Algorithmica
DOI 10.1007/s00453-009-9340-1

A Sequential Algorithm for Generating Random Graphs

Mohsen Bayati · Jeong Han Kim · Amin Saberi

Received: 31 October 2007 / Accepted: 23 June 2009

Abstract We present a nearly-linear time algorithm for counting and randomly generating simple graphs with a given degree sequence in a certain range. For degree sequence $(d_i)_{i=1}^n$ with maximum degree $d_{\max} = O(m^{1/4-\tau})$, our algorithm generates almost uniform random graphs with that degree sequence in time $O(md_{\max})$ where $m = \frac{1}{2}\sum_i d_i$ is the number of edges in the graph and τ is any positive constant. The fastest known algorithm for uniform generation of these graphs (McKay and Wormald in J. Algorithms 11(1):52–67, 1990) has a running time of $O(m^2 d_{\max}^2)$. Our method also gives an independent proof of McKay's estimate (McKay in Ars Combinatoria A 19:15–25, 1985) for the number of such graphs.

We also use sequential importance sampling to derive fully Polynomial-time Randomized Approximation Schemes (FPRAS) for counting and uniformly generating random graphs for the same range of $d_{\max} = O(m^{1/4-\tau})$.

Moreover, we show that for $d = O(n^{1/2-\tau})$, our algorithm can generate an asymptotically uniform d-regular graph. Our results improve the previous bound of $d = O(n^{1/3-\tau})$ due to Kim and Vu (Adv. Math. 188:444–469, 2004) for regular graphs.

Keywords Random graphs · Sequential importance sampling · FPRAS

M. Bayati
Department of Electrical Engineering, Stanford University, Stanford, CA 94305-9510, USA
e-mail: mohsen.bayati@hotmail.com

J.H. Kim (✉)
Department of Mathematics, Yonsei University, Yonsei, South Korea
e-mail: jehkim@yonsei.ac.kr

A. Saberi
Departments of Management Science and Engineering, Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA 94305, USA
e-mail: saberi@stanford.edu

Published online: 21 July 2009

1 Introduction

The focus of this paper is on generating random simple graphs (graphs with no multiple edges or self loop) with a given degree sequence. Random graph generation has been studied extensively as an interesting theoretical problem (see [10, 40] for detailed surveys). It has also become an important tool in a variety of real world applications including detecting motifs in biological networks [33] and simulating networking protocols on the Internet topology [1, 12, 17, 32, 38]. The best algorithm for this problem was given by McKay and Wormald [30] that uses certain switches on the configuration model and produces random graphs with uniform distribution in $O(m^2 d_{\max}^2)$ time. However, this running time can be slow for the networks with millions of edges. This has constrained practitioners to use simple heuristics that are non-rigorous and have often led to wrong conclusions [33, 34]. Our main contribution in this paper is to provide a nearly-linear time, fully polynomial randomized approximation scheme (FPRAS) for generating random graphs. An FPRAS provides an arbitrary close approximation in time that depends only polynomially on the input size and the desired error. (For precise definitions of FPRAS, see Definition 1 in Sect. 2.)

Recently, sequential importance sampling (SIS) has been suggested as a more suitable approach for designing fast graph generation algorithms [4, 10, 13, 28]. Chen et al. [13] used the SIS method to generate bipartite graphs with a given degree sequence. Later Blitzstein and Diaconis [10] used a similar approach for generating general graphs with given degrees. But these results are mostly empirical, and in a few cases SIS is shown to be slow [8]. However, the simplicity of these algorithms and their great performance in several instances suggest that a further study of the SIS method is necessary.

The Result Let $d_1, \ldots, d_n$ be non-negative integers with $\sum_{i=1}^n d_i = 2m$. Our algorithm for generating a graph with degree sequence $d_1, \ldots, d_n$ is a generalization of Steger and Wormald's algorithm for regular graphs [37]. It works as follows: start with an empty graph and sequentially add edges between the pairs of non-adjacent vertices. In every step of the procedure, the probability that an edge is added between two distinct vertices i and j is proportional to $\hat{d}_i \hat{d}_j (1 - d_i d_j / 4m)$ where $\hat{d}_i$ and $\hat{d}_j$ denote the remaining degrees of vertices i and j. The remaining degree of a vertex i is equal to d_i minus its current degree. We will show that this algorithm produces an asymptotically uniform sample with running time $O(m\, d_{\max})$ when $d_{\max} = O(m^{1/4-\tau})$ and τ is any positive constant. Then, we use SIS to obtain an FPRAS for any $\epsilon, \delta > 0$ with running time $O(m\, d_{\max} \epsilon^{-2} \log(1/\delta))$. The same result holds when the algorithm is used for generating bipartite graphs. Moreover, we show that for $d = O(n^{1/2-\tau})$, this algorithm can generate d-regular graphs with an asymptotically uniform distribution. Our results improve the bounds of Kim and Vu [27] and Steger and Wormald [37] for the regular graphs.

Related Work McKay and Wormald [29, 31] give asymptotic estimates for the number of graphs with $d_{\max} = O(m^{1/3-\tau})$. However, the error terms in their estimates are larger than what is needed to apply Jerrum, Valiant and Vazirani's [20, 21] reduction to achieve an asymptotically uniform sampling. Jerrum and Sinclair [19], however, use a random walk on the self-reducibility tree and give an FPRAS for uniformly sampling the graphs with $d_{\max} = o(m^{1/4})$. The running time of their algorithm is $O(m^3 n^2 \epsilon^{-2} \log(1/\delta))$ [36]. A different random walk that has been studied by [7, 20, 22], gives an FPRAS for the random generation of bipartite graphs with all degree sequences and general graphs with almost all degree sequences. However, the running time of all these algorithms is at least $O(n^4 m^3 d_{\max} \epsilon^{-2} \log(1/\delta))$. Other Markov chain methods are also studied in [15, 16, 18, 23].

McKay and Wormald also introduced an algorithm based on a certain switching technique on the configuration model that achieves the best performance [30]. It produces random graphs with uniform distribution (better than FPRAS) and has a faster running time. Their algorithm works for graphs with $d_{\max}^3 = O(m^2/\sum_i d_i^2)$ and $d_{\max}^3 = o(m + \sum_i d_i^2)$ with an average running time of $O(m + (\sum_i d_i^2)^2)$. This leads to an $O(n^2 d^4)$ average running time for d-regular graphs with $d = O(n^{1/3})$.

Very recently and independently from our work, Blanchet [9] has used McKay's estimate [29] and SIS technique to obtain an FPRAS with running time of $O(m^2 \epsilon^{-2} \times \log(1/\delta))$ for counting bipartite graphs with given degrees when $d_{\max} = o(m^{1/4})$. His work is based on defining an appropriate Lyapunov function as well as using Mckay's estimate.

Our Technical Contribution Our algorithm and its analysis are based on the beautiful works of Steger and Wormald [37] and Kim and Vu [26]. The technical contributions of our work beyond their analysis are as follows:

1. In both [26, 37] the output distribution of the proposed algorithms are asymptotically uniform. Here we use SIS technique to obtain an FPRAS.
2. Both [26, 37] use McKay's estimate [29] in their analysis. In this paper we give a combinatorial argument to control the failure probability of the algorithm and obtain a new proof for McKay's estimate.
3. We exploit the combinatorial structure and use a martingale tail inequality to show the concentration results for d-regular graphs with $d = O(n^{1/2-\tau})$ where the previous polynomial inequalities [25] do not work.

Other Applications and Extensions Our algorithm and its analysis provide more insight into the modern random graph models, such as the configuration model or the random graphs with a given *expected* degree sequence [14]. In these models, the probability of having an edge between vertices i and j of the graph is proportional to $d_i d_j$. However, one can use our analysis or McKay's formula [29] to see that in a random simple graph, this probability is proportional to $d_i d_j (1 - d_i d_j / 2m)$. We expect that by adding the correction term and using the concentration result of this paper, it is possible to obtain sandwiching theorems similar to [27].

In a follow up work, Bayati et al. [5] uses similar ideas to generate random graphs with large girth. These graphs are useful for designing high performance Low-Density Parity-Check (LDPC) codes (see [3]).

Organization of the Paper The rest of the paper has the following structure. The algorithm and the main results are stated in Sect. 2. In Sect. 3, we explain the intuition behind the weighted configuration model and our algorithm while also describing the SIS approach. Finally Sects. 4–7 are dedicated to the analysis and the proofs.

2 Our Algorithm

Suppose that n nonnegative integers $d_1, d_2, \dots, d_n$ with $\sum_{i=1}^n d_i = 2m$ are given. Assume that this sequence is also *graphical*. That is, there exists at least one simple graph with these degrees. We propose the following procedure for sampling (counting) an element (the number of elements) of the set $\mathcal{L}(\bar{d})$ of all labeled simple graphs G with vertices $V = \{v_1, v_2, \dots, v_n\}$ and degree sequence $\bar{d} = (d_1, d_2, \dots, d_n)$. Throughout this paper $m = \sum_{i=1}^n d_i/2$ is the number of edges in the graph, $d_{\max} = \max_{i=1}^n \{d_i\}$ and for the regular graphs, d refers to the degrees; i.e. $d_i = d$ for all $i = 1, \dots, n$. We denote the set of all d-regular graphs with n vertices by $\mathcal{L}(n, d)$.

Procedure A

Input: A graphical degree sequence $\bar{d} = (d_1, d_2, \dots, d_n)$.
Output: A graph G with degree sequence $\bar{d}$ or failure. An approximation N for the number of graphs with degree sequence $\bar{d}$ or 0.

1. Let E be a set of edges, $\hat{d} = (\hat{d}_1, \dots, \hat{d}_n)$ be an n-tuple of integers and P be a real number. Initialize them by $E =$ Empty set, $\hat{d} = \bar{d}$, and $P = 1$.
2. Choose two vertices $v_i, v_j \in V$ with probability proportional to $\hat{d}_i \hat{d}_j (1 - \frac{d_i d_j}{4m})$ among all pairs v_i, v_j with $i \neq j$ and $\{v_i, v_j\} \notin E$. Denote this probability by p_{ij}. Multiply P by p_{ij}, add $\{v_i, v_j\}$ to E and reduce each of $\hat{d}_i$, $\hat{d}_j$ by 1.
3. Repeat step (2) until no more edges can be added to E.
4. If $|E| < m$ report *failure* and output $N = 0$, otherwise output $G = (V, E)$ and $N = (m!\, P)^{-1}$.

Note that for the regular graphs the factors $1 - d_i d_j / 4m$ are redundant and Procedure A is the same as Steger and Wormald's [37] algorithm. The next two theorems characterize the output distribution of Procedure A.

Theorem 1 *For an arbitrary number $\tau > 0$ and for any degree sequence $\bar{d}$ with maximum degree of $O(m^{1/4-\tau})$, Procedure A can be implemented so that it terminates successfully with probability $(1 - o(1))$ in expected running time $O(m d_{\max})$. Furthermore, any graph G with degree sequence $\bar{d}$ is generated with a probability within $1 \pm o(1)$ factor of the uniform probability.*

For the regular graphs a similar result can be shown in a larger range for the degrees.

Theorem 2 *For an arbitrary number $\tau > 0$ and for $d = O(n^{1/2-\tau})$, Procedure A generates all graphs G in $\mathcal{L}(n, d)$ with probability within $1 \pm o(1)$ factor of the uniform probability, except for the graphs in a subset of size $o(|\mathcal{L}(n, d)|)$. In other words as $n \to \infty$, the output distribution of Procedure A converges to the uniform distribution in total variation distance.*

The results above show that the output distribution of Procedure A is close to uniform only when n is sufficiently large. Nevertheless, it is desirable to be close to uniform for all values of n. In order to do that, we find an FPRAS for calculating $|\mathcal{L}(\bar{d})|$ and also for randomly generating the elements of $\mathcal{L}(\bar{d})$.

Definition 1 An FPRAS for approximately counting graphs with degree sequence $\bar{d}$ is an algorithm that for any $\epsilon, \delta > 0$, outputs an estimate X for $|\mathcal{L}(\bar{d})|$ where $\mathbb{P}\{(1-\epsilon)|\mathcal{L}(\bar{d})| \le X \le (1+\epsilon)|\mathcal{L}(\bar{d})|\} \ge 1-\delta$, and has a running time polynomial in $m, 1/\epsilon, \log(1/\delta)$.

Similarly, an FPRAS for randomly generating graphs with degree sequence $\bar{d}$ is an algorithm that for any $\epsilon > 0$, has a running time polynomial in $m, 1/\epsilon$, and with probability at least 0.5, it outputs a graph from the set $\mathcal{L}(\bar{d})$ with probability within $\frac{1\pm\epsilon}{c}$ of the uniform where c is a constant.

Throughout this paper we assume $0 < \epsilon, \delta < 1$ and for convenience, we define a real valued random variable X to be an (ϵ, δ)-estimate for a number y if $\mathbb{P}\{(1-\epsilon)y \le X \le (1+\epsilon)y\} \ge 1-\delta$.

The following theorem summarizes our main result.

Theorem 3 *For an arbitrary number $\tau > 0$, degree sequence $\bar{d}$ with maximum degree of $O(m^{1/4-\tau})$, and any $\epsilon, \delta > 0$, the Algorithm CountGraphs of Sect. 3 is an FPRAS with an expected running time of $O(md_{\max}\epsilon^{-2}\log(1/\delta))$ for counting graphs with degree sequence $\bar{d}$. Moreover, the Algorithm GenerateGraph of Sect. 3 is an FPRAS with an expected running time of $O(md_{\max}\epsilon^{-2})$ for randomly generating graphs with degree sequence $\bar{d}$.*

Remark 1 For generating bipartite graphs, step (2) of Procedure A should be modified to:

2. Choose two vertices $v_i, v_j \in V$ with probability proportional to $\hat{d}_i\hat{d}_j(1-\frac{d_id_j}{2m})$ among all pairs v_i, v_j with $\{v_i, v_j\} \notin E$, and v_i, v_j not belonging to the same part of the graph. Denote this probability by p_{ij} and multiply P by p_{ij}. Add $\{v_i, v_j\}$ to E and reduce each of $\hat{d}_i, \hat{d}_j$ by 1.

Then corresponding versions of Theorems 1–3 can be shown.

3 Definitions and the Main Idea

Before explaining our approach let us quickly review the configuration model (see [6, 11, 35] for more details). Let $W = \bigcup_{i=1}^{n} W_i$ be a set of $2m = \sum_{i=1}^{n} d_i$ labeled mini-vertices with $|W_i| = d_i$. Consider a procedure that finds a random perfect matching $\mathcal{M}$ between mini-vertices by choosing pairs of mini-vertices sequentially and uniformly at random. Such a matching is also called a *configuration* on W. We can see that the number of all distinct configurations is equal to $(1/m!)\prod_{r=0}^{m-1}\binom{2m-2r}{2}$. Given a configuration $\mathcal{M}$, we can obtain a graph $G_{\mathcal{M}}$ with degree sequence $\bar{d}$ by combining the mini-vertices of each W_i to form a vertex v_i.

Note that the graph $G_{\mathcal{M}}$ might have self edge loops or multiple edges. In fact McKay and Wormald's estimate [31] shows that this happens with very high probability except when $d_{\max} = O(\log^{1/2} m)$. In order to fix this problem, Steger and Wormald [37] proposed that at any step one can only look at those pairs of mini-vertices that lead to simple graphs (denote these by *suitable pairs*) and pick one uniformly at random. For d-regular graphs when $d = O(n^{1/28-\tau})$ Steger and Wormald have shown that this approach asymptotically samples regular graphs with uniform distribution and Kim and Vu [26] have extended that to $d = O(n^{1/3-\tau})$.

3.1 Weighted Configuration Model

Unfortunately, when the degree sequence is not uniform, the above procedure may generate some graphs with a probability exponentially larger (or smaller) than uniform probability. In this paper, we will show that for non-regular degree sequences suitable pairs should be picked non-uniformly. In fact, Procedure A is a weighted configuration model where at any step a suitable pair $\{u, v\}$ with $u \in W_i$ and $v \in W_j$ is picked with probability proportional to $1 - d_i d_j / 4m$.

Here is a rough intuition behind Procedure A. Define the *execution tree* T of the configuration model as follows: Consider a rooted tree where its root (the vertex at level zero) corresponds to the empty matching in the beginning of the model and level r vertices correspond to all partial matchings that can be constructed after r steps. There is an edge in T between a partial matching $\mathcal{M}_r$ from level r to a partial matching $\mathcal{M}_{r+1}$ from level $r+1$ if $\mathcal{M}_r \subset \mathcal{M}_{r+1}$. Any path from the root to a leaf of T corresponds to one possible way of generating a random configuration.

Let us denote those partial matchings $\mathcal{M}_r$ whose corresponding partial graph $G_{\mathcal{M}_r}$ is simple by "valid" matchings and denote the remaining partial matchings by "invalid". Our goal is to sample valid leaves of the tree T uniformly at random. Steger and Wormald's improvement to the configuration model is to restrict the algorithm at step r to the valid children of $\mathcal{M}_r$ and picking one uniformly at random. This approach leads to an almost uniform generation for the regular graphs [26, 37] since the number of valid children for all partial matchings at level r of T, is almost equal. However, it is crucial to note that for non-regular degree sequences if the $(r+1)$th-edge matches two elements belonging to the vertices with larger degrees, the number of valid children for $\mathcal{M}_{r+1}$ will be smaller. Thus, there will be a bias towards graphs that have more of such edges.

In order to find a rough estimate of the bias, fix a graph G with degree sequence $\bar{d}$. Let $\mathrm{M}(G)$ be the set of all leaves $\mathcal{M}$ of the tree T that lead to graph G; i.e. those configurations $\mathcal{M}$ with $G_{\mathcal{M}} = G$. It is easy to see that $|\mathrm{M}(G)| = m! \prod_{i=1}^{n} d_i!$. Moreover, for exactly $(1-q_r)\,|\mathrm{M}(G)|$ of these leaves, a fixed edge $\{i, j\}$ of G appears in the first r steps of the path leading to them; i.e. $\{i, j\} \in \mathcal{M}_r$. Here $q_r = (m-r)/m$. Furthermore, we can show that for a typical matching after step r, the number of unmatched mini-vertices in each W_i is roughly $d_i q_r$. Thus the expected number of unsuitable pairs $\{u, v\}$ is about $\sum_{i \sim_G j} d_i d_j q_r^2 (1-q_r)$. Similarly, the expected number of unsuitable pairs corresponding to self edge loops is approximately $\sum_{i=1}^{n} \binom{d_i q_r}{2} \approx 2mq_r^2 \lambda(\bar{d})$ where $\lambda(\bar{d}) = \sum_{i=1}^{n} \binom{d_i}{2} / (\sum_{i=1}^{n} d_i)$. Therefore, defining $\gamma_G = \sum_{i \sim_G j} d_i d_j / 4m$ and using $\binom{2m-2r}{2} \approx 2m^2 q_r^2$ we can approximate $\mathbb{P}_{\mathrm{A}}(G)$, the probability of generating G with Procedure A by

$$\mathbb{P}_{\mathrm{A}}(G) \approx m! \left(\prod_{i=1}^{n} d_i!\right) \prod_{r=0}^{m-1} \frac{1}{2m^2 q_r^2 - 2m q_r^2 \lambda(\bar{d}) - 4m(1-q_r) q_r^2 \gamma_G}$$
$$\approx e^{\lambda(\bar{d})+\gamma_G} m! \left(\prod_{i=1}^{n} d_i!\right) \prod_{r=0}^{m-1} \frac{1}{\binom{2m-2r}{2}} \propto e^{\gamma_G}.$$

Hence, adding the edge $\{i, j\}$ roughly creates an $\exp(d_i d_j/4m)$ bias. To cancel that effect we need to reduce the probability of picking $\{i, j\}$ by $\exp(-d_i d_j/4m) \approx 1 - d_i d_j/4m$. We will rigorously prove the above argument in Sect. 4.

3.2 Obtaining a Fully Polynomial Randomized Approximation Scheme

The output distribution of Procedure A denoted by $\mathbb{P}_{\mathrm{A}}$ is asymptotically uniform. But when m is small, it is desirable to reduce the deviation of the output distribution from the uniform distribution. Note that it is not possible to use an accept/reject scheme to obtain uniform distribution since the probability $\mathbb{P}_{\mathrm{A}}(G)$ is not known for any given graph G. In fact, for an output G of Procedure A, the variable P is the probability of generating *one ordering* of the edges of G among all $m!$ possible permutations. Different orderings can have probabilities that vary exponentially which further complicates the calculation of $\mathbb{P}_{\mathrm{A}}(G)$.

However, we can use the Sequential Importance Sampling (SIS) method, similar to [13], to find very close estimates for $\mathbb{P}_{\mathrm{A}}(G)$ and $|\mathcal{L}(\bar{d})|$. Then with a simple accept/reject scheme we can obtain a distribution that is very close to the uniform distribution. For example if $\mathbb{P}_{\mathrm{A}}(G)|\mathcal{L}(\bar{d})| \geq 1$ then we can accept graph G with probability $(\mathbb{P}_{\mathrm{A}}(G)|\mathcal{L}(\bar{d})|)^{-1}$. This approach will be explained in more detail in this section.

3.2.1 FPRAS for Counting via SIS

Denote the set of all orderings $\mathcal{N}$ that lead to a graph in $\mathcal{L}(\bar{d})$ by $\mathcal{K}(\bar{d})$. Therefore, $|\mathcal{K}(\bar{d})| = m!|\mathcal{L}(\bar{d})|$. Let $\mathbb{Q}$ be the uniform distribution on $|\mathcal{K}(\bar{d})|$. Procedure A samples an ordering $\mathcal{N} \in \mathcal{K}(\bar{d})$ from a "trial distribution" $\mathbb{P}_{\mathrm{A}}$, where $\mathbb{P}_{\mathrm{A}}(\mathcal{N}) > 0$ for all $\mathcal{N} \in \mathcal{K}(\bar{d})$. Thus, we have

$$\mathbb{E}_{\mathbb{P}_{\mathrm{A}}}\left(\frac{1}{\mathbb{P}_{\mathrm{A}}}\right) = \sum_{\mathcal{N} \in \mathcal{K}(\bar{d})} \frac{1}{\mathbb{P}_{\mathrm{A}}(\mathcal{N})} \mathbb{P}_{\mathrm{A}}(\mathcal{N}) = |\mathcal{K}(\bar{d})|.$$

Hence, we can estimate $|\mathcal{K}(\bar{d})|$ by

$$\widehat{|\mathcal{K}(\bar{d})|} = \frac{1}{k} \sum_{i=1}^{k} \frac{1}{\mathbb{P}_{\mathrm{A}}(\mathcal{N}_i)}$$

from k i.i.d. samples $\mathcal{N}_1, \ldots, \mathcal{N}_k$ drawn from $\mathbb{P}_{\mathrm{A}}(\mathcal{N})$. Now in order to estimate $|\mathcal{L}(\bar{d})| = |\mathcal{K}(\bar{d})|/m!$ we can use

$$\widehat{|\mathcal{L}(\bar{d})|} = \frac{1}{k}\sum_{i=1}^{k} \frac{1}{m!\mathbb{P}_{\mathrm{A}}(\mathcal{N}_i)}.$$

Note that when an ordering $\mathcal{N}$ is the output of Procedure A then the number N, that is also an output of Procedure A, is equal to $\frac{1}{m!\mathbb{P}_{\mathrm{A}}(\mathcal{N})}$. Hence, we propose the following algorithm for estimating $|\mathcal{L}(\bar{d})|$.

Algorithm: CountGraphs

Input: A graphical degree sequence $\bar{d}$, positive numbers ϵ, δ, and an integer $k = k(\epsilon, \delta)$.
Output: An (ϵ, δ)-estimate X for the number of graphs with degree sequence $\bar{d}$.

1 Run Procedure A $k = k(\epsilon, \delta)$ times, and denote the corresponding values for the random variable N by $N_1, \ldots, N_k$.
2 Output $X = \frac{N_1 + \cdots + N_k}{k}$ as an estimate for $|\mathcal{L}(\bar{d})|$.

We will show in Sect. 8.1, that the variance of the random variable N is small enough and therefore, an integer $k = k(\epsilon, \delta) = O(\epsilon^{-2} \log(1/\delta))$ exists such that the Algorithm CountGraphs produces an (ϵ, δ)-estimate for $|\mathcal{L}(\bar{d})|$.

3.2.2 Approximating $\mathbb{P}_{\mathrm{A}}(G)$ with SIS

Similar to the above discussion, we will use SIS to find a very close approximation for $\mathbb{P}_{\mathrm{A}}(G)$ for each graph G. Recall that for any graph G, each ordering $\mathcal{N}$ of the edges of G is generated with probability $\mathbb{P}_{\mathrm{A}}(\mathcal{N})$ by Procedure A. Now let $\mathrm{S}(G)$ be the set of all $m!$ orderings of G. Therefore, the probability $\mathbb{P}_{\mathrm{A}}(G)$ is given by

$$\mathbb{P}_{\mathrm{A}}(G) = \sum_{\mathcal{N} \in \mathrm{S}(G)} \mathbb{P}_{\mathrm{A}}(\mathcal{N}). \tag{1}$$

Let $\mathbb{H}$ be the uniform distribution on the set $\mathrm{S}(G)$. Then (1) is equivalent to $\mathbb{P}_{\mathrm{A}}(G) = m!\, \mathbb{E}_{\mathbb{H}}(\mathbb{P}_{\mathrm{A}}(\mathcal{N}))$.

Therefore, we use $\mathbb{H}$ as trial distribution and draw ℓ i.i.d. samples $\mathcal{N}_1, \ldots, \mathcal{N}_\ell$ from $\mathbb{H}$. Then for each sample $\mathcal{N}_i$ we calculate $\mathbb{P}_{\mathrm{A}}(\mathcal{N}_i)$ and report

$$\widehat{\mathbb{P}_{\mathrm{A}}(G)} = \frac{m!}{\ell}\sum_{i=1}^{\ell} \mathbb{P}_{\mathrm{A}}(\mathcal{N}_i)$$

as an estimate for $\mathbb{P}_{\mathrm{A}}(G)$. This is given by Procedure B.

Note that the variable P at the end of step 3 is exactly $\mathbb{P}_{\mathrm{A}}(\mathcal{N})$ for an element $\mathcal{N} \in \mathrm{S}(G)$ that is sampled from distribution $\mathbb{H}$. Therefore, it is easy to see that $\mathbb{E}_{\mathrm{B}}(P) = \mathbb{E}_{\mathrm{H}}(\mathbb{P}_{\mathrm{A}}(\mathcal{N})) = \mathbb{P}_{\mathrm{A}}(G)/m!$ which makes P_G an unbiased estimate for $\mathbb{P}_{\mathrm{A}}(G)$. In Sect. 8.2, by controlling the variance of the random variable P, we will

Procedure B

Input: A graph G with degree sequence $\bar{d}$, and an integer $\ell = \ell(\epsilon, \delta)$.
Output: A real number P_G that is an (ϵ, δ)-estimate for $\mathbb{P}_{\mathrm{A}}(G)$.

1. Let E be a set of edges, $\hat{d} = (\hat{d}_1, \ldots, \hat{d}_n)$ be an n-tuple of integers, and P be a real number. Initialize them by $E =$ empty set, $\hat{d} = \bar{d}$, and $P = 1$.
2. Choose an edge $e = \{v_i, v_j\}$ of G among all those edges that are not in E, uniformly at random. Update P by

$$P = \frac{\hat{d}_i \hat{d}_j (1 - \frac{d_i d_j}{4m}) P}{\sum_{\substack{(v_r, v_s) \notin E \\ v_r \neq v_s}} \hat{d}_r \hat{d}_s (1 - \frac{d_r d_s}{4m})}.$$

   Add $\{v_i, v_j\}$ to E and reduce each of $\hat{d}_i$, $\hat{d}_j$ by 1.
3. Repeat step 2 until $|E| = m$.
4. Repeat steps 1 to 3 exactly $\ell = \ell(\epsilon, \delta)$ times and let $P_1, \ldots, P_\ell$ be the corresponding values for P. Output $P_G = m! \frac{P_1 + \cdots + P_\ell}{\ell}$ as an estimate for $m!\, \mathbb{E}_{\mathbb{H}}(\mathbb{P}_{\mathrm{A}}(\pi_G)) = \mathbb{P}_{\mathrm{A}}(G)$.

show the existence of an $\ell = \ell(\epsilon, \delta) = O(\epsilon^{-2} \log(1/\delta))$ such that the value of P_G is an (ϵ, δ)-estimate for $\mathbb{P}_{\mathrm{A}}(G)$.

3.2.3 FPRAS for Random Generation

Now that we can find (ϵ, δ)-estimates for both $|\mathcal{L}(\bar{d})|$ and $\mathbb{P}_{\mathrm{A}}(G)$ then an FPRAS for random generation is within reach. Algorithm GenerateGraph, given below provides such an FPRAS.

Algorithm: GenerateGraph

Input: A graphical degree sequence $\bar{d}$ and a positive numbers ϵ.
Output: A graph G with degree sequence $\bar{d}$.

1. Let $\epsilon' = \min(0.25, 1 - \frac{1}{\sqrt{1+\frac{\epsilon}{2}}}, \frac{1}{\sqrt{1-\frac{\epsilon}{2}}} - 1)$ and $\delta < 0.25$.
2. Run Algorithm CountGraph, to obtain X as an (ϵ', δ)-estimate for $|\mathcal{L}(\bar{d})|$.
3. Repeat Procedure A to obtain one successful outcome G.
4. Run Procedure B to obtain an (ϵ', δ)-estimate, P_G, for $\mathbb{P}_{\mathrm{A}}(G)$.
5. Report G with probability $\min(\frac{1}{\mathfrak{c} X P_G}, 1)$ and end. Otherwise go to step 3.

We will show in Sect. 4 that a universal constant $\mathfrak{c}$ exists (independent of all parameters $m, \bar{d}, \epsilon$) where the inequality $\mathfrak{c} X P_G \geq 1$ holds whenever $X \geq (1 - \epsilon')|\mathcal{L}(\bar{d})|$ and $P_G \geq (1 - \epsilon')\mathbb{P}_{\mathrm{A}}(G)$. Also note that we always assume $0 < \epsilon < 1$. Therefore, ϵ' is well defined.

4 Analysis

Let us fix a simple graph G with degree sequence $\bar{d}$. Recall the weighted configuration model from Sect. 3 which is equivalent to Procedure A. Denote the set of all perfect matchings on the mini-vertices of W that lead to G by $\mathrm{R}(G)$. Any two elements of $\mathrm{R}(G)$ can be obtained from one another by permuting the labels of the mini-vertices in any W_i. Due to this symmetry, all matchings in $\mathrm{R}(G)$ are generated with equal probability using Procedure A. In other words for a fixed element $\mathcal{M}$ in $\mathrm{R}(G)$ we have $\mathbb{P}_{\mathrm{A}}(G) = (\prod_{i=1}^{n} d_i!)\mathbb{P}_{\mathrm{A}}(\mathcal{M})$.

Now we will find $\mathbb{P}_{\mathrm{A}}(\mathcal{M})$. First note that there are $m!$ different orders for picking the edges of $\mathcal{M}$ sequentially. Moreover, different orderings can have different probabilities. Denote the set of these orderings by $\mathrm{S}(\mathcal{M})$. Thus

$$\mathbb{P}_{\mathrm{A}}(G) = \left(\prod_{i=1}^{n} d_i!\right) \sum_{\mathcal{N}\in \mathrm{S}(\mathcal{M})} \mathbb{P}_{\mathrm{A}}(\mathcal{N}).$$

For any ordering $\mathcal{N} = \{e_1, \ldots, e_m\}$ in the set $\mathrm{S}(\mathcal{M})$ and each r with $0 \le r \le m-1$ denote the probability of picking edge e_{r+1} at step $r+1$ of Procedure A by $\mathbb{P}(e_{r+1}|e_1, \ldots, e_r)$. Hence $\mathbb{P}_{\mathrm{A}}(\mathcal{N}) = \prod_{r=0}^{m-1} \mathbb{P}(e_{r+1}|e_1, \ldots, e_r)$ and each term $\mathbb{P}(e_{r+1}|e_1, \ldots, e_r)$ is given by

$$\mathbb{P}(e_{r+1} = \{i, j\}|e_1, \ldots, e_r) = \frac{(1 - d_i d_j/4m)}{\sum_{\{u,v\}\in E_r} d_u^{(r)} d_v^{(r)} (1 - d_u d_v/4m)} \tag{2}$$

where $d_i^{(r)}$ denotes the residual degree of vertex i at step $r+1$ and the set E_r consists of all possible edges after picking $e_1, \ldots, e_r$. Note that $d_i^{(r)}$ is also equal to the number of unmatched mini-vertices in W_i at step $r+1$. For the analysis we use the notations $\{i, j\}$ and $\{v_i, v_j\}$ interchangeably.

Denote the number of unsuitable pairs after choosing the edges in $\mathcal{N}_r = \{e_1, \ldots, e_r\}$ by $\Delta_r(\mathcal{N})$. Thus, the denominator of the right-hand side of (2) can be written as $\binom{2m-2r}{2} - \Psi_r(\mathcal{N})$ where $\Psi_r(\mathcal{N}) = \Delta_r(\mathcal{N}) + \sum_{\{u,v\}\in E_r} d_u^{(r)} d_v^{(r)} d_u d_v/4m$. This is because $\sum_{\{u,v\}\in E_r} d_u^{(r)} d_v^{(r)}$ is the number of the suitable pairs at step $r+1$, and is equal to $\binom{2m-2r}{2} - \Delta_r(\mathcal{N})$. The quantity $\Psi_r(\mathcal{N})$ can be also viewed as sum of the weights of the unsuitable pairs. Now using $1 - x = e^{-x+O(x^2)}$ for $0 \le x \le 1$, when $d_{\max} = O(m^{1/4-\tau})$ the expression for $\mathbb{P}_{\mathrm{A}}(G)$ is

$$\begin{aligned}\mathbb{P}_{\mathrm{A}}(G) &= \left(\prod_{i=1}^{n} d_i!\right) \left[\prod_{i\sim_G j} \left(1 - \frac{d_i d_j}{4m}\right)\right] \sum_{\mathcal{N}\in \mathrm{S}(\mathcal{M})} \prod_{r=0}^{m-1} \frac{1}{\binom{2m-2r}{2} - \Psi_r(\mathcal{N})} \\ &= \left(\prod_{i=1}^{n} d_i!\right) e^{-\gamma_G + o(1)} \sum_{\mathcal{N}\in \mathrm{S}(\mathcal{M})} \prod_{r=0}^{m-1} \frac{1}{\binom{2m-2r}{2} - \Psi_r(\mathcal{N})}\end{aligned}$$

where γ_G was defined in Sect. 3 to be $\gamma_G = \sum_{i\sim_G j} d_i d_j/4m$. The next step is to show that $\Psi_r(\mathcal{N})$ is sharply concentrated around a number $\psi_r(G)$, independent of the ordering $\mathcal{N}$. More specifically for

$$\psi_r(G) = (2m-2r)^2\left(\frac{\lambda(\bar{d})}{2m} + \frac{r\sum_{i\sim_G j}(d_i-1)(d_j-1)}{4m^3} + \frac{(\sum_{i=1}^n d_i^2)^2}{32m^3} + o(1)\right)$$

the following is true

$$\sum_{\mathcal{N}\in S(\mathcal{M})}\prod_{r=0}^{m-1}\frac{1}{\binom{2m-2r}{2}-\Psi_r(\mathcal{N})} = [1+o(1)]m!\prod_{r=0}^{m-1}\frac{1}{\binom{2m-2r}{2}-\psi_r(G)}. \tag{3}$$

The proof of this concentration result uses Kim and Vu's polynomial method [25] and is quite technical. It generalizes Kim and Vu's [26] calculations for the regular graphs to the general degree sequences. Section 7 is dedicated to this cumbersome analysis. But for the case of regular graphs, in Sect. 4.1, we will use a different technique based on Azuma's inequality to show concentration in a larger region.

The next step is to show that when $d_{\max} = O(m^{1/4-\tau})$,

$$\prod_{r=0}^{m-1}\frac{1}{\binom{2m-2r}{2}-\psi_r(G)} = \prod_{r=0}^{m-1}\frac{1}{\binom{2m-2r}{2}}e^{\lambda(\bar{d})+\lambda^2(\bar{d})+\gamma_G+o(1)}. \tag{4}$$

The proof of (4) is algebraic and is given in Sect. 7.2.

The above analysis can now be summarized in the following lemma.

Lemma 1 *For $d_{\max} = O(m^{1/4-\tau})$, Procedure A generates all graphs with degree sequence $\bar{d}$ with asymptotically equal probability. More specifically*

$$\sum_{\mathcal{N}\in S(\mathcal{M})}\mathbb{P}_A(\mathcal{N}) = \frac{m!}{\prod_{r=0}^{m}\binom{2m-2r}{2}}e^{\lambda(\bar{d})+\lambda^2(\bar{d})+o(1)}.$$

Now we can prove the first theorem.

Proof of Theorem 1 Lemma 1 shows that $\mathbb{P}_A(G)$ is asymptotically independent of G. Therefore, we only need to show Procedure A always succeeds with probability $1-o(1)$. We will show this in Sect. 5 by proving the following lemma.

Lemma 2 *For $d_{\max} = O(m^{1/4-\tau})$, the probability of failure of Procedure A is $o(1)$.*

Therefore, all graphs G are generated with asymptotically uniform probability. Note that this fact, combined with (3) will also give an independent proof of McKay's formula [29] for the number of graphs.

Finally we are left with the analysis of the running time which is summarized in the following lemma. The proof of this lemma is given in Sect. 6.

Lemma 3 *Procedure A can be implemented so that the expected running time is* $O(md_{\max})$ *for* $d_{\max} = O(m^{1/4-\tau})$.

This completes the proof of Theorem 1. □

Proof of Theorem 3 First we will prove that Algorithm CountGraphs is an FPRAS for the counting problem. This is shown by the following lemma.

Lemma 4 *For any* $\epsilon, \delta > 0$ *there exist* $k = k(\epsilon, \delta) = O(\epsilon^{-2}\log(1/\delta))$ *such that the output* X *of Algorithm CountGraphs is an* (ϵ, δ)*-estimate for* $|\mathcal{L}(\bar{d})|$.

Proof Since $\mathbb{E}_{\mathrm{A}}(N) = \mathcal{L}(\bar{d})$,

$$\begin{aligned}
&\mathbb{P}\big[(1-\epsilon)|\mathcal{L}(\bar{d})| < X < (1+\epsilon)|\mathcal{L}(\bar{d})|\big] \\
&\quad = \mathbb{P}\left(-\frac{\epsilon\mathbb{E}_{\mathrm{A}}(N)}{\sqrt{\frac{\mathrm{Var}_{\mathrm{A}}(N)}{k}}} < \frac{X - \mathbb{E}_{\mathrm{A}}(X)}{\sqrt{\frac{\mathrm{Var}_{\mathrm{A}}(N)}{k}}} < \frac{\epsilon\mathbb{E}_{\mathrm{A}}(N)}{\sqrt{\frac{\mathrm{Var}_{\mathrm{A}}(N)}{k}}}\right).
\end{aligned} \tag{5}$$

On the other hand, as a consequence of the Central Limit Theorem, when k goes to infinity, the quantity $\frac{X-\mathbb{E}_{\mathrm{A}}(X)}{\sqrt{\mathrm{Var}_{\mathrm{A}}(N)/k}}$ converges to a random variable Z which has a normal distribution with mean zero and variance 1. Therefore similar to the discussion given in [9], the inequality $\frac{\epsilon\mathbb{E}_{\mathrm{A}}(N)}{\sqrt{\mathrm{Var}_{\mathrm{A}}(N)/k}} > z_\delta$ guarantees that X is an (ϵ, δ)-estimate for $|\mathcal{L}(\bar{d})|$ where $\mathbb{P}(|Z| > z_\delta) = \delta$. This condition is equivalent to the following lower bound for the number of repetitions of Procedure A

$$k > z_\delta^2 \epsilon^{-2} \frac{\mathrm{Var}_{\mathrm{A}}(N)}{\mathbb{E}_{\mathrm{A}}(N)^2}.$$

Moreover, the tail of the normal distribution, $\mathbb{P}(|Z| > x)$, for very large values of x can be approximated by the quantity $ax^{-1}e^{-x^2/2}(2\pi)^{-1}$ where $a > 0$ is a constant. This means that the quantity z_δ^2 is of $O(\log(1/\delta))$. Therefore, if we show that the variance ratio $\mathrm{Var}_{\mathrm{A}}(N)/\mathbb{E}_{\mathrm{A}}(N)^2$ is bounded from above by a constant, then with $k = O(\log(1/\delta)\epsilon^{-2})$ repetitions, we can obtain an (ϵ, δ)-estimate. In fact we will prove the stronger statement

$$\frac{\mathrm{Var}_{\mathrm{A}}(N)}{\mathbb{E}_{\mathrm{A}}(N)^2} = o(1) \tag{6}$$

in Sect. 8.1. This finishes the proof of Lemma 4. □

Note that by Theorem 1, Procedure A uses $O(md_{\max})$ operations. Therefore the running time of Algorithm CountGraphs is $k(\epsilon, \delta)$ times $O(md_{\max})$ which is $O(md_{\max}\epsilon^{-2}\log(1/\delta))$. This shows that Algorithm CountGraphs is an FPRAS for estimating $|\mathcal{L}(\bar{d}|$.

Now we will prove that Algorithm GenerateGraph is an FPRAS for the random generation problem as well. First notice that if the ratio $\mathrm{Var}_{\mathrm{B}}(P)/\mathbb{E}_{\mathrm{B}}(P)^2$ is bounded from above by a constant, then similar calculations as in the proof of Lemma 4 for the tail of the normal distribution can be used to find $\ell = \ell(\epsilon, \delta) = O(\epsilon^{-2} \log(1/\delta))$ such that the output of Procedure B, P_G, is an (ϵ, δ)-estimate for $\mathbb{P}_{\mathrm{A}}(G)$. In fact we will show the stronger result

$$\frac{\mathrm{Var}_{\mathrm{B}}(P)}{\mathbb{E}_{\mathrm{B}}(P)^2} = o(1) \tag{7}$$

in Sect. 8.2. Therefore, (7) gives the following lemma.

Lemma 5 *For any $\epsilon, \delta > 0$ and a graph G with degree sequence $\bar{d}$, there exist $\ell = \ell(\epsilon, \delta) = O(\epsilon^{-2} \log(1/\delta))$ for Procedure B such that its output, P_G, is an (ϵ, δ)-estimate for $\mathbb{P}_{\mathrm{A}}(G)$.*

The next step in analyzing Algorithm GenerateGraph is to prove the existence of constant $\mathfrak{c}$ that is used in step 5.

Lemma 6 *There exists a constant $\mathfrak{c}$ such that for all parameters $m, \bar{d}, \epsilon$ and all graphs G with degree sequence $\bar{d}$, the inequality $\mathfrak{c} X P_G \geq 1$ holds whenever $X \geq (1-\epsilon')|\mathcal{L}(\bar{d})|$ and $P_G \geq (1-\epsilon')\mathbb{P}_{\mathrm{A}}(G)$.*

Proof By Theorem 1, $[1 - o(1)]|\mathcal{L}(\bar{d})|^{-1} \leq \mathbb{P}_{\mathrm{A}}(G) \leq [1 + o(1)]|\mathcal{L}(\bar{d})|^{-1}$. Let M be large enough such that for all $m > M$ the $o(1)$ terms are less than $1/2$. Now define

$$\mathfrak{d} = \min\left(\frac{1}{2}, \min_{m \leq M} \min_{G \in \mathcal{L}(\bar{d})} (\mathbb{P}_{\mathrm{A}}(G)|\mathcal{L}(\bar{d})|)\right)$$

and

$$\mathfrak{e} = \max\left(\frac{3}{2}, \max_{m \leq M} \max_{G \in \mathcal{L}(\bar{d})} (\mathbb{P}_{\mathrm{A}}(G)|\mathcal{L}(\bar{d})|)\right).$$

Therefore, $\mathfrak{d}$ and $\mathfrak{e}$ are positive and finite constants that are independent of all of the parameters $m, \bar{d}, \epsilon$ and $\mathfrak{d} \leq \mathbb{P}_{\mathrm{A}}(G)|\mathcal{L}(\bar{d})| \leq \mathfrak{e}$. Now when $X \geq (1-\epsilon')|\mathcal{L}(\bar{d})|$ and $P_G \geq (1-\epsilon')\mathbb{P}_{\mathrm{A}}(G)$,

$$\frac{\mathfrak{d}}{4} \leq \mathfrak{d}(1-\epsilon')^2 \leq P_G X.$$

This is because $\epsilon' \leq 0.25$. Therefore $\mathfrak{c} = 4/\mathfrak{d}$ suffices. □

Now we need to analyze the output distribution and the running time of Algorithm GenerateGraph. Consider one iteration of Algorithm GenerateGraph from step 1 to step 5. Let Ev_1 be the event that at least one of the fractions $\frac{X}{|\mathcal{L}(\bar{d})|}$ or $\frac{P_G}{\mathbb{P}_{\mathrm{A}}(G)}$ is not in the interval $[1-\epsilon', 1+\epsilon']$. Let Ev_2 be the event that a graph is reported in step 5. This means Ev_2^c is when "Otherwise go to step 3" is called. Therefore, $\mathbb{P}(Ev_1) \leq 2\delta < 0.5$ and $\mathbb{P}(Ev_1) + \mathbb{P}(Ev_2|Ev_1^c)\mathbb{P}(Ev_1^c) + \mathbb{P}(Ev_2^c|Ev_1^c)\mathbb{P}(Ev_1^c) = 1$.

For each graph $G \in \mathcal{L}(\bar{d})$ let $Ev_2(G)$ be the event that G is reported in step 5. Each graph G is reported with probability $\mathbb{P}(Ev_2(G)|Ev_1^c) = \mathbb{P}_{\mathrm{A}}(G)/(\mathfrak{c} X P_G)$ that satisfies

$$\frac{1-\epsilon/2}{\mathfrak{c}|\mathcal{L}(\bar{d})|} \leq \frac{1}{\mathfrak{c}(1+\epsilon')^2|\mathcal{L}(\bar{d})|} \leq \mathbb{P}(Ev_2(G)|Ev_1^c) \leq \frac{1}{\mathfrak{c}(1-\epsilon')^2|\mathcal{L}(\bar{d})|} \leq \frac{1+\epsilon/2}{\mathfrak{c}|\mathcal{L}(\bar{d})|}. \tag{8}$$

Note that the events $Ev_2(G)|Ev_1$ are not important and have low probability. Now we obtain

$$\mathbb{P}(Ev_2) \geq \mathbb{P}(Ev_1^c) \sum_{G\in\mathcal{L}(\bar{d})} \mathbb{P}(Ev_2(G)|Ev_1^c) \geq \frac{0.5(1-\epsilon/2)}{\mathfrak{c}}.$$

Therefore, the expected number of times that "Otherwise go to step 3" is called is $\mathbb{P}(Ev_2)^{-1} \leq 4\mathfrak{c}$. This means that the expected running time of Algorithm GenerateGraph is at most $4\mathfrak{c}$ times the expected running time of a successful run of Procedure A plus $4\mathfrak{c}$ times the expected running time of Procedure B plus the expected running time of Algorithm CountGraphs. The total number of operations can be written as

$$4\mathfrak{c}O(md_{\max}) + 4\mathfrak{c}O(md_{\max}\epsilon'^{-2}\log(1/\delta)) + O(md_{\max}\epsilon'^{-2}\log(1/\delta))$$

which is $O(md_{\max}\epsilon^{-2})$, since $\epsilon' \geq \min(\epsilon/4, 0.25)$ gives $\epsilon'^{-2} = O(\epsilon^{-2})$.

Notice that the probability that Algorithm GenerateGraph eventually reports a graph, in an iteration that Ev_1 did not occur, is at least $1 - \mathbb{P}(Ev_1) > 0.5$. Moreover, the probability that the reported graph is a fixed graph $G \in \mathcal{L}(\bar{d})$ satisfies

$$\sum_{i=0}^{\infty} \mathbb{P}(Ev_2^c)^i \mathbb{P}(Ev_2(G)|Ev_1^c)\mathbb{P}(Ev_1^c)$$
$$= \frac{\mathbb{P}(Ev_2(G)|Ev_1^c)\mathbb{P}(Ev_1^c)}{\mathbb{P}(Ev_2)} \in \left[\frac{1-\epsilon}{c'|\mathcal{L}(\bar{d})|}, \frac{1+\epsilon}{c'|\mathcal{L}(\bar{d})|}\right]$$

where $c' = \frac{\mathbb{P}(Ev_1^c)}{\mathfrak{c}\mathbb{P}(Ev_2)}$. This finishes the proof of Theorem 3. □

4.1 Concentration Inequality for Regular Graphs

The aim of this section is to prove Theorem 2. Recall that $\mathcal{L}(n,d)$ denotes the set of all simple d-regular graphs with $m = nd/2$ edges. Let $\mathbb{P}_{\mathrm{U}}$ be the uniform probability on $\mathcal{L}(n,d)$. Similar to the analysis of Procedure A for general degree sequences, let G be a fixed graph in $\mathcal{L}(n,d)$ and $\mathcal{M}$ be a fixed matching on W with $G_{\mathcal{M}} = G$. The main goal is to show that for $d = o(n^{1/2-\tau})$ the probability of generating G with Procedure A is at least $1 - o(1)$ times $\mathbb{P}_{\mathrm{U}}(G)$; i.e.

$$\mathbb{P}_{\mathrm{A}}(G) \geq (1-o(1))\mathbb{P}_{\mathrm{U}}(G). \tag{9}$$

For the moment, assume (9) is true. We will show that Theorem 2 follows. Later we will show why (9) holds.

Proof of Theorem 2 First, we will show that the total variation distance between the probability measures $\mathbb{P}_{\mathrm{A}}$ and $\mathbb{P}_{\mathrm{U}}$, $d_{\mathrm{TV}}(\mathbb{P}_{\mathrm{A}}, \mathbb{P}_{\mathrm{U}}) \equiv \sup_{\mathcal{S}\subset\mathcal{L}(n,d)} |\mathbb{P}_{\mathrm{A}}(\mathcal{S}) - \mathbb{P}_{\mathrm{U}}(\mathcal{S})|$ is $o(1)$. We will use the following upper bound on the total variation distance

$$d_{\mathrm{TV}}(\mathbb{P}_{\mathrm{A}}, \mathbb{P}_{\mathrm{U}}) \leq \sum_{G\in\mathcal{L}(n,d)} |\mathbb{P}_{\mathrm{A}}(G) - \mathbb{P}_{\mathrm{U}}(G)|.$$

Therefore, we have the upper-bound

$$\begin{aligned}
&\sum_{G\in\mathcal{L}(n,d)} |\mathbb{P}_{\mathrm{A}}(G) - \mathbb{P}_{\mathrm{U}}(G)| \\
&= \sum_{\substack{G\in\mathcal{L}(n,d)\\ \mathbb{P}_{\mathrm{A}}\geq\mathbb{P}_{\mathrm{U}}}} \Big(\mathbb{P}_{\mathrm{A}}(G) - \mathbb{P}_{\mathrm{U}}(G)\Big) + \sum_{\substack{G\in\mathcal{L}(n,d)\\ \mathbb{P}_{\mathrm{A}}<\mathbb{P}_{\mathrm{U}}}} |\mathbb{P}_{\mathrm{A}}(G) - \mathbb{P}_{\mathrm{U}}(G)| \\
&= \sum_{G\in\mathcal{L}(n,d)} \Big(\mathbb{P}_{\mathrm{A}}(G) - \mathbb{P}_{\mathrm{U}}(G)\Big) + 2\sum_{\substack{G\in\mathcal{L}(n,d)\\ \mathbb{P}_{\mathrm{A}}<\mathbb{P}_{\mathrm{U}}}} |\mathbb{P}_{\mathrm{A}}(G) - \mathbb{P}_{\mathrm{U}}(G)| \\
&\overset{\text{(a)}}{\leq} 2\sum_{\substack{G\in\mathcal{L}(n,d)\\ \mathbb{P}_{\mathrm{A}}<\mathbb{P}_{\mathrm{U}}}} |\mathbb{P}_{\mathrm{A}}(G) - \mathbb{P}_{\mathrm{U}}(G)| \\
&\overset{\text{(b)}}{\leq} 2\,o(1) \sum_{\substack{G\in\mathcal{L}(n,d)\\ \mathbb{P}_{\mathrm{A}}<\mathbb{P}_{\mathrm{U}}}} \mathbb{P}_{\mathrm{U}}(G) \leq o(1).
\end{aligned}$$

Here (a) uses $\sum_{G\in\mathcal{L}(n,d)} \mathbb{P}_{\mathrm{A}}(G) \leq 1$ and $\sum_{G\in\mathcal{L}(n,d)} \mathbb{P}_{\mathrm{U}}(G) = 1$. To see why (b) holds, note that $\mathbb{P}_{\mathrm{U}}(G) - \mathbb{P}_{\mathrm{A}}(G) \leq o(1)\mathbb{P}_{\mathrm{U}}(G)$ which is equivalent to inequality (9).

Now, $d_{\mathrm{TV}}(\mathbb{P}_{\mathrm{A}}, \mathbb{P}_{\mathrm{U}}) = o(1)$ implies that $\mathbb{P}_{\mathrm{A}}(G) \leq (1 + o(1))\mathbb{P}_{\mathrm{U}}(G)$ except for graphs G in a subset of $\mathcal{L}(n,d)$ with size $o(|\mathcal{L}(n,d)|)$. This finishes the proof of Theorem 2. □

4.1.1 *Proof of inequality* (9)

In order to prove inequality (9) we prove the following equivalent inequality

$$(d!)^n \sum_{\mathcal{N}\in \mathrm{S}(\mathcal{M})} \mathbb{P}(\mathcal{N}) \geq \frac{1 - o(1)}{|\mathcal{L}(n,d)|}. \tag{10}$$

Our proof of inequality (10) builds upon the steps in [27]. First define $\mu_r = \mu_r^{(1)} + \mu_r^{(2)}$ where

$$\begin{aligned}
\mu_r^{(1)} &= \frac{(2m-2r)^2(d-1)}{4m} \\
\mu_r^{(2)} &= \frac{(2m-2r)^2(d-1)^2 r}{4m^2}.
\end{aligned}$$

Let $m_1 = \frac{m}{d^2\omega}$ where ω goes to infinity very slowly; e.g. $\omega = O(\log^\delta n)$ for some small $\delta > 0$. The following summarizes the analysis of Kim and Vu [27] for $d = O(n^{1/3-\tau})$

$$
\begin{aligned}
|\mathcal{L}(n,d)|(d!)^n \sum_{\mathcal{N}\in \mathrm{S}(\mathcal{M})} \mathbb{P}(\mathcal{N}) &\overset{(c)}{=} \frac{1-o(1)}{m!} \sum_{\mathcal{N}\in \mathrm{S}(\mathcal{M})} \prod_{r=0}^{m-1} \frac{\binom{2m-2r}{2} - \mu_r}{\binom{2m-2r}{2} - \Delta_r(\mathcal{N})} \\
&\overset{(d)}{\geq} \frac{1-o(1)}{m!} \sum_{\mathcal{N}\in \mathrm{S}(\mathcal{M})} \prod_{r=0}^{m_1} \left(1 + \frac{\Delta_r(\mathcal{N}) - \mu_r}{\binom{2m-2r}{2} - \Delta_r(\mathcal{N})}\right) \\
&\overset{(e)}{\geq} \Big(1-o(1)\Big) \prod_{r=0}^{m_1} \left(1 - 3\frac{T_r^{(1)} + T_r^{(2)}}{(2m-2r)^2}\right) \\
&\overset{(f)}{\geq} \Big(1-o(1)\Big) \exp\left(-3e \sum_{r=0}^{m_1} \frac{T_r^{(1)} + T_r^{(2)}}{(2m-2r)^2}\right). \qquad (11)
\end{aligned}
$$

Here we explain these steps in more detail. Our main focus will be on step (e) which is the main step. For the rest, we provide a brief description and a reference to [27]. Step (c) follows from (3.5) of [27] and writing McKay-Wormald's estimate [31] for $|\mathcal{L}(n,d)|$ as a multiple of the product $\prod_{r=0}^{m-1}[\binom{2m-2r}{2} - \mu_r]$. Similarly, step (d) follows from the algebraic calculations in page 455 of [27].

The important step (e) follows from a sharp concentration. For simplicity write Δ_r instead of $\Delta_r(\mathcal{N})$ and break Δ_r into two terms $\Delta_r^{(1)} + \Delta_r^{(2)}$. Here $\Delta_r^{(1)}$ and $\Delta_r^{(2)}$ denote the number of unsuitable pairs in step r corresponding to the self edge loops and to the double edges respectively. For $p_r = r/m$, $q_r = 1 - p_r$ Kim and Vu [27] used their polynomial concentration inequality [25] to derive two bounds $T_r^{(1)}$, $T_r^{(2)}$ and to show that with very high probability $|\Delta_r^{(1)} - \mu_r^{(1)}| < T_r^{(1)}$ and $|\Delta_r^{(2)} - \mu_r^{(2)}| < T_r^{(2)}$. More precisely for some constants c_1, c_2 the bounds are

$$
T_r^{(1)} = c_1 \log^2 n \sqrt{nd^2 q_r^2 (2dq_r + 1)}, \qquad T_r^{(2)} = c_2 \log^3 n \sqrt{nd^3 q_r^2 (d^2 q_r + 1)}.
$$

Now it is easy to see that for each $i \in \{1,2\}$ the bound $T_r^{(i)}$ and the quantity $\mu_r^{(i)}$ are $o((2m-2r)^2)$. This validates the step (e).

Finally, the step (f) is straightforward using $1 - x \geq e^{-ex}$ for $0 \leq x \leq 1$.

The rest of the proof focuses on showing that the right-hand side of inequality (11) is at least $1 - o(1)$. Kim and Vu show that for $d = O(n^{1/3-\tau})$ the exponent in (11) is $o(1)$. Using similar calculations as (3.13) in [27] it can be shown that for $d = O(n^{1/2-\tau})$ and $m_2 = (m\log^3 n)/d$

$$
\sum_{r=0}^{m_1} \frac{T_r^{(1)}}{(2m-2r)^2} = o(1), \qquad \sum_{r=m_2}^{m_1} \frac{T_r^{(2)}}{(2m-2r)^2} = o(1).
$$

But unfortunately the summation $\sum_{r=0}^{m_2} \frac{T_r^{(2)}}{(2m-2r)^2}$ is $\Omega(d^3/n)$. In fact it turns out that the random variable $\Delta_r^{(2)}$ has large variance for $d = O(n^{1/2-\tau})$.

Let us explain the main difficulty for moving from $d = O(n^{1/3-\tau})$ to $d = O(n^{1/2-\tau})$. Note that $\Delta_r^{(2)}$ is defined on a random subgraph $G_{\mathcal{N}_r}$ of graph G which has exactly r edges. Both [37] and [26, 27] have approximated the subgraph $G_{\mathcal{N}_r}$ with G_{p_r} in which each edge of G appears independently with probability $p_r = r/m$. But when $d = O(n^{1/2-\tau})$, this approximation causes the variance of $\Delta_r^{(2)}$ to become exponentially large.

In order to fix the problem, we modify $\Delta_r^{(2)}$ before moving to G_{p_r}. It can be shown via simple algebraic calculations that: $\Delta_r^{(2)} - \mu_r^{(2)} = X_r - Y_r$ where

$$X_r = \sum_{u \sim_{G_{\mathcal{N}_r}} v} [d_u^{(r)} - q_r(d-1)][d_v^{(r)} - q_r(d-1)],$$

$$Y_r = q_r(d-1) \sum_u [(d_u^{(r)} - q_r d)^2 - d p_r q_r].$$

This modification is critical since the equality $\Delta_r^{(2)} - \mu_r^{(2)} = X_r - Y_r$ does not hold in G_{p_r}.

The next task is to find a new bound $\hat{T}_r^{(2)}$ such that $|X_r - Y_r| < \hat{T}_r^{(2)}$ with very high probability and $\sum_{r=0}^{m_2} \frac{\hat{T}_r^{(2)}}{(2m-2r)^2} = o(1)$. It is easy to see that in G_{p_r} both X_r and Y_r have zero expected value.

At this time we will move to G_{p_r} and show that X_r and Y_r are sharply concentrated around zero. It is easy to see that with probability at least $1/n$, the subgraph G_{p_r} has exactly r edges. This is in fact Lemma 21 which is proven in Sect. 7. Therefore, X_r and Y_r will be sharply concentrated around 0 in $G_{\mathcal{N}_r}$ as well. In the following we will show the concentration of X_r in G_{p_r}. The concentration of Y_r can be shown similarly.

Consider the edge exposure martingale (page 94 of [2]) for G_{p_r} that examines the edges of G in the order $e_1, \dots, e_m$. In particular for any $0 \le \ell \le r$ define $Z_\ell^r = \mathbb{E}(X_r \,|\, e_1, \dots, e_\ell)$. Therefore, Z_m^r is just the value of X_r and Z_0^r is its expected value $\mathbb{E}(X_r)$ in G_{p_r}. To simplify the notation, let us drop the index r from $Z_\ell^r, d_u^{(r)}, p_r$ and q_r.

The next step is to bound the martingale difference $|Z_i - Z_{i-1}|$ and use a martingale concentration inequality. In order to bound the quantity $|Z_i - Z_{i-1}|$, assume that $e_i = \{u, v\}$. The difference between Z_i and Z_{i-1} is in the terms involving e_i in the summation $\sum_{u' \sim_{G_p} v'} [d_{u'} - q(d-1)][d_{v'} - q(d-1)]$. But e_i only participates in d_u and d_v. Thus, for any u' where $u' \sim_{G_p} u$, the term $[d_{u'} - q(d-1)][d_u - q(d-1)]$ appears in both Z_i and Z_{i-1}. The value of $d_{u'} - q(d-1)$ is unchanged by revealing the status of e_i, but the value of $d_u - q(d-1)$ can fluctuate by at most 1. Moreover, if $e_i \in G_p$ then an extra term $[d_u - q(d-1)][d_v - q(d-1)]$ is also added to Z_i. This means we have

$$|Z_i - Z_{i-1}| \le \big|(d_u - (d-1)q)(d_v - (d-1)q)\big| + \Bigg|\sum_{u' \sim_{G_p} u} (d_{u'} - (d-1)q)\Bigg| + \Bigg|\sum_{v' \sim_{G_p} v} (d_{v'} - (d-1)q)\Bigg|. \quad (12)$$

Bounding the above difference should be done carefully since the standard worst case bounds are weak for our purpose.

First, we start by a useful observation. For a typical ordering $\mathcal{N}$ of the edges of G, the residual degrees, $d_u,\ d_v,\ d_{u'},\ d_{v'}$ are roughly $dq \pm \sqrt{dq}$. We will make this more precise. For any vertex $\bar{u} \in G$ consider the event

$$L_{\bar{u}} = \left\{|d_{\bar{u}} - dq| \le c \log^{1/2} n (dq)^{1/2}\right\}$$

where $c > 0$ is a large constant.

Lemma 7 *For all* $0 \le r \le m_2$ *we have* $\mathbb{P}(L_{\bar{u}}^c) = o(\frac{1}{m^4})$.

Proof Note that in the G_p model the residual degree of a vertex $\bar{u}$, $d_{\bar{u}}$, is sum of d independent Bernoulli random variables with mean q. Two generalizations of Chernoff inequality (Theorems A.1.11, A.1.13 in page 267 of [2]) state that for $a > 0$ and $X_1, \ldots, X_d$ i.i.d. Bernoulli(q) random variables:

$$\mathbb{P}(X_1 + \cdots + X_d - qd \ge a) < e^{-\frac{a^2}{2qd} + \frac{a^3}{2(qd)^2}}$$

$$\mathbb{P}(X_1 + \cdots + X_d - qd < -a) < e^{-\frac{a^2}{2qd}}$$

Applying these two for $a = \sqrt{12qd \log n}$ proves Lemma 7. □

To finish bounding the martingale difference we look at the last two terms in the right-hand side of (12). For the vertex u consider the event

$$K_u = \left\{\left|\sum_{u' \sim_{G_p} u} (d_{u'} - (d-1)q)\right| \le c\left[(dq)^{3/2} + qd + dq^{1/2}\right]\log n\right\}$$

where $c > 0$ is a large constant. We will use the following lemma to show that the complement of K_u has very low probability.

Lemma 8 *For all* $0 \le r \le m_2$ *the event* K_u^c *has probability* $o(\frac{d}{m^4})$.

Proof For any vertex u let $N_G(u) \subset V(G)$ denote the neighbors of u in G. Consider the subsets

$$A_G(u),\ B_G(u),\ C_G(u) \subset E(G)$$

where $A_G(u)$ consists of the edges that are adjacent to u, $B_G(u)$ has those edges with both endpoints in $N_G(u)$, and $C_G(u)$ contains those edges with exactly one endpoint in $N_G(u)$ and one endpoint outside $N_G(u) \cup \{u\}$. For any edge e of G let $t_e = 1_{\{e \notin G_p\}}$. Then we can write

$$
\begin{aligned}
&\sum_{u'\sim_{G_p}u}(d_{u'}-(d-1)q)\\
&=\sum_{u'\in N_{G_p}(u)}\ \sum_{e\in A_G(u')\setminus A_G(u)}(t_e-q)\\
&=\sum_{u'\in N_G(u)}\ \sum_{e\in A_G(u')\setminus A_G(u)}(t_e-q)-\sum_{u'\in N_G(u)\setminus N_{G_p}(u)}\ \sum_{e\in A_G(u')\setminus A_G(u)}(t_e-q)\\
&=\underbrace{\sum_{e\in C_G(u)}(t_e-q)}_{\text{(i)}}+2\underbrace{\sum_{e\in B_G(u)}(t_e-q)}_{\text{(ii)}}-\underbrace{\sum_{u'\in N_G(u)\setminus N_{G_p}(u)}(d'_u-1-q(d-1))}_{\text{(iii)}}.
\end{aligned}
$$

Here each of (i) and (ii) is a sum of $O(d^2)$ i.i.d. Bernoulli(q) random variables minus their expectations. Therefore similar to Lemma 7, both (i) and (ii) can be shown to be $O(\sqrt{12qd^2\log n})$ with a probability at least $1-o(1/m^4)$. For (iii) we can say

$$
\sum_{u'\in N_G(u)\setminus N_{G_p}(u)}(d'_u-1-q(d-1))\le d_u\max_{u'\in N_G(u)\setminus N_{G_p}(u)}(|d_{u'}-1-q(d-1)|).
$$

Now using Lemma 7 for d_u and each term $d_{u'}-1-q(d-1)$ we can say (iii) is $O([dq+\sqrt{12qd\log n}]\sqrt{12qd\log n})$ with a probability at least $1-o(d/m^4)$. These finish the proof of Lemma 8. □

The final step in bounding the martingale difference is to apply Lemmas 7, 8 and the union bound to event $L=\bigcap_{r=0}^{m_2}\bigcap_{u=1}^{n}(L_u\cap K_u)$ and obtain $\mathbb{P}(L^c)=o(1/m^2)$.

Hence for the martingale difference we have

$$
|Z_i-Z_{i-1}|\mathbf{1}_L\le O(dq+dq^{1/2}+(dq)^{3/2})\log n.
$$

Note that Azuma's inequality cannot be used directly, since the martingale difference $|Z_i-Z_{i-1}|$ can be large outside the set L. But the complement of L has very low probability and we can use the following variation of Azuma's inequality.

Proposition 1 (Kim [24]) *Consider a martingale* $\{Y_i\}_{i=0}^n$ *adaptive to a filtration* $\{\mathcal{B}_i\}_{i=0}^n$. *If for all* k *there are* $A_{k-1}\in\mathcal{B}_{k-1}$ *such that* $\mathbf{E}[e^{\omega Y_k}|\mathcal{B}_{k-1}]\mathbf{1}_{A_{k-1}}\le C_k$ *for all* $k=1,2,\ldots,n$ *with* $C_k\ge 1$ *for all* k, *then*

$$
\mathbf{P}(Y-\mathbf{E}[Y]\ge\lambda)\le e^{-\lambda\omega}\prod_{k=1}^{n}C_k+\mathbf{P}\left(\bigcup_{k=0}^{n-1}A_k\right)
$$

Proof of Theorem 2 Applying the above proposition for a large enough constant $c'>0$ gives

$$
\mathbb{P}\left(|X_r|>c'\sqrt{6r\log^3 n\left(dq+d(q)^{1/2}+(dq)^{3/2}\right)^2}\right)\le e^{-3\log n}+\mathbb{P}(L^c)=o\left(\frac{1}{m^2}\right).
$$

Now using the fact that G_p has r edges with probability at least $1/n$, the same event in the random model $G_{\mathcal{N}_r}$ has probability $o(1/m)$. A similar bound holds for Y_r since the martingale difference for Y_r is $O(|dq(d_u - qd)|) = O((dq)^{3/2}\log^{1/2} n))$ using Lemma 7.

Therefore defining $\hat{T}_r^{(2)} = c'(dq + d(q)^{1/2} + (dq)^{3/2})\sqrt{6r\log^3 n}$, we only need to show

$$\sum_{r=0}^{m_2} \frac{(dq + d(q)^{1/2} + (dq)^{3/2})\sqrt{6r\log^3 n}}{(2m-2r)^2} = o(1).$$

But using $ndq = 2m - 2r$ we have

$$\begin{aligned}
&\sum_{r=0}^{m_2} \frac{(dq + dq^{1/2} + (dq)^{3/2})\sqrt{6r\log^3 n}}{n^2 d^2 q^2} \\
&\quad = \sum_{r=0}^{m_2} O\left(\frac{d^{1/2}\log^{1.5} n}{n^{1/2}(2m-2r)} + \frac{d\log^{1.5} n}{(2m-2r)^{3/2}} + \frac{d^{1/2}\log^{1.5} n}{n(2m-2r)^{1/2}}\right) \\
&\quad = O\left(\frac{d^{1/2}\log(nd)}{n^{1/2}} + \frac{d}{(n\log^3 n)^{1/2}} + \frac{d}{n^{1/2}}\right)\log^{1.5} n = o(1)
\end{aligned}$$

for $d = O(n^{1/2-\tau})$. □

5 Probability of Failure of Procedure A

In this section we will prove Lemma 2 from Sect. 4. First we present the following remark.

Remark 2 Lemma 1 gives an upper bound for the number of simple graphs with degree sequence $\bar{d}$ independently from all known formulas for $|\mathcal{L}(\bar{d})|$. If $d_{\max} = O(m^{1/4-\tau})$ then

$$|\mathcal{L}(\bar{d})| \leq e^{-\lambda(\bar{d})-\lambda^2(\bar{d})+o(1)} \frac{\prod_{r=0}^{m}\binom{2m-2r}{2}}{m!\prod_{i=1}^{n} d_i!}.$$

In this section we will show that the above inequality is in fact an equality. This is done by proving that the probability of failure of Procedure A is very small.

First we will characterize the degree sequence of the partial graph that is generated up to the time of failure. Then we apply the upper bound of Remark 2 to derive an upper bound on the probability of failure and show that it is $o(1)$.

Lemma 9 *If Procedure A fails in step s then $2m - 2s \leq d_{\max}^2 + 1$.*

Proof Procedure A fails when there is no suitable pair left to choose. If the failure occurs in step s then the number of unsuitable edges is equal to the total number of possible pairs, that is $\binom{2m-2s}{2}$. On the other hand, it can be easily shown that the number of unsuitable edges at step s is at most $d_{\max}^2(2m-2s)/2$ (see Corollary 3.1 in [37] for more detail). Therefore $2m-2s \leq d_{\max}^2+1$. □

Failure in step s means there are some W_i's which have unmatched mini-vertices ($d_i^{(s)} \neq 0$). Let us call them "unfinished" W_i's. Since the algorithm fails, any two unfinished W_i's should be already connected. Hence there are at most $d_{\max}$ of them. This is because for all i: $|W_i| = d_i \leq d_{\max}$. The main goal is now to show that this scenario is a very rare event. Without loss of generality assume that $W_1, W_2, \ldots, W_k$ are all the unfinished sets. The argument given above shows $k \leq d_{\max}$. Moreover, by construction $k \leq 2m-2s$. The algorithm up to this step has created a partial matching $\mathcal{M}_s$ where graph $G_{\mathcal{M}_s}$ is simple and has degree sequence $\bar{d}^{(s)} = (d_1 - d_1^{(s)}, \ldots, d_k - d_k^{(s)}, d_{k+1}, \ldots, d_n)$. Let $A_{d_1^{(s)},\ldots,d_k^{(s)}}$ denote the above event of failure. Hence

$$\mathbb{P}(\text{fail}) = \sum_{2m-2s=2}^{d_{\max}^2+1} \sum_{k=1}^{\max(d_{\max},2m-2s)} \sum_{i_1,\ldots,i_k=1}^{n} \mathbb{P}_{\mathrm{A}}(A_{d_1^{(s)},\ldots,d_k^{(s)}}). \tag{13}$$

The following lemma is the central part of the proof.

Lemma 10 *The probability of the event that Procedure A fails in step s and the vertices $v_1, \ldots, v_k$ are the only unfinished vertices*; *i.e.* $d_i^{(s)} \neq 0$ $i = 1, \ldots, k$, *is at most*

$$(1+o(1))\, \frac{d_{\max}^{k(k-1)} \prod_{i=1}^{k} d_i^{d_i^{(s)}}}{m^{\binom{k}{2}} (2m)^{2m-2s}} \binom{2m-2s}{d_1^{(s)}, \ldots, d_k^{(s)}}.$$

Proof Following the above notation, the event that we are considering is denoted by $A_{d_1^{(s)},\ldots,d_k^{(s)}}$. Note that graph $G_{\mathcal{M}_s}$ should have a clique of size k on vertices $v_1, \ldots, v_k$. Therefore, the number of such graphs should be less than $|\mathcal{L}(\bar{d}_k^{(s)})|$ where $\bar{d}_k^{(s)}=(d_1 - d_1^{(s)} - (k-1), \ldots, d_k - d_k^{(s)} - (k-1), d_{k+1}, \ldots, d_n)$. Thus, $\mathbb{P}_{\mathrm{A}}(A_{d_1^{(s)},\ldots,d_k^{(s)}})$ is at most $|\mathcal{L}(\bar{d}_k^{(s)})|\ \mathbb{P}_{\mathrm{A}}(G_{\mathcal{M}_s})$. On the other hand, we can use Remark 1 to derive an upper bound for $|\mathcal{L}(\bar{d}_k^{(s)})|$ because $m-s$ and k are small relative to m and it is easy to show that $d_{\max} = O([s - \binom{k}{2}]^{1/4-\tau})$. The result of these steps is

$$\mathbb{P}_{\mathrm{A}}(A_{d_1^{(s)},\ldots,d_k^{(s)}}) \leq \left(\frac{(2s - k(k-1))! \exp[-\lambda(\bar{d}_k^{(s)}) - \lambda^2(\bar{d}_k^{(s)}) + o(1)]}{[s - \binom{k}{2}]! 2^{s-\binom{k}{2}} \prod_{i=1}^{n} (d_i^{(s)})!} \right) \mathbb{P}_{\mathrm{A}}(G_{\mathcal{M}_s}).$$

The next step is to bound $\mathbb{P}_{\mathrm{A}}(G_{\mathcal{M}_s})$. We can use the same methodology as in the beginning of Sect. 4 to derive

$$\begin{aligned}\mathbb{P}_{\mathrm{A}}(G_{\mathcal{M}_s}) &= \frac{\prod_{i=1}^{n} d_i!}{(\prod_{i=1}^{k}[d_i^{(s)}]!)} \sum_{\mathcal{N}_s \in S(\mathcal{M}_s)} \mathbb{P}_{\mathrm{A}}(\mathcal{N}_s) \\ &= s!\exp\left(-\frac{\sum_{i\sim_{G_s} j} d_i d_j}{4m} + o(1)\right)\prod_{r=0}^{s-1}\frac{1}{\binom{2m-2r}{2} - \psi_r(G_{\mathcal{M}_s})} \\ &= s!\exp\left(\frac{m}{s}\lambda(\bar{d}) + \frac{m^2}{s^2}\lambda^2(\bar{d}) + o(1)\right)\prod_{r=0}^{s-1}\frac{1}{\binom{2m-2r}{2}}.\end{aligned}$$

Similar to ψ_r, the quantity $\psi_r(G_{\mathcal{M}_s})$ is an approximation for the expected value of Ψ_r conditioned on obtaining $G_{\mathcal{M}_s}$ at step s. Now using the simple algebraic approximation

$$\begin{aligned}\frac{m}{s}\lambda(\bar{d}) + \frac{m^2}{s^2}\lambda^2(\bar{d}) - \lambda(\bar{d}_k^{(s)}) - \lambda^2(\bar{d}_k^{(s)}) &= O\big(\lambda(\bar{d})[\lambda(\bar{d}) - \lambda(\bar{d}_k^{(s)})]\big) \\ &= O\left(\frac{d_{\max}^4}{m^2}\right) = o(1)\end{aligned}$$

the following is true

$$\begin{aligned}\mathbb{P}_{\mathrm{A}}(A_{d_1^{(s)},\ldots,d_k^{(s)}}) &\le e^{o(1)}\frac{[2s-k(k-1)]!\,(2m-2s)!s!2^{\binom{k}{2}}\prod_{i=1}^{k} d_i!}{[s-\binom{k}{2}]!(2m)!\prod_{i=1}^{k}[(d_i^{(s)})!(d_i-k-d_i^{(s)}+1)!]} \\ &\le e^{o(1)}\frac{\prod_{i=1}^{k} d_i^{d_i^{(s)}+k-1}}{\prod_{\ell=2s+1}^{2m}\ell\prod_{j=1}^{\binom{k}{2}}(2s-2j+1)}\binom{2m-2s}{d_1^{(s)},\ldots,d_k^{(s)}}. \qquad (14)\end{aligned}$$

The next step is to use $m-s=O(d_{\max}^2)$ and $k=O(d_{\max})$ to show that $\prod_{j=1}^{\binom{k}{2}}(2s-2j+1) \ge m^{\binom{k}{2}}$ and $(1/m^{2m-2s})\prod_{\ell=2s+1}^{2m}\ell \ge e^{-O(d_{\max}^4/m)}$. These two facts combined with (14) finish the proof of Lemma 10. □

Now we are ready to prove the main result of this section.

Proof of Lemma 2 First, we show that the event of failure has a negligible probability when there is only one unfinished vertex left, i.e., when $k=1$. In this case Lemma 10 simplifies to $\mathbb{P}_{\mathrm{A}}(A_{d_1^{(s)}}) = O((\frac{D}{m})^{2m-2s})$. Therefore, summing over all possibilities of $k=1$ gives

$$\sum_{2m-2s=2}^{d_{\max}^2+1}\sum_{i=1}^{n}\mathbb{P}_{\mathrm{A}}(A_{d_i^{(s)}}) = O\left(\sum_{2m-2s=2}^{d_{\max}^2+1}\frac{d_{\max}^{2m-2s-1}}{m^{2m-2s-1}}\right) = O\left(\frac{d_{\max}}{m}\right) = o(1).$$

For $k > 1$ we use Lemma 10 differently. Using $d_{\max}^{k(k-1)}/m^{\binom{k}{2}} \leq d_{\max}^2/m$ and (13) we have

$$\mathbb{P}(\text{fail}) \leq o(1)$$

$$+ \frac{e^{o(1)} d_{\max}^2}{m} \sum_{2m-2s=2}^{d_{\max}^2+1} \frac{\overbrace{\sum_{k=2}^{\max(d_{\max}, 2m-2s)} \sum_{i_1,\dots,i_k=1}^{n} \prod_{i=1}^{k} d_i^{d_i^{(s)}} \binom{2m-2s}{d_1^{(s)}, \dots, d_k^{(s)}}}^{\text{(a)}}}{(2m)^{2m-2s}}.$$

Now note that the double sum (a) is at most $(d_1 + \cdots + d_n)^{2m-2s} = (2m)^{2m-2s}$ since $\sum_{i=1}^{k} d_i^{(s)} = 2m - 2s$. Therefore

$$\mathbb{P}(\text{fail}) \leq o(1) + e^{o(1)} \frac{d_{\max}^2}{m} \sum_{2m-2s=2}^{d_{\max}^2+1} 1 = O\left(\frac{d_{\max}^4}{m}\right) = o(1). \qquad \square$$

6 Running Time of Procedure A

In this section we prove Lemma 3.

Proof of Lemma 3 Our proof is very similar to the analysis of Steger and Wormald [37]. They use a non-trivial data structure and algorithm to efficiently choose a pair of vertices $v_i \in V$ and $v_j \in V$ with probabilities proportional to $\hat{d}_i$ and $\hat{d}_j$ respectively. They explain their methods for regular graphs but they only use the fact that the maximum degree is bounded. We include their analysis in Sect. 6.1 for the sake of completeness.

We need to add a few steps to their method. After choosing vertices v_i and v_j with the above probabilities, toss a biased coin that comes head with probability $1 - d_i d_j / 4m$. Accept the pair $\{v_i, v_j\}$ if the coin shows head, $i \neq j$, and $\{v_i, v_j\} \notin E$. Add $\{v_i, v_j\}$ to E and reduce each of $\hat{d}_i$, $\hat{d}_j$ by 1. Otherwise reject the pair $\{v_i, v_j\}$ and repeat. The expected number of repeats is bounded by a constant because $d_{\max} = O(m^{1/4-\tau})$ and therefore $1 - d_i d_j / 4m > 1/2$.

Efficient calculation of P is also straightforward. Note that

$$p_{ij} = \frac{(1 - d_i d_j / 4m) d_i^{(r)} d_j^{(r)}}{\binom{2m-2r}{2} - \Psi_r(\mathcal{N})}.$$

Therefore, p_{ij} can be easily calculated from $\binom{2m-2r}{2} - \Psi_r(\mathcal{N})$. At the beginning of Procedure A we have

$$\binom{2m}{2} - \Psi(\mathcal{N}_0) = \binom{2m}{2} - \sum_u \binom{d_u}{2} - \frac{(\sum_u d_u^2)^2 - \sum_u d_u^4}{8m}$$

which can be calculated with $O(n)$ operations. Now we show that in step $r+1$, p_{ij} can be updated from step r with $O(d_{\max})$ operations. This is because by choosing a pair $\{v_i, v_j\}$ at step $r+1$:

$$\begin{aligned}
&\left[\binom{2m-2r-2}{2} - \Psi_{r+1}(\mathcal{N})\right] - \left[\binom{2m-2r}{2} - \Psi_r(\mathcal{N})\right] \\
&\quad = \sum_{(v_a,v_b)\in E_{r+1}} d_a^{(r+1)} d_b^{(r+1)}\left(1-\frac{d_a d_b}{4m}\right) - \sum_{(v_a,v_b)\in E_r} d_a^{(r)} d_b^{(r)}\left(1-\frac{d_a d_b}{4m}\right) \\
&\quad = -d_i^{(r)} d_j^{(r)}\left(1-\frac{d_i d_j}{4m}\right) - \sum_{(v_{i'},v_i)\in E_r} d_{i'}^{(r)}\left(1-\frac{d_i d_{i'}}{4m}\right) \\
&\qquad - \sum_{(v_{j'},v_j)\in E_r} d_{j'}^{(r)}\left(1-\frac{d_j d_{j'}}{4m}\right) \\
&\quad = -d_i^{(r)} d_j^{(r)}\left(1-\frac{d_i d_j}{4m}\right) + \Xi_{i,r} + \Xi_{j,r} + \frac{d_i + d_j}{4m}\Omega_r + O_{i,r} + O_{j,r}
\end{aligned}$$

where $\Xi_{i,r} = \sum_{v_{i'}\sim_{G_{\mathcal{N}_r}} v_i} d_{i'}^{(r)}(1 - \frac{d_i d_{i'}}{4m})$, $\Omega_r = \sum_{i'=1}^{n} d_{i'}^{r} d_{i'}$, and $O_{i,r} = d_i^{(r)}(1 - d_i^2/4m) - (2m-2r)$. It is clear to see from $\Omega_{r+1} - \Omega_r = -d_i - d_j$ that Ω_r can be updated at each step by only one operation, and the calculation of $O_{i,r}$, $O_{j,r}$ takes constant time. Moreover, each of $\Xi_{i,r}$, $\Xi_{j,r}$ is a summation with at most $d_{\max}$ terms. We will show in the next section that it is possible to find neighbors of v_i and v_j in $G_{\mathcal{N}_r}$ with $O(d_{\max})$ operations. Therefore $\Xi_{i,r}$, $\Xi_{j,r}$ can be calculated with $O(d_{\max})$ operations. Thus the running time of the new implementation of Procedure A is $O(md_{\max})$ for general degree sequences. Now using Lemma 2, the running time of Procedure A is of $O(md_{\max})$. □

6.1 Steger and Wormald's Method for Choosing a Suitable Pair

Steger and Wormald's (SW) [37] implementation has three phases and uses the configuration model.

In the first phase, the algorithm puts all of the mini-vertices in an array L where all of the matched mini-vertices are kept in the front. It is also assumed that the members of each pair of matched mini-vertices will be two consecutive elements of L. There is another array I that keeps location of each mini-vertex inside array L. Then two elements of L (selected uniformly at random) can be checked for suitability in time $O(d_{\max})$. This is because from I we can find the neighbors of the selected elements in the partially constructed graph $G_{\mathcal{N}_r}$. Note that in our modification (Procedure A), the pair is accepted with probability $1 - d_i d_j/4m$ when they belong to W_i, W_j. This also completes the above argument for updating $\Psi_r(\mathcal{N})$ with $O(d_{\max})$ operations. Repeat the above till a suitable pair is found then update L and I.

Phase 1 ends when the number of remaining mini-vertices falls below $2d_{\max}^2$. Hence using Corollary 3.1 in [37], throughout phase 1 the number of suitable pairs is more than half of the total number of available pairs. Therefore, the expected number of repetitions in the above process is at most 2. This means the expected running time of phase 1 is $O(md_{\max})$.

Phase 2 starts when the number of available mini-vertices is less than $2d_{\max}^2$ and finishes when the number of available vertices is at least $2d_{\max}$. In this phase instead of choosing the mini-vertices, choose a pair of vertices of $G_{\mathcal{N}_r}$ (two random set W_i, W_j in the configuration model) from the set of vertices that are not fully matched. Repeat the above till v_i, v_j is not already an edge in $G_{\mathcal{N}_r}$. Again the expected number of repetitions is at most 2. Now randomly choose one mini-vertex in each selected W_i. If both of the mini-vertices are not matched yet add the edge, otherwise pick another two mini-vertices. The expected number of repetitions here is at most $O(d_{\max}^2)$ and hence the expected running time of the phase 2 is at most $O(d_{\max}^4)$.

Phase 3 starts when the number of available vertices (not fully matched W_i's) is less than $2d_{\max}$. We can construct a graph H, in time $O(d_{\max}^2)$, that indicates the set of all possible connections. Now choose an edge $\{v_i, v_j\}$ of H uniformly at random and accept it with probability $\hat{d}_i\hat{d}_j/d_{\max}^2$. Again, the expected number of repetitions will be at most $O(d_{\max}^2)$. Update H in constant time and repeat the above till H is empty. Therefore the expected running time of phase 3 is also $O(d_{\max}^4)$.

Hence, the total running time for $d_{\max} = O(m^{1/4-\tau})$ will be $O(md_{\max})$.

7 Generalizing Kim and Vu's Analysis

The aim of this section is to show (3) via generalization of Kim and Vu's analysis [26]. Let us define

$$f(\mathcal{N}) = \prod_{r=0}^{m-1} \frac{\binom{2m-2r}{2} - \psi_r(G)}{\binom{2m-2r}{2} - \Psi_r(\mathcal{N})}$$

then (3) is equivalent to

$$\mathbb{E}(f(\mathcal{N})) = 1 + o(1) \tag{15}$$

where the expectation is with respect to the uniform distribution on the set $S(\mathcal{M})$ of all $m!$ orderings of the matching $\mathcal{M}$. Proof of (15) is done by partitioning the set $S(\mathcal{M})$ into smaller subsets and looking at the deviation of f on each set separately. The partition is explained in Sect. 7.3. But before that we need to define some notation.

7.1 Definitions

In Sect. 4 we saw that the probability of choosing an edge between W_i and W_j at step $r+1$ of Procedure A is equal to $(1 - \frac{d_i d_j}{4m})[\binom{2m-2r}{2} - \Psi_r(\mathcal{N})]^{-1}$ where

$$\Psi_r(\mathcal{N}) = \sum_{\{v_i,v_j\}\notin E_r} d_i^{(r)} d_j^{(r)} + \sum_{\{v_i,v_j\}\in E_r} d_i^{(r)} d_j^{(r)} \frac{d_i d_j}{4m}.$$

To simplify the notation, throughout the rest of this section, we will use Ψ_r and Δ_r to denote $\Psi_r(\mathcal{N})$ and $\Delta_r(\mathcal{N})$ respectively. We will also use the notation $\{v_i, v_j\}$ and $\{i, j\}$ interchangeably. Moreover, the notation $\{i, j\}$ includes the cases of $i = j$ as well.

For our analysis we need to write $\Psi_r = \Delta_r + \Lambda_r$ where

$$\Delta_r = \binom{2m-2r}{2} - \sum_{\{i,j\} \in E_r} d_i^{(r)} d_j^{(r)},$$

$$\Lambda_r = \sum_{\substack{\{i,j\} \\ i \neq j}} d_i^{(r)} d_j^{(r)} \frac{d_i d_j}{4m} - \sum_{\substack{\{i,j\} \notin E_r \\ i \neq j}} d_i^{(r)} d_j^{(r)} \frac{d_i d_j}{4m}.$$

Notice that Δ_r counts the number of possibilities for creating a self loop ($i = j$) or making double edges. We distinguish between these two cases by an extra index. That is

$$\Delta_r^{(1)} = \sum_{i=1}^{n} \binom{d_i^{(r)}}{2} = \text{\# of self loops}, \quad \text{and}$$

$$\Delta_r^{(2)} = \Delta_r - \Delta_r^{(1)} = \text{\# of double edges}.$$

Note that since all the existing pairs are suitable, the only type of multiple pairs that can be created at step $r + 1$ are double pairs. Moreover,

$$\begin{aligned} 4m\Lambda_r &= \sum_{\substack{\{i,j\} \\ i \neq j}} d_i^{(r)} d_j^{(r)} d_i d_j - \sum_{\substack{\{i,j\} \notin E_r \\ i \neq j}} d_i^{(r)} d_j^{(r)} d_i d_j \\ &= \frac{(\sum_{i=1}^{n} d_i^{(r)} d_i)^2 - \sum_{i=1}^{n} (d_i^{(r)})^2 d_i^2}{2} - \sum_{\substack{\{i,j\} \notin E_r \\ i \neq j}} d_i^{(r)} d_j^{(r)} d_i d_j. \end{aligned}$$

We distinguish between these three summations by adding a numerical index to Λ_r; i.e.

$$\Lambda_r^{(1)} = \sum_{i=1}^{n} d_i^{(r)} d_i, \qquad \Lambda_r^{(2)} = \sum_{i=1}^{n} (d_i^{(r)})^2 d_i^2, \qquad \Lambda_r^{(3)} = \sum_{\substack{\{i,j\} \notin E_r \\ i \neq j}} d_i^{(r)} d_j^{(r)} d_i d_j.$$

Hence,

$$\Lambda_r = \frac{(\Lambda_r^{(1)})^2 - \Lambda_r^{(2)}}{8m} - \frac{\Lambda_r^{(3)}}{4m}.$$

The following simple bounds will be very useful throughout Sect. 7.

Lemma 11 *For all r the following equations hold*:

(i) $\Delta_r \leq \frac{(2m-2r)d_{\max}^2}{2}$

(ii) $\Lambda_r^{(1)} \le d_{\max}(2m-2r)$
(iii) $\Lambda_r \le \frac{(2m-2r)^2 d_{\max}^2}{8m}$.

Proof (i) At step r there are $2m-2r$ mini-vertices left and for each $u \in W_i$ there are at most $d_{\max}-1$ mini-vertices in W_i that u can connect to. Hence, $\Delta_r^{(1)} \le \frac{(2m-2r)(d_{\max}-1)}{2}$. Similarly u can connect to at most $(d_{\max}-1)^2$ mini-vertices in some W_j with $i \ne j$ to create a double edge. Thus, $\Delta_r^{(2)} \le \frac{(2m-2r)(d_{\max}-1)^2}{2}$. Now using $\Delta_r = \Delta_r^{(1)} + \Delta_r^{(2)}$ the proof of (i) is clear.

(ii) $\Lambda_r^{(1)} \le d_{\max} \sum_u d_u^{(r)} = d_{\max}(2m-2r)$.

(iii) It follows from the definition of Λ_r that

$$\Lambda_r = \sum_{\{i,j\}\in E_r} d_i^{(r)} d_j^{(r)} \frac{d_i d_j}{4m} \le \frac{d_{\max}^2}{4m} \sum_{\{i,j\}\in E_r} d_i^{(r)} d_j^{(r)} \le \frac{d_{\max}^2}{4m} \binom{2m-2r}{2}.$$ □

In order to define ψ_r we look at a slightly similar model. Recall that $G_{\mathcal{N}_r}$ is the partial graph that is constructed up to step r. Imposing the uniform distribution on $S(\mathcal{M})$, graph $G_{\mathcal{N}_r}$ turns to a random subgraph of G that has exactly r edges. We can approximate this graph by a different random subgraph of G. This is done, by selecting each edge of G independently with probability $p_r = r/m$ and denoting the resulted graph by G_{p_r}. Now using G_{p_r} as an approximation to $G_{\mathcal{N}_r}$, we are ready evaluate quantities $\mathbb{E}_{p_r}(\Delta_r^{(1)})$, $\mathbb{E}_{p_r}(\Delta_r^{(2)})$, $\mathbb{E}_{p_r}(\Lambda_r^{(1)})$, $\mathbb{E}_{p_r}(\Lambda_r^{(2)})$, and $\mathbb{E}_{p_r}(\Lambda_r^{(3)})$. Throughout this section we often use the notations $\Delta_{p_r}^{(i)}$, $\Lambda_{p_r}^{(i)}$, and Ψ_{p_r} to emphasis that the model is G_{p_r} instead of $G_{\mathcal{N}_r}$.

Lemma 12 *For each r the following equations hold*:

$$\text{(i)}\quad \mathbb{E}_{p_r}(\Delta_r^{(1)}) = \frac{(2m-2r)^2}{2}\left(\frac{\sum_{i=1}^n \binom{d_i}{2}}{2m^2}\right) = \frac{(2m-2r)^2}{2}\left(\frac{\lambda(\bar{d})}{m}\right)$$

$$\text{(ii)}\quad \mathbb{E}_{p_r}(\Delta_r^{(2)}) = \frac{(2m-2r)^2}{2}\left(\frac{r\sum_{i\sim_G j}(d_i-1)(d_j-1)}{2m^3}\right)$$

$$\text{(iii)}\quad \mathbb{E}_{p_r}(\Lambda_r^{(1)}) = (2m-2r)\frac{\sum_{i=1}^n d_i^2}{2m}$$

$$\text{(iv)}\quad \mathbb{E}_{p_r}(\Lambda_r^{(2)}) = (2m-2r)^2\frac{\sum_{i=1}^n d_i^4}{4m^2} + 2r(2m-2r)\frac{\sum_{i=1}^n d_i^3}{4m^2}$$

$$\text{(v)}\quad \mathbb{E}_{p_r}(\Lambda_r^{(3)}) = \frac{(2m-2r)^2}{2}\left(\frac{r\sum_{i\sim_G j} d_i d_j (d_i-1)(d_j-1)}{2m^3}\right).$$

Proof (i) In the random model of G_{p_r} each edge has a probability of $\frac{r}{m}$ to be chosen. Let X_i the number of unsuitable edges that connect two mini-vertices of W_i at

$(r+1)$th step of creating $\mathcal{N}$. Hence, X_i is equal to the number of unordered tuples $\{j,i,k\}$ where $\{j,i\},\{i,k\} \in E(G) \setminus E(G_{\mathcal{N}_r})$ which gives

$$\Delta_r^{(1)} = \sum_{i=1}^{n} X_i. \tag{16}$$

On the other hand for a fixed i, the number of tuples $\{j,i,k\}$ where $\{j,i\},\{i,k\} \in E(G)$ is exactly $\binom{d_i}{2}$, and with probability $(1-\frac{r}{m})^2$ the edges $\{j,i\},\{i,k\}$ do not belong to $E(G_{\mathcal{N}_r})$. Thus, the equality $\mathbb{E}(X_i) = (1-\frac{r}{m})^2\binom{d_i}{2}$ holds and it can be used in (16) to complete the proof of (i).

(ii) Define Y_{ij} to be the number of unsuitable edges between W_i and W_j at $(r+1)$th step of creating $\mathcal{N}$. It is not hard to see that Y_{ij} also counts the number of unordered tuples $\{k,i,j,l\}$ where $\{i,j\} \in E(G_{\mathcal{N}_r})$ but $\{k,i\},\{j,l\} \in E(G) \setminus E(G_{\mathcal{N}_r})$. Hence,

$$\Delta_r^{(2)} = \sum_{i \sim_G j} Y_{ij}. \tag{17}$$

On the other hand for a fixed $i \sim_G j$, the number of tuples $\{k,i,j,l\}$ where $\{k,i\},\{j,l\}$ belong to $E(G)$ is exactly $(d_i-1)(d_j-1)$. Moreover, the edges $\{k,i\},\{j,l\}$ do not belong to $E(G_{\mathcal{N}_r})$ with probability $(1-\frac{r}{m})^2$, and the edge $\{i,j\}$ belongs to $E(G_{\mathcal{N}_r})$ with probability $\frac{r}{m}$. This gives the equality $\mathbb{E}(Y_{ij}) = \frac{r}{m}(1-\frac{r}{m})^2(d_i-1)(d_j-1)$ which can be used with (17) to complete the proof of (ii).

(iii) The proof directly follows from $\mathbb{E}(d_i^{(r)}) = (1-\frac{r}{m})d_i$.

(iv) Since each $d_i^{(r)}$ is a summation of d_i Bernoulli i.i.d. random variables. We can show

$$\mathbb{E}[(d_i^{(r)})^2] = \left(1-\frac{r}{m}\right)^2 d_i^2 + \frac{r}{m}\left(1-\frac{r}{m}\right)d_i$$

which proves (iv).

(v) The proof is similar to (ii), except we are using the following instead of (17)

$$\Lambda_r^{(3)} = \sum_{i \sim_G j} \frac{d_i d_j}{4m} Y_{ij}.$$

□

The next step is to define ψ_r as an approximation to $\mathbb{E}_{p_r}(\Psi_r)$. For that we will use Lemma 12 and the following two estimates

$$\mathbb{E}_{p_r}\left(\frac{\Lambda_r^{(2)}}{8m}\right) = \frac{(2m-2r)^2}{2}\left[O\left(\frac{d_{\max}^3}{m^2}\right) + O\left(\frac{d_{\max}^2}{m^2}\frac{2r}{2m-2r}\right)\right],$$

$$\mathbb{E}_{p_r}\left(\frac{\Lambda_r^{(3)}}{4m}\right) = \frac{r(2m-2r)^2}{2} O\left(\frac{d_{\max}^4}{m^3}\right).$$

Note that here we used the bound

$$\sum_{i=1}^{n} d_i^s = \sum_{i \sim_G j} (d_i^{s-1} + d_j^{s-1}) = O(m d_{\max}^{s-1})$$

that will be repeatedly used in this section.

Now $\mathbb{E}_{p_r}(\Psi_r)$ is given by the following expression

$$\mathbb{E}_{p_r}(\Psi_r) = \frac{(2m-2r)^2}{2}\Bigg[\frac{\lambda(\bar{d})}{m} + \frac{r\sum_{i\sim_G j}(d_i-1)(d_j-1)}{2m^3} + \frac{(\sum_{i=1}^{n} d_i^2)^2}{16m^3} + O\left(\frac{r d_{\max}^4}{m^3} + \frac{r d_{\max}^2}{(m-r)m^2}\right)\Bigg]. \tag{18}$$

Definition 2 The expected value of Ψ_r is denoted by ψ_r. i.e. $\psi_r = \mathbb{E}_{p_r}(\Psi_r)$.

The following lemma is equivalent to (18).

Lemma 13 *For all r,*

$$\psi_r = \frac{(2m-2r)^2}{2}\left(\frac{\lambda(\bar{d})}{m} + \frac{r\sum_{i\sim_G j}(d_i-1)(d_j-1)}{2m^3} + \frac{(\sum_{i=1}^{n} d_i^2)^2}{16m^3} + \varsigma_r\right)$$

where $\varsigma_r = O(\frac{r d_{\max}^4}{m^3} + \frac{r d_{\max}^2}{(m-r)m^2})$.

It is also straightforward to show that the following upper bound holds for ψ_r.

Lemma 14 *For all r the quantity ψ_r is bounded above by $O(\frac{d_{\max}^2(2m-2r)^2}{2m})$.*

Now we are ready to prove (4).

7.2 Algebraic Proof of (4)

For simplicity, we define χ_G to be $\sum_{i\sim_G j}(d_i-1)(d_j-1)$. Therefore,

$$\begin{aligned}
&\prod_{r=0}^{m-1} \frac{\binom{2m-2r}{2}}{\binom{2m-2r}{2} - \psi_r} \\
&= \prod_{r=0}^{m-1}\left(1 + \frac{\psi_r}{\binom{2m-2r}{2} - \psi_r}\right) \\
&= \prod_{r=0}^{m-1}\left(1 + \frac{\frac{\lambda(\bar{d})}{m} + \frac{r\sum_{i\sim_G j}(d_i-1)(d_j-1)}{2m^3} + \frac{(\sum_i d_i^2)^2}{16m^3} + \varsigma_r}{1 - \frac{1}{2m-2r} - O(\frac{d_{\max}^2}{m})}\right)
\end{aligned}$$

$$= \exp\left[\sum_{r=0}^{m-1} \log\left(1 + \frac{\frac{\lambda(\bar{d})}{m} + \frac{r\chi_G}{2m^3} + \frac{(\sum_i d_i^2)^2}{16m^3} + \varsigma_r}{1 - \frac{1}{2m-2r} - O(\frac{d_{\max}^2}{m})}\right)\right]$$

$$= \exp\left[\sum_{r=0}^{m-1} \log\left(1 + \frac{\lambda(\bar{d})}{m} + \frac{r\chi_G}{2m^3} + \frac{(\sum_i d_i^2)^2}{16m^3} + O\left(\frac{d_{\max}^4}{m^2} + \frac{r d_{\max}^2}{(m-r)m^2}\right)\right)\right]$$

$$= \exp\left[\sum_{r=0}^{m-1} \left(\frac{\lambda(\bar{d})}{m} + \frac{r\chi_G}{2m^3} + \frac{(\sum_i d_i^2)^2}{16m^3} + O\left(\frac{d_{\max}^4}{m^2} + \frac{r d_{\max}^2}{(m-r)m^2}\right)\right)\right] \quad (19)$$

$$= \exp\left[\lambda(\bar{d}) + \frac{m(m-1)\chi_G}{4m^3} + \frac{(\sum_i d_i^2)^2}{16m^2} + O\left(\frac{d_{\max}^4}{m} + \frac{d_{\max}^2}{m}\log(2m)\right)\right]$$

$$= \exp\left[\lambda(\bar{d}) + \frac{\chi_G}{4m} + \frac{(\sum_i d_i^2)^2}{16m^2} + o(1)\right]$$

$$= \exp\left[\lambda(\bar{d}) + \frac{\sum_{i\sim_G j} d_i d_j}{4m} - \frac{\sum_{i\sim_G j}(d_i + d_j)}{4m} + \frac{1}{4} + \frac{(\sum_i d_i^2)^2}{16m^2} + o(1)\right] \quad (20)$$

$$= \big(1 + o(1)\big)\exp\left[\lambda(\bar{d}) + \lambda^2(\bar{d}) + \frac{\sum_{i\sim_G j} d_i d_j}{4m}\right] \quad (21)$$

where (19) uses $\log(1+x) = x - O(x^2)$ and (20) uses $d_{\max} = O(m^{1/4-\tau})$. The bound $\frac{\psi_r}{(2m-2r)^2} = O(\frac{d_{\max}^2}{m})$ was used a few times as well. □

7.3 Partitioning the Set of Orderings $S(\mathcal{M})$

In order to prove (15), we need to study the large deviation behavior of function f on the set $S(\mathcal{M})$. For that we partition the set $S(\mathcal{M})$ in four "major" steps. At each step, one subset of $S(\mathcal{M})$ will be removed from it.

Step 1. Consider those orderings $\mathcal{N} \in S(\mathcal{M})$ where at any state during the algorithm, the number of unsuitable edges does not exceed a constant (strictly less than 1) fraction of the number of all available edges. More specifically, for a small number $0 < \tau \le 1/3$ let

$$S^*(\mathcal{M}) = \left\{\mathcal{N} \in S(\mathcal{M}) \mid \Psi_r(\mathcal{N}) \le (1 - \tau/4)\binom{2m-2r}{2} : \forall 0 \le r \le m-1\right\}.$$

Then the first element of the partition will be $S(\mathcal{M}) \setminus S^*(\mathcal{M})$.

Step 2. Consider those orderings $\mathcal{N}$ from the set $S^*(\mathcal{M})$ for which $\Psi_r(\mathcal{N}) - \psi_r > T_r(\log n)^{1+\delta}$ for all $0 \le r \le m-1$. The function T_r will be defined in Sect. 7.4 and δ is a small positive constant. For example $\delta < 0.1$ works. Denote the set of all such $\mathcal{N}$ by $\mathcal{A}$.

Step 3. From the set $S^*(\mathcal{M}) \setminus \mathcal{A}$, remove those elements with $\Psi_r(\mathcal{N}) > 0$ for some r with $2m - 2r \le (\log n)^{1+2\delta}$. Put these elements in the set $\mathcal{B}$.

Step 4. The last element of the partition is the remaining subset $\mathcal{C} = S^*(\mathcal{M}) \setminus (\mathcal{A} \cup \mathcal{B})$.

The journey towards proving (3) is divided into these five parts

$$\mathbb{E}(f(\mathcal{N})1_{\mathcal{A}}) = o(1), \tag{22}$$

$$\mathbb{E}(f(\mathcal{N})1_{\mathcal{B}}) = o(1), \tag{23}$$

$$\mathbb{E}(f(\mathcal{N})1_{\mathcal{C}}) \leq 1 + o(1), \tag{24}$$

$$\mathbb{E}(f(\mathcal{N})1_{\mathcal{C}}) \geq 1 - o(1), \tag{25}$$

$$\mathbb{E}(f(\mathcal{N})1_{S(\mathcal{M})\setminus S^*(\mathcal{M})}) = o(1). \tag{26}$$

These parts will be all proved is Sect. 7.5. The hardest of these proofs is for (22) which is carried out by partitioning the set $\mathcal{A}$ into further subsets and using Vu's inequality on them. The remaining proofs for (23)–(26) are based on the standard combinatorial and algebraic bounds.

7.4 More Notation

In order to prove (22) we need more notation. Remember from Sect. 7.3 that $\delta > 0$ is a very small constant. Let $\omega = (\log n)^{\delta}$. Let $\lambda_0 = \omega \log n$ and $\lambda_i = 2^i \lambda_0$ for $i = 1, 2, \ldots, L$. L is such that $\lambda_{L-1} < cd_{\max}^2 \log n \leq \lambda_L$ where c is a large constant that is specified later.

Definition 3 Let $q_r = (1 - r/m)$, $p_r = 1 - q_r$ $\forall 0 \leq r \leq m-1$. Then let

$$\beta_r(\lambda) = c\sqrt{\lambda(md_{\max}^2 q_r^2 + \lambda^2)(d_{\max}^2 q_r + \lambda)},$$

$$\gamma_r(\lambda) = c\sqrt{\lambda(md_{\max}^2 q_r^3 + \lambda^3)(d_{\max}^2 q_r^2 + \lambda^2)},$$

$$\nu_r = 8md_{\max}^2 q_r^3.$$

Now the function T_r for all $0 \leq r \leq m-1$ is defined by

$$T_r(\lambda) = \begin{cases} 3\beta_r(\lambda) + 2\min(\gamma_r(\lambda), \nu_r) & \text{if } 2m - 2r \geq \omega\lambda, \\ \lambda^2/\omega & \text{otherwise.} \end{cases}$$

The intuition behind this definition will become clear when we use Vu's concentration inequality in Sect. 7.5.2. Note that inequalities $\alpha_r(\lambda) \leq \beta_r(\lambda)$ and $\zeta_r(\lambda) \leq \beta_r(\lambda)$ hold and we will use them in Sect. 7.5 to simplify the computations. Moreover, with the above definition, since $\lambda_i = 2\lambda_{i-1}$, the following relation holds between $T_r(\lambda_i)$, $T_r(\lambda_{i-1})$.

$$T_r(\lambda_i) \leq 8T_r(\lambda_{i-1}). \tag{27}$$

Now we will subpartition $\mathcal{A}$ and $\mathcal{B}$. Define subsets $A_0 \subseteq A_1 \subseteq \cdots \subseteq A_L \subseteq S^*(\mathcal{M})$ by

$$A_i = \{\mathcal{N} \in S^*(\mathcal{M}) \mid \Psi_r(\mathcal{N}) - \psi_r < T_r(\lambda_i),\ \forall 0 \leq r \leq m-1\}.$$

Moreover, define A_∞ by $A_\infty = S^*(\mathcal{M}) \setminus \bigcup_{i=0}^{L} A_i$. Then we have

$$\mathcal{A} = A_\infty \cup \left(\bigcup_{i=1}^{L} A_i \setminus A_{i-1} \right).$$

Since the main objective of partitioning $\mathcal{A}$ is to prove (22), we are only interested in finding upper bounds for $f(\mathcal{N}) = \prod_{r=0}^{m-1}(1 + \frac{\Psi_r(\mathcal{N})-\psi_r}{\binom{2m-2r}{2}-\Psi_r(\mathcal{N})})$. Therefore, the cases with $\Psi_r(\mathcal{N}) \le \psi_r$ are not troublesome.

Let K be an integer such that $2^{K-1} < (\log n)^{2+\delta} + 1 \le 2^K$. Next step is to consider a chain of subsets $B_0 \subseteq B_1 \subseteq \cdots \subseteq B_K \subseteq A_0$ that are defined by

$$B_j = \{\mathcal{N} \in A_0 \mid \Psi_r(\mathcal{N}) < 2^j,\ \forall r \ge (2m - \omega\lambda_0)/2\}.$$

It is not hard to see that the set $\mathcal{C}$ that was defined in step 4 in Sect. 7.3 is equal to the set B_0. Note that T_r's are chosen such that for all $r \ge (2m - \omega\lambda_0)/2$ we have $T_r(\lambda_0) = \lambda_0 \log n$ and by Lemma 14, for all $r \ge (2m - \omega\lambda_0)/2$ we have $\psi_r = o(1)$. Thus, for all such r and all elements of A_0,

$$\Psi_r < \lambda_0 \log n + \psi_r < 2^K.$$

This shows that $A_0 = (\bigcup_{j=0}^{K} B_j) \cup \mathcal{C}$ and also $\mathcal{B} = \bigcup_{j=1}^{K} B_j$.

7.5 Proofs of (22), (23) and (24)

In this section we will bound the expected value $\mathbb{E}(f(\mathcal{N}))$ on the sets A_∞, $\mathcal{C}$, and on each of the sets of the form $A_i \setminus A_{i-1}$ and $B_j \setminus B_{j-1}$.

Lemma 15 *For all* $1 \le i \le L$:

(a) $\mathbb{P}(A_i \setminus A_{i-1}) \le e^{-\Omega(\lambda_i)}$.
(b) *For all* $\mathcal{N}$ *in* $A_i \setminus A_{i-1}$ *we have* $f(\mathcal{N}) \le e^{o(\lambda_i)}$.

Lemma 16 *For a large enough constant* c:

(a) $\mathbb{P}(A_\infty) \le e^{-cd_{\max}^2 \log n}$.
(b) *For all* $\mathcal{N}$ *in* A_∞ *we have* $f(\mathcal{N}) \le e^{4d_{\max}^2 \log n}$.

Lemma 17 *For all* $1 \le j \le K$:

(a) $\mathbb{P}(B_j \setminus B_{j-1}) \le e^{-\Omega(2^{j/2} \log n)}$
(b) *For all* $\mathcal{N}$ *in* $B_j \setminus B_{j-1}$ *we have* $f(\mathcal{N}) \le e^{O(2^{3j/4})}$.

Lemma 18 *For all* $\mathcal{N} \in \mathcal{C}$ *we have* $f(\mathcal{N}) \le 1 + o(1)$.

Now it is easy to see that (22) follows from Lemmas 15 and 16. Note that by the definition of K we have $2^{K/4} \ll \log n$ which gives $2^{3j/4} \ll 2^{j/2} \log n$. Thus, we can deduce (23) from Lemma 17. Finally, (24) is consequence of Lemma 18.

Proof of Lemma 15 uses Vu's concentration inequality but for the other three lemmas, typical algebraic and combinatorial bounds are sufficient. Throughout the rest of this section we present a quick introduction to Vu's concentration inequality. Then we prove the above lemmas.

7.5.1 *Vu's Concentration Inequality*

Proofs of Lemmas 15(a) and 16(a) use a very strong concentration inequality proved by Vu [39] which is a generalized version of an earlier result by Kim and Vu [25]. Consider independent random variables $t_1, t_2, \dots, t_n$ with arbitrary distribution in $[0, 1]$. Let $Y(t_1, t_2, \dots, t_n)$ be a polynomial of degree k and coefficients in $(0, 1]$. For any multi-set A of elements $t_1, t_2, \dots, t_n$ let $\partial_A Y$ denote the partial derivative of Y with respect to variables in A. For example if $Y = t_1 + t_1^3 t_2^2$ and $A = \{t_1, t_1\}$, $B = \{t_1, t_2\}$ then

$$\partial_A Y = \frac{\partial^2}{\partial t_1^2} Y = 6 t_1 t_2^2, \qquad \partial_B Y = \frac{\partial^2}{\partial t_1 \partial t_2} Y = 6 t_1^2 t_2.$$

For all $0 \le j \le k$, let $\mathbb{E}_j(Y) = \max_{|A| \ge j} \mathbb{E}(\partial_A Y)$. Define parameters c_k, d_k recursively as follows: $c_1 = 1$, $d_1 = 2$, $c_k = 2k^{1/2}(c_{k-1} + 1)$, $d_k = 2(d_{k-1} + 1)$.

Theorem 4 (Vu) *Take a polynomial Y as defined above. For any collection of positive numbers $\mathcal{E}_0 > \mathcal{E}_1 > \cdots > \mathcal{E}_k = 1$ and λ satisfying*:

(a) $\mathcal{E}_j \ge \mathbb{E}_j(Y)$, *and*
(b) $\mathcal{E}_j / \mathcal{E}_{j+1} \ge \lambda + 4j \log n,\ 0 \le j \le k - 1$

the following is true

$$\mathbb{P}\left(|Y - \mathbb{E}(Y)| \ge c_k \sqrt{\lambda \mathcal{E}_0 \mathcal{E}_1}\right) \le d_k e^{-\lambda/4}.$$

7.5.2 *Proof of part (a) of Lemmas 15 and 16*

In order to show part (a) of Lemma 15 we prove the stronger property

$$\mathbb{P}(A_{i-1}^c) \le e^{-\Omega(\lambda_i)}. \tag{28}$$

This property combined with $\lambda_L \ge c d_{\max}^2 \log n$ proves part (a) of Lemma 16 as well. From (27) we have

$$A_{i-1}^c \subseteq \left\{ \Psi_r - \psi_r \ge \frac{T_r(\lambda_i)}{8} \right\}.$$

Hence, in order to show (28) it is sufficient to show the following two lemmas.

Lemma 19 *For all r such that $2m - 2r \ge \omega \lambda_i$*:

$$\mathbb{P}\left(|\Psi_r(\mathcal{N}) - \psi_r| \ge \frac{3\beta_r(\lambda_i) + 2\min(\gamma_r(\lambda_i), \nu_r)}{8}\right) \le e^{-\Omega(\lambda_i)}.$$

Lemma 20 *For any r such that $2m - 2r < \omega\lambda_i$ we have*

$$\mathbb{P}(\Psi_r(\mathcal{N}) - \psi_r \geq \lambda_i^2/\omega) \leq e^{-\Omega(\lambda_i)}.$$

Now we focus on Lemma 19. For each variable Δ_r, Λ_r, Ψ_r denote their analogues quantity in G_{p_r} by Δ_{p_r}, Λ_{p_r}, Ψ_{p_r}.

Lemma 21 *For all r we have $\mathbb{P}_{p_r}(\{|E(G_{p_r})| = r\}) \geq \frac{1}{n}$.*

Proof let $f(m,r) = \mathbb{P}_{p_r}(\{|E(G_{p_r})| = r\})$ then it can be seen that

$$\frac{f(m,r+1)}{f(m,r)} = \frac{(1+1/r)^r}{(1+\frac{1}{m-r-1})^{m-r}} \leq 1 \quad \forall r \leq (m-1)/2.$$

Hence, the minimum of $f(m,r)$ is around $r = m/2$. Using Stirling's approximation we can get $f(m,r) \geq \frac{1}{\sqrt{2m}} \geq \frac{1}{n}$. □

By Lemma 21, with probability at least $1/n$, G_{p_r} has exactly r edges. Hence, using $\lambda_i \gg \log n$, for proving Lemma 19 we only need to show

$$\mathbb{P}\left(|\Psi_{p_r} - \psi_r| \geq \frac{3\beta_r(\lambda_i) + 2\min(\gamma_r(\lambda_i), \nu_r)}{8}\right) \leq e^{-\Omega(\lambda_i)}. \tag{29}$$

In order to prove (29) we define

$$\alpha_r(\lambda) = c\sqrt{\lambda(md_{\max}q_r^2 + \lambda^2)(d_{\max}q_r + \lambda)},$$

$$\zeta_r(\lambda) = c\frac{d_{\max}^2}{m}\sqrt{\lambda(md_{\max}q_r^2 + \lambda^2)(q + \lambda)}.$$

It is flashforward that $\alpha_r(\lambda),\ \zeta_r(\lambda) \leq \beta_r(\lambda)$. Therefore, (29) is the result of the following lemma. Throughout the rest of the proof we fix r, i and remove all sub-indices r, i for simplicity.

Lemma 22 *For all p we have*:

(i) $\mathbb{P}(|\Delta_p^{(1)} - \mathbb{E}(\Delta_p^{(1)})| \geq \frac{\alpha}{8}) \leq e^{-\Omega(\lambda)}$
(ii) $\mathbb{P}(|\Delta_p^{(2)} - \mathbb{E}(\Delta_p^{(2)})| \geq \frac{\min(\beta+\gamma,\beta+\nu)}{8}) \leq e^{-\Omega(\lambda)}$
(iii) $\mathbb{P}(|\frac{(\Lambda_p^{(1)})^2 - \Lambda_p^{(2)}}{8m} - \frac{\mathbb{E}(\Lambda_p^{(1)})^2 - \mathbb{E}(\Lambda_p^{(2)})}{8m}| \geq \frac{\zeta}{8}) \leq e^{-\Omega(\lambda)}$
(iv) $\mathbb{P}(|\frac{\Lambda_p^{(3)}}{4m} - \frac{\mathbb{E}(\Lambda_p^{(3)})}{4m}| \geq \frac{\min(\beta+\gamma,\beta+\nu)}{8}) \leq e^{-\Omega(\lambda)}$.

Proof (i) Similar to Kim and Vu's proof, for each edge e of G consider a random variable t_e which is equal to 0 when e is present in G_p and 1 otherwise. These t_e's will be i.i.d. Bernoulli with mean q. Now note that

$$\Delta_p^{(1)} = \sum_u \sum_{u \in e \cap f,\ e \neq f} t_e t_f$$

and

$$\mathbb{E}(\varDelta_p^{(1)}) = \sum_u \binom{d_u}{2} q^2 \le m d_{\max} q^2.$$

For each t_e we have

$$\mathbb{E}(\partial_{t_e} \varDelta_p^{(1)}) = \mathbb{E}\Bigg(\sum_{f : f \cap e \neq \emptyset} t_f \Bigg) \le 2(d_{\max} - 1)q < 2 d_{\max} q.$$

Moreover, any partial second order derivative is at most 1. Hence,

$$\mathbb{E}_0(\varDelta_p^{(1)}) \le \max(m d_{\max} q^2, 2 d_{\max} q, 1),$$
$$\mathbb{E}_1(\varDelta_p^{(1)}) \le \max(2 d_{\max} q, 1) \quad \text{and}$$
$$\mathbb{E}_2(\varDelta_p^{(1)}) \le 1.$$

Now set $\mathcal{E}_0 = 4 m d_{\max} q^2 + 4\lambda^2$, $\mathcal{E}_1 = 2 d_{\max} q + 2\lambda$, and $\mathcal{E}_2 = 1$. Then since $\lambda \gg \log m$, the conditions of Theorem 4 are fulfilled. On the other hand, for c sufficiently large in the definition of α, $c_2\sqrt{\lambda \mathcal{E}_0 \mathcal{E}_1} \le \alpha/8$.

(ii) We need to prove the following statements

$$\mathbb{P}\Bigg(|\varDelta_p^{(2)} - \mathbb{E}(\varDelta_p^{(2)})| \ge \frac{\beta + \gamma}{8} \Bigg) \le e^{-\Omega(\lambda)}, \tag{30}$$

$$\mathbb{P}\Bigg(|\varDelta_p^{(2)} - \mathbb{E}(\varDelta_p^{(2)})| \ge \frac{\beta + \nu}{8} \Bigg) \le e^{-\Omega(\lambda)}. \tag{31}$$

Consider the same random variables t_e from part (i). Let Q be the set of all paths of length 3 in G. Then

$$\varDelta_p^{(2)} = \sum_{\{e,f,g\} \in Q} t_e t_g (1 - t_f) = \sum_{\{e,f,g\} \in Q} t_e t_g - \sum_{\{e,f,g\} \in Q} t_e t_f t_g.$$

Now let $Y_1 = \sum_{\{e,f,g\} \in Q} t_e t_g / 4$ and $Y_2 = \sum_{\{e,f,g\} \in Q} t_e t_f t_g$. Similar to part (i) we have

$$\mathbb{E}_0(Y_1) \le \max(m d_{\max}^2 q^2/4, d_{\max}^2 q/2, 1),$$
$$\mathbb{E}_1(Y_1) \le \max(d_{\max}^2 q/2, 1) \quad \text{and} \quad \mathbb{E}_2(Y_1) \le 1.$$

Therefore, set $\mathcal{E}_0 = m d_{\max}^2 q^2/2 + 2\lambda^2$, $\mathcal{E}_1 = d_{\max}^2 q/2 + 2\lambda$, and $\mathcal{E}_2 = 1$. These satisfy the conditions of Theorem 4. Again by considering c large enough we have

$$\mathbb{P}(|Y_1 - \mathbb{E}(Y_1)| \ge \beta/32) \le e^{-\Omega(\lambda)}. \tag{32}$$

For Y_2 we have

$$\mathbb{E}_0(Y_2) \le \max(md_{\max}^2 q^3, 2d_{\max}^2 q^2, 2d_{\max} q, 1),$$
$$\mathbb{E}_1(Y_2) \le \max(2d_{\max}^2 q^2, 2d_{\max} q, 1)$$

and

$$\mathbb{E}_2(Y_2) \le \max(2d_{\max} q, 1) \quad \text{and} \quad \mathbb{E}_3(Y_2) = 1.$$

As before, set $\mathcal{E}_0 = 2md_{\max}^2 q^3 + 3\lambda^3$, $\mathcal{E}_1 = 2d_{\max}^2 q^2 + 2\lambda^2$, and $\mathcal{E}_2 = 2d_{\max} q + \lambda$, $\mathcal{E}_3 = 1$ to obtain

$$\mathbb{P}\left(|Y_2 - \mathbb{E}(Y_2)| \ge \frac{\gamma}{8}\right) \le e^{-\Omega(\lambda)}. \tag{33}$$

Combining (32) and (33), (30) is proved. Finally, (31) is the result of (32) and the following,

$$|\Delta_p^{(2)} - \mathbb{E}(\Delta_p^{(2)})| \le |4Y_1 - 4\mathbb{E}(Y_1)| + \mathbb{E}(Y_2) \le |4Y_1 - 4\mathbb{E}(Y_1)| + md_{\max}^2 q^3$$
$$\le |4Y_1 - 4\mathbb{E}(Y_1)| + \frac{\nu}{8}.$$

(iii) Here we will prove

$$\mathbb{P}\left(\left|\frac{(\Lambda_p^{(1)})^2}{8m} - \frac{\mathbb{E}(\Lambda_p^{(1)})^2}{8m}\right| \ge c_1 d_{\max}^2 q \sqrt{\lambda(\lambda + mq)}\right) \le e^{-\Omega(\lambda)}, \quad \text{and} \tag{34}$$

$$\mathbb{P}\left(\left|\frac{\Lambda_p^{(2)}}{8m} - \mathbb{E}\left(\frac{\Lambda_p^{(2)}}{8m}\right)\right| \ge \frac{c_1 d_{\max}^2}{m}\sqrt{\lambda(md_{\max} q^2 + 2\lambda^2)(q + \lambda)}\right) \le e^{-\Omega(\lambda)}. \tag{35}$$

Note that by making c in the definition of ζ large enough, (34) and (35) together give us (iii). First we prove (34). Write

$$\frac{\Lambda_p^{(1)}}{2d_{\max}} = \sum_{e=\{u,v\}\in E(G)} \frac{d_u + d_v}{2d_{\max}} t_e$$

which is a polynomial with coefficients in (0, 1]. As before

$$\mathbb{E}_0\left(\frac{\Lambda_p^{(1)}}{2d_{\max}}\right) \le \max(mq, 1), \qquad \mathbb{E}_1\left(\frac{\Lambda_p^{(1)}}{2d_{\max}}\right) \le 1.$$

Now set $\mathcal{E}_0 = \lambda + mq$ and $\mathcal{E}_1 = 1$. Thus,

$$\mathbb{P}\left(\left|\frac{\Lambda_p^{(1)}}{2d_{\max}} - \mathbb{E}\left(\frac{\Lambda_p^{(1)}}{2d_{\max}}\right)\right| \le c_1\sqrt{\lambda(\lambda + mq)}\right) \le d_1 e^{-\Omega(\lambda)}. \tag{36}$$

By Lemma 11(ii) we have $\Lambda_p^{(1)} \le 2md_{\max}q$. Hence, inequality $|(\Lambda_p^{(1)})^2 - \mathbb{E}(\Lambda_p^{(1)})^2| \ge 8c_1 md_{\max}^2 q\sqrt{\lambda(\lambda+mq)}$ gives

$$\left|\frac{\Lambda_p^{(1)}}{2d_{\max}} - \mathbb{E}\left(\frac{\Lambda_p^{(1)}}{2d_{\max}}\right)\right| \ge c_1\sqrt{\lambda(\lambda+mq)}.$$

Now using (36), (34) is trivial.

The proof of (35) is similar to the proofs in (i) and (ii). We start with the following polynomial representation for $\Lambda_p^{(2)}$

$$\begin{aligned}\frac{\Lambda_p^{(2)}}{2d_{\max}^2} &= \sum_{i=1}^n \frac{d_i^2}{2d_{\max}^2}\left(\sum_{e=(i,.)} t_e\right)^2 \\ &= \sum_{i=1}^n \frac{d_i^2}{2d_{\max}^2}\left(\sum_{e=(i,.)} t_e\right) + 2\sum_{i=1}^n \frac{d_i^2}{2d_{\max}^2}\sum_{e\cap f=i} t_e t_f.\end{aligned}$$

Then we represent the right-hand side by $Z_1 + Z_2$ where

$$Z_1 = \sum_{i=1}^n \frac{d_i^2}{2d_{\max}^2}\left(\sum_{e=(i,.)} t_e\right) \quad \text{and} \quad Z_2 = 2\sum_{i=1}^n \frac{d_i^2}{2d_{\max}^2}\sum_{e\cap f=i} t_e t_f.$$

The next step is to use Vu's inequality for both Z_1 and Z_2 separately. The concentration for Z_2 is less sharp and it will dominate the concentration for $Z_1 + Z_2$. For Z_1 the inequalities

$$\mathbb{E}_0(Z_1) \le \max(mq, 1), \qquad \mathbb{E}_1(Z_2) \le 1$$

show that the same $\mathcal{E}_0, \mathcal{E}_1$ as in (36) can be used to obtain the inequality

$$\mathbb{P}\left(\left|\frac{2d_{\max}^2 Z_1}{8m} - \mathbb{E}\left(\frac{2d_{\max}^2 Z_1}{8m}\right)\right| \le c_2\frac{d_{\max}^2}{m}\sqrt{\lambda(\lambda+mq)}\right) \le d_2 e^{-\Omega(\lambda)}. \tag{37}$$

Now for Z_2 the bounds on the partial derivatives are given by $\mathbb{E}_0(Z_2) \le \max(\frac{md_{\max}q^2}{2}, q, 1)$, $\mathbb{E}_1(Y_1) \le \max(q, 1)$, and $\mathbb{E}_2(Y_1) = 1$. Therefore, $\mathcal{E}_0 = md_{\max}q^2 + 2\lambda^2$ and $\mathcal{E}_1 = q + \lambda$, $\mathcal{E}_2 = 1$ satisfy the conditions of Theorem 4 and we obtain the inequality

$$\begin{aligned}\mathbb{P}\left(\left|\frac{2d_{\max}^2 Z_2}{8m} - \mathbb{E}\left(\frac{2d_{\max}^2 Z_2}{8m}\right)\right| \le c_3\frac{d_{\max}^2}{m}\sqrt{\lambda(md_{\max}q^2 + 2\lambda^2)(q+\lambda)}\right) \\ \le d_2 e^{-\Omega(\lambda)}.\end{aligned} \tag{38}$$

The final inequality (35) can now be shown by combining equations (37) and (38).

(iv) This case is treated exactly the same as (ii) because we have the following

$$\frac{\Lambda_p^{(3)}}{d_{\max}^2} = \sum_{\{e,f,g\}\in R,\ e=\{u,v\}} \frac{d_u d_v}{d_{\max}^2} t_e t_g (1 - t_f).$$

□

Proof of Lemma 20 Using Lemma 11(iii) and the definition of Ψ, from $\Psi_p \geq \lambda^2/\omega$ we can get

$$\Delta_p \geq \frac{\lambda^2}{\omega} - \Lambda_p \geq \frac{\lambda^2}{\omega} - m d_{\max}^2 q^2 > \frac{\lambda^2}{\omega} - \frac{d_{\max}^2 \omega^2 \lambda^2}{4m} \tag{39}$$

$$> \frac{\lambda^2}{2\omega} \tag{40}$$

where (39) uses $2mq = 2m - 2r < \omega\lambda$ and (40) holds since $d_{\max}^2 \omega^3 \ll m$.

Since $2m - 2r$ is small then G_p is very dense. Let us consider its complement G_q which is sparse. Let $N_0(u) = N(u) \cup \{u\}$. Then using

$$\Delta_{p_r} \leq \sum_u d_{G_q}(u) \sum_{v \in N_0(u)} d_{G_q}(v)$$

and $\Delta_p \geq \lambda^2/2\omega$, one of the following statements should hold:

(a) G_q has more than $\omega^2\lambda/4$ edges.
(b) For some u, $\sum_{v \in N_0(u)} d_{G_q}(v) \geq \lambda/\omega^3$.

If (a) holds, since $2mq \leq \omega\lambda$ then

$$\mathbb{P}\left(G_q \text{ has more than } \frac{\omega^2\lambda}{4} \text{ edges}\right) \leq \binom{m}{\frac{\omega^2\lambda}{4}} q^{\frac{\omega^2\lambda}{4}} \leq \left(\frac{4mqe}{\omega^2\lambda}\right)^{\frac{\omega^2\lambda}{4}}$$
$$\leq e^{-\frac{\omega^2\lambda}{4}(\log\omega - 1 - \log 2)} = e^{-\Omega(\lambda)}.$$

If (b) holds then the number of edges in G that contribute to $\sum_{v \in N_0(u)} d_{G_q}(v)$ is at most $d_{\max}^2$ and each edge can contribute at most twice. Hence,

$$\mathbb{P}\left(\sum_{v \in N_0(u)} d_{G_q}(v) \geq \lambda/\omega^3\right) \leq \binom{d_{\max}^2}{\frac{\lambda}{2\omega^3}} q^{\frac{\lambda}{2\omega^3}}$$
$$\leq \left(\frac{2d_{\max}^2 q\omega^3 e}{\lambda}\right)^{\frac{\lambda}{2\omega^3}} \leq \left(\frac{d_{\max}^2 \omega^4 e}{m}\right)^{\frac{\lambda}{2\omega^3}}$$
$$= e^{-\frac{\lambda}{2\omega^3}(\log m - \log(d_{\max}^2\omega^4) - 1)} \leq e^{-\Omega(\frac{\lambda}{\omega^3}\log m)} = e^{-\Omega(\lambda)}.$$

Note that we need δ in the definition of ω to be small enough such that $\log m \gg \omega^3$ and for $\delta < 1$ this is true. □

7.5.3 Proof of part (b) of Lemmas 15 and 16

Note that

$$f(\mathcal{N}) = \prod_{r=0}^{m-1}\left(1 + \frac{\Psi_r(\mathcal{N}) - \psi_r}{\binom{2m-2r}{2} - \Psi_r(\mathcal{N})}\right)$$

and since $\Psi_r(\mathcal{N}) \le (1-\tau/4)\binom{2m-2r}{2}$ for $\mathcal{N} \in S^*(\mathcal{M})$ then

$$f(\mathcal{N}) \le \prod_{r=0}^{m-1} \left(1 + \frac{16/\tau \max(\Psi_r(\mathcal{N}) - \psi_r, 0)}{(2m-2r)^2}\right).$$

Proof of Lemma 15(b) Using $1 + x \le e^x$ we only need to show

$$\sum_{r=0}^{m-1} \frac{\max(\Psi_r(\mathcal{N}) - \psi_r, 0)}{(2m-2r)^2} \le o(\lambda).$$

To simplify the notation, let $g(r) = \frac{\max(\Psi_r(\mathcal{N})-\psi_r,0)}{(2m-2r)^2}$. Note that $0 \le g(r) \le 1$ which gives $\sum_{2m-2r=2}^{\lambda/\omega^{1/2}} g(r) = o(\lambda)$. Hence, we only need to show $\sum_{2m-2r=\lambda/\omega^{1/2}}^{2m-2} g(r) = o(\lambda)$. Also note that the numerator of $g(r)$ is at most $T_r(\lambda)$. Therefore, using the definition of $T_r(\lambda)$,

$$\sum_{2m-2r=\lambda/\omega^{1/2}}^{2m-2} g(r) \le \sum_{2m-2r=\lambda/\omega^{1/2}}^{\omega\lambda} \frac{\lambda^2}{(2m-2r)^2\omega} + \sum_{2m-2r=\omega\lambda}^{\omega\lambda^2} \frac{3\beta_r(\lambda) + 2\nu_r}{(2m-2r)^2}$$
$$+ \sum_{2m-2r=\omega\lambda^2}^{2m-2} \frac{3\beta_r(\lambda) + 2\gamma_r(\lambda)}{(2m-2r)^2}.$$

Therefore, it suffices to show

$$\sum_{2m-2r=\lambda/\omega^{1/2}}^{\omega\lambda} \frac{\lambda^2}{(2m-2r)^2\omega} + \sum_{2m-2r=\omega\lambda}^{\omega\lambda^2} \frac{3\beta_r(\lambda) + 2\nu_r}{(2m-2r)^2}$$
$$+ \sum_{2m-2r=\omega\lambda^2}^{2m-2} \frac{3\beta_r(\lambda) + 2\gamma_r(\lambda)}{(2m-2r)^2} = o(\lambda).$$

A series of elementary inequalities will now be used to bound these three summations. We will use $q_r = \frac{2m-2r}{2m}$ to obtain

$$\sum_{2m-2r=2}^{2m-2} \frac{(\lambda m d_{\max}^4 q_r^3)^{1/2}}{(2m-2r)^2} = \frac{\lambda^{1/2} d_{\max}^2}{2m\sqrt{2}} \sum_{2m-2r=2}^{2m-2} \frac{1}{\sqrt{2m-2r}}$$
$$= O\left(\frac{\lambda^{1/2} d_{\max}^2}{m} \int_{x=2}^{2m} \frac{1}{\sqrt{x}} dx\right) = O\left(\frac{\lambda^{1/2} d_{\max}^2}{\sqrt{m}}\right) = o(\lambda),$$

$$\sum_{2m-2r=2}^{2m-2} \frac{(\lambda^2 m d_{\max}^2 q_r^2)^{1/2}}{(2m-2r)^2} = \frac{\lambda d_{\max}}{2m^{1/2}} \sum_{2m-2r=2}^{2m-2} \frac{1}{2m-2r} = O\left(\frac{\lambda d_{\max}}{m^{1/2}} \log m\right) = o(\lambda),$$

$$\sum_{2m-2r=2}^{2m-2} \frac{(\lambda^3 d_{\max}^2 q_r)^{1/2}}{(2m-2r)^2} = \frac{\lambda^{3/2} d_{\max}}{(2m)^{1/2}} \sum_{2m-2r=2}^{2m-2} \frac{1}{(2m-2r)^{3/2}}$$

$$= O\left(\frac{\lambda^{3/2} d_{\max}}{m^{1/2}}\right) = o(\lambda), \quad \text{and}$$

$$\sum_{2m-2r=\omega\lambda}^{2m-2} \frac{\lambda^2}{(2m-2r)^2} \leq \lambda^2 \int_{x=\omega\lambda}^{2m} x^{-2} dx = o(\lambda).$$

Furthermore, we can show the following bounds

$$\sum_{2m-2r=2}^{2m-2} \frac{(\lambda^3 m d_{\max}^2 q_r^3)^{1/2}}{(2m-2r)^2} = \frac{\lambda^{3/2} d_{\max}}{2m\sqrt{2}} \sum_{2m-2r=2}^{2m-2} \frac{1}{\sqrt{2m-2r}} = O\left(\frac{\lambda^{3/2} d_{\max}}{\sqrt{m}}\right),$$

$$\sum_{2m-2r=2}^{2m-2} \frac{\lambda^2 d_{\max} q_r}{(2m-2r)^2} = O\left(\frac{\lambda^2 d_{\max} \log m}{2m}\right),$$

$$\sum_{2m-2r=\omega\lambda^2}^{2m-2} \frac{\lambda^3}{(2m-2r)^2} = O\left(\lambda^3 \int_{x=\omega\lambda^2}^{\infty} x^{-2} dx\right) = O\left(\frac{\lambda^3}{\omega\lambda^2}\right) = o(\lambda), \tag{41}$$

and

$$\sum_{2m-2r=2}^{\omega\lambda^2} \frac{m d_{\max}^2 q_r^3}{(2m-2r)^2} = \sum_{2m-2r=2}^{\omega\lambda^2} \frac{d_{\max}^2 (2m-2r)}{8m^2} = O\left(\frac{\omega^2 \lambda^4 d_{\max}^2}{m^2}\right). \tag{42}$$

Remark 3 All previous equations are of order $o(\lambda)$, since $\lambda \leq \lambda_L = O(d_{\max}^2 \log n)$ and $d_{\max} = o(m^{\frac{1}{4}-\tau})$. Note that we also used $\sqrt{A+B} \leq \sqrt{A} + \sqrt{B}$ to find upper bounds for β_r, γ_r.

□

Proof of Lemma 16(b) Similar to proof of Lemma 15(b) we will show

$$f(\mathcal{N}) \leq \prod_{r=m-d_{\max}^2+1}^{m} \frac{\binom{2m-2r}{2} - \psi_r}{\binom{2m-2r}{2} - \Psi_r(\mathcal{N})} \prod_{r=0}^{m-d_{\max}^2} \left(1 + 16/\tau \frac{\max(\Psi_r(\mathcal{N}) - \psi_r, 0)}{(2m-2r)^2}\right)$$

$$\leq \binom{2d_{\max}^2}{2}^{d_{\max}^2} \cdot \prod_{r=0}^{m-d_{\max}^2} \left(1 + 16/\tau \frac{\Psi_r}{(2m-2r)^2}\right)$$

$$\leq (2d_{\max}^4)^{d_{\max}^2} \cdot \prod_{r=0}^{m-d_{\max}^2} \left(1 + 16/\tau \frac{d_{\max}^2}{2m-2r}\right) \tag{43}$$

$$
\begin{aligned}
&\le e^{d_{\max}^2 \log(2d_{\max}^4)+3\sum_{i=d_{\max}^2+1}^{m} \frac{d_{\max}^2}{i}} \\
&\le e^{d_{\max}^2 (\log(2d_{\max}^4)+3\log d_{\max}+\log m)} \\
&\le e^{4d_{\max}^2 \log n} \qquad (44)
\end{aligned}
$$

where (43) use Lemma 11, and (44) uses $m \le nd_{\max}/2$ and $d_{\max} \ll m^{1/3} \le n^{1/2}$. □

Proof of Lemma 18 By the definition of $\mathcal{C}$: $\sum_{2m-2r=2}^{\omega\lambda_0} g(r) = 0$. Thus, we only need to show that if $\Psi_r(\mathcal{N}) - \psi_r \le T_r(\lambda_0)$ for all r with $2m - 2r \ge \omega\lambda_0$ then

$$\sum_{2m-2r=\omega\lambda_0}^{m} g(r) = o(1).$$

For that it is sufficient to prove

$$\sum_{2m-2r=\omega\lambda_0}^{m} \frac{T_r(\lambda_0)}{(2m-2r)^2} = o(1).$$

The proof is similar to the proof of Lemma 15(b) with a slight modification. Instead of using (41) and (42) we use

$$\sum_{2m-2r=\omega\lambda_0^3}^{2m-2} \frac{\lambda_0^3}{(2m-2r)^2} = O\left(\lambda_0^3 \int_{\omega\lambda_0^3}^{\infty} x^{-2} dx\right) = O\left(\frac{\lambda_0^3}{\omega\lambda_0^3}\right) = o(1),$$

and

$$\sum_{2m-2r=2}^{\omega\lambda_0^3} \frac{m d_{\max}^2 q_r^3}{(2m-2r)^2} = \sum_{2m-2r=2}^{\omega\lambda_0^3} \frac{(2m-2r) d_{\max}^2}{m^2} = O\left(\frac{d_{\max}^2 \omega^2 \lambda_0^6}{m^2}\right) = o(1).$$ □

For the other equations in the proof of Lemma 15(b) let $\lambda = \lambda_0$ and they will be $o(1)$.

Proof of Lemma 17 (a) We have $2m - 2r \le \omega\lambda_0 \ll (\log n)^2$. This means proving the bound only for one r is enough. Similar to the proof of Lemma 20, from $\Psi_p \ge 2^{j-1}$ we get $\Delta_p \ge 2^{j-2}$. Thus, one of the following statements hold:

(i) G_q has more than $2^{j/2-2}$ edges
(ii) For some u, $\sum_{v \in N_0(u)} d_{G_q}(v) \ge 2^{j/2-1}$

and rest of the proof will be exactly as in Lemma 20.

(b) By the definition of B_j

$$\sum_{2m-2r=2}^{\omega\lambda_0} g(r) \le \sum_{2m-2r=2}^{\omega\lambda_0} \frac{2^j}{(2m-2r)^2} = O(2^j).$$ □

7.6 Proof of (25)

From Lemma 19, for all r with $2m-2r \geq \omega\lambda_0$,

$$\mathbb{P}(|\Psi_r - \psi_r| \geq \alpha_r(\lambda_0) + \beta_r(\lambda_0) + (1 + d_{\max}^2/4m)\gamma_r(\lambda_0) + \zeta_r(\lambda_0)) = o(1). \quad (45)$$

Let $\mathcal{N}$ be an ordering with $|\Psi_r - \psi_r| \leq \alpha_r(\lambda_0) + \beta_r(\lambda_0) + (1 + d_{\max}^2/4m)\gamma_r(\lambda_0) + \zeta_r(\lambda_0)$ for all $2m-2r \geq \omega\lambda_0$. Then

$$f(\mathcal{N}) \geq \prod_{2m-2r=\omega\lambda_0^3}^{2m-2} \left(1 - (16/\tau)\frac{\alpha_r(\lambda_0) + \beta_r(\lambda_0) + \gamma_r(\lambda_0) + \zeta_r(\lambda_0)}{(2m-2r)^2}\right)$$
$$\times \prod_{2m-2r=2}^{\omega\lambda_0^3} \left(1 - (16/\tau)\frac{\psi_r}{(2m-2r)^2}\right). \quad (46)$$

In Sect. 7.5 it was shown that $\frac{3}{\tau}\sum_{2m-2r=\omega\lambda_0^3}^{2m-2} \frac{\alpha_r(\lambda_0)+\beta_r(\lambda_0)+\gamma_r(\lambda_0)+\zeta_r(\lambda_0)}{(2m-2r)^2} = o(1)$. Now one can use $1 - x \geq e^{-2x}$ when $0 \leq x \leq 1/2$ to see that the first product in the right-hand side of (46) is $1 - o(1)$. The second product is also $1 - o(1)$ because of $\omega\lambda_0^3 d^2 = o(m)$ and the bound $\psi_r = O[(2m-2r)^2 \frac{d_{\max}^2}{m}]$ given Lemma 14. These, together with (45) finish the proof of (25). In fact they show the stronger statement $\mathbb{E}(f(\mathcal{N})1_{S^*(\mathcal{M})}) > 1 - o(1)$.

Remark 4 The proofs of this section and Sect. 7.5 yield the following corollary which will be used in Sect. 7.7.

Corollary 1 *For sufficiently large c in the definition of λ_L,*

$$\mathbb{E}\left(\exp\left[\frac{1}{\tau^2}\sum_{r=0}^{m-1} \frac{\max(\Psi_r(\mathcal{N}) - \psi_r, 0)}{(2m-2r)^2}\right]\right) = 1 + o(1) \quad (47)$$

Proof Bounds of Sect. 7.5 show that the contribution of the sets $A_i \setminus A_{i-1}$ and $B_j \setminus B_{j-1}$ are all $o(1)$ and the contribution of $\mathcal{C}$ is $1 + o(1)$. The contribution of A_∞ also is $o(1)$ by taking the constant c large enough. □

7.7 Proof of (26)

In this section we deal with those orderings $\mathcal{N}$ for which the condition

$$\Psi_r(\mathcal{N}) \leq (1 - \tau/4)\binom{2m-2r}{2} \quad (*)$$

is violated for some r. If this happens for some r then from Lemma 11(iii) and $d_{\max}^4 = o(m)$ we have

$$\Delta_r(\mathcal{N}) \geq \Psi_r(\mathcal{N}) - \frac{d_{\max}^2}{8m}(2m-2r)^2$$
$$> \Psi_r(\mathcal{N}) - \tau/4\binom{2m-2r}{2} > (1-\tau/2)\binom{2m-2r}{2}.$$

On the other hand using Lemma 11(i) we have $\Delta_r(\mathcal{N}) \leq \frac{d_{\max}^2(2m-2r)}{2}$. So for $2m-2r \geq \frac{d_{\max}^2}{2-\tau}$ we have $\Delta_r(\mathcal{N}) \leq (1-\tau/2)\binom{2m-2r}{2}$. Thus condition $(*)$ is violated only for r very close to m. Let $S_t(\mathcal{M})$, $t = 1, \ldots, \frac{d_{\max}^2}{2-\tau}$, be the set of all ordering $\mathcal{N}$ for which $(*)$ fails for the first time at $r = m - t$. We will use $\sum_{t=1}^{\infty} \frac{1}{m^{\tau t}} = o(1)$ to prove (26). In particular we show

$$\mathbb{E}(f(\mathcal{N})1_{S_t}) \leq O\left(\frac{1}{m^{\tau t}}\right).$$

Note that $\binom{2m-2r}{2} - \Psi_r(\mathcal{N}) = \sum_{\{i,j\} \in E_r} d_i^{(r)} d_j^{(r)} (1 - \frac{d_i d_j}{4m}) \geq (m-r)(1-\frac{d_{\max}^2}{4m})$ since at step r there should be at least $m-r$ suitable edges to complete the ordering $\mathcal{N}$. Hence using $d_{\max} = O(m^{\frac{1}{4}-\tau})$ we have

$$\frac{\binom{2m-2r}{2}}{\binom{2m-2r}{2} - \Psi_r(\mathcal{N})} \leq 2m - 2r - 1 + O\left(\frac{d_{\max}^4}{m}\right) \leq 2m - 2r. \tag{48}$$

This gives

$$\prod_{r=m-t}^{m-1} \frac{\binom{2m-2r}{2}}{\binom{2m-2r}{2} - \Psi_r(\mathcal{N})} \leq 2^t t! \leq 2t^t$$

and since t is the first place that $(*)$ is violated, then

$$\prod_{r=0}^{m-t-1} \frac{\binom{2m-2r}{2} - \psi_r}{\binom{2m-2r}{2} - \Psi_r(\mathcal{N})} \leq \exp\left[\frac{16}{\tau} \sum_{r=0}^{m-1} \frac{\max(\Psi_r(\mathcal{N}) - \psi_r, 0)}{(2m-2r)^2}\right].$$

Thus,

$$f(\mathcal{N})1_{S_t} = 1_{S_t} \prod_{r=0}^{m-1} \frac{\binom{2m-2r}{2} - \psi_r}{\binom{2m-2r}{2} - \Psi_r(\mathcal{N})}$$
$$\leq 2t^t 1_{S_t} \exp\left[\frac{16}{\tau} \sum_{r=0}^{m-1} \frac{\max(\Psi_r(\mathcal{N}) - \psi_r, 0)}{(2m-2r)^2}\right].$$

Now using Hölder's inequality

$$\mathbb{E}(f(\mathcal{N})1_{S_t}) \le 2t^t \mathbb{E}\left(1_{S_t} \exp\left[\frac{16}{\tau}\sum_{r=0}^{m-t-1}\frac{\max(\Psi_r(\mathcal{N})-\psi_r,0)}{(2m-2r)^2}\right]\right)$$
$$\le 2t^t \mathbb{E}(1_{S_t})^{1-\tau/2}\mathbb{E}\left(1_{S_t} \exp\left[\frac{32}{\tau^2}\sum_{r=0}^{m-t-1}\frac{\max(\Psi_r(\mathcal{N})-\psi_r,0)}{(2m-2r)^2}\right]\right)^{\tau/2}.$$

But using Corollary 1, the second term in the above product is $1+o(1)$ and we only need to show

$$2t^t\mathbb{P}(S_t)^{1-\tau/2} \le (1+o(1))\frac{1}{m^{\tau t}}.$$

Let $r = m - t$ and $\Gamma(u) = N_{G_{\mathcal{N}_r}}(u)$ be the set of all neighbors of u in $G_{\mathcal{N}_r}$. Note that

$$\Delta_r(\mathcal{N}) = \frac{1}{2}\sum_u d_u^{(r)} \sum_{v\in\Gamma(u)\cup\{u\}} (d_v^{(r)} - 1_{u=v})$$

and

$$\binom{2m-2r}{2} = \frac{1}{2}\sum_u d_u^{(r)}\sum_v (d_v^{(r)} - 1_{u=v}).$$

Now $\Delta_r(\mathcal{N}) > (1-\tau/2)\binom{2m-2r}{2} > (1-\tau)\binom{2m-2r}{2}$ implies that a vertex u with $d_u^{(r)} > 0$ exists and

$$\sum_{v\in\Gamma(u)\cup\{u\}} (d_v^{(r)} - 1_{u=v}) > (1-\tau)\sum_v (d_v^{(r)} - 1_{u=v}).$$

Equivalently

$$\sum_{v\notin\Gamma(u)\cup\{u\}} d_v^{(r)} \le \tau\sum_v (d_v^{(r)} - 1_{u=v}) \le \tau(2m-2r-1) \le 2\tau t. \tag{49}$$

Any of the last t edges of $\mathcal{N}$ that have at least one endpoint outside of $\Gamma(u)$, contributes at least once to the left-hand side of (49). So there are at most $2\tau t$ such edges. Let $k = d_u - |\Gamma(u)|$ and let ℓ be the number of edges that are entirely in $\Gamma(u)$. Then we should have $k \ge 1$ and $\ell \ge (1-2\tau)i$. Thus, the probability that $d_u^{(r)} > 0$ and $\sum_{v\notin\Gamma(u)\cup\{u\}} d_v^{(r)} \le 2\tau t$, for a fixed vertex u is upper bounded by

$$\sum_{k\ge 1,\ \ell\ge(1-2\tau)t} \frac{\binom{d_u}{k}\binom{\binom{d_u-k}{2}}{\ell}\binom{m-d_u-\binom{d_u-k}{2}}{t-k-\ell}}{\binom{m}{t}}.$$

Hence,

$$\mathbb{P}(S_t) \le \sum_u \sum_{k\ge 1,\ \ell\ge(1-2\tau)t} \frac{\binom{d_u}{k}\binom{\binom{d_u-k}{2}}{\ell}\binom{m-d_u-\binom{d_u-k}{2}}{t-k-\ell}}{\binom{m}{t}}.$$

Now using

$$\binom{d_u}{k} \le \frac{d_u^k}{k!}, \quad \binom{\binom{d_u-k}{2}}{\ell} \le \frac{(d_u^2/2)^\ell}{\ell!},$$

$$\binom{m - d_u - \binom{d_u-k}{2}}{t-k-\ell} \le \frac{m^{t-k-\ell}}{(t-k-\ell)!}$$

for $t = O(d_{\max}^2) = o(m^{1/2})$ we have

$$\binom{m}{t} = \big(1 + o(1)\big)\frac{m^t}{t!}.$$

This means

$$\begin{aligned}
\mathbb{P}(S_t) &\le (1+o(1)) \sum_u \sum_{k\ge 1,\ \ell \ge (1-2\tau)t} \frac{\frac{d_u^k}{k!}\frac{(d_u^2/2)^\ell}{\ell!}\frac{m^{t-k-\ell}}{(t-k-\ell)!}}{\frac{m^t}{t!}} \\
&= (1+o(1)) \sum_u \sum_{k\ge 1,\ \ell \ge (1-2\tau)t} \frac{(d_u/m)^k (d_u^2/2m)^\ell t!}{k!\ell!(t-k-\ell)!} \\
&\le (1+o(1)) 2\tau t \sum_u (d_u/m)(d_u^2/2m)^{(1-2\tau)t}\binom{t}{2\tau t} \\
&\le (1+o(1)) t \frac{d_{\max}}{m} \sum_u (d_u^2/2m)^{(1-2\tau)t} 2^t \qquad (50) \\
&\le (1+o(1)) t 2^{2t/3} \frac{d_{\max}}{m} \sum_u \left(\frac{d_u^2}{m}\right)^{(1-2\tau)t} \qquad (51) \\
&\le \big(1+o(1)\big) 2t 2^{2t/3} \left(\frac{d_{\max}^2}{m}\right)^{(1-2\tau)t} \qquad (52)
\end{aligned}$$

where (50) and (51) are based on $\tau \le 1/3$ and $\binom{a}{b} \le 2^a$. Moreover, (52) uses $\sum_u d_u^k = \sum_{u\sim_G v}(d_u^{k-1} + d_v^{k-1}) \le 2md_{\max}^{k-1}$. Now we can use $t \le \frac{d_{\max}^2}{2-\tau}$, $d_{\max} \le m^{\frac{1}{4}-\tau}$, and $\tau \le 1/3$ to get

$$\begin{aligned}
2t^t \mathbb{P}(S_t)^{1-\tau/2} &\le (1+o(1)) 4t \left(\frac{2^{2-\tau/3}}{2-\tau}\frac{d_{\max}^{4-5\tau+2\tau^2}}{m^{1-2.5\tau+\tau^2}}\right)^t \\
&\le (1+o(1)) 4t \left(\frac{d_{\max}^{4-5\tau+2\tau^2}}{m^{1-2.5\tau+\tau^2}}\right)^t \\
&\le (1+o(1)) 4t \left(m^{-2.75\tau+3.5\tau^2-2\tau^3}\right)^t \\
&\le O(m^{-\tau t}).
\end{aligned}$$

□

8 Bounding the Variance of the SIS Estimate

In this section we will prove two variance bounds from Sect. 4. We will borrow some notation and results from Sect. 7.

8.1 Proof of (6)

It is easy to see that instead of proving (6) directly, we can consider the equivalent formulation $\mathbb{E}_{\mathrm{A}}(N^2)/\mathbb{E}_{\mathrm{A}}(N)^2 \le 1 + o(1)$. For the numerator we have

$$\mathbb{E}_{\mathrm{A}}(N^2) = \sum_G \sum_{\mathcal{N}} \left(\frac{1}{m!\,\mathbb{P}_{\mathrm{A}}(\mathcal{N})}\right)^2 \mathbb{P}_{\mathrm{A}}(\mathcal{N}) = \sum_G \sum_{\mathcal{N}} \frac{1}{(m!)^2\,\mathbb{P}_{\mathrm{A}}(\mathcal{N})}.$$

On the other, we have the following estimate from the analysis of Theorem 1,

$$|\mathcal{L}(\bar{d})| = \frac{[1+o(1)]\prod_{r=0}^{m-1}\left[\binom{2m-2r}{2} - \psi_r\right]}{m!\prod_{i=1}^n d_i!}.$$

Therefore,

$$\begin{aligned}
\frac{\mathbb{E}_{\mathrm{A}}(N^2)}{\mathbb{E}_{\mathrm{A}}(N)^2} &= \frac{\sum_G \sum_{\mathcal{N}} \frac{1}{(m!)^2\,\mathbb{P}_{\mathrm{A}}(\mathcal{N})}}{|\mathcal{L}(\bar{d})|^2} \\
&= \frac{\sum_G \sum_{\mathcal{N}} \prod_{r=0}^{m-1}\left[\frac{\binom{2m-2r}{2} - \Psi_r(\mathcal{N})}{\binom{2m-2r}{2} - \psi_r}\right]}{m!|\mathcal{L}(\bar{d})|} \\
&= \frac{\sum_G \mathbb{E}(g(\mathcal{N}))}{|\mathcal{L}(\bar{d})|} \qquad (53)
\end{aligned}$$

where $g(\mathcal{N}) = \prod_{r=0}^{m-1} \frac{\binom{2m-2r}{2} - \Psi_r(\mathcal{N})}{\binom{2m-2r}{2} - \psi_r}$ and the expectation $\mathbb{E}$ is with respect to the uniform distribution on the set of all $m!$ orderings, $S(\mathcal{M})$. The goal is now to show that if $G \in \mathcal{L}(\bar{d})$ then

$$\mathbb{E}(g(\mathcal{N})) \le 1 + o(1). \qquad (54)$$

Note that (53) and (54) finish the proof. Thus, we only need to prove (54).

Proof of (54) Before starting the proof it is important to see that $g(\mathcal{N}) = f(\mathcal{N})^{-1}$ and the aim of Sect. 7 was to show that $\mathbb{E}(f(\mathcal{N})) = 1 + o(1)$. In this section we will show that the concentration results of Sect. 7 are strong enough to bound the variance of $g(\mathcal{N})$ as well.

Recall the definitions for variables λ_i and $T(\lambda_i)$ from Sect. 7. Here we will consider a different partitioning of the set $S(\mathcal{M})$. Define subsets $F_0 \subseteq F_1 \subseteq \cdots \subseteq F_L \subseteq S(\mathcal{M})$ as follows:

$$F_i = \{\mathcal{N} \in S(\mathcal{M}) \mid \psi_r - \Psi_r(\mathcal{N}) < T_r(\lambda_i) : \ \forall 0 \le r \le m - \omega\lambda_i/2\}$$

and $F_\infty = S(\mathcal{M}) \setminus \bigcup_{i=0}^{L} F_i$. The following two lemmas are equivalent versions of Lemmas 15, 18.

Lemma 23 *For all* $1 \le i \le L$:

(a) $\mathbb{P}(F_i \setminus F_{i-1}) \le e^{-\Omega(\lambda_i)}$.
(b) *For all* $\mathcal{N}$ *in* $F_i \setminus F_{i-1}$ *we have* $g(\mathcal{N}) \le e^{o(\lambda_i)}$.

Lemma 24 *If* $\mathcal{N} \in \mathcal{F}_0$ *then* $g(\mathcal{N}) \le 1 + o(1)$.

Proof of these lemmas is similar to the proofs for Lemmas 15 and 18, and the only extra information that is required is

$$\sum_{2m-2r=2}^{\omega\lambda} g(\mathcal{N}) \le 2\frac{\psi_r}{\frac{(2m-2r)^2}{2}} = O\left(\frac{\omega\lambda d_{\max}^2}{m}\right).$$

Then for Lemma 23 we use $\frac{\omega\lambda d_{\max}^2}{m} = o(\lambda)$ and for Lemma 24 we use $\frac{\omega\lambda_0 d_{\max}^2}{m} = o(1)$. The combination of these two lemmas gives $\mathbb{E}(g(\mathcal{N})) \le 1 + o(1)$. □

8.2 Proof of (7)

Similar to Sect. 8.1 we will use lemmas from Sect. 7. The main technical point in this section is a new result which exploits the combinatorial structure of the model to obtain a tighter bound than in Sect. 7.

Equation (7) is equivalent to

$$\frac{\mathbb{E}_{\mathrm{B}}(P^2)}{\mathbb{E}_{\mathrm{B}}(P)^2} < 1 + o(1).$$

First notice that

$$\frac{\mathbb{E}_{\mathrm{B}}(P^2)}{\mathbb{E}_{\mathrm{B}}(P)^2} = \frac{m!\,\sum_{\mathcal{N}} \mathbb{P}_{\mathrm{B}}(\mathcal{N})^2}{\mathbb{P}_{\mathrm{B}}(G)^2} = \frac{\mathbb{E}(f(\mathcal{N})^2)}{\mathbb{E}(f(\mathcal{N}))^2}.$$

Therefore, all we need to show is $\mathbb{E}(f(\mathcal{N})^2) = 1 + o(1)$.

Consider the same partitioning of the set $S(\mathcal{M})$ as in Sect. 7. It is straightforward to see that Lemmas 15, 16, 17, and 18 give us the following stronger results as well

$$\mathbb{E}(f(\mathcal{N})^2 1_{\mathcal{A}}) = o(1),$$
$$\mathbb{E}(f(\mathcal{N})^2 1_{\mathcal{B}}) = o(1),$$
$$\mathbb{E}(f(\mathcal{N})^2 1_{\mathcal{C}}) \le 1 + o(1).$$

Thus, the only missing part is the following

$$\mathbb{E}(f(\mathcal{N})^2 1_{S^*(\mathcal{M}) \setminus S^*(\mathcal{M})}) = o(1) \tag{55}$$

which we will prove by using the combinatorial properties of the model.

Proof of (55) Recall that $S^*(\mathcal{M}) \setminus S^*(\mathcal{M})$ consists of those orderings $\mathcal{N}$ that violate the condition

$$\Psi_r(\mathcal{N}) \leq (1-\tau/4)\binom{2m-2r}{2} \qquad (*)$$

for some r. If this happens for some r then from Lemma 11(iii) and $d_{\max}^4 = o(m)$ we have

$$\begin{aligned}\Delta_r(\mathcal{N}) &\geq \Psi_r(\mathcal{N}) - \frac{d_{\max}^2}{8m}(2m-2r)^2 \\ &> \Psi_r(\mathcal{N}) - \tau/4\binom{2m-2r}{2} > (1-\tau/2)\binom{2m-2r}{2}.\end{aligned}$$

On the other hand using Lemma 11(i) from Sect. 7: $\Delta_r(\mathcal{N}) \leq \frac{d_{\max}^2(2m-2r)}{2}$. So for $2m-2r \geq \frac{d_{\max}^2}{2-\tau}$ we have $\Delta_r(\mathcal{N}) \leq (1-\tau/2)\binom{2m-2r}{2}$. Thus condition $(*)$ is violated only for r very close to m. For these values of r we use the following combinatorial lemma to find an upper bound for $f(\mathcal{N})$.

Lemma 25 *For all r such that $2m-2r \leq \frac{d_{\max}^2}{2-\tau}$,*

$$\frac{\binom{2m-2r}{2} - \psi_r}{\binom{2m-2r}{2} - \Psi_r(\mathcal{N})} \leq 2d_{\max}.$$

Proof Let n_r be the number of available vertices (v_i's with $W_i \neq 0$) at step $r+1$. Without loss of generality assume that all such vertices are $v_1, \ldots, v_{n_r}$. For each $1 \leq i \leq n_r$ let $\tilde{d}_i^{(r)}$ be the number of neighbors of v_i among $v_1, \ldots, v_{n_r}$ at step $r+1$. Then the number of suitable pairs at step $r+1$ is at least $1/2\sum_{i=1}^{n_r}(n_r - 1 - \tilde{d}_i^{(r)})d_i^{(r)}$. Now consider the cases $n_r \geq 2d_{\max}$ or $n_r < 2d_{\max}$ separately.

1. For $n_r \geq 2d_{\max}$ the number of suitable pairs at step $r+1$ is at least $1/2\sum_{i=1}^{n_r}(d_{\max})d_i^{(r)} = d_{\max}(m-r)$. Therefore,

$$\frac{\binom{2m-2r}{2} - \psi_r}{\binom{2m-2r}{2} - \Psi_r(\mathcal{N})} \leq \frac{(m-r)(2m-2r-1)}{d_{\max}(m-r)(1-\frac{d_{\max}^2}{4m})} \leq 2d_{\max}.$$

Here we used $d_{\max}^2 = o(m)$ and $(2m-2r) \leq \frac{d_{\max}^2}{2-\tau} \leq 3d_{\max}^2/5$.

2. For $n_r < 2d_{\max}$ we use $n_r \geq 1 + \tilde{d}_i^{(r)} + d_i^{(r)}$ to show that the number of suitable pairs is at least

$$1/2\sum_{i=1}^{n_r}(n_r - 1 - \tilde{d}_i^{(r)})d_i^{(r)} \geq 1/2\sum_{i=1}^{n_r}(d_i^{(r)})^2 \geq 1/2\frac{(\sum_{i=1}^{n_r} d_i^{(r)})^2}{n_r}.$$

Hence,

$$\frac{\binom{2m-2r}{2}-\psi_r}{\binom{2m-2r}{2}-\Psi_r(\mathcal{N})} \leq \frac{(m-r)(2m-2r-1)}{\frac{(m-r)(2m-2r)}{n_r}(1-\frac{d_{\max}^2}{4m})}$$
$$\leq n_r \frac{1-\frac{1}{d_{\max}^2}}{1-o(1)} \leq 2d_{\max}. \qquad \square$$

Lemma 25 gives

$$\prod_{r=m-t}^{m-1} \frac{\binom{2m-2r}{2}-\psi_r}{\binom{2m-2r}{2}-\Psi_r(\mathcal{N})} \leq 2^t d_{\max}^t.$$

From here we will closely follow the steps taken in Sect. 7.7. Since t is the first place that $(*)$ is violated

$$\prod_{r=0}^{m-t-1} \frac{\binom{2m-2r}{2}-\psi_r}{\binom{2m-2r}{2}-\Psi_r(\mathcal{N})} \leq \exp\left[\frac{16}{\tau}\sum_{r=0}^{m-1}\frac{\max(\Psi_r(\mathcal{N})-\psi_r,0)}{(2m-2r)^2}\right].$$

So,

$$f(\mathcal{N})1_{S_t} = 1_{S_t}\prod_{r=0}^{m-1} \frac{\binom{2m-2r}{2}-\psi_r}{\binom{2m-2r}{2}-\Psi_r(\mathcal{N})}$$
$$\leq 2^t d_{\max}^t 1_{S_t} \exp\left[\frac{16}{\tau}\sum_{r=0}^{m-1}\frac{\max(\Psi_r(\mathcal{N})-\psi_r,0)}{(2m-2r)^2}\right].$$

Now using Hölder's inequality

$$\mathbb{E}(f(\mathcal{N})^2 1_{S_t})$$
$$\leq 2^{2t} d_{\max}^{2t}\mathbb{E}\left(1_{S_t}\exp\left[\frac{32}{\tau}\sum_{r=0}^{m-t-1}\frac{\max(\Psi_r(\mathcal{N})-\psi_r,0)}{(2m-2r)^2}\right]\right)$$
$$\leq 2^{2t} d_{\max}^{2t}\mathbb{E}(1_{S_t})^{1-\tau/2}\mathbb{E}\left(1_{S_t}\exp\left[\frac{64}{\tau}\sum_{r=0}^{m-t-1}\frac{\max(\Psi_r(\mathcal{N})-\psi_r,0)}{(2m-2r)^2}\right]\right)^{\tau/2}.$$

From Corollary 1 the second term in the above product is $1+o(1)$ and we only need to show

$$2^{2t} d_{\max}^{2t}\mathbb{P}(S_t)^{1-\tau/2} \leq \Big(1+o(1)\Big)\frac{1}{m^{\tau t}}.$$

Now using the bound given by (52) for $\mathbb{P}(S_t)$ we have

$$2^{2t} d_{\max}^{2t} 2t^t\mathbb{P}(S_t)^{1-\tau/2} \leq (1+o(1))2t\left(\frac{2^{4-\tau/3}}{2-\tau}\frac{d_{\max}^{4-5\tau+2\tau^2}}{m^{1-2.5\tau+\tau^2}}\right)^t$$

$$\leq (1+o(1))2t\left(4\frac{d_{\max}^{4-5\tau+2\tau^2}}{m^{1-2.5\tau+\tau^2}}\right)^t$$

$$\leq (1+o(1))2t\left(4m^{-2.75\tau+3.5\tau^2-2\tau^3}\right)^t$$

$$\leq O(m^{-\tau t}). \qquad \square$$

Acknowledgements We would like to thank Joe Blitzstein, Persi Diaconis, Adam Guetz, Milena Mihail, Alistair Sinclair, Eric Vigoda and Ying Wang for insightful discussions and useful comments on earlier version of this paper. We also thank the anonymous referees for their great comments and suggestions. M. Bayati was supported by Microsoft Technical Computing Iniciative. J.H. Kim was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government(MOST) (No. R16-2007-075-01000-0) and the second stage of the Brain Korea 21 Project in 2007. A. Saberi thanks the support of NSF.

References

1. Alderson, D., Doyle, J., Willinger, W.: Toward and optimization-driven framework for designing and generating realistic Internet topologies. HotNets (2002)
2. Alon, N., Spencer, J.: The Probabilistic Method. Wiley, New York (1992)
3. Amraoui, A., Montanari, A., Urbanke, R.: How to find good finite-length codes: from art towards science. Preprint, cs.IT/0607064 (2006)
4. Bassetti, F., Diaconis, P.: Examples comparing importance sampling and the Metropolis algorithm (2005)
5. Bayati, M., Montanari, A., Saberi, A.: Generating random graphs with large girth. In: ACM-SIAM Symposium on Discrete Algorithms (SODA) (2009)
6. Bender, E.A., Canfield, E.R.: The asymptotic number of labeled graphs with given degree sequence. J. Comb. Theory Ser. A **24**(3), 296–307 (1978)
7. Bezáková, I., Bhatnagar, N., Vigoda, E.: Sampling binary contingency tables with a greedy start. In: Symposium on Discrete Algorithms (SODA) (2006)
8. Bezáková, I., Sinclair, A., Štefankovič, D., Vigoda, E.: Negative examples for sequential importance sampling of binary contingency tables. In: Proceedings of Annual European Symposium, vol. 14. Lecture Notes in Computer Science, vol. 4168, pp. 136–147. Springer, Berlin (2006)
9. Blanchet, J.: Efficient importance sampling for binary contingency tables. Ann. Appl. Probab. **19**(3), 949–982 (2009)
10. Blitzstein, J., Diaconis, P.: A sequential importance sampling algorithm for generating random graphs with prescribed degrees. Ann. Appl. Probab. (2005, submitted)
11. Bollobás, B.: A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. Eur. J. Comb. **1**(4), 311–316 (1980)
12. Bu, T., Towsley, D.: On distinguishing between Internet power law topology generator. In: INFOCOM (2002)
13. Chen, Y., Diaconis, P., Holmes, S., Liu, J.S.: Sequential Monte Carlo methods for statistical analysis of tables. J. Am. Stat. Assoc. **100**, 109–120 (2005)
14. Chung, F., Lu, L.: Connected components in random graphs with given expected degree sequence. Ann. Comb. **6**(2), 125–145 (2002)
15. Cooper, C., Dyer, M., Greenhill, C.: Sampling regular graphs and peer-to-peer network. Comb. Probab. Comput. **16** (2007)
16. Diaconis, P., Gangolli, A.: Rectangular arrays with fixed margins. In: Discrete Probability and Algorithms, Minneapolis, MN, 1993. IMA Volumes in Mathematics and Its Applications, vol. 72, pp. 15–41. Springer, New York (1995)
17. Faloutsos, M., Faloutsos, P., Faloutsos, C.: On power-law relationships of the Internet topology. In: SIGCOM (1999)
18. Gkantsidis, C., Mihail, M., Zegura, E.: The Markov chain simulation method for generating connected power law random graphs. Alenex (2003)

19. Jerrum, M., Sinclair, A.: Approximate counting, uniform generation and rapidly mixing Markov chains. Inf. Comput. **82**(1), 93–133 (1989)
20. Jerrum, M., Sinclair, A.: Fast uniform generation of regular graphs. Theor. Comput. Sci. **73**(1), 91–100 (1990)
21. Jerrum, M., Valiant, L., Vazirani, V.: Random generation of combinatorial structures from a uniform distribution. Theor. Comput. Sci. **43**, 169–188 (1986)
22. Jerrum, M., Sinclair, A., McKay, B.: When is a graphical sequence stable? In: Random Graphs, Poznań, 1989, vol. 2, pp. 101–115. Wiley-Interscience, New York (1992)
23. Kannan, R., Tetali, P., Vempala, S.: Simple Markov chain algorithms for generating bipartite graphs and tournaments. Random Struct. Algorithms **14**, 293–308 (1999)
24. Kim, J.H.: On Brooks' theorem for sparse graphs. Comb. Probab. Comput. **4**, 97–132 (1995)
25. Kim, J.H., Vu, V.H.: Concentration of multivariate polynomials and its applications. Combinatorica **20**(3), 417–434 (2000)
26. Kim, J.H., Vu, V.H.: Generating random regular graphs. In: STOC 2003, pp. 213–222
27. Kim, J.H., Vu, V.: Sandwiching random graphs. Adv. Math. **188**, 444–469 (2004)
28. Knuth, D.: Mathematics and computer science: coping with finiteness. Science **194**(4271), 1235–1242 (1976)
29. McKay, B.: Asymptotics for symmetric 0-1 matrices with prescribed row sums. Ars Combinatoria A **19**, 15–25 (1985)
30. McKay, B., Wormald, N.C.: Uniform generation of random regular graphs of moderate degree. J. Algorithms **11**(1), 52–67 (1990)
31. McKay, B., Wormald, N.C.: Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$. Combinatorica **11**(4), 369–382 (1991)
32. Medina, A., Matta, I., Byers, J.: On the origin of power laws in Internet topologies. ACM Comput. Commun. Rev. **30**(2), 18–28 (2000)
33. Milo, R., ShenOrr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U.: Network motifs: simple building blocks of complex networks. Science **298**, 824–827 (2002)
34. Milo, R., Kashtan, N., Itzkovitz, S., Newman, M., Alon, U.: On the uniform generation of random graphs with prescribed degree sequences. http://arxiv.org/PS_cache/cond-mat/pdf/0312/0312028.pdf (2004)
35. Molloy, M., Reed, B.: A critical point for random graphs with a given degree sequence. Random Struct. Algorithms **6**(2–3), 161–179 (1995)
36. Sinclair, A.: Personal communication (2006)
37. Steger, A., Wormald, N.C.: Generating random regular graphs quickly. Comb. Probab. Comput. **8**(4), 377–396 (1997) (English summary in Random Graphs and Combinatorial Structures, Oberwolfach)
38. Tangmunarunkit, H., Govindan, R., Jamin, S., Shenker, S., Willinger, W.: Network topology generators: degree based vs. structural. In: ACM SIGCOM (2002)
39. Vu, V.H.: Concentration of non-Lipschitz functions and applications, Probabilistic methods in combinatorial optimization. Random Struct. Algorithms **20**(3), 267–316 (2002)
40. Wormald, N.C.: Models of random regular graphs. In: Surveys in Combinatorics. Canterbury. London Mathematical Society Lecture Note Series, vol. 265, pp. 239–298. Cambridge University Press, Cambridge (1999)