneracy in statistical models of social networks. Tech. rep., Center for Statistics and the Social Sciences, Univ. Washington. Available at www.csss.washington.edu/Papers.
[24] Holland, P.W. and Leinhardt, S. (1976). Local structure in social networks. *Sociol. Method.* 1–45.

[25] Holland, P.W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *J. Amer. Statist. Assoc.* **76** 33–65. MR0608176
[26] Hollway, J. and Koskinen, J. (2016). Multilevel embeddedness: The case of the global fisheries governance complex. *Soc. Netw.* **44** 281–294.
[27] Hollway, J., Lomi, A., Pallotti, F. and Stadtfeld, C. (2017). Multilevel social spaces: The network dynamics of organizational fields. *Netw. Sci.* **5** 187–212.
[28] Hunter, D.R. (2007). Curved exponential family models for social networks. *Soc. Netw.* **29** 216–230. https://doi.org/10.1016/j.socnet.2006.08.005
[29] Hunter, D.R., Goodreau, S.M. and Handcock, M.S. (2008). Goodness of fit of social network models. *J. Amer. Statist. Assoc.* **103** 248–258. MR2394635 https://doi.org/10.1198/016214507000000446
[30] Hunter, D.R. and Handcock, M.S. (2006). Inference in curved exponential family models for networks. *J. Comput. Graph. Statist.* **15** 565–583. MR2291264 https://doi.org/10.1198/106186006X133069
[31] Hunter, D.R., Krivitsky, P.N. and Schweinberger, M. (2012). Computational statistical methods for social network models. *J. Comput. Graph. Statist.* **21** 856–882. MR3005801 https://doi.org/10.1080/10618600.2012.732921
[32] Jin, J. (2015). Fast community detection by SCORE. *Ann. Statist.* **43** 57–89. MR3285600 https://doi.org/10.1214/14-AOS1265
[33] Jonasson, J. (1999). The random triangle model. *J. Appl. Probab.* **36** 852–867. MR1737058 https://doi.org/10.1239/jap/1032374639
[34] Kontorovich, L. and Ramanan, K. (2008). Concentration inequalities for dependent random variables via the martingale method. *Ann. Probab.* **36** 2126–2158. MR2478678 https://doi.org/10.1214/07-AOP384
[35] Krivitsky, P.N. (2012). Exponential-family random graph models for valued networks. *Electron. J. Stat.* **6** 1100–1128. MR2988440 https://doi.org/10.1214/12-EJS696
[36] Krivitsky, P.N. and Kolaczyk, E.D. (2015). On the question of effective sample size in network modeling: An asymptotic inquiry. *Statist. Sci.* **30** 184–198. MR3353102 https://doi.org/10.1214/14-STS502
[37] Lauritzen, S., Rinaldo, A. and Sadeghi, K. (2018). Random networks, graphical models and exchangeability. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 481–508. MR3798875 https://doi.org/10.1111/rssb.12266
[38] Lazega, E. and Snijders, T.A.B., eds. (2016). *Multilevel Network Analysis for the Social Sciences*. Cham: Springer.
[39] Lei, J. and Rinaldo, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43** 215–237. MR3285605 https://doi.org/10.1214/14-AOS1274
[40] Leskovec, J., Lang, K.J., Dasgupta, A. and Mahoney, M.W. (2009). Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Math.* **6** 29–123. MR2736090
[41] Lomi, A., Robins, G. and Tranmer, M. (2016). Introduction to multilevel social networks. *Soc. Netw.* **44** 266–268.
[42] Lusher, D., Koskinen, J. and Robins, G. (2013). *Exponential Random Graph Models for Social Networks*. Cambridge, UK: Cambridge Univ. Press.
[43] Molloy, M. and Reed, B. (2002). *Graph Colouring and the Probabilistic Method. Algorithms and Combinatorics* **23**. Berlin: Springer. MR1869439 https://doi.org/10.1007/978-3-642-04016-0
[44] Mossel, E., Neeman, J. and Sly, A. (2015). Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields* **162** 431–461. MR3383334 https://doi.org/10.1007/s00440-014-0576-6
[45] Nowicki, K. and Snijders, T.A.B. (2001). Estimation and prediction for stochastic blockstructures. *J. Amer. Statist. Assoc.* **96** 1077–1087. MR1947255 https://doi.org/10.1198/016214501753208735

[46] Pattison, P. and Robins, G. (2002). Neighborhood-based models for social networks. In *Sociological Methodology* (R.M. Stolzenberg, ed.) **32** 301–337. Boston, MA: Blackwell Publishing.

[47] Priebe, C.E., Sussman, D.L., Tang, M. and Vogelstein, J.T. (2015). Statistical inference on errorfully observed graphs. *J. Comput. Graph. Statist.* **24** 930–953. MR3432923 https://doi.org/10.1080/10618600.2014.951049

[48] Rapoport, A. (1953). Spread of information through a population with socio-structural bias. I. Assumption of transitivity. *Bull. Math. Biophys.* **15** 523–533. MR0058955 https://doi.org/10.1007/bf02476440

[49] Rapoport, A. (1953). Spread of information through a population with socio-structural bias. II. Various models with partial transitivity. *Bull. Math. Biophys.* **15** 535–546. MR0058956 https://doi.org/10.1007/bf02476441

[50] Rinaldo, A., Fienberg, S.E. and Zhou, Y. (2009). On the geometry of discrete exponential families with application to exponential random graph models. *Electron. J. Stat.* **3** 446–484. MR2507456 https://doi.org/10.1214/08-EJS350

[51] Rinaldo, A., Petrović, S. and Fienberg, S.E. (2013). Maximum likelihood estimation in the β-model. *Ann. Statist.* **41** 1085–1110. MR3113804 https://doi.org/10.1214/12-AOS1078

[52] Rohe, K., Chatterjee, S. and Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. MR2893856 https://doi.org/10.1214/11-AOS887

[53] Rohe, K., Qin, T. and Fan, H. (2014). The highest dimensional stochastic blockmodel with a regularized estimator. *Statist. Sinica* **24** 1771–1786. MR3308662

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

[55] Schweinberger, M. (2020). Supplement to "Consistent structure estimation of exponential-family random graph models with block structure." https://doi.org/10.3150/19-BEJ1153SUPP.

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

[57] Schweinberger, M., Krivitsky, P.N., Butts, C.T. and Stewart, J. (2019). Exponential-family models of random graphs: Inference in finite-, super-, and infinite-population scenarios. *Statist. Sci.* To appear.

[58] Schweinberger, M. and Luna, P. (2018). HERGM: Hierarchical exponential-family random graph models. *J. Stat. Softw.* **85** 1–39.

[59] Schweinberger, M. and Stewart, J. (2019). Concentration and consistency results for canonical and curved exponential-family models of random graphs. *Ann. Statist.* To appear.

[60] Shalizi, C.R. and Rinaldo, A. (2013). Consistency under sampling of exponential random graph models. *Ann. Statist.* **41** 508–535. MR3099112 https://doi.org/10.1214/12-AOS1044

[61] Slaughter, A.J. and Koehly, L.M. (2016). Multilevel models for social networks: Hierarchical Bayesian approaches to exponential random graph modeling. *Soc. Netw.* **44** 334–345.

[62] Snijders, T.A.B. (2010). Conditional marginalization for exponential random graph models. *J. Math. Sociol.* **34** 239–252.

[63] Snijders, T.A.B., Pattison, P.E., Robins, G.L. and Handcock, M.S. (2006). New specifications for exponential random graph models. *Sociol. Method.* **36** 99–153.

[64] Stephens, M. (2000). Dealing with label switching in mixture models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 795–809. MR1796293 https://doi.org/10.1111/1467-9868.00265

[65] Stewart, J., Schweinberger, M., Bojanowski, M. and Morris, M. (2019). Multilevel network data facilitate statistical inference for curved ERGMs with geometrically weighted terms. *Soc. Netw.* **59** 98–119.

[66] Wang, P., Robins, G., Pattison, P. and Lazega, E. (2013). Exponential random graph models for multilevel networks. *Soc. Netw.* **35** 96–115.

[67] Wang, Y., Fang, H., Yang, D., Zhao, H. and Deng, M. (2018). Network clustering analysis using mixture exponential-family random graph models and its application in genetic interaction data. *IEEE/ACM Trans. Comput. Biol. Bioinform.* https://doi.org/10.1109/TCBB.2017.2743711.

[68] Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge Univ. Press.

[69] Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks. I. An introduction to Markov graphs and p. *Psychometrika* **61** 401–425. MR1424909 https://doi.org/10.1007/BF02294547

[70] Yan, T., Leng, C. and Zhu, J. (2016). Asymptotics in directed exponential random graph models with an increasing bi-degree sequence. *Ann. Statist.* **44** 31–57. MR3449761 https://doi.org/10.1214/15-AOS1343

[71] Yan, T., Qin, H. and Wang, H. (2016). Asymptotics in undirected random graph models parameterized by the strengths of vertices. *Statist. Sinica* **26** 273–293. MR3468353

[72] Yan, T., Zhao, Y. and Qin, H. (2015). Asymptotic normality in the maximum entropy models on graphs with an increasing number of parameters. *J. Multivariate Anal.* **133** 61–76. MR3282018 https://doi.org/10.1016/j.jmva.2014.08.013

[73] Zappa, P. and Lomi, A. (2015). The analysis of multilevel networks in organizations: Models and empirical tests. *Organ. Res. Methods* **18** 542–569.

[74] Zhang, A.Y. and Zhou, H.H. (2016). Minimax rates of community detection in stochastic block models. *Ann. Statist.* **44** 2252–2280. MR3546450 https://doi.org/10.1214/15-AOS1428

[75] Zhao, Y., Levina, E. and Zhu, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *Ann. Statist.* **40** 2266–2292. MR3059083 https://doi.org/10.1214/12-AOS1036

Received March 2018 and revised March 2019