Adv. Appl. Prob. **44,** 139–165 (2012)
Printed in Northern Ireland

JOINT VERTEX DEGREES IN THE INHOMOGENEOUS RANDOM GRAPH MODEL $\mathcal{G}(n, \{p_{ij}\})$

KAISHENG LIN * AND
GESINE REINERT,** *University of Oxford*

Abstract

In a random graph, counts for the number of vertices with given degrees will typically be dependent. We show via a multivariate normal and a Poisson process approximation that, for graphs which have independent edges, with a possibly inhomogeneous distribution, only when the degrees are large can we reasonably approximate the joint counts as independent. The proofs are based on Stein's method and the Stein–Chen method with a new size-biased coupling for such inhomogeneous random graphs, and, hence, bounds on the distributional distance are obtained. Finally, we illustrate that apparent (pseudo-)power-law-type behaviour can arise in such inhomogeneous networks despite not actually following a power-law degree distribution.

Keywords: Stein's method; size-biased coupling; vertex degree; inhomogeneous random graph; power law

2010 Mathematics Subject Classification: Primary 60F05
Secondary 05C80; 90B15

1. Introduction

It has been observed in many real-world networks that, when plotting the observed number of vertices of degree k against k, on a log-log scale the plots tend to look linear. This so-called *scale-free behaviour* (see, e.g. [7]) motivated the *scale-free network model* introduced in [1], yielding a probability distribution for the number of vertices of large degree which is scale free.

Some issues arise when trying to assess the vertex degree distribution from a single network. The log-log scale lends itself to overinterpretation; Solow *et al.* [16] discussed a good number of pitfalls when trying to test for power law using such plots. Moreover, Stumpf *et al.* [18] have shown that, when sampling from a scale-free network, the sampled network will not in general be scale free. In addition, the total number of vertices in the network is fixed, and, hence, counts for different degrees will be dependent. We shall see in this paper that even edges occur independently, the dependence between degree counts depends on the size of the degrees under consideration—only when the degrees are large can we reasonably approximate the joint counts as independent. We establish these facts by proving a multivariate normal approximation, with possibly nondiagonal asymptotic covariance matrix, as well as a Poisson process approximation, with independent coordinates. We give bounds for these approximations which depend on the size of the degrees under consideration. Finally, we shall illustrate that apparent (pseudo-)power-law-type behaviour can arise in networks which are constructed using independent edges, and do not follow a power-law behaviour.

Received 16 November 2009; revision received 5 August 2011.
* Email address: k.lin@wolfson.oxon.org
** Postal address: Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK.
Email address: reinert@stats.ox.ac.uk

The degree of a vertex is one of the fundamental summaries for random graphs, and, hence, the degree distribution is a natural object to study. In a general random graph $\mathcal{G}^n$ on a set V of n vertices, the degree of a vertex v, denoted by $D(v)$, is defined as the number of vertices adjacent to v. The most basic model of a random graph is that of the Bernoulli graph $\mathcal{G}(n, p)$, introduced by Erdös and Rényi [8]. A survey of the Poisson approximation for the distribution of the kth largest degree for large k in the Bernoulli model $\mathcal{G}(n, p)$, as well as of both a Poisson approximation and a normal approximation for the number of vertices of a given degree, can be found in [4], with bounds on the distributional distance. For the joint distribution of degrees in the Bernoulli model $\mathcal{G}(n, p)$, McKay and Wormald [12] gave an approximation with simpler models derived from a binomial distribution and used this for univariate normal approximations.

While Bernoulli random graphs typically do not model real-world networks well, in [6] a mixture model for Bernoulli random graphs was shown to be suitable for some biological networks. Under the name *stochastic block model*, a similar mixture model has proven successful in the area of social network analysis; see [14]. Here we use the inhomogeneous model $\mathcal{G}(n, \{p_{ij}\})$ as a submodel of $\mathcal{G}^n$, consisting of all graphs in which the edges occur independently, and, for $i, j \in V$, the probability that vertices i and j are adjacent is p_{ij}. This general model not only includes Bernoulli random graphs, but also mixtures of Bernoulli random graphs, Newman–Moore–Watts–Strogatz small-world networks as defined in [13], and the exponential random graph model, which is defined by assuming in $\mathcal{G}(n, \{p_{ij}\})$ that $p_{ij} = \exp(\theta_i + \theta_j)/\{1 + \exp(\theta_i + \theta_j)\}$, where $\{\theta_i,\ i \in V\}$ are parameters of the model. For fairly general random graph models which include a Barabási–Albert scale-free model, but do not quite cover the class $\mathcal{G}(n, \{p_{ij}\})$ in full generality, Bollobás *et al.* [5] gave a univariate mixed Poisson approximation for the number of vertices with a given degree. There is a lack of results for multivariate approximations, despite the need to understand log-log plots. In addition, networks consist of a finite number of vertices, and, depending on the complexity, the distribution of vertices with a fixed degree may be far from the asymptotic regime; thus, bounds on the distributional approximations are required.

In order to understand log-log plots of the number of vertices with degree k versus k, we consider the degree-count sequence $\boldsymbol{W} := (W_i,\ 0 \le i \le n-1)$, where W_i counts the number of vertices having degree exactly i. The definitions of both sequences $\boldsymbol{D} := (D(v),\ v \in V)$ and $\boldsymbol{W}$ can be related by introducing the index set

$$\Gamma := \{(v, i) \colon v \in V; 0 \le i \le n-1\}, \tag{1.1}$$

and defining, for $(v, i) \in \Gamma$, the Bernoulli random variables $X_{(v,i)} := \mathbf{1}(D(v) = i)$, where $\mathbf{1}(\cdot)$ is the indicator function. Then $D(v) = \sum_{i=0}^{n-1} i X_{(v,i)}$, and $W_i = \sum_{v \in V} X_{(v,i)}$. Other interesting statistics may be also obtained by this setting. For instance, one may define the random variable $Z_k = \sum_{i \ge k} W_i$ as the number of vertices having degree at least k, for $0 \le k \le n-1$, and consider the sequence $\boldsymbol{Z} := (Z_k,\ 0 \le k \le n-1)$.

In the flavour of probability theory, as the sequences $\boldsymbol{D}$, $\boldsymbol{W}$, and $\boldsymbol{Z}$ are deterministic functions of the collection $\boldsymbol{X} := \{X_{(v,i)},\ (v, i) \in \Gamma\}$, the σ-fields $\sigma(\boldsymbol{D})$, $\sigma(\boldsymbol{W})$, and $\sigma(\boldsymbol{Z})$, generated by $\boldsymbol{D}$, $\boldsymbol{W}$, and $\boldsymbol{Z}$, respectively, are all contained in the σ-field $\sigma(\boldsymbol{X})$ generated by $\boldsymbol{X}$. The collection $\boldsymbol{X}$ in turn can be represented by the point process Ξ, defined by $\Xi := \sum_{\alpha \in \Gamma} \delta_\alpha X_\alpha$, where δ_α is the point measure at α, that is, for a set B, $\delta_\alpha(B) = 1$ if $\alpha \in B$ and $\delta_\alpha(B) = 0$ otherwise.

For the degree-count sequence $\boldsymbol{W} = (W_i,\ 0 \le i \le n-1)$ in $\mathcal{G}(n, p)$, two results in particular have inspired the current work. Barbour *et al.* [2, Theorem 5.E, p. 97, and Theorem 5.F, p. 99] gave univariate Poisson approximations for the distribution of Z_k and W_k, and Goldstein and Rinott [9] proved a multivariate normal approximation for the joint distribution of any subsequence $(W_{d_1}, W_{d_2}, \dots, W_{d_m})$ of $\boldsymbol{W}$. Both results use Stein's method; in the context of Poisson approximations this method is usually called the *Stein–Chen method*. The applications of Stein's method in these two papers use a coupling construction to compute bounds on the errors made in the distributional approximations. For graph degree counts, for any $\alpha \in \Gamma$, a new graph model $\mathcal{G}^\alpha(n, p)$ is constructed, conditional on the model $\mathcal{G}(n, p)$, such that the distribution of $\mathcal{G}^\alpha(n, p)$ is the same as the conditional distribution of $\mathcal{G}(n, p)$ given $X_\alpha = 1$; this coupling is a special case of a *size-biased coupling*. The difference between the degree counts in $\mathcal{G}(n, p)$ and in $\mathcal{G}^\alpha(n, p)$ is then used ingeniously to give a bound on the distance to the target distribution.

In Section 2 we construct such a coupling in the inhomogeneous model $\mathcal{G}(n, \{p_{ij}\})$, generalizing the existing construction for the homogeneous model. This coupling will be the main tool for our distributional approximations, which we derive in Section 3. Firstly, in Theorem 3.1, we provide a multivariate normal approximation for the joint counts of vertices with predescribed degrees. The bound depends on the chosen degrees, and on the heterogeneity of the underlying graph. The approximating normal distribution has nondiagonal covariance matrix in general, and, hence, in the normal limit the counts will often not be independent.

The multivariate normal approximation is suitable when the degrees under consideration are not too far away from the centre of the degree distribution. For large degrees, a compound Poisson approximation is more appropriate. Indeed, Theorem 3.2 gives a Poisson point process approximation for the *M-truncated* point process Ξ_M defined by

$$\Xi_M := \sum_{\alpha \in \Gamma_M} \delta_\alpha X_\alpha, \tag{1.2}$$

where, for $0 \le M \le n-1$, we set $\Gamma_M := \{(v, i) \colon v \in V;\ M \le i \le n-1\}$.

Using the invariant property of the total variation distance in functional transformations of point processes, we obtain, from Theorem 3.2, in Corollary 3.1 a multivariate compound Poisson approximation for the M-truncated degree sequence $\boldsymbol{D}_M := (D(v)\,\mathbf{1}(D(v) \ge M), v \in V)$ in $\mathcal{G}(n, \{p_{ij}\})$. The result shows that counts for large vertex degrees are asymptotically independent when the edge probabilities are not too heterogeneous. All these results also contain a bound in distributional distance. This bound depends on the size of the degrees under consideration, and on the number of vertices, as well as on the heterogeneity in the edge probabilities.

We illustrate our results using simulations for a Bernoulli random graph as well as several classes of inhomogeneous random graphs. Finally, we show that the log-log plots for vertex degrees can appear to be power-law like, without following a power law, when the edge probabilities are small.

The proofs are postponed until Section 4.

2. A size-biased coupling for vertex degrees in $\mathcal{G}(n, \{p_{ij}\})$

The size-biased distribution of a collection of variables $\boldsymbol{X}$ relates to a sampling procedure where the probability of an item to be included in the sample is directly proportional to its size. Formally, it can be defined as follows; see, for example, [9].

Definition 2.1. Let $\mathcal{I}$ be an arbitrary index set, and let $\boldsymbol{X} = \{X_\alpha \colon \alpha \in \mathcal{I}\}$ be a collection of nonnegative random variables with means $\mathrm{E}\, X_\alpha = \lambda_\alpha > 0$. For $\beta \in \mathcal{I}$, we say that $\boldsymbol{X}^\beta = \{X_\alpha^\beta \colon \alpha \in \mathcal{I}\}$ has *the $\boldsymbol{X}$-size-biased distribution in the βth coordinate* if $\mathrm{E}\, X_\beta G(\boldsymbol{X}) = \lambda_\beta\, \mathrm{E}\, G(\boldsymbol{X}^\beta)$ for all functions G such that the expectations exist.

We can also apply Definition 2.1 to a vector $\boldsymbol{W} = (W_1, \ldots, W_m) \in \mathbb{R}^m$ by identifying the vector with the collection $\{W_j,\ j \in \mathcal{I}\}$ with $\mathcal{I} = \{1, \ldots, m\}$.

A construction of $(\boldsymbol{X}, \boldsymbol{X}^\beta)$ for each $\beta \in \mathcal{I}$ on a joint probability space is called a *size-biased coupling*. For any subset $B \subset \mathcal{I}$, we set $X_B = \sum_{\alpha \in B} X_\alpha$, and $\lambda_B = \mathrm{E}\, X_B$. Goldstein and Rinott [9] gave the following mixture construction of a size-biased coupling for $\boldsymbol{X}$ in 'coordinate' B. Suppose that $\lambda_B < \infty$ and that, for $\beta \in B$, we have a variable $\boldsymbol{X}^\beta$ which has the $\boldsymbol{X}$-size-biased distribution in coordinate β as in Definition 2.1. Then the random variable $\boldsymbol{X}^B$ which is obtained as the mixture of the distributions $\boldsymbol{X}^\beta$, $\beta \in B$, with weights $\lambda_\beta / \lambda_B$, satisfies $\mathrm{E}\, X_B G(\boldsymbol{X}) = \lambda_B\, \mathrm{E}\, G(\boldsymbol{X}^B)$.

The application of this construction for coupling variables for the degree-count sequence $\boldsymbol{W} = (W_{d_i},\ 1 \le i \le m)$ has been carried out in [9], for distinct and fixed d_i (where $i = 1, \ldots, m$). The idea is, for a given degree d_i and vertex v, we force v to have degree d_i. If the degree of v was equal to d_i in the first place, no adjustment is necessary. If the degree $D(v)$ of v in the original graph was larger than d_i, then $D(v) - d_i$ edges are chosen at random from the edges which include v as one endpoint, and are removed. If $D(v) < d_i$ then $d_i - D(v)$ edges of the form $\{u, v\}$ are added to the graph, where the vertices u are chosen uniformly at random from the $n - 1 - D(v)$ vertices not adjacent to v. Randomizing over $v \in V$ then gives a size-biased version $\boldsymbol{W}^i$ of $\boldsymbol{W}$ for the graph $\mathcal{G}(n, p)$.

For the inhomogeneous model $\mathcal{G}(n, \{p_{ij}\})$, we use the index set Γ in (1.1), which covers all possible combinations between vertices and their degrees, and we write $A_i = V \times \{i\}$ for $i \in \{0, \ldots, n-1\}$. For $i \in \{0, \ldots, n-1\}$, we construct $(\boldsymbol{X}, \boldsymbol{X}^\beta)$ for $\beta \in A_i$, with $\boldsymbol{X} = \{X_{(v,i)} \colon (v, i) \in \Gamma\}$ and $X_{(v,i)} = \mathbf{1}(D(v) = i)$, as detailed below; we call the resulting graph $\mathcal{G}^\beta(n, \{p_{ij}\})$.

Then, we construct $(\boldsymbol{W}, \boldsymbol{W}^i)$ by using a random index I in A_i, i.e. a random vertex, which has the probability mass function $\mathrm{P}(I = \beta) = \lambda_\beta / \lambda_{A_i}$, independently of all other random variables in the system; we call the resulting graph $\mathcal{G}^i(n, \{p_{ij}\})$.

To describe the detailed construction of $\mathcal{G}^\beta(n, \{p_{ij}\})$, let E denote the (potential) edge set of the graph model, and define, for an edge $\{a, b\} \in E$, the Bernoulli random variables $X_{\{a,b\}} := \mathbf{1}(a \sim b)$ in $\mathcal{G}(n, \{p_{ij}\})$ (and similarly $X^\beta_{\{a,b\}}$ in $\mathcal{G}^\beta(n, \{p_{ij}\})$), where $a \sim b$ denotes the event that a is adjacent to b. We also use the following notation: $N(v)$ is the random neighbourhood of vertex v in $\mathcal{G}(n, \{p_{ij}\})$ (similarly $N^\beta(v)$ in $\mathcal{G}^\beta(n, \{p_{ij}\})$), and $\boldsymbol{x}_i = \boldsymbol{x}_i(v)$ is an i-set (i.e. a set with i elements) of $V_v := V \setminus \{v\}$.

Construction 2.1. *For each $\beta = (v, i) \in A_i$, conditional on $\mathcal{G}(n, \{p_{ij}\})$, the following assertions hold.*

- *If $D(v) = d = i$, let $\mathcal{G}^{(v,i)}(n, \{p_{ij}\}) = \mathcal{G}(n, \{p_{ij}\})$, that is, set $X^{(v,i)}_{\{a,b\}} = X_{\{a,b\}}$ for all $\{a, b\} \in E$.*
- *If $D(v) = d > i$ and $N(v) = \boldsymbol{x}_d$, then we choose $\boldsymbol{x}_i \subset \boldsymbol{x}_d$ with probability proportional to*

$$f^+(\boldsymbol{x}_i \mid \boldsymbol{x}_d) := \sum_{j=0}^{i} \frac{1}{\binom{d-j}{i-j}} \mathrm{P}(|N(v) \cap \boldsymbol{x}_i| = j,$$
$$|N(v) \cap (V_v \setminus \boldsymbol{x}_d)| = i - j \mid D(v) = i), \qquad (2.1)$$

and delete all the edges between v and the vertices in $\boldsymbol{x}_d \setminus \boldsymbol{x}_i$. That is, with probability proportional to (2.1), we set $X^{(v,i)}_{\{v,x\}} = 0$ for x in $\boldsymbol{x}_d \setminus \boldsymbol{x}_i$, and $X^{(v,i)}_{\{a,b\}} = X_{\{a,b\}}$ for all $\{a, b\}$ elsewhere.

- *If $D(v) = d < i$ and $N(v) = \boldsymbol{x}_d$, then we choose $\boldsymbol{x}_i \supset \boldsymbol{x}_d$ with probability proportional to*

$$f^-(\boldsymbol{x}_i \mid \boldsymbol{x}_d) := \sum_{j=i}^{n-1} \frac{1}{\binom{j-d}{j-i}} \mathrm{P}(|N(v) \cup \boldsymbol{x}_i| = j,$$
$$|N(v) \cup (V_v \setminus \boldsymbol{x}_d)| = n - 1 - j + i \mid D(v) = i), \quad (2.2)$$

and add all the edges between v and the vertices in $\boldsymbol{x}_i \setminus \boldsymbol{x}_d$. That is, with probability proportional to (2.2), we set $X^{(v,i)}_{\{v,x\}} = 1$ for x in $\boldsymbol{x}_i \setminus \boldsymbol{x}_d$, and $X^{(v,i)}_{\{a,b\}} = X_{\{a,b\}}$ for all $\{a, b\}$ elsewhere.

In Section 4 we shall prove that (2.1) and (2.2) are indeed probabilities.

Lemma 2.1. *We have*

$$\sum_{\{\boldsymbol{x}_i \,:\, \boldsymbol{x}_i \subset \boldsymbol{x}_d\}} f^+(\boldsymbol{x}_i \mid \boldsymbol{x}_d) = 1 \quad \textit{and} \quad \sum_{\{\boldsymbol{x}_i \,:\, \boldsymbol{x}_i \supset \boldsymbol{x}_d\}} f^-(\boldsymbol{x}_i \mid \boldsymbol{x}_d) = 1.$$

Remark 2.1. Note that, in all cases, Construction 2.1 indeed yields $N^{(v,i)}(v) = \boldsymbol{x}_i$ if $N(v) = \boldsymbol{x}_d$.

Remark 2.2. For $\mathcal{G}(n, \{p_{ij}\})$, we can define the complement graph $\mathcal{G}(n, \{q_{ij}\})$ with $q_{ij} = 1 - p_{ij}$ such that (u, v) is an edge in $\mathcal{G}(n, \{q_{ij}\})$ if and only if it is not an edge in $\mathcal{G}(n, \{p_{ij}\})$. Then Construction 2.1 for $d < i$ in $\mathcal{G}(n, \{p_{ij}\})$ is the same as the construction for $d > i$ in $\mathcal{G}(n, \{q_{ij}\})$.

We shall show in Section 4 that the distribution of $\mathcal{G}^\beta(n, \{p_{ij}\})$ is indeed the same as the conditional distribution of $\mathcal{G}(n, \{p_{ij}\})$ given $X_\beta = 1$, yielding a construction of $(\boldsymbol{X}, \boldsymbol{X}^\beta)$ for $\beta \in A_i$, which in turn gives a construction of $(\boldsymbol{W}, \boldsymbol{W}^i)$ via $\mathcal{G}^i(n, \{p_{ij}\})$ using the random index $I \in A_i$ with $\mathrm{P}(I = \beta) = \lambda_\beta / \lambda_{A_i}$. In the next section, we shall use Construction 2.1 in $\mathcal{G}(n, \{p_{ij}\})$ to obtain a multivariate normal approximation for the degree-count sequence $\boldsymbol{W}$, and a compound Poisson approximation for the truncated degree sequence $\boldsymbol{D}_M$.

3. Distributional approximation for degree counts in $\mathcal{G}(n, \{p_{ij}\})$

3.1. Multivariate normal approximation for joint counts

For a multivariate normal approximation, we generalize the argument from [9], which is based on Stein's method; see [17]. Let $V = \{1, \ldots, n\}$, let d_i, $i = 1, \ldots, p$, be distinct numbers in $\{0, 1, \ldots, n-1\}$, and let $W_{d_i} = \sum_{v=1}^n \mathbf{1}(D(v) = d_i)$ denote the number of vertices of degree d_i in $\mathcal{G}(n, \{p_{ij}\})$. Denote by $\boldsymbol{W} = (W_{d_1}, \ldots, W_{d_p})$ the vector of degree counts. As $D(v)$ has a Poisson-binomial distribution which is cumbersome to write explicitly, we abbreviate

$$q_{v,d} := \mathrm{P}(D(v) = d).$$

Let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)$ denote the expectation vector of $\boldsymbol{W}$; for $i = 1, \ldots, p$,

$$\lambda_i := \mathrm{E}\, W_{d_i} = \sum_{v=1}^n \mathrm{P}(D(v) = d_i) = \sum_{v=1}^n q_{v,d_i}.$$

We also abbreviate, for $v = 1, \ldots, n$,

$$\mu_v := \mathrm{E}\, D(v) = \sum_u p_{u,v},$$

https://doi.org/10.1239/aap/1331216648 Published online by Cambridge University Press

where, and as everywhere else, we use the convention that $p_{v,v} = 0$ in $\mathcal{G}(n, \{p_{ij}\})$. Then

$$\mathrm{var}(D(v)) = \sum_u p_{u,v}(1 - p_{u,v}) \le \mu_v. \tag{3.1}$$

For vertices $v_1, v_2, \ldots, v_m$, denote by $\mathcal{G}^{(v_1,\ldots,v_m)}(n-m, \{p_{ij}\})$ the random graph $\mathcal{G}(n, \{p_{ij}\})$ with vertices $v_1, v_2, \ldots, v_m$ and all their edges removed. For this graph, let $D^{(v_1,\ldots,v_m)}(w)$ be the degree of vertex w, where $w \notin \{v_1, \ldots, v_m\}$. We let

$$q_{w,d}^{(v_1,\ldots,v_m)} := \mathrm{P}(D^{(v_1,\ldots,v_m)}(w) = d).$$

It is straightforward to calculate that the entries of $\Sigma = (\sigma_{i,j})$, the covariance matrix of $\boldsymbol{W}$, are

$$\begin{aligned}\sigma_{ij} &= \mathbf{1}(i=j)\lambda_i - \sum_v q_{v,d_i} q_{v,d_j} \\ &\quad + \sum_v \sum_{w \ne v} [p_{wv} q_{v,d_i-1}^{(w)} q_{w,d_j-1}^{(v)} + (1 - p_{wv}) q_{v,d_i}^{(w)} q_{w,d_j}^{(v)} - q_{v,d_i} q_{w,d_j}].\end{aligned}$$

Using the notation from [9], for smooth functions $h : \mathbb{R}^p \to \mathbb{R}$, we let Dh denote the vector of first partial derivatives of h, and in general let D^k denote the kth derivative of h; $\|h\|$ denotes the supremum norm. We also abbreviate

$$\bar{p}_v = \frac{1}{n-1} \sum_w p_{wv}.$$

Define $\Sigma_0 = (\sigma_{ij}^0)$ with σ_{ij}^0 given by

$$\sigma_{ij}^0 = \mathbf{1}(i=j)\lambda_i - \sum_v q_{v,d_i} q_{v,d_j} + \sum_{v,w} \sqrt{\bar{p}_v}\sqrt{\bar{p}_w} (q_{v,d_i-1} - q_{v,d_i})(q_{w,d_j-1} - q_{w,d_j}).$$

Finally, let

$$M = \max\left\{\sum_v \mu_v; \sum_v \mu_v^3\right\}.$$

The following result gives a bound for the distance between the distribution of our degree count vector $\boldsymbol{W}$ to a multivariate normal distribution with the same mean as $\boldsymbol{W}$, but with covariance matrix Σ_0. The proof can be found in Section 4.

Theorem 3.1. *For any function $h : \mathbb{R}^p \to \mathbb{R}$ having bounded mixed partial derivatives up to order 3,*

$$\begin{aligned}|\mathrm{E}h(\Sigma_0^{-1/2}(\boldsymbol{W} - \boldsymbol{\lambda})) - Nh| &\le \frac{p^2\tau^2}{2}\|D^2 h\|(40p^2\sqrt{M} + S) \\ &\quad + \frac{p^5\tau^3}{3}\|D^3 h\|\left\{2M + \sum_{i=1}^p \lambda_i (d_i + 1)^2\right\}.\end{aligned}$$

Here

$$\begin{aligned}S &= 4p^2 \sum_{v,w} p_{wv}^2 + \sum_{v;w\ne v} |p_{wv} - \sqrt{\bar{p}_v}\sqrt{\bar{p}_w}| \sum_{i,j=1}^p |(q_{v,d_i-1} - q_{v,d_i})(q_{w,d_j-1} - q_{w,d_j})| \\ &\quad + \sum_v \bar{p}_v \sum_{i,j=1}^p |q_{v,d_i-1} - q_{v,d_i}||q_{v,d_j-1} - q_{v,d_j}|\end{aligned}$$

https://doi.org/10.1239/aap/1331216648 Published online by Cambridge University Press

and

$$\tau = \left[\sum_v \min_i q_{v,d_i}\left(1 - \sum_i q_{v,d_i}\right)\right]^{-1/2}.$$

Remark 3.1. For every $v = 1, \ldots, n$, the degree $D(v)$ can be approximated by a Poisson distribution with parameter μ_v. From [2, Equation (1.23), p. 8],

$$d_{\mathrm{TV}}(\mathcal{L}(D(v)), \mathrm{Po}(\mu_v)) \le \frac{1}{\mu_v}(1 - \mathrm{e}^{-\mu_v}) \sum_u p_{u,v}^2 \le \min\left(1; \frac{1}{\mu_v}\right) \sum_u p_{u,v}^2. \tag{3.2}$$

Here d_{TV} denotes the total variation distance; for two probability measures μ and ν on the same probability space with σ-algebra $\mathcal{B}$, we define

$$d_{\mathrm{TV}}(\mu, \nu) = \sup_{B \in \mathcal{B}} |\mu(B) - \nu(B)|.$$

The Poisson approximation is good, for example, when $p_{u,v} \approx c/(n-1) = \pi$ for all u, v and some constant c; then $\mu_v \approx c = O(1)$, and $\Sigma_u p_{u,v}^2 \approx c^2/(n-1)$. The distributional regime where the normal approximation is plausible is when all degrees are moderate, $\mu_v = O(1)$ for all v, so that $M = O(n)$; then it is reasonable to think of $\lambda_i \asymp n$ and $q_{v,d_i-1}^{(u)} = O(1)$ as well as $q_{v,d_i} = O(1)$. In this regime, with fixed p, $\tau \asymp n^{-1/2}$, and if $\sum_i d_i^2 = O(1)$, this yields an overall bound of the order $n^{-1/2}$.

Remark 3.2. The term S arises from the variance approximation;

$$\sum_v \sum_w |p_{wv} - \sqrt{\bar{p}_v}\sqrt{\bar{p}_w}| \sum_{i,j=1}^{p} |(q_{v,d_i-1} - q_{v,d_i})(q_{w,d_j-1} - q_{w,d_j})|$$

vanishes when all $p_{wv} = \pi$ are equal.

Remark 3.3. In the case that $p_{u,v} = c/n$ for all $u \ne v$, putting $q_d = q_{v,d}$, the approximating covariance simplifies to

$$\sigma_{ij}^0 = \mathbf{1}(i = j) n q_{d_i} + n q_{d_i} q_{d_j} \left[\frac{(n-1)(d_i - c)(d_j - c)}{nc(1 - d_i/n)(1 - d_j/n)} - 1\right].$$

Under the regime that d_i and d_j are typical degrees, so that q_{d_i} and q_{d_i} are moderate, this expression will not in general tend to 0 for $i \ne j$ as $n \to \infty$; the covariance does not in general vanish, and the degree counts will be asymptotically dependent.

Remark 3.4. While our bounds are for smooth test functions h only, they could be generalized to nonsmooth test functions along the lines of [15]. Corresponding work is in progress, [3], and to avoid duplicate work, we restrict ourselves to smooth test functions.

3.2. Poisson process approximation for the truncated degree sequences in $\mathcal{G}(n, \{p_{ij}\})$

Construction 2.1 of $(\boldsymbol{X}, \boldsymbol{X}^\beta)$ allows us to assess the distribution of the M-truncated degree sequence $\boldsymbol{D}_M := (D(v)\,\mathbf{1}(D(v) \ge M),\ v \in V)$ for an arbitrary integer $M \ge 0$. To this end, define Γ_M by $\{1, \ldots, n\} \times \{M, \ldots, n-1\}$, a subset of Γ, and restrict the definitions of $\boldsymbol{X}$ and $\boldsymbol{X}^\beta$ to Γ_M to have $\boldsymbol{X}_M := \{X_{(v,i)} : (v, i) \in \Gamma_M\}$ and $\boldsymbol{X}_M^\beta := \{X_{(v,i)}^\beta : (v, i) \in \Gamma_M\}$, such that $\mathcal{L}(\boldsymbol{X}_M^\beta) = \mathcal{L}(\boldsymbol{X}_M \mid X_\beta = 1)$ for $\beta \in \Gamma_M$. Construction 2.1 can be used to derive a Poisson process approximation, with respect to the total variation distance, for the point process Ξ_M defined in (1.2), where the target Poisson point process Θ_M on Γ_M has intensity $\boldsymbol{\lambda}_M = (\mathrm{E}\, X_\alpha,\ \alpha \in \Gamma_M)$.

https://doi.org/10.1239/aap/1331216648 Published online by Cambridge University Press

Theorem 3.2. *In $\mathcal{G}(n, \{p_{ij}\})$, we have*

$$d_{\mathrm{TV}}(\mathcal{L}(\Xi_M), \mathrm{Po}(\boldsymbol{\lambda}_M)) \le B_{M,1} + B_{M,2},$$

where

$$B_{M,1} = \sum_{v \in V} (\mathrm{P}(D(v) \ge M))^2 \quad \text{and} \quad B_{M,2} = 4 \sum_{v \in V} \sum_{u \in V_v} \mathrm{P}(D(v) \ge M)\, \mathrm{P}(D^{(v)}(u) \ge M-1).$$

Since the total variation distance between the two processes also serves as an upper bound of the total variation distance between deterministic functions of the two processes, that is, $d_{\mathrm{TV}}(\mathcal{L}(f(\Xi_M)), \mathcal{L}(f(\Theta_M))) \le d_{\mathrm{TV}}(\mathcal{L}(\Xi_M), \mathcal{L}(\Theta_M))$, where $\Theta_M \sim \mathrm{Po}(\boldsymbol{\lambda}_M)$ and f is any deterministic function, we assess the distribution of $\boldsymbol{D}_M$ by taking the function f on point measures ξ on Γ_M as $f(\xi) := (\sum_{i=0}^{n-1} i\xi((v,i)),\ v \in V)$. In this way, the target distribution $\mathcal{L}(f(\Theta_M))$ gives rise to a multivariate compound Poisson approximation for $\boldsymbol{D}_M$, in the next corollary. The result justifies the independence assumption among large vertex degrees as used when interpreting log-log plots for vertex degrees when the degrees are observed not in independent graphs, but in the same graph. It also bounds the departure from an independent point process in terms of the degree threshold M.

Corollary 3.1. *In the model $\mathcal{G}(n, \{p_{ij}\})$, let $\boldsymbol{D}_M$ denote the M-truncated degree sequence $(D(v)\,\mathbf{1}(D(v) \ge M),\ v \in V)$, and let $\boldsymbol{Y}_M$ denote the compound Poisson vector $(Y_{v,M},\ v \in V)$, in which all components are independent and $Y_{v,M} = \sum_{i=M}^{n-1} iY_{v,i}$ with $Y_{v,i} \sim \mathrm{Po}(\mathrm{E}\, X_{(v,i)})$. Then, with the $B_{M,i}$s from Theorem 3.2,*

$$d_{\mathrm{TV}}(\mathcal{L}(\boldsymbol{D}_M), \mathcal{L}(\boldsymbol{Y}_M)) \le B_{M,1} + B_{M,2}.$$

Remark 3.5. Corollary 3.1 is consistent with Theorem 3.13 of [5], where, in a fairly general submodel of $\mathcal{G}(n, \{p_{ij}\})$, it was shown that the distribution of an individual vertex degree converges to a mixed Poisson distribution. In contrast, Corollary 3.1 not only applies in the multivariate case, but it also provides an explicit error bound on the distance.

Remark 3.6. Using $\mu_u^{(v)} = \sum_{x \in V \setminus \{u,v\}} p_{ux}$, we obtain, from (3.2),

$$\mathrm{P}(D(v) \ge M) \le \mathrm{Po}(\mu_v)\{[M, n-1]\} + \frac{1 - \mathrm{e}^{-\mu_v}}{\mu_v} \sum_{x \in V_v} p_{vx}^2$$

and

$$\mathrm{P}(D^{(v)}(u) \ge M-1) \le \mathrm{Po}(\mu_u^{(v)})\{[M-1, n-2]\} + \frac{1 - \mathrm{e}^{-\mu_u^{(v)}}}{\mu_u^{(v)}} \sum_{x \in V \setminus \{u,v\}} p_{ux}^2.$$

This yields an upper bound for the quantities in Theorem 3.2; we can use the Poisson distribution as a guideline for a good choice of M. These probabilities could be further bounded using Proposition A.2.3 of [2].

3.3. Simulations for the correlation between counts

We now illustrate the dependence structure in four different random graph models, all on $n = 100$ vertices, with independent edges. We estimate the correlations from 10 000 samples of graphs for each model. The models are as follows.

(M1) The first model is the Bernoulli random graph with $p_{u,v} = p = 1/n$. This graph is at criticality; some, but not all, realizations may yield a giant component (see [4, Chapter 6]).

(M2) In this model $p_{u,v} = \frac{1}{5}$ if $0 < |u-v| \pmod{100} \le 10$ and $p_{u,v} = \frac{1}{80}$ if $|u-v| \pmod{100} > 10$. This is a modified Newman–Moore–Watts small-world model (see [13]) with 100 vertices; two vertices at most distance 10 away from each other are connected with probability $\frac{1}{5}$, and two vertices more than distance k away from each other are connected with probability $\frac{1}{80}$.

(M3) Here $p_{u,v} = \min(u, v)/n$ for $u \neq v$; the smaller of the two vertices determines the probability.

(M4) This model is motivated by Rasch-type models; for $u < v$, we set $p_{uv} = \alpha_u(3)\alpha_v(10)$, with

$$\alpha_u(i) = \begin{cases} \dfrac{1}{i\sqrt{n}}, & u \le n/2, \\ \dfrac{i}{\sqrt{n}}, & u > n/2. \end{cases}$$

Figure 1 shows the correlations between the degree counts in the four models; except for model (M3), there is an appreciable correlation even far away from the diagonal.

Figure 2 shows the degree count correlations, firstly between degree counts for k and $k+1$, and secondly for degree counts of an asymptotically normally distributed degree count and successive degree counts; the quantile-quantile plots are given for reassurance. We observe a

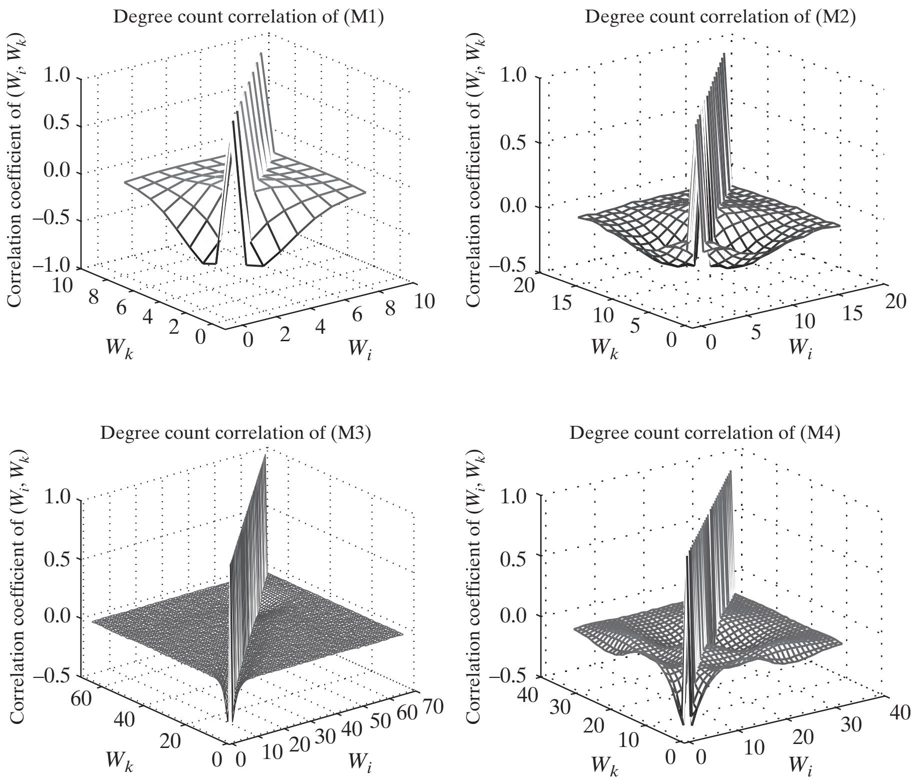

FIGURE 1: Degree count correlation in models (M1)–(M4).

https://doi.org/10.1239/aap/1331216648 Published online by Cambridge University Press

strong negative correlation for counts, but then close to zero correlation with counts of large degrees.

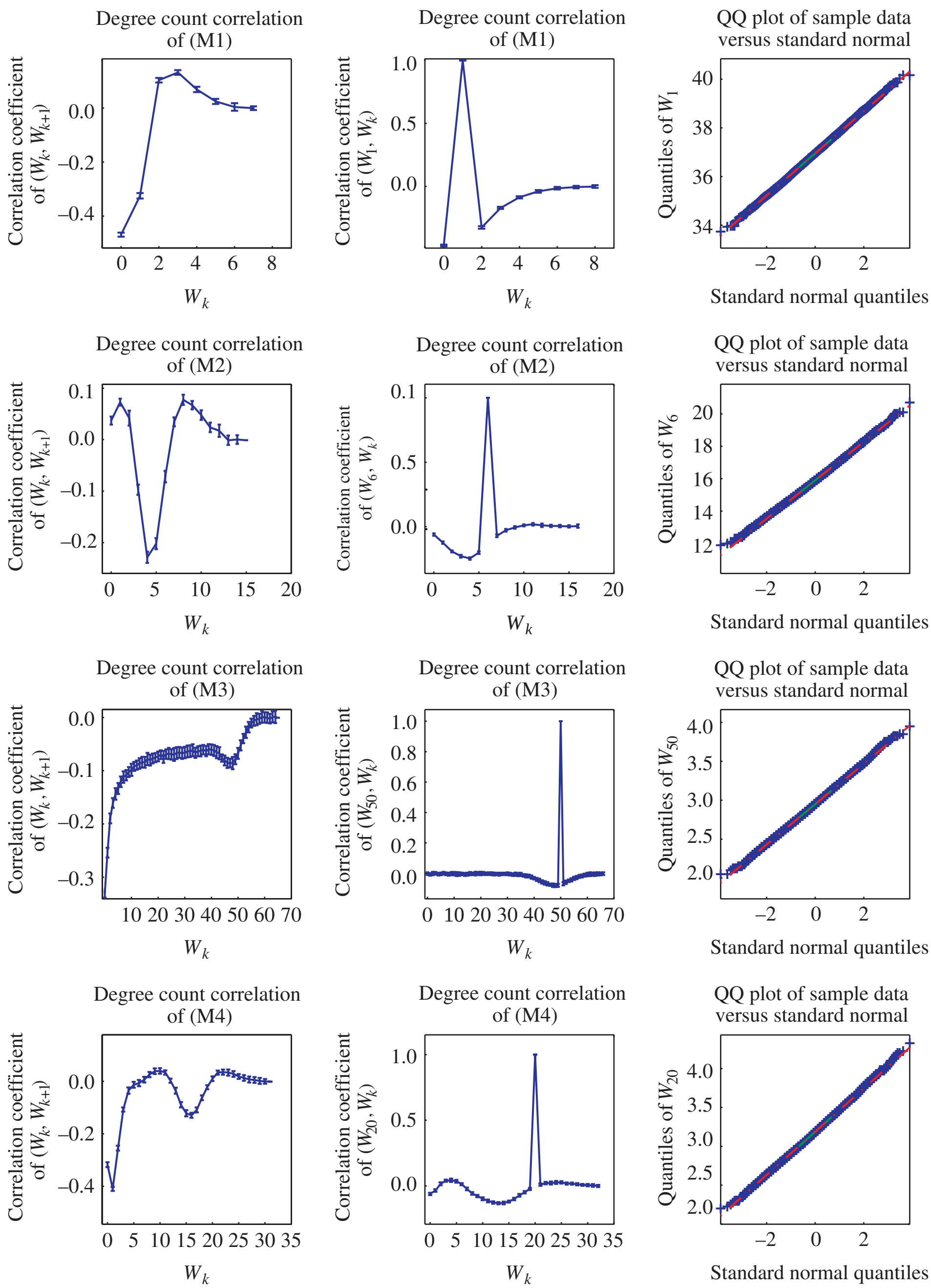

FIGURE 2: Degree count correlation and QQ plots in models (M1)–(M4).

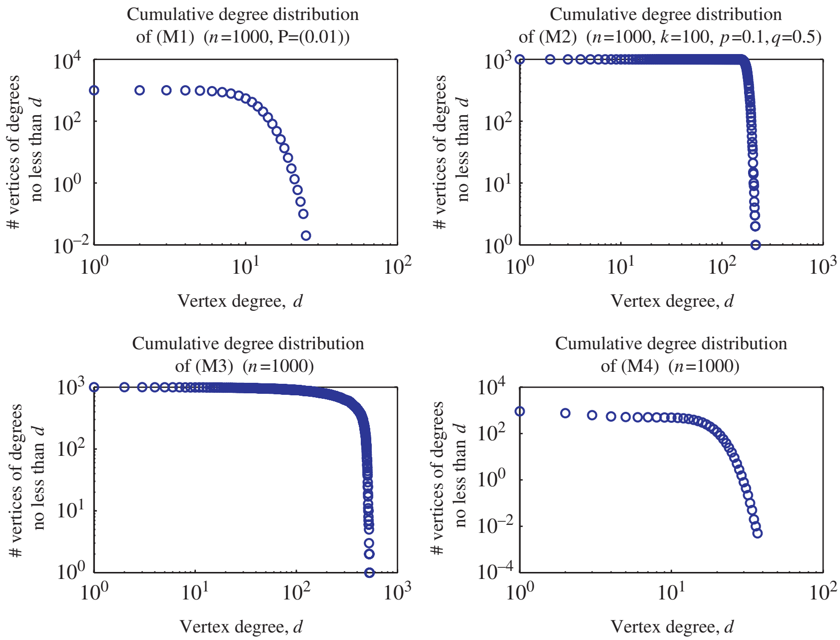

FIGURE 3: Power law; log-log scale.

3.4. Simulations for power-law-type behaviour

Using models (M1)–(M4), but now with $n = 1000$ vertices, in Figure 3 we plot the number of vertices of degree no less than d versus d itself, on a log-log scale. Despite the networks being created using independent edges, the plots seem to display a sharp linear decline, which could easily be mis-interpreted as displaying a power-law behaviour.

These simulations confirm the pseudo-power-law phenomena, and, therefore, raise the issue that, without rigorous analysis, simulation-based claims of detecting power-law-type behaviour, or scale-free behaviour, could in fact be unreliable and misleading. The vertex degree distribution may not be a suitable visual method for distinguishing different network models. In contrast, our distributional results help assess the joint distribution of vertex degrees under a fairly general null model.

4. Proofs

In this section we provide proofs for the size-biased coupling construction given in Construction 2.1, as well as for Theorem 4.1, Theorem 3.1, and Theorem 3.2. First we prove Lemma 2.1.

Proof of Lemma 2.1. We first rewrite $f^+(\boldsymbol{x}_i \mid \boldsymbol{x}_d)$ and $f^-(\boldsymbol{x}_i \mid \boldsymbol{x}_d)$ by writing out the set $N(v)$ in terms of those vertices which remain fixed in the construction, and those which get added or removed, respectively. Here $\boldsymbol{x}_i$, $\boldsymbol{x}_d$, $\boldsymbol{y}_j$, $\boldsymbol{z}_{(i-j)}$, and $\boldsymbol{z}_{(n-1-j+i)}$ are all subsets of V_v. We have

$$f^+(\boldsymbol{x}_i \mid \boldsymbol{x}_d) = \sum_{j=0}^{i} \sum_{\{\boldsymbol{y}_j\,:\,\boldsymbol{y}_j \subset \boldsymbol{x}_i\}} \sum_{\substack{\{\boldsymbol{z}_{(i-j)}\,: \\ \boldsymbol{z}_{(i-j)} \subset V_v \setminus \boldsymbol{x}_d\}}} \frac{1}{\binom{d-j}{i-j}} \mathrm{P}(N(v) = \boldsymbol{y}_j \cup \boldsymbol{z}_{(i-j)} \mid D(v) = i) \quad (4.1)$$

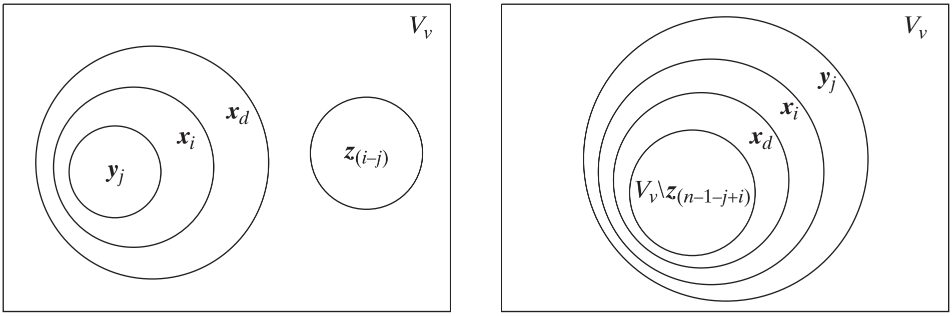

FIGURE 4: The set diagram for (4.1) (*left*) and (4.2) (*right*).

and

$$f^-(\boldsymbol{x}_i \mid \boldsymbol{x}_d) = \sum_{j=i}^{n-1} \sum_{\{\boldsymbol{y}_j : \boldsymbol{y}_j \supset \boldsymbol{x}_i\}} \sum_{\substack{\{z_{(n-1-j+i)}: \\ z_{(n-1-j+i)} \supset V_v \setminus \boldsymbol{x}_d\}}} \frac{1}{\binom{j-d}{j-i}} \mathrm{P}(N(v) = \boldsymbol{y}_j \cap z_{(n-1-j+i)} \mid D(v) = i). \tag{4.2}$$

Figure 4 illustrates the set relations in (4.1) and (4.2). See Lemma 4.3.2 of [11] for more details.

Now, from (4.1), since, for any $\boldsymbol{x}_d$ and $\boldsymbol{y}_j$, there are $\binom{d-j}{i-j}$ choices of $\boldsymbol{x}_i$, we have

$$\sum_{\{\boldsymbol{x}_i : \boldsymbol{x}_i \subset \boldsymbol{x}_d\}} f^+(\boldsymbol{x}_i \mid \boldsymbol{x}_d) = \frac{1}{\mathrm{P}(D(v) = i)} \times \sum_{j=0}^{i} \sum_{\{\boldsymbol{y}_j : \boldsymbol{y}_j \subset \boldsymbol{x}\}} \sum_{\{z_{(i-j)} : z_{(i-j)} \subset V_v \setminus \boldsymbol{x}_d\}} \mathrm{P}(N(v) = \boldsymbol{y}_j \cup z_{(i-j)}).$$

Let $\boldsymbol{X}_i = \{\boldsymbol{x}_i : \boldsymbol{x}_i \subset V_v\}$. Note that $\boldsymbol{\omega}_i \in \boldsymbol{X}_i$ if and only if $\boldsymbol{\omega}_i$ can be uniquely decomposed as $\boldsymbol{\omega}_i = \boldsymbol{y}_j \cup z_{(i-j)}$ such that $\boldsymbol{y}_j \subset \boldsymbol{x}_d$ and $z_{(i-j)} \subset V_v \setminus \boldsymbol{x}_d$. Thus,

$$\sum_{\{\boldsymbol{x}_i : \boldsymbol{x}_i \subset \boldsymbol{x}_d\}} f^+(\boldsymbol{x}_i \mid \boldsymbol{x}_d) = \frac{1}{\mathrm{P}(D(v) = i)} \sum_{\boldsymbol{\omega}_i \in \boldsymbol{X}_i} \mathrm{P}(N(v) = \boldsymbol{\omega}_i) = 1,$$

as required. Similarly, from (4.2) we find that

$$\sum_{\{\boldsymbol{x}_i : \boldsymbol{x}_i \supset \boldsymbol{x}_d\}} f^-(\boldsymbol{x}_i \mid \boldsymbol{x}_d) = \frac{1}{\mathrm{P}(D(v) = i)} \times \sum_{j=i}^{n-1} \sum_{\{\boldsymbol{y}_j : \boldsymbol{y}_j \supset \boldsymbol{x}_d\}} \sum_{\substack{\{z_{(n-1-j+i)}: \\ z_{(n-1-j+i)} \supset V_v \setminus \boldsymbol{x}_d\}}} \mathrm{P}(N(v) = \boldsymbol{y}_j \cap z_{(n-1-j+i)}).$$

Note that, $\boldsymbol{\omega}_i \in \boldsymbol{X}_i$ if and only if $\boldsymbol{\omega}_i$ can be uniquely written as $\boldsymbol{\omega}_i = \boldsymbol{y}_j \cap z_{(n-1-j+i)}$ such that $\boldsymbol{y}_j \supset \boldsymbol{x}_d$ and $z_{(n-1-j+i)} \supset V_v \setminus \boldsymbol{x}_d$. Thus,

$$\sum_{\{\boldsymbol{x}_i : \boldsymbol{x}_i \supset \boldsymbol{x}_d\}} f^-(\boldsymbol{x}_i \mid \boldsymbol{x}_d) = \frac{1}{\mathrm{P}(D(v) = i)} \sum_{\boldsymbol{\omega}_i \in \boldsymbol{X}_i} \mathrm{P}(N(v) = \boldsymbol{\omega}_i) = 1,$$

as required.

Proof of Construction 2.1. The goal of this proof is to show that, for $\beta = (v, i) \in \Gamma$,

$$\mathcal{L}(\boldsymbol{X}^{\beta}) = \mathcal{L}(\boldsymbol{X} \mid X_{\beta} = 1), \tag{4.3}$$

where $\boldsymbol{X} = \{X_{(v,i)} \colon (v, i) \in \Gamma\}$ and $\boldsymbol{X}^{\beta} = \{X^{\beta}_{(v,i)} \colon (v, i) \in \Gamma\}$. Indeed, we shall show that the distribution of the constructed model $\mathcal{G}^{\beta}(n, \{p_{ij}\})$ is the same as the conditional distribution of the original model $\mathcal{G}(n, \{p_{ij}\})$ given $X_{\beta} = 1$, that is, with $\boldsymbol{X}_{\text{edge}} = \{X_{\{a,b\}} \colon \{a, b\} \in E\}$ and $\boldsymbol{X}^{\beta}_{\text{edge}} = \{X^{\beta}_{\{a,b\}} \colon \{a, b\} \in E\}$, we need to show that, for all $\boldsymbol{w} := (w_{a,b},\ \{a, b\} \in E) \in \{0, 1\}^{|E|}$,

$$\mathrm{P}(\boldsymbol{X}^{(v,i)}_{\text{edge}} = \boldsymbol{w}) = \mathrm{P}(\boldsymbol{X}_{\text{edge}} = \boldsymbol{w} \mid X_{(v,i)} = 1). \tag{4.4}$$

The desired equation (4.3) then follows because $\boldsymbol{X}$ is a function of $\boldsymbol{X}_{\text{edge}}$.

By the definition of $X_{(v,i)}$, the right-hand side of (4.4) (denoted RHS) is 0 when $\sum_{x \in V_v} w_{v,x} \neq i$, and, by construction, in that case the left-hand side of (4.4) (denoted LHS) is 0 also. Assume that $\sum_{x \in V_v} w_{v,x} = i$. Then

$$\begin{aligned}
\text{RHS} &:= \mathrm{P}(\boldsymbol{X}_{\text{edge}} = \boldsymbol{w} \mid X_{(v,i)} = 1) \\
&= \mathrm{P}(\boldsymbol{X}_{\text{edge}} = \boldsymbol{w} \mid D(v) = i) \\
&= \mathrm{P}(X_{\{a,b\}} = w_{a,b} \text{ for all } a \neq v,\ b \neq v;\ N(v) = \{x \in V_v \colon w_{v,x} = 1\} \mid D(v) = i) \\
&= \mathrm{P}(X_{\{a,b\}} = w_{a,b} \text{ for all } a \neq v,\ b \neq v)\, \mathrm{P}(N(v) = \boldsymbol{x}_i \mid D(v) = i),
\end{aligned} \tag{4.5}$$

where the last equality follows from the independence of the edges, as the condition $D(v) = i$ only affects $(X_{\{v,x\}},\ x \in V_v)$, but not edges which do not contain v. On the other hand,

$$\begin{aligned}
\text{LHS} &= \mathrm{P}(X_{\{a,b\}} = w_{a,b} \text{ for all } a \neq v,\ b \neq v;\ X^{(v,i)}_{\{v,x\}} = w_{v,x} \text{ for } x \in V_v) \\
&= \mathrm{P}(X_{\{a,b\}} = w_{a,b} \text{ for all } a \neq v,\ b \neq v)\, \mathrm{P}(X^{(v,i)}_{\{v,x\}} = w_{v,x} \text{ for } x \in V_v),
\end{aligned} \tag{4.6}$$

as the construction of $\boldsymbol{X}^{(v,i)}_{\text{edge}}$ from $\boldsymbol{X}_{\text{edge}}$ affects only the edges with v as one of its endpoints, and the edges are independent. Note that, in (4.6),

$$\mathrm{P}(X^{(v,i)}_{\{v,x\}} = w_{v,x} \text{ for all } x \in V_v) = \mathrm{P}(N^{(v,i)}(v) = \boldsymbol{x}_i).$$

Hence, to conclude that (4.6) equals (4.5), it remains to show that

$$\mathrm{P}(N^{(v,i)}(v) = \boldsymbol{x}_i) = \mathrm{P}(N(v) = \boldsymbol{x}_i \mid D(v) = i). \tag{4.7}$$

Indeed,

$$\begin{aligned}
&\mathrm{P}(N^{(v,i)}(v) = \boldsymbol{x}_i) \\
&\quad = q_{v,i}\, \mathrm{P}(N(v) = \boldsymbol{x}_i \mid D(v) = i) \\
&\qquad + \sum_{\{d \colon d > i\}} q_{v,d} \sum_{\{\boldsymbol{x}_d \colon \boldsymbol{x}_d \supset \boldsymbol{x}_i\}} \mathrm{P}(N(v) = \boldsymbol{x}_d \mid D(v) = d)\, \mathrm{P}(N^{(v,i)}(v) = \boldsymbol{x}_i \mid N(v) = \boldsymbol{x}_d) \\
&\qquad + \sum_{\{d \colon d < i\}} q_{v,d} \sum_{\{\boldsymbol{x}_d \colon \boldsymbol{x}_d \subset \boldsymbol{x}_i\}} \mathrm{P}(N(v) = \boldsymbol{x}_d \mid D(v) = d)\, \mathrm{P}(N^{(v,i)}(v) = \boldsymbol{x}_i \mid N(v) = \boldsymbol{x}_d),
\end{aligned} \tag{4.8}$$

where the three terms corresponding to the coupling construction given in Construction 2.1. Now, we calculate the sums over $\{\boldsymbol{x}_d : \boldsymbol{x}_d \supset \boldsymbol{x}_i\}$ and over $\{\boldsymbol{x}_d : \boldsymbol{x}_d \subset \boldsymbol{x}_i\}$ separately. In fact, for the first case $d > i$ and $\boldsymbol{x}_d \supset \boldsymbol{x}_i$, it follows from (2.1) and (4.1) that

$$\begin{aligned}\sum_{\{\boldsymbol{x}_d\,:\,\boldsymbol{x}_d\supset\boldsymbol{x}_i\}} &\mathrm{P}(N(v)=\boldsymbol{x}_d \mid D(v)=d)\,\mathrm{P}(N^{(v,i)}(v)=\boldsymbol{x}_i \mid N(v)=\boldsymbol{x}_d)\\ &=\sum_{j=0}^{i}\sum_{\{\boldsymbol{x}_d\,:\,\boldsymbol{x}_d\supset\boldsymbol{x}_i\}}\sum_{\{\boldsymbol{y}_j\,:\,\boldsymbol{y}_j\subset\boldsymbol{x}_i\}}\sum_{\substack{\{\boldsymbol{z}_{(i-j)}\,:\\ \boldsymbol{z}_{(i-j)}\subset V_v\setminus\boldsymbol{x}_d\}}}\frac{1}{\binom{d-j}{i-j}}\frac{\mathrm{P}(N(v)=\boldsymbol{x}_d)\,\mathrm{P}(N(v)=\boldsymbol{y}_j\cup\boldsymbol{z}_{(i-j)})}{\mathrm{P}(D(v)=d)\,\mathrm{P}(D(v)=i)},\end{aligned}\tag{4.9}$$

where

$$\mathrm{P}(N(v)=\boldsymbol{x}_d)=\bigg[\bigg(\prod_{x\in\boldsymbol{x}_d}p_{vx}\bigg)\bigg(\prod_{x\in V_v\setminus\boldsymbol{x}_d}(1-p_{vx})\bigg)\bigg].$$

Since $\boldsymbol{y}_j \subset \boldsymbol{x}_i \subset \boldsymbol{x}_d$ and $\boldsymbol{z}_{(i-j)} \subset V_v \setminus \boldsymbol{x}_d$, we have $V_v \setminus (\boldsymbol{y}_j \cup \boldsymbol{z}_{(i-j)}) \supset \boldsymbol{x}_d \setminus \boldsymbol{x}_i$ (see Figure 4 for reference). Therefore,

$$\begin{aligned}&\mathrm{P}(N(v)=\boldsymbol{x}_d)\,\mathrm{P}(N(v)=\boldsymbol{y}_j\cup\boldsymbol{z}_{(i-j)})\\ &\quad=\bigg[\bigg(\prod_{x\in\boldsymbol{x}_i}p_{vx}\bigg)\bigg(\prod_{x\in V_v\setminus\boldsymbol{x}_d}(1-p_{vx})\bigg)\bigg(\prod_{x\in\boldsymbol{x}_d\setminus\boldsymbol{x}_i}(1-p_{vx})\bigg)\bigg]\\ &\quad\quad\times\bigg[\bigg(\prod_{x\in\boldsymbol{x}_d\setminus\boldsymbol{x}_i}p_{vx}\bigg)\bigg(\prod_{x\in\boldsymbol{y}_j\cup\boldsymbol{z}_{(i-j)}}p_{vx}\bigg)\bigg(\prod_{x\in V_v\setminus[\boldsymbol{y}_j\cup\boldsymbol{z}_{(i-j)}\cup(\boldsymbol{x}_d\setminus\boldsymbol{x}_i)]}(1-p_{vx})\bigg)\bigg]\\ &\quad=\mathrm{P}(N(v)=\boldsymbol{x}_i)\,\mathrm{P}(N(v)=\boldsymbol{y}_j\cup\boldsymbol{z}_{(i-j)}\cup(\boldsymbol{x}_d\setminus\boldsymbol{x}_i)).\end{aligned}$$

Hence, from (4.9), we have, for $d > i$,

$$\begin{aligned}\sum_{\{\boldsymbol{x}_d\,:\,\boldsymbol{x}_d\supset\boldsymbol{x}_i\}} &\mathrm{P}(N(v)=\boldsymbol{x}_d \mid D(v)=d)\,\mathrm{P}(N^{(v,i)}(v)=\boldsymbol{x}_i \mid N(v)=\boldsymbol{x}_d)\\ &=\frac{\mathrm{P}(N(v)=\boldsymbol{x}_i)}{\mathrm{P}(D(v)=i)}\frac{1}{\mathrm{P}(D(v)=d)}\\ &\quad\times\sum_{j=0}^{i}\sum_{\{\boldsymbol{x}_d\,:\,\boldsymbol{x}_d\supset\boldsymbol{x}_i\}}\sum_{\{\boldsymbol{y}_j\,:\,\boldsymbol{y}_j\subset\boldsymbol{x}_i\}}\sum_{\substack{\{\boldsymbol{z}_{(i-j)}\,:\\ \boldsymbol{z}_{(i-j)}\subset V_v\setminus\boldsymbol{x}_d\}}}\frac{1}{\binom{d-j}{i-j}}\mathrm{P}(N(v)=\boldsymbol{y}_j\cup\boldsymbol{z}_{(i-j)}\cup(\boldsymbol{x}_d\setminus\boldsymbol{x}_i)).\end{aligned}\tag{4.10}$$

Since $\boldsymbol{y}_j \cup \boldsymbol{z}_{(i-j)}$ is an i-set (i.e. a set with i elements), $(\boldsymbol{x}_d \setminus \boldsymbol{x}_i)$ is a $(d-i)$-set, and $(\boldsymbol{y}_j \cup \boldsymbol{z}_{(i-j)}) \cap (\boldsymbol{x}_d \setminus \boldsymbol{x}_i) = \varnothing$, it follows that $\boldsymbol{y}_j \cup \boldsymbol{z}_{(i-j)} \cup (\boldsymbol{x}_d \setminus \boldsymbol{x}_i)$ is a d-set in $\boldsymbol{X}_d$. Conversely, for any d-set $\boldsymbol{\omega}_d \in \boldsymbol{X}_d$, since $\boldsymbol{x}_i$ is fixed, we can decompose $\boldsymbol{\omega}_d$ as

$$\boldsymbol{\omega}_d = \boldsymbol{y}_j \cup \boldsymbol{z}_{(i-j)} \cup (\boldsymbol{x}_d \setminus \boldsymbol{x}_i),\tag{4.11}$$

such that $\boldsymbol{y}_j \subset \boldsymbol{x}_i$, $\boldsymbol{x}_d \supset \boldsymbol{x}_i$, and $\boldsymbol{z}_{(i-j)} \subset V_v \setminus \boldsymbol{x}_d$. Referring to Figure 4, it is easy to show that, for any $\boldsymbol{\omega}_d \in \boldsymbol{X}_d$, there are $\binom{d-j}{i-j}$ solutions $(\hat{\boldsymbol{y}}_j, \hat{\boldsymbol{x}}_d, \hat{\boldsymbol{z}}_{(i-j)})$ to decompose $\boldsymbol{\omega}_d$ as (4.11) (see Lemma 4.3.4 of [11] for more details). Thus, it follows that

$$\begin{aligned}
&\sum_{j=0}^{i} \sum_{\{\boldsymbol{x}_d\,:\,\boldsymbol{x}_d \supset \boldsymbol{x}_i\}} \sum_{\{\boldsymbol{y}_j\,:\,\boldsymbol{y}_j \subset \boldsymbol{x}_i\}} \sum_{\{\boldsymbol{z}_{(i-j)}\,:\,\boldsymbol{z}_{(i-j)} \subset V_v \setminus \boldsymbol{x}_d\}} \frac{1}{\binom{d-j}{i-j}} \mathrm{P}(N(v) = \boldsymbol{y}_j \cup \boldsymbol{z}_{(i-j)} \cup (\boldsymbol{x}_d \setminus \boldsymbol{x}_i)) \\
&\quad = \sum_{\boldsymbol{\omega}_d \in \boldsymbol{X}_d} \mathrm{P}(N(v) = \boldsymbol{\omega}_d) \\
&\quad = \mathrm{P}(D(v) = d),
\end{aligned}$$

and from (4.10) we have, for $d > i$,

$$\sum_{\{\boldsymbol{x}_d\,:\,\boldsymbol{x}_d \supset \boldsymbol{x}_i\}} \mathrm{P}(N(v) = \boldsymbol{x}_d \mid D(v) = d)\, \mathrm{P}(N^{(v,i)}(v) = \boldsymbol{x}_i \mid N(v) = \boldsymbol{x}_d) = \frac{\mathrm{P}(N(v) = \boldsymbol{x}_i)}{\mathrm{P}(D(v) = i)}. \tag{4.12}$$

The case $d < i$ and $\boldsymbol{x}_d \subset \boldsymbol{x}_i$ is treated similarly, giving

$$\sum_{\{\boldsymbol{x}_d\,:\,\boldsymbol{x}_d \subset \boldsymbol{x}_i\}} \mathrm{P}(N(v) = \boldsymbol{x}_d \mid D(v) = d)\, \mathrm{P}(N^{(v,i)}(v) = \boldsymbol{x}_i \mid N(v) = \boldsymbol{x}_d) = \frac{\mathrm{P}(N(v) = \boldsymbol{x}_i)}{\mathrm{P}(D(v) = i)}. \tag{4.13}$$

See [11] for details. Combining (4.8), (4.12), and (4.13),

$$\mathrm{P}(N^{(v,i)}(v) = \boldsymbol{x}_i) = \frac{\mathrm{P}(N(v) = \boldsymbol{x}_i)}{\mathrm{P}(D(v) = i)} = \mathrm{P}(N(v) = \boldsymbol{x}_i \mid D(v) = i),$$

as required in (4.7) to complete the proof.

The proof of Theorem 3.1 is based on the following theorem, which is similar to Theorem 1.2 of [9], but gives a multivariate normal approximation with respect to an alternative covariance matrix Σ_0 for which $\|\Sigma_0\|$ is straightforward to bound and which is close to Σ. We use the notation from [9]. For a vector $\boldsymbol{b} \in \mathbb{R}^p$, we let $\|\boldsymbol{b}\| = \max_{1 \le i \le p} |b_i|$. More generally, for an array $A = (a_{i,j})$, the notation $\|\cdot\|$ is its maximal absolute value. For an array $A(\mathbf{w}) = \{a_i(\mathbf{w})\}$ of functions, $\|A\| = \sup_{\mathbf{w}} \max_i |a_i(\mathbf{w})|$.

Theorem 4.1. *Let $\boldsymbol{W} = (W_i,\ 1 \le i \le p)$ be a random vector in $\mathbb{R}^p$ with nonnegative components. Let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p) = \mathrm{E}\,\boldsymbol{W}$, and assume that* $\mathrm{var}(\boldsymbol{W}) = \Sigma = (\sigma_{ij})$ *exists. Let $\Sigma_0 = (\sigma^0_{ij})$ be a positive definite (invertible) $p \times p$ matrix. For each $i = 1, \ldots, p$, let $(\boldsymbol{W}, \boldsymbol{W}^i)$ be a random vector defined on a joint probability space with $\boldsymbol{W}^i$ having the $\boldsymbol{W}$-size-biased distribution in the ith coordinate. Let $h\colon \mathbb{R}^p \to \mathbb{R} \in C_b^3$, and let $Nh = \mathrm{E}\,h(\boldsymbol{Z})$, where $\boldsymbol{Z}$ denotes a standard normal variable in $\mathbb{R}^p$. Then*

$$\begin{aligned}
&|\mathrm{E}h(\Sigma_0^{-1/2}(\boldsymbol{W} - \boldsymbol{\lambda})) - Nh| \\
&\quad \le \frac{p^2}{2} \|\Sigma_0^{-1/2}\|^2 \|D^2 h\| \sum_{i=1}^{p} \sum_{j=1}^{p} \lambda_i \sqrt{\mathrm{var}\,\mathrm{E}[W_j^i - W_j \mid \boldsymbol{W}]} \\
&\qquad + \frac{1}{2} \frac{p^3}{3} \|\Sigma_0^{-1/2}\|^3 \|D^3 h\| \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{k=1}^{p} \lambda_i\, \mathrm{E}\,|(W_j^i - W_j)(W_k^i - W_k)| \\
&\qquad + \frac{p^2}{2} \|\Sigma_0^{-1/2}\|^2 \|D^2 h\| \sum_{i=1}^{p} \sum_{j=1}^{p} |\sigma^0_{ij} - \sigma_{ij}|.
\end{aligned}$$

https://doi.org/10.1239/aap/1331216648 Published online by Cambridge University Press

Proof. The proof follows closely the lines of the proof of Theorem 1.2 of [9]. The only difference is that instead of their decomposition (18), we use

$$
\begin{aligned}
&\mathrm{E}[h(\Sigma_0^{-1/2}(\boldsymbol{W}-\boldsymbol{\lambda})) - Nh] \\
&\quad = -\mathrm{E}\bigg[\sum_{i,j=1}^{p}[\lambda_i(W_j^i - W_j) - \sigma_{ij}]\frac{\partial^2}{\partial w_i \partial w_j} f(\boldsymbol{W})\bigg] && (4.14)\\
&\qquad - \mathrm{E}\bigg[\sum_{i,j=1}^{p}(\sigma_{ij} - \sigma_{ij}^0)\frac{\partial^2}{\partial w_i \partial w_j} f(\boldsymbol{W})\bigg] && (4.15)\\
&\qquad - \mathrm{E}\bigg[\sum_{i,j,k=1}^{p}\lambda_i(W_j^i - W_j)(W_k^i - W_k)\int_0^1 (1-t)\frac{\partial^3}{\partial w_i \partial w_j \partial w_k} f[\boldsymbol{W} + t(\boldsymbol{W}^i - \boldsymbol{W})] \\
&\qquad\qquad \times (W_j^i - W_j)(W_k^i - W_k)\,\mathrm{d}t\bigg]. && (4.16)
\end{aligned}
$$

The bounds for (4.14) and (4.16) are as in [9]; for (4.15), we obtain the bound

$$\frac{p^2}{2}\|\Sigma_0^{-1/2}\|^2\|D^2h\|\sum_{i=1}^{p}\sum_{j=1}^{p}|\sigma_{ij} - \sigma_{ij}^0|.$$

This completes the proof.

Proof of Theorem 3.1. The proof is based on Theorem 4.1 and the size-biased coupling construction given in Construction 2.1. For $i = 1, \dots, p$, let $d_i \in \{0, \dots, n-1\}$ and $\beta = (v, d_i) \in A_i$. We use Construction 2.1 to obtain $\mathcal{G}^{\beta}(n, \{p_{ij}\})$. Now we randomize over the first component v of β, but keep i fixed; we pick I such that $\mathrm{P}(I = v) = q_{v,d_i}/\lambda_i$; we call the resulting graph $\mathcal{G}^i(n, \{p_{ij}\})$. Denote by $W_{d_j}^{d_i}$ the number of vertices of degree d_j in $\mathcal{G}^i(n, \{p_{ij}\})$. If $I = v$ and $D(v)$ is the degree of v in $\mathcal{G}(n, \{p_{ij}\})$, then $|W_{d_j}^{d_i} - W_{d_j}| \le |D(v) - d_i| + 1$, because at most $|D(v) - d_i|$ edges are added or removed; and the degree of v is fixed to equal d_i in the size-biased distribution. Hence,

$$\mathrm{E}\,|(W_{d_j}^{d_i} - W_{d_j})(W_{d_k}^{d_i} - W_{d_k})| \le \mathrm{E}(D(I) + d_i + 1)^2 \le 2(d_i+1)^2 + 2\,\mathrm{E}\,D(I)^2.$$

Now, for d_i, the vertex $I_1 = v$ is chosen proportional to q_{v,d_i}; hence,

$$\mathrm{E}\,D(I)^2 = \frac{1}{\lambda_i}\sum_{v=1}^{n} q_{v,d_i}(\operatorname{var} D(v) + (\mathrm{E}\,D(v))^2),$$

and we use (3.1) to bound the variance.

The bound on $\sqrt{\operatorname{var}\mathrm{E}[W_{d_j}^{d_i} - W_{d_j} \mid \boldsymbol{W}]}$ is straightforward and follows the lines of [9]. First note that $\operatorname{var}\mathrm{E}[W_{d_j}^{d_i} - W_{d_j} \mid \boldsymbol{W}] \le \operatorname{var}\mathrm{E}[W_{d_j}^{d_i} - W_{d_j} \mid \mathcal{G}(n, \{p_{ij}\})]$. With the notation (2.1) and (2.2), we abbreviate

$$
\begin{aligned}
\alpha_{i,j}(u, v) &= \alpha(u, v) \\
&= \frac{q_{v,d_i}}{\lambda_i}\sum_{d > d_i}\sum_{\{\boldsymbol{x}_d \colon u \in \boldsymbol{x}_d\}} \mathbf{1}(N(v) = \boldsymbol{x}_d)\{\mathbf{1}(D(u) = d_j + 1) - \mathbf{1}(D(u) = d_j)\} \\
&\quad \times \sum_{\boldsymbol{x}_{d_i} \subset \boldsymbol{x}_d; u \notin \boldsymbol{x}_{d_i}} f^{+}(\boldsymbol{x}_{d_i} \mid \boldsymbol{x}_d)
\end{aligned}
$$

and

$$\begin{aligned}\beta_{i,j}(u, v) &= \beta(u, v)\\ &= \frac{q_{v,d_i}}{\lambda_i} \sum_{d<d_i} \sum_{\{\boldsymbol{x}_d\,:\,u\notin \boldsymbol{x}_d\}} \mathbf{1}(N(v) = \boldsymbol{x}_d)\{\mathbf{1}(D(u) = d_j - 1) - \mathbf{1}(D(u) = d_j)\}\\ &\quad \times \sum_{\boldsymbol{x}_{d_i}\supset \boldsymbol{x}_d;\, u\in \boldsymbol{x}_{d_i}} f^-(\boldsymbol{x}_{d_i} \mid \boldsymbol{x}_d).\end{aligned} \tag{4.17}$$

We note that

$$|\alpha(u, v)| \le \frac{q_{v,d_i}}{\lambda_i} \quad \text{and} \quad |\beta(u, v)| \le \frac{q_{v,d_i}}{\lambda_i}, \tag{4.18}$$

and $\alpha(u, v) = 0$ when $u \not\sim v$; $\beta(u, v) = 0$ when $u \sim v$. With this notation,

$$\begin{aligned}&\mathrm{E}[W_{d_j}^{d_i} - W_{d_j} \mid \mathcal{G}(n, \{p_{ij}\})]\\ &\quad = \sum_v \sum_{u\neq v} \alpha(u, v) + \sum_v \sum_{u\neq v} \beta(u, v)\\ &\qquad + \frac{1}{\lambda_i} \sum_v q_{v,d_i}\, \mathbf{1}(D(v) \neq d_i)\, \mathbf{1}(i = j) - \frac{1}{\lambda_i} \sum_v q_{v,d_i}\, \mathbf{1}(D(v) = d_j)\, \mathbf{1}(i \neq j).\end{aligned}$$

This gives

$$\begin{aligned}\operatorname{var} \mathrm{E}[W_{d_j}^{d_i} - W_{d_j} \mid \boldsymbol{W}] \le 4\Bigg\{&\operatorname{var}\Big(\sum_v \sum_{u\neq v} \alpha(u, v)\Big) &(4.19)\\ &+ \operatorname{var}\Big(\sum_v \sum_{u\neq v} \beta(u, v)\Big) &(4.20)\\ &+ \mathbf{1}(i = j) \operatorname{var}\Big(\frac{1}{\lambda_i} \sum_v q_{v,d_i}\, \mathbf{1}(D(v) \neq d_i)\Big) &(4.21)\\ &+ \mathbf{1}(i \neq j) \operatorname{var}\Big(\frac{1}{\lambda_i} \sum_v q_{v,d_i}\, \mathbf{1}(D(v) = d_j)\Big)\Bigg\}. &(4.22)\end{aligned}$$

Firstly, for (4.19), with (4.18) and using the fact that $\alpha(u, v) = \alpha(u, v)\, \mathbf{1}(u \sim v)$,

$$\begin{aligned}&\lambda_i^2 \operatorname{var}(\alpha(u, v)) \le \mathrm{E}(\alpha(u, v))^2 \le p_{u,v} q_{v,d_i}^2,\\ &\lambda_i^2 |\operatorname{cov}(\alpha(u, v); \alpha(u', v))| \le 2p_{u,v} p_{u',v} q_{v,d_i}^2, \qquad u' \neq u, v,\\ &\lambda_i^2 |\operatorname{cov}(\alpha(u, v); \alpha(u, v'))| \le 2p_{u,v} p_{u,v'} q_{v,d_i} q_{v',d_i}, \qquad v' \neq u, v,\\ &\lambda_i^2 |\operatorname{cov}(\alpha(u, v); \alpha(u', u))| \le 2p_{u,v} p_{u',u} q_{v,d_i} q_{u,d_i}, \qquad u' \neq u, v,\\ &\lambda_i^2 |\operatorname{cov}(\alpha(u, v); \alpha(v, v'))| \le 2p_{u,v} p_{v,v'} q_{v,d_i} q_{v',d_i}, \qquad v' \neq u, v.\end{aligned}$$

The contributions of these terms to the overall variance can hence be bounded by

$$\begin{aligned}&\sum_v \sum_{u\neq v} p_{u,v} q_{v,d_i}^2 + 2 \sum_{u,u',v \text{ distinct}} p_{u,v} p_{u',v} q_{v,d_i}^2 + 2 \sum_{u,v,v' \text{ distinct}} p_{u,v} p_{u,v'} q_{v,d_i} q_{v',d_i}\\ &+ 2 \sum_{u,u',v \text{ distinct}} p_{u,v} p_{u',u} q_{v,d_i} q_{u,d_i} + 2 \sum_{u,v,v' \text{ distinct}} p_{u,v} p_{v,v'} q_{v,d_i} q_{v',d_i}\\ &\quad \le 9M.\end{aligned}$$

Here we used the crude bound that $q_{v,d} \leq 1$. For $u' \neq u, v$ and $v' \neq u, u', v$, let $C = C(u, u', v, v') = \mathbf{1}(u \not\sim u', u \not\sim v', v \not\sim v', u' \not\sim v)$. Then

$$\mathrm{P}(C(u, u', v, v') = 1) \geq 1 - (p_{u,u'} + p_{u,v'} + p_{v,v'} + p_{u',v}).$$

Put $\lambda(u, v) = \mathrm{E}\,\alpha(u, v)$; then, for u, u', v, v' mutually different,

$$\begin{aligned}\operatorname{cov}(\alpha(u, v), \alpha(u', v')) &= \mathrm{E}[(\alpha(u, v) - \lambda(u, v))(\alpha(u', v') - \lambda(u', v')) \mid C = 1]\,\mathrm{P}(C = 1)\\ &\quad + \mathrm{E}[(\alpha(u, v) - \lambda(u, v))(\alpha(u', v') - \lambda(u', v')) \mid C = 0]\,\mathrm{P}(C = 0).\end{aligned}$$

As $\alpha(u, v)| \leq \mathbf{1}(u \sim v) q_{v,d_i}/\lambda_i$ by (4.18) and as the edge indicators are independent, we can bound

$$\begin{aligned}&|\mathrm{E}[(\alpha(u, v) - \lambda(u, v))(\alpha(u', v') - \lambda(u', v')) \mid C = 0]|\,\mathrm{P}(C = 0)\\ &\quad\leq 4\frac{q_{v,d_i} q_{v',d_i}}{\lambda_i^2} p_{u,v} p_{u',v'} (p_{u,u'} + p_{u,v'} + p_{v,v'} + p_{u',v}).\end{aligned}$$

The contribution of this term to the overall variance can hence be bounded by

$$\frac{4}{\lambda_i^2} \sum_{u,u',v,v' \text{ distinct}} q_{v,d_i} q_{v',d_i} p_{u,v} p_{u',v'} (p_{u,u'} + p_{u,v'} + p_{v,v'} + p_{u',v}) \leq \frac{16}{\lambda_i^2} M.$$

Now $\alpha(u, v)$ is a random variable which depends only on $\{\mathbf{1}(u \sim v), \mathbf{1}(w \sim u), \mathbf{1}(w \sim v),\ w \neq u, v\}$, and, thus, it follows that, conditional on $C = 1$, $\alpha(u, v)$ and $\alpha(u', v')$ are independent. Moreover,

$$\begin{aligned}&\mathrm{E}[\alpha(u, v) - \lambda(u, v) \mid C = 1]\\ &\quad= \frac{q_{v,d_i}}{\lambda_i} \sum_{d > d_i} \sum_{\{\boldsymbol{x}_d :\, u \in \boldsymbol{x}_d\}} \sum_{\boldsymbol{x}_{d_i} \subset \boldsymbol{x}_d;\, u \notin \boldsymbol{x}_{d_i}} f^+(\boldsymbol{x}_{d_i} \mid \boldsymbol{x}_d)\\ &\qquad\times \{\mathrm{E}[\mathbf{1}(N(v) = \boldsymbol{x}_d)\{\mathbf{1}(D(u) = d_j + 1) - \mathbf{1}(D(u) = d_j)\} \mid C = 1]\\ &\qquad\quad - \mathrm{E}[\mathbf{1}(N(v) = \boldsymbol{x}_d)\{\mathbf{1}(D(u) = d_j + 1) - \mathbf{1}(D(u) = d_j)\}]\}.\end{aligned}$$

Regrouping the terms and conditioning on $u \sim v$ give

$$\begin{aligned}&\mathrm{E}[\mathbf{1}(N(v) = \boldsymbol{x}_d)\,\mathbf{1}(u \sim v)\,\mathbf{1}(D(u) = d_j + 1) \mid C = 1]\\ &- \mathrm{E}[\mathbf{1}(N(v) = \boldsymbol{x}_d)\,\mathbf{1}(u \sim v)\,\mathbf{1}(D(u) = d_j + 1)]\\ &\quad= p_{u,v}\{[\mathrm{P}(N^{(u,u',v')}(v) = \boldsymbol{x}_d \setminus \{u\}) - \mathrm{P}(N^{(u)}(v) = \boldsymbol{x}_d \setminus \{u\})]\,\mathrm{P}(D^{(u',v,v')}(u) = d_j)\\ &\qquad - \mathrm{P}(N^{(u)}(v) = \boldsymbol{x}_d \setminus \{u\})[\mathrm{P}(D^{(v)}(u) = d_j) - \mathrm{P}(D^{(u',v,v')}(u) = d_j)]\}.\end{aligned}$$

By conditioning on whether or not $u' \sim v$ and $v' \sim v$,

$$\begin{aligned}&|\mathrm{P}(N^{(u,u',v')}(v) = \boldsymbol{x}_d \setminus \{u\}) - \mathrm{P}(N^{(u)}(v) = \boldsymbol{x}_d \setminus \{u\})|\\ &\quad\leq (p_{u',v} + p_{v',v})\{\mathrm{P}(N^{(u,u',v')}(v) = \boldsymbol{x}_d \setminus \{u\}) + \mathrm{P}(N^{(u,u',v')}(v) = \boldsymbol{x}_d \setminus \{u, u'\})\\ &\qquad + \mathrm{P}(N^{(u,u',v')}(v) = \boldsymbol{x}_d \setminus \{u, v'\}) + \mathrm{P}(N^{(u,u',v')}(v) = \boldsymbol{x}_d \setminus \{u, u', v'\})\}.\end{aligned}$$

https://doi.org/10.1239/aap/1331216648 Published online by Cambridge University Press

Similarly, $|\mathrm{P}(D^{(v)}(u) = d_j) - \mathrm{P}(D^{(u',v,v')}(u) = d_j)| \le p_{u,u'} + p_{u,v'}$. Hence,

$$\begin{aligned} &|\mathrm{E}[\mathbf{1}(N(v) = \boldsymbol{x}_d)\,\mathbf{1}(u \sim v)\,\mathbf{1}(D(u) = d_j + 1) \mid C = 1] \\ &\quad - \mathrm{E}[\mathbf{1}(N(v) = \boldsymbol{x}_d)\,\mathbf{1}(u \sim v)\,\mathbf{1}(D(u) = d_j + 1)]| \\ &\qquad \le p_{u,v}\{(p_{u',v} + p_{v',v})\,\mathrm{P}(N^{(u,u',v')}(v) \in A) \\ &\qquad\qquad + (p_{u,u'} + p_{u,v'})\,\mathrm{P}(N^{(u)}(v) = \boldsymbol{x}_d \setminus \{u\})\}, \end{aligned}$$

where $A = \{\boldsymbol{x}_d \setminus \{u\}, \boldsymbol{x}_d \setminus \{u, u'\}, \boldsymbol{x}_d \setminus \{u, v'\}, \boldsymbol{x}_d \setminus \{u, u', v'\}\}$. Also, we have $|\mathrm{P}(D^{(v)}(u) = d_j - 1) - \mathrm{P}(D^{(u',v,v')}(u) = d_j - 1)| \le p_{u,u'} + p_{u,v'}$; replacing $d_j + 1$ by d_j in the above argument also yields

$$\begin{aligned} &|\mathrm{E}[\mathbf{1}(N(v) = \boldsymbol{x}_d)\,\mathbf{1}(u \sim v)\,\mathbf{1}(D(u) = d_j) \mid C = 1] \\ &\quad - \mathrm{E}[\mathbf{1}(N(v) = \boldsymbol{x}_d)\,\mathbf{1}(u \sim v)\,\mathbf{1}(D(u) = d_j)]| \\ &\qquad \le p_{u,v}\{(p_{u',v} + p_{v',v})\,\mathrm{P}(N^{(u,u',v')}(v) \in A) \\ &\qquad\qquad + (p_{u,u'} + p_{u,v'})\,\mathrm{P}(N^{(u)}(v) = \boldsymbol{x}_d \setminus \{u\})\}. \end{aligned}$$

Using these bounds,

$$\begin{aligned} &|\mathrm{E}[\alpha(u, v) - \lambda(u, v) \mid C = 1]| \\ &\quad \le \frac{q_{v,d_i}}{\lambda_i} \sum_{d > d_i} \sum_{\{\boldsymbol{x}_d \,:\, u \in \boldsymbol{x}_d\}} \sum_{\substack{\boldsymbol{x}_{d_i} \subset \boldsymbol{x}_d \\ u \notin \boldsymbol{x}_{d_i}}} f^{+}(\boldsymbol{x}_{d_i} \mid \boldsymbol{x}_d)\{p_{u,v}(p_{u',v} + p_{v',v})\,\mathrm{P}(N^{(u,u',v')}(v) \in A) \\ &\qquad\qquad + 2p_{u,v}(p_{u,u'} + p_{u,v'})\,\mathrm{P}(N^{(u)}(v) = \boldsymbol{x}_d \setminus \{u\})\} \\ &\quad \le 8p_{u,v}(p_{u,u'} + p_{u,v'} + p_{u',v} + p_{v',v})\frac{q_{v,d_i}}{\lambda_i}. \end{aligned}$$

Recalling that, conditional on $C = 1$, $\alpha(u, v)$ and $\alpha(u', v')$ are independent, we obtain

$$\begin{aligned} &|\mathrm{E}[(\alpha(u, v) - \lambda(u, v))(\alpha(u', v') - \lambda(u', v')) \mid C = 1]| \\ &\qquad \le 64p_{u,v}p_{u',v'}(p_{u,u'} + p_{u,v'} + p_{u',v} + p_{v',v})^2\frac{q_{v,d_i}q_{v',d_i}}{\lambda_i^2}. \end{aligned}$$

The contribution of this term to the overall variance can hence be bounded by

$$\frac{64}{\lambda_i^2} \sum_{u,u',v,v' \text{ distinct}} p_{u,v}p_{u',v'}(p_{u,u'} + p_{u,v'} + p_{u',v} + p_{v',v})^2 q_{v,d_i}q_{v',d_i} \le \frac{256}{\lambda_i^2}M.$$

Thus, for (4.19),

$$\mathrm{var}\biggl(\sum_{v} \sum_{u \ne v} \alpha(u, v)\biggr) \le \frac{281}{\lambda_i^2}M. \tag{4.23}$$

A similar argument holds for (4.20), involving $\beta(u, v)$; recall (4.17). Firstly,

$$\begin{aligned} \mathrm{var}(\beta(u, v)) \le \frac{q_{v,d_i}^2}{\lambda_i^2}\,\mathrm{E}\biggl(&\sum_{d < d_i} \sum_{\{\boldsymbol{x}_d \,:\, u \notin \boldsymbol{x}_d\}} \mathbf{1}(N(v) = \boldsymbol{x}_d)\{\mathbf{1}(D(u) = d_j - 1) - \mathbf{1}(D(u) = d_j)\} \\ &\times \mathbf{1}(u \notin \boldsymbol{x}_d) \sum_{\boldsymbol{x}_{d_i} \supset \boldsymbol{x}_d;\, u \in \boldsymbol{x}_{d_i}} f^{-}(\boldsymbol{x}_{d_i} \mid \boldsymbol{x}_d)\biggr)^2. \end{aligned}$$

We bound the probability that vertex u is picked to be added to the neighbours of v; if $N(v) = \boldsymbol{x}_d$,

$$\mathrm{P}(u \text{ picked} \mid N(v) = \boldsymbol{x}_d) = \mathbf{1}(u \notin \boldsymbol{x}_d) \sum_{\boldsymbol{x}_{d_i} \supset \boldsymbol{x}_d; u \in \boldsymbol{x}_{d_i}} f^{-}(\boldsymbol{x}_{d_i} \mid \boldsymbol{x}_d) \le p_{u,v} \frac{q^{(u)}_{v,d_i-1}}{q_{v,d_i}}. \tag{4.24}$$

Note that from (4.24) it also follows that

$$|\beta(u, v)| \le p_{u,v} \frac{q^{(u)}_{v,d_i-1}}{q_{v,d_i}}. \tag{4.25}$$

With (4.25) we have

$$\sum_{v} \sum_{u \neq v} \operatorname{var}(\beta(u, v)) \le \sum_{u,v} p^2_{u,v} \left(\frac{q^{(u)}_{v,d_i-1}}{\lambda_i}\right)^2 \le \frac{1}{\lambda_i^2} \sum_{u,v} p^2_{u,v} \le \frac{1}{\lambda_i^2} M.$$

Similarly as above, we obtain

$$\sum_{u,u',v \text{ distinct}} \operatorname{cov}(\beta(u, v), \beta(u', v)) \le 2 \sum_{u,u',v \text{ distinct}} p_{u,v} p_{u',v} \frac{q^{(u)}_{v,d_i-1} q^{(u')}_{v,d_i-1}}{\lambda_i^2} \le \frac{2}{\lambda_i^2} M$$

and

$$\sum_{u,v,v' \text{ distinct}} \operatorname{cov}(\beta(u, v), \beta(u, v')) \le 2 \sum_{u,v,v' \text{ distinct}} p_{u,v} p_{u,v'} \frac{q^{(u)}_{v,d_i-1} q^{(u)}_{v',d_i-1}}{\lambda_i^2} \le \frac{2}{\lambda_i^2} M$$

as well as

$$\sum_{u,u',v \text{ distinct}} \operatorname{cov}(\beta(u, v), \beta(u', u)) \le 2 \sum_{u,u',v \text{ distinct}} p_{u,v} p_{u',u} \frac{q^{(u)}_{v,d_i-1} q^{(u')}_{u,d_i-1}}{\lambda_i^2} \le \frac{2}{\lambda_i^2} M$$

and, similarly,

$$\sum_{u,v,v' \text{ distinct}} \operatorname{cov}(\beta(u, v), \beta(v, v')) \le \frac{2}{\lambda_i^2} M.$$

Now assume that u, v, u', and v' are all distinct. We refine the definition of β; for $t = 1, 2$, define $\beta^{(t)}$ by

$$\beta^{(1)}(u, v) = \frac{q_{v,d_i}}{\lambda_i} \sum_{d<d_i} \sum_{\{\boldsymbol{x}_d \,:\, u \notin \boldsymbol{x}_d\}} \mathbf{1}(N(v) = \boldsymbol{x}_d; D(u) = d_j - 1) \sum_{\boldsymbol{x}_{d_i} \supset \boldsymbol{x}_d; u \in \boldsymbol{x}_{d_i}} f^{-}(\boldsymbol{x}_{d_i} \mid \boldsymbol{x}_d),$$

$$\beta^{(2)}(u, v) = \frac{q_{v,d_i}}{\lambda_i} \sum_{d<d_i} \sum_{\{\boldsymbol{x}_d \,:\, u \notin \boldsymbol{x}_d\}} \mathbf{1}(N(v) = \boldsymbol{x}_d; D(u) = d_j) \sum_{\boldsymbol{x}_{d_i} \supset \boldsymbol{x}_d; u \in \boldsymbol{x}_{d_i}} f^{-}(\boldsymbol{x}_{d_i} \mid \boldsymbol{x}_d),$$

so that $\beta(u, v) = \beta^{(1)}(u, v) - \beta^{(2)}(u, v)$. Then

$$\begin{aligned}\operatorname{cov}(\beta(u, v), \beta(u', v')) &= \operatorname{cov}(\beta^{(1)}(u, v), \beta^{(1)}(u', v')) - \operatorname{cov}(\beta^{(1)}(u, v), \beta^{(2)}(u', v')) \\ &\quad - \operatorname{cov}(\beta^{(2)}(u, v), \beta^{(1)}(u', v')) + \operatorname{cov}(\beta^{(2)}(u, v), \beta^{(2)}(u', v')).\end{aligned}$$

https://doi.org/10.1239/aap/1331216648 Published online by Cambridge University Press

Now,

$$\begin{aligned}\mathrm{E}\,\beta^{(1)}(u,v) = \frac{q_{v,d_i}}{\lambda_i}\sum_{d<d_i}\sum_{\{\boldsymbol{x}_d\,:\,u\notin\boldsymbol{x}_d\}} &\{p_{u,v}\,\mathrm{P}(D^{(v)}(u)=d_j-2)\,\mathrm{P}(N^{(u)}(v)=\boldsymbol{x}_d\setminus\{u\})\\ &+(1-p_{u,v})\,\mathrm{P}(D^{(v)}(u)=d_j-1)\,\mathrm{P}(N^{(u)}(v)=\boldsymbol{x}_d)\}\\ &\times\mathrm{P}(u\text{ picked}\mid N(v)=\boldsymbol{x}_d)\end{aligned}$$

and

$$\begin{aligned}&\mathrm{E}\,\beta^{(1)}(u,v)\beta^{(1)}(u',v')\\ &\quad=\frac{q_{v,d_i}q_{v',d_i}}{\lambda_i^2}\sum_{d<d_i}\sum_{d'<d_i}\sum_{\{\boldsymbol{x}_d\,:\,u\notin\boldsymbol{x}_d\}}\sum_{\{\boldsymbol{y}_{d'}\,:\,u'\notin\boldsymbol{y}_{d'}\}}\mathrm{P}(u\text{ picked}\mid N(v)=\boldsymbol{x}_d)\\ &\qquad\times\mathrm{P}(u'\text{ picked}\mid N(v')=\boldsymbol{y}_{d'})\\ &\qquad\times\mathrm{P}(N(v)=\boldsymbol{x}_d,\ N(v')=\boldsymbol{y}_{d'},\ D(u)=d_j-1,\ D(u')=d_j-1).\end{aligned}$$

Now

$$\begin{aligned}&\operatorname{cov}(\beta^{(1)}(u,v),\beta^{(1)}(u',v'))\\ &\quad=\frac{q_{v,d_i}q_{v',d_i}}{\lambda_i^2}\sum_{d<d_i}\sum_{d'<d_i}\sum_{\{\boldsymbol{x}_d\,:\,u\notin\boldsymbol{x}_d\}}\sum_{\{\boldsymbol{y}_{d'}\,:\,u'\notin\boldsymbol{y}_{d'}\}}\mathrm{P}(u\text{ picked}\mid N(v)=\boldsymbol{x}_d)\\ &\qquad\times\mathrm{P}(u'\text{ picked}\mid N(v')=\boldsymbol{y}'_d)\\ &\qquad\times\{\mathrm{P}(N(v)=\boldsymbol{x}_d,\ D(u)=d_j-1,\ N(v')=\boldsymbol{y}_{d'},\ D(u')=d_j-1)\\ &\qquad\quad-\mathrm{P}(N(v)=\boldsymbol{x}_d,\ D(u)=d_j-1)\,\mathrm{P}(N(v')=\boldsymbol{y}_{d'},\ D(u')=d_j-1)\}.\end{aligned}$$

Conditioning on whether or not $C=1$ and using independence on $C=1$ give

$$\begin{aligned}&\mathrm{P}(N(v)=\boldsymbol{x}_d,\ D(u)=d_j-1,\ N(v')=\boldsymbol{y}_{d'},\ D(u')=d_j-1)\\ &-\mathrm{P}(N(v)=\boldsymbol{x}_d,\ D(u)=d_j-1)\,\mathrm{P}(N(v')=\boldsymbol{y}_{d'},\ D(u')=d_j-1)\\ &\quad=\mathrm{P}(C=1)\{\mathrm{P}(N^{u',v'}(v)=\boldsymbol{x}_d,\ D^{u',v'}(u)=d_j-1)\\ &\qquad\qquad\times\mathrm{P}(N^{u,v}(v')=\boldsymbol{y}_{d'},\ D^{u,v}(u')=d_j-1)\\ &\qquad\qquad-\mathrm{P}(N(v)=\boldsymbol{x}_d,\ D(u)=d_j-1)\,\mathrm{P}(N(v')=\boldsymbol{y}_{d'},\ D(u')=d_j-1)\}\\ &\qquad+\mathrm{P}(C=0)\{\mathrm{P}(N(v)=\boldsymbol{x}_d,\ D(u)=d_j-1,\ N(v')=\boldsymbol{y}_{d'},\ D(u')=d_j-1\mid C=0)\\ &\qquad\qquad-\mathrm{P}(N(v)=\boldsymbol{x}_d,\ D(u)=d_j-1)\,\mathrm{P}(N(v')=\boldsymbol{y}_{d'},\ D(u')=d_j-1)\}.\end{aligned}$$

Again, with conditioning on C,

$$\begin{aligned}&\mathrm{P}(N^{u',v'}(v)=\boldsymbol{x}_d,\ D^{u',v'}(u)=d_j-1)-\mathrm{P}(N(v)=\boldsymbol{x}_d,\ D(u)=d_j-1)\\ &\quad=\mathrm{P}(C=0)\{\mathrm{P}(N^{u',v'}(v)=\boldsymbol{x}_d,\ D^{u',v'}(u)=d_j-1)\\ &\qquad\qquad-\mathrm{P}(N(v)=\boldsymbol{x}_d,\ D(u)=d_j-1\mid C=0)\}.\end{aligned}$$

https://doi.org/10.1239/aap/1331216648 Published online by Cambridge University Press

Using a similar bound for (u, v) and (u', v') interchanged allows us to conclude that

$$
\begin{aligned}
&\operatorname{cov}(\beta^{(1)}(u, v), \beta^{(1)}(u', v')) \\
&\quad = \mathrm{P}(C = 0)\frac{q_{v,d_i} q_{v',d_i}}{\lambda_i^2} \sum_{d<d_i} \sum_{d'<d_i} \sum_{\{\boldsymbol{x}_d : u \notin \boldsymbol{x}_d\}} \sum_{\{\boldsymbol{y}_{d'} : u' \notin \boldsymbol{y}_{d'}\}} \mathrm{P}(u \text{ picked} \mid N(v) = \boldsymbol{x}_d) \\
&\qquad \times \mathrm{P}(u' \text{ picked} \mid N(v') = \boldsymbol{y}'_d) \\
&\qquad \times [\mathrm{P}(C = 1)\{\mathrm{P}(N^{u',v'}(v) = \boldsymbol{x}_d,\ D^{u',v'}(u) = d_j - 1) \\
&\qquad\qquad - \mathrm{P}(N(v) = \boldsymbol{x}_d,\ D(u) = d_j - 1 \mid C = 0)\} \\
&\qquad\quad \times \mathrm{P}(N^{u,v}(v') = \boldsymbol{y}_{d'},\ D^{u,v}(u') = d_j - 1) \\
&\qquad\quad + \mathrm{P}(C = 1)\,\mathrm{P}(N(v) = \boldsymbol{x}_d,\ D(u) = d_j - 1) \\
&\qquad\qquad \times \{\mathrm{P}(N^{u,v}(v') = \boldsymbol{y}_{d'},\ D^{u,v}(u') = d_j - 1) \\
&\qquad\qquad\quad - \mathrm{P}(N(v') = \boldsymbol{y}_{d'},\ D(u') = d_j - 1 \mid C = 0)\} \\
&\qquad\quad + \{\mathrm{P}(N(v) = \boldsymbol{x}_d,\ D(u) = d_j - 1,\ N(v') = \boldsymbol{y}_{d'},\ D(u') = d_j - 1 \mid C = 0) \\
&\qquad\qquad - \mathrm{P}(N(v) = \boldsymbol{x}_d,\ D(u) = d_j - 1)\,\mathrm{P}(N(v') = \boldsymbol{y}_{d'},\ D(u') = d_j - 1)\}].
\end{aligned}
$$

From (4.24) and bounding $\mathrm{P}(C = 1) \le 1$ as well as $q_{v,d} \le 1$, we obtain

$$
\begin{aligned}
&|\operatorname{cov}(\beta^{(1)}(u, v), \beta^{(1)}(u', v'))| \\
&\quad \le \mathrm{P}(C = 0)\frac{p_{u,v} p_{u',v'}}{\lambda_i^2} \sum_{d<d_i} \sum_{d'<d_i} \sum_{\{\boldsymbol{x}_d : u \notin \boldsymbol{x}_d\}} \sum_{\{\boldsymbol{y}_{d'} : u' \notin \boldsymbol{y}_{d'}\}} \mathrm{P}(u \text{ picked} \mid N(v) = \boldsymbol{x}_d) \\
&\qquad \times \mathrm{P}(u' \text{ picked} \mid N(v') = \boldsymbol{y}'_d) \\
&\qquad \times [\{\mathrm{P}(N^{u',v'}(v) = \boldsymbol{x}_d,\ D^{u',v'}(u) = d_j - 1) \\
&\qquad\quad + \mathrm{P}(N(v) = \boldsymbol{x}_d,\ D(u) = d_j - 1 \mid C = 0)\} \\
&\qquad\quad \times \mathrm{P}(N^{u,v}(v') = \boldsymbol{y}_{d'},\ D^{u,v}(u') = d_j - 1) \\
&\qquad\quad + \mathrm{P}(N(v) = \boldsymbol{x}_d,\ D(u) = d_j - 1) \\
&\qquad\qquad \times \{\mathrm{P}(N^{u,v}(v') = \boldsymbol{y}_{d'},\ D^{u,v}(u') = d_j - 1) \\
&\qquad\qquad\quad + \mathrm{P}(N(v') = \boldsymbol{y}_{d'},\ D(u') = d_j - 1 \mid C = 0)\} \\
&\qquad\quad + \{\mathrm{P}(N(v) = \boldsymbol{x}_d,\ D(u) = d_j - 1,\ N(v') = \boldsymbol{y}_{d'},\ D(u') = d_j - 1 \mid C = 0) \\
&\qquad\qquad + \mathrm{P}(N(v) = \boldsymbol{x}_d,\ D(u) = d_j - 1)\,\mathrm{P}(N(v') = \boldsymbol{y}_{d'},\ D(u') = d_j - 1)\}].
\end{aligned}
$$

As all the sums over $d < d_i$, $d' < d_i$, $\{\boldsymbol{x}_d : u \notin \boldsymbol{x}_d\}$, and $\{\boldsymbol{y}_{d'} : u' \notin \boldsymbol{y}_{d'}\}$ are bounded by 1,

$$|\operatorname{cov}(\beta^{(1)}(u, v), \beta^{(1)}(u', v'))| \le \frac{6}{\lambda_i^2}\,\mathrm{P}(C = 0)\,p_{u,v}\,p_{u',v'}.$$

We can bound the covariances $|\operatorname{cov}(\beta^{(1)}(u, v), \beta^{(2)}(u', v'))|$ and $|\operatorname{cov}(\beta^{(2)}(u, v), \beta^{(2)}(u', v'))|$ similarly. Thus, we obtain

$$\sum_{v} \sum_{u \ne v} \sum_{u' \ne u,v} \sum_{v' \ne u,u',v} \operatorname{cov}(\beta(u, v), \beta(u', v')) \le \frac{96}{\lambda_i^2} M. \tag{4.26}$$

For (4.21), use the bound

$$\operatorname{var}\bigg(\frac{1}{\lambda_i}\sum_v q_{v,d_i}\,\mathbf{1}(D(v)\neq d_i)\bigg) \\ \leq \frac{1}{\lambda_i^2}\sum_v q_{v,d_i}^2 + \frac{1}{\lambda_i^2}\sum_v\sum_{u\neq v} q_{v,d_i}q_{u,d_i}\operatorname{cov}(\mathbf{1}(D(v)\neq d_i),\mathbf{1}(D(u)\neq d_i)).$$

Using the facts that $D^{(u)}(v)$ and $D^{(v)}(u)$ are uncorrelated and that

$$|q_{v,d_i}^{(u)} - q_{v,d_i}| \leq \mathrm{P}(D^{(u)}(v)\neq D(v)) = p_{u,v}, \tag{4.27}$$

we obtain, for (4.21),

$$\begin{aligned}\operatorname{var}\bigg(\frac{1}{\lambda_i}\sum_v q_{v,d_i}\,\mathbf{1}(D(v)\neq d_i)\bigg) &\leq \frac{1}{\lambda_i^2}\sum_v q_{v,d_i}^2 + \frac{3}{\lambda_i^2}\sum_v\sum_{u\neq v} q_{v,d_i}q_{u,d_i}p_{u,v} \\ &\leq \frac{3}{\lambda_i^2}\sum_v\sum_u q_{v,d_i}q_{u,d_i}p_{u,v} \\ &\leq \frac{3}{\lambda_i^2}M, \end{aligned}\tag{4.28}$$

where we have used the convention that $p_{v,v}=0$. Similarly, for (4.22), when $i\neq j$,

$$\operatorname{var}\bigg(\frac{1}{\lambda_i}\sum_v q_{v,d_i}\,\mathbf{1}(D(v)=d_j)\bigg) \leq \frac{3}{\lambda_i^2}\sum_v\sum_u q_{v,d_i}q_{u,d_i}p_{u,v} \leq \frac{3}{\lambda_i^2}M. \tag{4.29}$$

Combining the bounds (4.23) with (4.26), (4.28), and (4.29), we obtain

$$\operatorname{var}\mathrm{E}[W_{d_j}^{d_i} - W_{d_j} \mid \boldsymbol{W}] \leq \frac{1568}{\lambda_i^2}M,$$

with M as given in the statement of Theorem 3.1. The sum over all i and j yields an additional factor p^2.

The next step is bounding $\|\Sigma_0^{-1/2}\|$. Define $b_v(i)=\sqrt{q_{v,d_i}}$, and let $\boldsymbol{b}_v=(b_v(1),\dots,b_v(p))^\top$ and $\boldsymbol{D}_v=\operatorname{diag}(b_v(i))$. Furthermore, let $\boldsymbol{a}$ denote the $p\times 1$ vector with entries $a_j=\sum_w\sqrt{\bar{p}_w}(q_{w,d_j-1}-q_{w,d_j})$. This gives $\Sigma_0=\sum_v \boldsymbol{D}_v(\boldsymbol{I}_p-\boldsymbol{b}_v\boldsymbol{b}_v^\top)D_v+\boldsymbol{a}\boldsymbol{a}^\top$, where $\boldsymbol{I}_p$ is the $p\times p$ identity matrix. For any matrix $\boldsymbol{A}$, let $\rho_1(\boldsymbol{A})\leq\cdots\leq\rho_p(\boldsymbol{A})$ denote the eigenvalues of $\boldsymbol{A}$ in increasing order. Letting $\boldsymbol{B}_v=\boldsymbol{I}_p-\boldsymbol{b}_v\boldsymbol{b}_v^\top$, by Weyl's theorem (see [10, Theorem 4.3.1]),

$$\rho_1(\Sigma_0)\geq\sum_v\rho_1(\boldsymbol{D}_v\boldsymbol{B}_v\boldsymbol{D}_v)+\rho_1(\boldsymbol{a}\boldsymbol{a}^\top)\geq\sum_v\rho_1(\boldsymbol{D}_v\boldsymbol{B}_v\boldsymbol{D}_v).$$

It can be shown that the eigenvalues of $\boldsymbol{B}_v$ are 1, with multiplicity $p-1$, and $\rho_1(\boldsymbol{B}_v)=1-\boldsymbol{b}_v^\top\boldsymbol{b}_v$, the least eigenvalue corresponding to the eigenvector $\boldsymbol{b}_v$. Now, using the Rayleigh–Ritz characterization of eigenvalues (see [10, Theorem 4.2.2]),

$$\begin{aligned}\rho_1(\boldsymbol{D}_v\boldsymbol{B}_v\boldsymbol{D}_v) &= \min_{\boldsymbol{x}} \frac{\boldsymbol{x}^\top\boldsymbol{D}_v\boldsymbol{B}_v\boldsymbol{D}_v\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{x}}\\ &= \min_{\boldsymbol{y}} \frac{\boldsymbol{y}^\top\boldsymbol{B}_v\boldsymbol{y}}{\boldsymbol{y}^\top\boldsymbol{D}_v^{-2}\boldsymbol{y}}\\ &\geq \frac{\rho_1(\boldsymbol{B}_v)}{\rho_p(\boldsymbol{D}_v^{-2})}\\ &= \min_i q_{v,d_i}\bigg(1-\sum_i q_{v,d_i}\bigg).\end{aligned}$$

It therefore follows that

$$\|\Sigma_0^{-1/2}\| \leq \rho_p(\Sigma_0^{-1/2}) = \frac{1}{\rho_1(\Sigma_0^{-1/2})} \leq \bigg[\sum_v \min_i q_{v,d_i}\bigg(1-\sum_i q_{v,d_i}\bigg)\bigg]^{-1/2} =: \tau.$$

Finally, we bound $\sum_{i,j}|\sigma_{ij}-\sigma_{ij}^0|$. With (4.27),

$$\begin{aligned}&\sum_{i,j}|\sigma_{ij}-\sigma_{ij}^0|\\ &\quad= \sum_{i,j}\bigg|\sum_{v,w}\{p_{wv}(1-p_{wv})(q_{v,d_i-1}^{(w)}-q_{v,d_i})(q_{w,d_j-1}^{(v)}-q_{w,d_j})\\ &\qquad\qquad - \sqrt{\bar p_v}\sqrt{\bar p_w}(q_{v,d_i-1}-q_{v,d_i})(q_{w,d_j-1}-q_{w,d_j})\}\bigg|\\ &\quad\leq 4p^2\sum_{v,w}p_{wv}^2 + \sum_{v;w\neq v}|p_{wv}-\sqrt{\bar p_v}\sqrt{\bar p_w}|\sum_{i,j}|(q_{v,d_i-1}-q_{v,d_i})(q_{w,d_j-1}-q_{w,d_j})|\\ &\qquad + \sum_v \bar p_v\sum_{i,j}|q_{v,d_i-1}-q_{v,d_i}||q_{v,d_j-1}-q_{v,d_j}|\\ &\quad=: S.\end{aligned}$$

Collecting the bounds gives the result.

Proof of Theorem 3.2. Recall that $q_{v,i}$ denotes $\mathrm{E}\,X_{(v,i)} = \mathrm{P}(D(v)=i)$. Based on Theorem 10.B of [2], we obtain

$$\begin{aligned}d_{\mathrm{TV}}(\mathcal{L}(\Xi_M),\mathrm{Po}(\boldsymbol{\lambda}_M)) = \sum_{v\in V}\sum_{i=M}^{n-1}q_{v,i}^2 + \bigg\{&\sum_{v\in V}\sum_{i=M}^{n-1}q_{v,i}\sum_{j=M,j\neq i}^{n-1}\mathrm{E}\,|X_{(v,j)}-X_{(v,j)}^{(v,i)}|\\ &+\sum_{v\in V}\sum_{i=M}^{n-1}q_{v,i}\sum_{u\in V_v}\sum_{j=M}^{n-1}\mathrm{E}\,|X_{(u,j)}-X_{(u,j)}^{(v,i)}|\bigg\}.\end{aligned}$$

Note that, the first summand in the bracket is for the case $u=v$ for $j\neq i$, and the second summand in the bracket is for $u\in V_v$ for all j, covering all (u,j) except the case $(u,j)=(v,i)$. Since $X_{(v,j)}^{(v,i)}=0$ for $j\neq i$ and $\mathrm{E}\,X_{(v,j)}=q_{v,j}$, the first summand in the bracket is equal to

$\sum_{v\in V}\sum_{i=M}^{n-1} q_{v,i}\sum_{j=M,\, j\neq i}^{n-1} q_{v,j}$. This, together with the summand outside the bracket, yields $\sum_{v\in V}[\sum_{i=M}^{n-1} q_{v,i}]^2$.

Let $\delta(u,v,i,j) = \mathrm{E}\,|X_{(u,j)} - X_{(u,j)}^{(v,i)}|$ for $u \in V_v$. Then Construction 2.1 gives

$$\begin{aligned}
&\delta(u,v,i,j)\\
&\quad= \sum_{\{d:\, d>i\}} q_{v,d} \sum_{\boldsymbol{x}_d\subset V_v} \mathrm{P}(N(v)=\boldsymbol{x}_d \mid D(v)=d)\,\mathrm{E}[|X_{(u,j)} - X_{(u,j)}^{(v,i)}| \mid N(v)=\boldsymbol{x}_d]\\
&\qquad+ \sum_{\{d:\, d<i\}} q_{v,d} \sum_{\boldsymbol{x}_d\subset V_v} \mathrm{P}(N(v)=\boldsymbol{x}_d \mid D(v)=d)\,\mathrm{E}[|X_{(u,j)} - X_{(u,j)}^{(v,i)}| \mid N(v)=\boldsymbol{x}_d].
\end{aligned}$$

Let $\delta_1(u,v,i,j)$ and $\delta_2(u,v,i,j)$ respectively denote the above two terms. We first calculate $\delta_1(u,v,i,j)$. Note that if $D(v)=d>i$ and $N(v)=\boldsymbol{x}_d$, then, by Construction 2.1, $|X_{(u,j)} - X_{(u,j)}^{(v,i)}| = 1$ if and only if vertex u satisfies (i) $u\in \boldsymbol{x}_d$, (ii) $D(u)=j+1$ or j in $\mathcal{G}(n,\{p_{ij}\})$, and (iii) $u \nsim v$ in $\mathcal{G}^{(v,i)}(n,\{p_{ij}\})$, namely, u is not adjacent to v in the graph. Hence, $\delta_1(u,v,i,j)$ can be written as

$$\begin{aligned}
\delta_1(u,v,i,j) = &\sum_{\{d:\, d>i\}} q_{v,d}\\
&\times \sum_{\{\boldsymbol{x}_d\subset V_v:\, u\in\boldsymbol{x}_d\}} \mathrm{P}(N(v)=\boldsymbol{x}_d \mid D(v)=d)\\
&\qquad\qquad \times \mathrm{E}[\mathbf{1}\{D(u)=j+1 \text{ or } j\}\,\mathbf{1}^{(v,i)}\{u\nsim v\} \mid N(v)=\boldsymbol{x}_d].
\end{aligned}$$

Conditioning on the event '$D(u)=j+1$ or j',

$$\begin{aligned}
&\mathrm{E}[\mathbf{1}\{D(u)=j+1 \text{ or } j\}\,\mathbf{1}^{(v,i)}\{u\nsim v\} \mid N(v)=\boldsymbol{x}_d]\\
&\quad= \mathrm{P}(D(u)=j+1 \text{ or } j \mid N(v)=\boldsymbol{x}_d)\\
&\qquad\times \mathrm{E}[\mathbf{1}^{(v,i)}\{u\nsim v\} \mid N(v)=\boldsymbol{x}_d,\ D(u)=j+1 \text{ or } j].
\end{aligned}$$

Since the neighbourhood of vertex v does not contain any relevant information for the degree of vertex u once we know whether or not $u\sim v$, we have

$$\begin{aligned}
\delta_1(u,v,i,j) = &\sum_{\{d:\, d>i\}} q_{v,d}\\
&\times \sum_{\{\boldsymbol{x}_d\subset V_v:\, u\in\boldsymbol{x}_d\}} \mathrm{P}(N(v)=\boldsymbol{x}_d \mid D(v)=d)\\
&\quad\times \mathrm{P}(D(u)=j+1 \text{ or } j \mid X_{\{u,v\}}=1)\,\mathrm{P}(X_{\{u,v\}}^{(v,i)}=0 \mid N(v)=\boldsymbol{x}_d).
\end{aligned} \tag{4.30}$$

If $d>i$ and $u\in\boldsymbol{x}_d$, then $\mathrm{P}(X_{\{u,v\}}^{(v,i)}=0 \mid N(v)=\boldsymbol{x}_d)$ is the probability that the edge $\{u,v\}$ is deleted in Construction 2.1 and, hence, by (2.1),

$$\begin{aligned}
\mathrm{P}(X_{\{u,v\}}^{(v,i)}=0 \mid N(v)=\boldsymbol{x}_d) &= \sum_{\{\boldsymbol{x}_i\subset V_v:\, \boldsymbol{x}_i\subset\boldsymbol{x}_d,\, u\notin\boldsymbol{x}_i\}} \mathrm{P}(N^{(v,i)}(v)=\boldsymbol{x}_i \mid N(v)=\boldsymbol{x}_d)\\
&= \sum_{\{\boldsymbol{x}_i\subset V_v:\, \boldsymbol{x}_i\subset\boldsymbol{x}_d, u\notin\boldsymbol{x}_i\}} f^+(\boldsymbol{x}_i \mid \boldsymbol{x}_d).
\end{aligned}$$

For (4.30), we then obtain

$$\delta_1(u, v, i, j) = \sum_{\{d\,:\,d>i\}} \sum_{\{\boldsymbol{x}_d \subset V_v\,:\,u \in \boldsymbol{x}_d\}} \sum_{\{\boldsymbol{x}_i \subset V_v\,:\,\boldsymbol{x}_i \subset \boldsymbol{x}_d,\,u \notin \boldsymbol{x}_i\}} f^{+}(\boldsymbol{x}_i \mid \boldsymbol{x}_d) \times \mathrm{P}(N(v) = \boldsymbol{x}_d)\,\mathrm{P}(D(u) = j+1 \text{ or } j \mid X_{\{u,v\}} = 1). \tag{4.31}$$

Next, we find $\delta_2(u, v, i, j)$ by a similar argument as for $\delta_1(u, v, i, j)$. If $D(v) = d < i$ and $N(v) = \boldsymbol{x}_d$, then, by Construction 2.1, $|X_{(u,j)} - X_{(u,j)}^{(v,i)}| = 1$ if and only if vertex u satisfies (i) $u \notin \boldsymbol{x}_d$, (ii) $D(u) = j - 1$ or j in $\mathcal{G}(n, \{p_{ij}\})$, and (iii) $u \sim v$ in $\mathcal{G}^{(v,i)}(n, \{p_{ij}\})$. Hence, following the same argument as for $\delta_1(u, v, i, j)$, we arrive at

$$\delta_2(u, v, i, j) = \sum_{\{d\,:\,d<i\}} \sum_{\{\boldsymbol{x}_d \subset V_v\,:\,u \notin \boldsymbol{x}_d\}} \sum_{\{\boldsymbol{x}_i \subset V_v\,:\,\boldsymbol{x}_i \supset \boldsymbol{x}_d,\,u \in \boldsymbol{x}_i\}} f^{-}(\boldsymbol{x}_i \mid \boldsymbol{x}_d) \times \mathrm{P}(N(v) = \boldsymbol{x}_d)\,\mathrm{P}(D(u) = j-1 \text{ or } j \mid X_{\{u,v\}} = 0). \tag{4.32}$$

Now, the sum of $\delta_1(u, v, i, j)$ at (4.31) and $\delta_2(u, v, i, j)$ at (4.32) gives $\delta(u, v, i, j)$, and, hence, $d_{\mathrm{TV}}(\mathcal{L}(\Xi_M), \mathrm{Po}(\boldsymbol{\lambda}_M)) \le b_{M,1} + b_{M,2} + b_{M,3}$, where

$$b_{M,1} = \sum_{v \in V} \left[\sum_{i=M}^{n-1} \mathrm{P}(D(v) = i) \right]^2 =: B_{M,1},$$

$$b_{M,2} = \sum_{v \in V} \sum_{i=M}^{n-1} \sum_{u \in V_v} \sum_{j=M}^{n-1} \sum_{\{d\,:\,d>i\}} \sum_{\{\boldsymbol{x}_d \subset V_v\,:\,u \in \boldsymbol{x}_d\}} \sum_{\{\boldsymbol{x}_i \subset V_v\,:\,\boldsymbol{x}_i \subset \boldsymbol{x}_d,\,u \notin \boldsymbol{x}_i\}} f^{+}(\boldsymbol{x}_i \mid \boldsymbol{x}_d) \times \mathrm{P}(D(v) = i)\,\mathrm{P}(N(v) = \boldsymbol{x}_d)\,\mathrm{P}(D(u) = j+1 \text{ or } j \mid X_{\{u,v\}} = 1),$$

and

$$b_{M,3} = \sum_{v \in V} \sum_{i=M}^{n-1} \sum_{u \in V_v} \sum_{j=M}^{n-1} \sum_{\{d\,:\,d<i\}} \sum_{\{\boldsymbol{x}_d \subset V_v\,:\,u \notin \boldsymbol{x}_d\}} \sum_{\{\boldsymbol{x}_i \subset V_v\,:\,\boldsymbol{x}_i \supset \boldsymbol{x}_d,\,u \in \boldsymbol{x}_i\}} f^{-}(\boldsymbol{x}_i \mid \boldsymbol{x}_d) \times \mathrm{P}(D(v) = i)\,\mathrm{P}(N(v) = \boldsymbol{x}_d)\,\mathrm{P}(D(u) = j-1 \text{ or } j \mid X_{\{u,v\}} = 0).$$

In particular, removing the constraints of vertex u on choosing $\boldsymbol{x}_d$ and $\boldsymbol{x}_i$, and using Lemma 2.1, we obtain

$$\begin{aligned} b_{M,2} &\le 2 \sum_{v \in V} \sum_{u \in V_v} \sum_{i=M}^{n-1} \mathrm{P}(D(v) = i)\,\mathrm{P}(D(v) > i)\,\mathrm{P}(D(u) \ge M \mid X_{\{u,v\}} = 1) \\ &\le 2 \sum_{v \in V} \sum_{u \in V_v} \mathrm{P}(D(v) \ge M)\ \mathrm{P}(D(u) \ge M \mid X_{\{u,v\}} = 1) \\ &=: \tfrac{1}{2} B_{M,2}, \end{aligned}$$

and, similarly,

$$b_{M,3} \le 2 \sum_{v \in V} \sum_{u \in V_v} \mathrm{P}(D(v) \ge M)\,\mathrm{P}(D(u) \ge M - 1 \mid X_{\{u,v\}} = 0).$$

Now, with both $\mathrm{P}(D(u) \ge M \mid X_{\{u,v\}} = 1)$ and $\mathrm{P}(D(u) \ge M - 1 \mid X_{\{u,v\}} = 0)$ equal to $\mathrm{P}(D^{(v)}(u) \ge M - 1)$, the assertion follows.

Acknowledgements

The authors would like to thank the anonymous referee for many helpful comments. GR was supported in part by the EPSRC EP/D003/05/1 Amorphous Computing Grant, and by the BBSRC and EPSRC through OCISB.

References

[1] Barabasi, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science* **286,** 509–512.
[2] Barbour, A. D., Holst, L. and Janson, S. (1992). *Poisson Approximation.* Oxford University Press.
[3] Bartroff, J. and Goldstein, L. (2009). A Berry-Esseen bound with applications to the number of multinomial cells of given occupancy and the number of graph vertices of given degree. Preprint.
[4] Bollobás, B. (2001). *Random Graphs*. Cambridge University Press.
[5] Bollobás, B., Janson, S. and Riordan, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* **31,** 3–122.
[6] Daudin, J-J., Picard, F. and Robin, S. (2008). A mixture model for random graphs. *Statist. Comput.* **18,** 173–183.
[7] Dorogovtsev, S. N. and Mendes, J. F. F. (2003). *Evolution of Networks*. Oxford University Press.
[8] Erdös, P. and Rényi, A. (1959). On random graphs. *Publ. Math. Debrecen* **6,** 290–297.
[9] Goldstein, L. and Rinott, Y. (1996). Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Prob.* **33,** 1–17.
[10] Horn, R. A. and Johnson, C. R. (1985). *Matrix Analysis*. Cambridge University Press.
[11] Lin, K. (2008). Motif counts, clustering coefficients, and vertex degrees in models of random networks. Doctoral Thesis, University of Oxford.
[12] McKay, B. D. and Wormald, N. C. (1997). The degree sequence of a random graph. I. The models. *Random Structures Algorithms* **11,** 97–117.
[13] Newman, M. E. J., Moore, C. and Watts, D. J. (2000). Mean-field solution of the small-world network model. *Phys. Rev. Lett.* **84,** 3201–3204.
[14] Nowicki, K. and Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *J. Amer. Statist. Assoc.* **96,** 1077–1087.
[15] Rinott, Y. and Rotar, V. (1996). A multivariate CLT for local dependence with $n^{-1/2}\log n$ rate and applications to multivariate graph related statistics. *J. Multivariate Anal.* **56,** 333–350.
[16] Solow, A. R., Costello, C. J. and Ward, M. (2003). Testing the power law model for discrete size data. *Amer. Nat.* **162,** 685–689.
[17] Stein, C. (1986). *Approximate Computation of Expectations* (Inst. Math. Statist. Lecture Notes—Monogr. Ser. **7**). Institute of Mathematical Statistics, Hayward, CA.
[18] Stumpf, M. P. H., Wiuf, C. and May, R. M. (2005). Subnets of scale-free networks are not scale-free: sampling properties of networks. *Proc. Nat. Acad. Sci. USA* **12,** 4221–4224.