Generating random regular graphs quickly

A. STEGER[1] and N. C. WORMALD[2 †]

[1] Institut für Informatik, Technische Universität München,
D-80290 München, Germany

[2] Department of Mathematics and Statistics, University of Melbourne
Parkville, VIC 3052, Australia

Received

We present a practical algorithm for generating random regular graphs. For all d growing as a small power of n, the d-regular graphs on n vertices are generated approximately uniformly at random, in the sense that all d-regular graphs on n vertices have in the limit the same probability as $n \to \infty$. The expected runtime for these d's is $\mathcal{O}(nd^2)$.

1. Introduction

There are various algorithms known for generating graphs with n vertices of given degrees uniformly at random. Unfortunately, none of them is of practical use for all degree sequences, even for those with all degrees equal. In this paper we examine an algorithm which, although it does not generate uniformly at random, is provably close to a uniform generator when the degrees are relatively small. Moreover, it is easy to implement and quite fast in practice.

The most interesting case is the regular one, when all degrees are equal to $d = d(n)$ say. Moreover, methods for the regular case of this problem usually extend to arbitrary degree sequences, although the analysis can become more complicated and it may be needed to impose restrictions on the variation in the degrees (such as is analyzed by Jerrum et al. [4]). The first algorithm for generating d-regular graphs uniformly at random was implicit in the paper of Bollobás [2] and also in the approaches to counting regular graphs by Bender and Canfield [1] and in [13] (see also [14] for explicit algorithms). The *configuration* or *pairing* model of random d-regular graphs is as follows. Start with nd points (nd even) in n groups, and choose a random pairing of the points. Then create a graph with an edge from i to j if there is a pair containing points in the i'th and j'th groups. If no duplicated edge or loop (i.e., a pair of points in the same group) occurs,

† Research supported by the Australian Research Council

the resulting d-regular graphs occur uniformly at random. For graphs on n vertices this takes expected time of the order of $nde^{(d^2-1)/4}$ per graph, at least for d up to $n^{1/3}$, so is not polynomial time unless $d = \mathcal{O}(\sqrt{\log n})$. In [9] a polynomial expected time uniform generation algorithm was given for $d = \mathcal{O}(n^{1/3})$, but the expected running time per graph is $\mathcal{O}(n^2 d^4)$. This is rather complicated to implement. Also, it applies to arbitrary degree sequences. Another algorithm was given specifically for regular graphs which reduces the time to $\mathcal{O}(nd^3)$, however this is prohibitively difficult to implement, and again it only applies for $d = \mathcal{O}(n^{1/3})$.

The need for generating such graphs can also be met by simpler algorithms which do not generate the graphs uniformly at random. For example, Tinhofer [12] gives one. However, these algorithms are not easy to analyse, and the resulting probability distribution can be virtually unknown. As discussed in [12], one can achieve uniformity by an accept/reject procedure, but the inherent difficulties in analysis mean that no such algorithms are yet known which are of practical use for uniform generation.

On the other hand, Jerrum and Sinclair [5] provided an approximately uniform generation algorithm, which runs in time polynomial in n and $1/\epsilon$, where all graphs have probabilites varying by a factor of at most $1+\epsilon$. They do not precisely analyse the running time of there algorithm, nor do they claim that their algorithm is of practical use.

There are two ways to describe the algorithm in the present paper. It can be regarded as a modification of the pairing algorithm described above. First, we define two points to be *suitable* if they lie in different groups and no currently existing pair contains points in the same two groups. Our algorithm is the following.

Algorithm 1

1 Start with nd points $\{1, 2, \ldots, nd\}$ (nd even) in n groups. Put $U = \{1, 2, \ldots, nd\}$. (U denotes the set of unpaired points.)

2 Repeat the following until no suitable pair can be found: Choose two random points i and j in U, and if they are suitable, pair i with j and delete i and j from U.

3 Create a graph G with edge from vertex r to vertex s if and only if there is a pair containing points in the r'th and s'th groups. If G is d-regular, output it, otherwise return to Step 1.

However, the algorithm actually arose from extensions of the algorithm examined in [11]. There, one begins with n vertices and continually selects a random edge to add, subject to keeping all vertices of degree at most d (and no multiple edges). It was shown in [11] that for fixed d and dn even, this almost surely produces a d-regular graph (as $n \to \infty$). In [6] this process is modified by selecting the edges non-uniformly. One interesting choice of the non-uniformity studied there gives the following algorithm.

Algorithm 2

1 Start with a graph G with n vertices $\{1, 2, \ldots, n\}$ and no edges.

2 Repeat the following until the set S is empty: Let S denote the set of pairs of vertices of G which are non-adjacent and both have degree at most $d-1$. Choose a random

pair $\{u, v\}$ in S with probability proportional to $(d - d(u))(d - d(v))$ where d denotes the degree in G. Add the edge $\{u, v\}$ to G.

3 If G is d-regular, output it, otherwise return to Step 1.

Note that the probabilities of edges in Algorithm 2 are exactly the probabilities of points being chosen between the corresponding groups in Algorithm 1, so the two algorithms are equivalent.

Our results in Section 2 show that Algorithm 2 generates graphs with nearly uniform probability distribution, in the sense that as $n \to \infty$, provided d does not grow too quickly with n, the probabilities of *all* graphs are asymptotically equal. For larger d, but still not very large, we show the algorithm generates d-regular graphs with a distribution which is close to uniform, in the sense that the probability of any event is different from that in the uniform space by $o(1)$. In Section 4 we show that the expected time for Step 2 of Algorithm 1 is $\mathcal{O}(nd^2 + d^4)$ for all d. Moreover, the probability it produces a regular graph (which is then accepted at Step 3) is $1 - o(1)$ for $d = o((n/(\log n)^3)^{\frac{1}{11}})$, and so the number of repetitions of Step 2 required is $\mathcal{O}(1)$ for such d. In fact, it is quite possibly bounded for all $d \leq n/2$.

1.1. Notation and Preliminary Results. We use $\log x$ to denote the natural logarithm, and assume $d \geq 2$ throughout the paper. By $\mathcal{R}(n, d)$ we denote the set of all labelled d-regular graphs on n vertices.

We can view Step 2 of Algorithm 2 as the random selection of a sequence $x_1, \ldots, x_k$ of edges on n vertices. We call such a sequence a *path* because it is a possible path for the course of the algorithm. For each path $\mathcal{P}$, we can consider the probability that one run of Step 2 of Algorithm 2 produces $\mathcal{P}$. This induces a probability space whose elements are paths, which we call the A-*model*. Exactly the same space is produced by Algorithm 1. Note that the edges in such a path determine a maximal graph with no vertices of degree greater than d.

We need to compare this with the pairing model, which can be defined similarly to Algorithm 1, but with Step 2 accepting two points i and j even when they are not suitable. By the U-*model* we mean the probability space of paths induced by one run of Step 2 in this modified algorithm. (Actually, pairs of points are chosen in Step 2, and the corresponding edges should be determined by the mechanism in Step 3 of Algorithm 1, at which stage loops and multiple edges may be formed.) Note that in practice one could restart the algorithm as soon as unsuitable points were found, but for reasons of taste in this analysis we permit Step 2 to run its course in the modified algorithm until all points are paired.

Let G be a d-regular graph. By $Paths(G)$ we denote the set of all orderings of the set of edges of G; that is, "paths" whose edges are precisely the edges of G. Then

$$|Paths(G)| = \left(\tfrac{1}{2}nd\right)!.$$

For a path $\mathcal{P} \in Paths(G)$ we use

$$\mathbf{P}_A[\mathcal{P}] \qquad \text{resp.} \qquad \mathbf{P}_U[\mathcal{P}]$$

to denote the probability of $\mathcal{P}$ in the A-model resp. in the U-model. Note that in the

U-model all paths $\mathcal{P} \in Paths(G)$ have, in fact, the same probability. Namely,

$$\mathbf{P}_U[\mathcal{P}] = \frac{2^{\frac{nd}{2}} \cdot (d!)^n}{(nd)!}.$$

On the other hand, denote the subgraph of G consisting of the first m edges of $\mathcal{P}$ by $G_m(\mathcal{P})$. If a path $\mathcal{P} \in Paths(G)$ has edges $x_1, \ldots, x_{dn/2}$ where edge x_i joins vertices u_i and v_i then

$$\mathbf{P}_A[\mathcal{P}] = \prod_{m=0}^{nd/2-1} \frac{(d - d_m(u_i))(d - d_m(v_i))}{\sum (d - d_m(u))(d - d_m(v))} \tag{1}$$

where the sum in the denominator is over all $u \neq v \in V(G)$ such that $\{u, v\} \notin E(G_m)$, and d_m denotes the degree in G_m. For comparison, in the pairing model for generating regular graphs uniformly, after m pairs have been added there are $nd - 2m$ unmatched points remaining. Also, the number of pairs corresponding to a valid edge $\{u, v\}$ is $(d - d_m(u))(d - d_m(v))$. Hence, the probability in the U-model can be written as

$$\mathbf{P}_U[\mathcal{P}] = \prod_{m=0}^{nd/2-1} \frac{(d - d_m(u_i))(d - d_m(v_i))}{\binom{nd-2m}{2}}. \tag{2}$$

For a subgraph H of G we denote by $\Delta(H)$ the difference in the normalizing denominators in (1) and (2) for the factors due to $G_m = H$. That is, we let

$$\Delta(H) = \Delta^{(1)}(H) + \Delta^{(2)}(H),$$

where

$$\Delta^{(1)}(H) := \sum_{v \in V(G)} \binom{d - d_H(v)}{2}. \tag{3}$$

and

$$\Delta^{(2)}(H) := \sum_{\{u,v\} \in E(H)} (d - d_H(u))(d - d_H(v)) \tag{4}$$

For brevity we use $\Delta_m(\mathcal{P})$ for $\Delta(G_m(\mathcal{P}))$. Observe from (1) and (2) and by considering the number of acceptable pairs of points in Step 2 of Algorithm 1 that

$$\frac{\mathbf{P}_A[\mathcal{P}]}{\mathbf{P}_U[\mathcal{P}]} = \prod_{m=0}^{\frac{nd}{2}-1} \frac{\binom{nd-2m}{2}}{\binom{nd-2m}{2} - \Delta_m(\mathcal{P})}. \tag{5}$$

We will find it useful to know the value of $\mathbf{P}_U[G]$ approximately. The following estimate of the size of $|\mathcal{R}(n, d)|$ was found first in the special case of bounded d by Bender and Canfield [1]. Bollobás [2] proved it for all $d \leq \sqrt{2 \log n} - 1$ by considering the probability of obtaining no loops or mulitiple edges in the pairing model. Then McKay analysed the same model for substantially larger d.

Theorem 1.1. *(McKay [8]) Assume $d = o(n^{1/3})$. Then*

$$|\mathcal{R}(n, d)| = (1 + o(1)) \cdot e^{-\frac{1}{4}(d^2-1)} \cdot \frac{(nd)!}{(\frac{nd}{2})! \cdot 2^{\frac{nd}{2}} \cdot (d!)^n}.$$

Correction terms are required if the upper bound on d is relaxed further — see McKay and Wormald [10] for the formula when $d = o(\sqrt{n})$.

The formula above was obtained by estimating the probability in the configuration model of having no loops or multiple edges, so although it is working back-to-front we can deduce the following.

Corollary 1.1. *Assume $d = o(n^{1/3})$. Then*

$$\mathbf{P}_U[G] = (1 + o(1)) \cdot e^{-\frac{1}{4}(d^2-1)}/|\mathcal{R}(n,d)|.$$

Proof. We just observe that

$$\mathbf{P}_U[G] = \sum_{\mathcal{P} \in Paths(G)} \mathbf{P}_U[\mathcal{P}]$$

and

$$\mathbf{P}_U[\mathcal{P}] = 2^{\frac{nd}{2}}(d!)^n/(nd)! \qquad \text{for all paths } \mathcal{P} \in Paths(G).$$

□

2. Main Results

Our first result shows that the probabilities of all graphs generated by Algorithm 1 are asymptotically equal, provided d grows slowly enough with n.

Theorem 2.1. *Assume that $d = O(n^{\frac{1}{28}})$. Then there exists a function $f(n,d) = o(1)$ such that all d-regular graphs G satisfy*

$$\left| \mathbf{P}_A[G] - \frac{1}{|\mathcal{R}(n,d)|} \right| < \frac{f(n,d)}{|\mathcal{R}(n,d)|}.$$

Our proof actually yields the same result for a slightly larger d. But we believe it is true even for much more quickly growing d.

The next result is useful if Algorithm 1 is to be used for estimating probabilities by simulation. The conclusion of this theorem is equivalent to the assertion that the total variation distance between the distribution of graphs given by Algorithm 1, and the uniform distribution, goes to 0 as $n \to \infty$.

Theorem 2.2. *Assume that $d = o((n/(\log n)^3)^{\frac{1}{11}})$. Then there exists a function $f(n,d)$ $= o(1)$ and a subset $\mathcal{X} \subseteq \mathcal{R}(n,d)$ such that*

$$\mathbf{P}_A[G] = (1 + \mathcal{O}(f(n,d))) \cdot \frac{1}{|\mathcal{R}(n,d)|} \qquad \textit{for all graphs } G \in \mathcal{X}$$

and

$$|\mathcal{X}| = (1 - f(n,d)) \cdot |\mathcal{R}(n,d)|.$$

Observe that Theorem 2.2 implies that one can experimentally determine the probability of an event $\mathcal{E}$ defined on the set of d-regular graphs by simulation using Algorithm 1. We formulate this in a slightly more general setting:

Corollary 2.1. *Assume that $d = o((n/(\log n)^3)^{\frac{1}{11}})$ and let X be a bounded function of graphs defined on the sets $\mathcal{R}(n,d)$ for all n and d. Then*

$$\sum_{G\in\mathcal{R}(n,d)} X(G)\cdot \mathbf{P}_U[G] \;-\; \sum_{G\in\mathcal{R}(n,d)} X(G)\cdot \mathbf{P}_A[G] \;=\; o(1).$$

That is, one can compute the expectation to within $o(1)$ error by simulation using Algorithm 1.

For practical use of the algorithm it is comforting to know the following.

Theorem 2.3. *Assume that $d = o((n/(\log n)^3)^{\frac{1}{11}})$. Then the probability that Step 2 of Algorithm 1 produces a regular graph is asymptotic to 1 as $n\to\infty$.*

This result is in a similar direction to the main theorem of [11], which was only proved for bounded d, the process in that paper being rather different.

3. Proof

3.1. Outline. Fix a d-regular graph G on n vertices. In the U-model $\mathbf{P}_U[\mathcal{P}]$ is constant for all paths $\mathcal{P}\in Paths(G)$. This is obviously not true in the A-model. Here the probability $\mathbf{P}_A[\mathcal{P}]$ depends strongly on the *order* of the edges. We can however estimate an expected or average probability by considering a path chosen uniformly at random from all possible paths. For now just assume that $av(G)$ denotes such an estimate. In Section 3.3 we will assign a particular value to $av(G)$.

Our proof strategy is as follows. We partition the set $Paths(G)$ into suitable subsets

$$Paths(G) = \mathcal{M}_1(G)\cup\mathcal{M}_2(G)\cup\mathcal{A}v(G)$$

such that *i*) the sets $\mathcal{M}_i(G)$, which contain in some sense "misbehaving" paths, are "small" and *ii*) paths in $\mathcal{A}v(G)$ have a probability which is roughly equal to $av(G)$. We now make these ideas precise.

Lemma 3.1. *Let $d = d(n)$ and let $g(n,d)$ be a positive function with $g(n,d)\to\infty$. Assume there exists a positive function $av(G)$ such that for all d-regular graphs G there is a partition of the set $Paths(G)$ as above such that:*

$$\mathbf{P}_A[\mathcal{P}] \le e^{g(n,d)}\cdot av(G) \qquad \text{for all } \mathcal{P}\in Paths(G), \tag{6}$$

$$|\mathcal{M}_k(G)| = o(e^{-g(n,d)})\cdot |Paths(G)|, \tag{7}$$

and

$$\mathbf{E}(\mathbf{P}_A[\mathcal{P}]) = (1+o(1))\cdot av(G) \tag{8}$$

for $\mathcal{P}$ chosen uniformly at random from $\mathcal{A}v(G)$. Then

$$\mathbf{P}_A[G] = (1+o(1)) \cdot av(G) \cdot |Paths(G)|. \tag{9}$$

Proof.

$$\begin{aligned}
\mathbf{P}_A[G] &= \sum_{\mathcal{P}\in Paths(G)} \mathbf{P}_A[\mathcal{P}] \\
&= \sum_{\mathcal{P}\in \mathcal{M}_1(G)\cup\mathcal{M}_2(G)} \mathbf{P}_A[\mathcal{P}] + \sum_{\mathcal{P}\in \mathcal{A}v(G)} \mathbf{P}_A[\mathcal{P}] \\
&= o(|Paths(G)|) \cdot av(G) + (1+o(1)) \cdot av(G) \cdot |\mathcal{A}v(G)| \\
&= (1+o(1)) \cdot av(G) \cdot |Paths(G)|.
\end{aligned}$$

□

As we know that $|Paths(G)| = \left(\frac{nd}{2}\right)!$, Theorems 2.1 and 2.2 will follow easily from Lemma 3.1 if we can show that $av(G)$ is independent of the graph G and has the correct order of magnitude. This we will show in Section 3.3.

Throughout the remaining part of this section we assume that G is an arbitrary, but fixed d-regular graph.

3.2. Some Useful Lemmas. In this section we collect some helpful lemmas and facts that will eventually enable us to bound the ratio given in (5) for a randomly chosen path $\mathcal{P} \in Paths(G)$.

Lemma 3.2. *Let H be a subgraph of G and let $m = |E(H)|$. Then*

$$\Delta^{(1)}(H) \le \frac{1}{2}(d-1)(nd-2m).$$

Proof. By Jensen's inequality the sum in (3) is bounded from above by $\frac{nd-2m}{d} \cdot \binom{d}{2}$. □

Lemma 3.3. *Let H be a subgraph of G and let $m = |E(H)|$. Then*

$$\Delta^{(2)}(H) \le \frac{1}{2}(d-1)^2(nd-2m).$$

Proof.

$$\sum_{\{u,v\}\in E(H)} (d-d_H(u)) \cdot (d-d_H(v)) \quad = \quad \frac{1}{2}\sum_x \sum_{u:\{x,u\}\in E(G)\setminus E(H)} \sum_{v\in\Gamma_H(u)} d_{G\setminus H}(v).$$

Observe that $d_{G\setminus H}(v) \le d-1$ for all $v \in \Gamma_H(u)$ and that there at most $d-1$ such v's. □

Corollary 3.1. *All paths $\mathcal{P}$ in $Paths(G)$ satisfy*

$$\Delta_m(\mathcal{P}) \le \frac{1}{2}d^2(nd-2m) \qquad \text{for all } 0 \le m \le \tfrac{1}{2}nd.$$

The following tail bound on the sum of independent random variables is due to Hoeffding [3].

Hoeffding's inequality. *Let $X_1, \ldots, X_n$ be independent random variables with $0 \le X_i \le 1$ for all $1 \le i \le n$, and let $X = \sum_{i=1}^n X_i$, $p = \frac{1}{n}\mathbf{E}X$ and $q = 1 - p$. Then for $0 \le t < q$*

$$\mathbf{P}[X - pn \ge tn] \le \left(\left(\frac{p}{p+t} \right)^{p+t} \left(\frac{q}{q-t} \right)^{q-t} \right)^n .$$

A corollary of this will turn out to be very useful.

Corollary 3.2. *Let $X_1, \ldots, X_n$ be independent variables with $0 \le X_i \le 1$ for all $1 \le i \le n$, and let $X := \sum_{i=1}^n X_i$. Then*

$$\mathbf{P}[X > (1+\delta)\mathbf{E}X] \le \left(\frac{e^\delta}{(1+\delta)^{1+\delta}} \right)^{\mathbf{E}X} \le \begin{cases} e^{-\frac{1}{4}\delta^2 \mathbf{E}X} & \textit{if } \delta \le \frac{4}{5} \\ e^{-\frac{1}{4}\delta \mathbf{E}X} & \textit{otherwise} \end{cases}$$

and

$$\mathbf{P}[X < (1-\delta)\mathbf{E}X] \le e^{-\frac{1}{2}\delta^2 \mathbf{E}X}.$$

Proof. The first factor in Hoeffding's inequality can be written as $(1+\delta)^{-(1+\delta)p}$ and the second as $(1 + (\delta p)/(1 - p - \delta p))^{1-p-\delta p} < e^{\delta p}$. From here, the first inequality can be verified by comparing logarithms (noting that we can assume $t \le q$ because $X \le n$ always). The second is by McDiarmid [7, (5.6)], which he attributes in the binomial case to Angluin and Valiant. ☐

In fact by [7, (5.5)], it is possible to improve the factor $1/4$ in the exponent in the case $\delta \le 4/5$ to $1/3$, but we do not need this.

3.3. Average Paths. Our aim in this section is to estimate $\Delta_m(\mathcal{P})$ for a (uniformly) randomly chosen path $\mathcal{P} \in Paths(G)$. Observe that this just corresponds to the value of $\Delta_m(H)$ for a random subgraph H of G with m edges. To simplify calculations, however, we don't choose m edges randomly from G, but instead choose edges with probability $p = p(m) = \frac{2m}{nd}$. We denote such a random subgraph by RG_p. Note that we are *not* claiming that the expectation of Δ_m, which we are about to calculate, is the same for both models. In the following we simply calculate that expectation with respect to the model where the edges of G are chosen with probability $p = p(m)$. Later we draw conclusions about $\Delta_m(\mathcal{P})$ from this calculation.

To simplify notation we abbreviate in the rest of the paper $\mathbf{E}\Delta(RG_p)$ where $p = \frac{2m}{nd}$ by μ_m. Similarly, we let $\mu_m^{(i)} := \mathbf{E}\Delta^{(i)}(RG_p)$ for $p = \frac{2m}{nd}$ and $i = 1, 2$.

Lemma 3.4.

$$\mu_m^{(1)} = \frac{1}{2}(nd - 2m)^2 \cdot \frac{d-1}{nd}.$$

Proof. For every path of length 2 in G we introduce a 0/1 variable X_i which is equal to 1 iff both edges of the path do *not* belong to RG_p. As $\mathbf{E}X_i = (1-p)^2$ and there are $n \cdot \binom{d}{2}$ such paths we have

$$\mu_m^{(1)} = \mathbf{E}\textstyle\sum_i X_i = n \cdot \binom{d}{2} \cdot (1-p)^2 = n \cdot \binom{d}{2} \cdot (1 - \frac{2m}{nd})^2 = \frac{1}{2}(nd-2m)^2 \frac{d-1}{nd}.$$

□

Lemma 3.5.

$$\mu_m^{(2)} = (nd-2m)^2 \cdot \frac{m(d-1)^2}{n^2d^2}.$$

Proof. For every (ordered) tuple (x,u,v,y) such that $\{x,u\}, \{u,v\}$, and $\{v,y\}$ are edges of G we introduce a 0/1 variable Y_i which is equal to one iff the edge $\{u,v\}$ belongs to RG_p and both edges $\{x,u\}$ and $\{v,y\}$ do *not* belong to RG_p.

As $\mathbf{E}Y_i = p(1-p)^2$ and there are $nd(d-1)^2$ such tuples we have

$$2\mu_m^{(2)} = \mathbf{E}\textstyle\sum_i Y_i \;=\; nd(d-1)^2 \cdot p(1-p)^2 \;=\; nd(d-1)^2 \cdot \frac{2m}{nd}(1-\frac{2m}{nd})^2.$$

□

Corollary 3.3.

$$\mu_m = \frac{1}{2}(nd-2m)^2 \cdot \left(\frac{d-1}{nd} + \frac{2(d-1)^2m}{n^2d^2}\right).$$

After these preliminaries we are now in a position to precisely define $av(G)$, the probability of an "average" path in the A-model. Namely, we assume that the value Δ_m occurring in the denominators of the edge probabilities of the path is exactly equal to μ_m. To make the formula slightly nicer, we define it in terms of the probability of a path in the U-model, which we know is independent of the graph G and the actual path $\mathcal{P} \in Paths(G)$:

$$av(G) \;:=\; \mathbf{P}_U[\mathcal{P}] \cdot \prod_{m=0}^{\frac{nd}{2}-1} \frac{\binom{nd-2m}{2}}{\binom{nd-2m}{2} - \mu_m}. \tag{10}$$

Lemma 3.6. *The value $av(G)$ is independent of the graph G and, if $d = o(n^{1/3})$ then*

$$av(G) = (1+o(1)) \cdot e^{\frac{1}{4}(d^2-1)} \cdot \mathbf{P}_U[\mathcal{P}] = (1+o(1)) \cdot e^{\frac{1}{4}(d^2-1)} \cdot \frac{\mathbf{P}_U[G]}{|Paths(G)|}.$$

Proof. The independence of $av(G)$ from G follows immediately from the fact that the values μ_m are independent of G .

$$\begin{aligned}
\frac{av(G)}{\mathbf{P}_U[\mathcal{P}]} &= \prod_{m=0}^{\frac{nd}{2}-1} \left(1 + \frac{\mu_m}{\binom{nd-2m}{2} - \mu_m}\right) \\
&= \prod_{m=0}^{\frac{nd}{2}-1} \left(1 + \left(\frac{d-1}{nd} + \frac{2(d-1)^2m}{n^2d^2}\right) \Big/ \left(1 - \frac{1}{nd-2m} - \mathcal{O}(\frac{d}{n})\right)\right)
\end{aligned}$$

$$= \exp\left(\sum_{m=0}^{\frac{nd}{2}-1} \log\left(1+\left(\frac{d-1}{nd}+\frac{2(d-1)^2m}{n^2d^2}\right)\Big/\left(1-\frac{1}{nd-2m}-\mathcal{O}(\frac{d}{n})\right)\right)\right)$$

As $\log(1+x) = x - \mathcal{O}(x^2)$ for all $x \le 1$, we can bound the sum for all $d = o(n^{1/3})$ as follows:

$$\begin{aligned}
&\sum_{m=0}^{\frac{nd}{2}-1} \log\left(1+\left(\frac{d-1}{nd}+\frac{2(d-1)^2m}{n^2d^2}\right)\Big/\left(1-\frac{1}{nd-2m}-\mathcal{O}(\frac{d}{n})\right)\right)\\
&= \sum_{m=0}^{\frac{nd}{2}-1} \log\left(1+\frac{d-1}{nd}+\frac{2(d-1)^2m}{n^2d^2}+\mathcal{O}\left(\frac{d}{n}\cdot\left(\frac{1}{nd-2m}+\frac{d}{n}\right)\right)\right)\\
&= \sum_{m=0}^{\frac{nd}{2}-1} \left(\frac{d-1}{nd}+\frac{2(d-1)^2m}{n^2d^2}+\mathcal{O}\left(\frac{d^2}{n^2}+\frac{d}{n(nd-2m)}\right)\right)\\
&= \frac{1}{2}(d-1)+\frac{(d-1)^2}{n^2d^2}\left(\frac{nd}{2}-1\right)\frac{nd}{2}+\mathcal{O}\left(\frac{nd\cdot d^2}{n^2}+\frac{d}{n}\log(nd)\right)\\
&= \frac{1}{2}(d-1)+\frac{1}{4}(d-1)^2+o(1)\\
&= \frac{1}{4}(d^2-1)+o(1).
\end{aligned}$$

□

3.4. The Misbehaving Sets. In this section we define the sets of "misbehaving" paths. They will be characterized in terms of the deviation of their value $\Delta_m(\mathcal{P})$ from μ_m. Recall that we defined μ_m to represent, in a somewhat vague sense, the value of $\Delta_m(\mathcal{P})$ for a random path in $Paths(G)$. So we expect that "most" paths have values $\Delta_m(\mathcal{P})$ which are "close" to μ_m. In this section we will quantify this. Let

$\mathcal{M}_1(G) :=$ set of all paths $\mathcal{P} \in Paths(G)$ such that

$$|\mu_m^{(1)} - \Delta_m^{(1)}(\mathcal{P})| \ge \max\{d^4\sqrt{24\mu_m^{(1)}\cdot\log n}\,, 24d^5\log n\}$$

for some $0 \le m \le \frac{1}{2}nd-1$,

$\mathcal{M}_2(G) :=$ set of all paths $\mathcal{P} \in Paths(G)$ such that

$$|\mu_m^{(2)} - \Delta_m^{(2)}(\mathcal{P})| \ge \max\{d^5\sqrt{24\mu_m^{(2)}\cdot\log n}\,, 24d^6\log n\}$$

for some $0 \le m \le \frac{1}{2}nd-1$, and

Lemma 3.7. *For n sufficiently large,*

$$|\mathcal{M}_1(G)| \le e^{-5d^2\log n}\cdot|Paths(G)|.$$

Proof. We consider an arbitrary but fixed $0 \le m \le \frac{1}{2}nd-1$. Clearly, it suffices to show that for this m we have

$$\mathbf{P}[|\Delta_m^{(1)}(\mathcal{P}) - \mu_m^{(1)}| \ge \max\{d^4\sqrt{24\mu_m^{(1)}\cdot\log n}\,, 24d^5\log n\}] \le e^{-5d^2\log n}/(nd). \qquad (11)$$

If we choose a path $\mathcal{P} \in Paths(G)$ randomly, then $G_m(\mathcal{P})$ corresponds to a random subgraph of G with m edges. That is, if we denote with RG_m the random subgraph of G with m edges, we observe that (11) is equivalent to

$$\mathbf{P}[|\Delta^{(1)}(RG_m) - \mu_m^{(1)}| \geq \max\{d^4\sqrt{24\mu_m^{(1)} \cdot \log n}\,, 24d^5 \log n\}] \leq e^{-5d^2 \log n}/(nd).$$

Instead of the random graph RG_m we will work with the random subgraph RG_p of G with edge probability $p = \frac{2m}{nd}$. Observe that in this model every subgraph of G with m edges is equally likely and

$$\mathbf{P}[|E(RG_p)| = m] \;=\; \binom{\frac{1}{2}nd}{m} \cdot p^m \cdot (1-p)^{\frac{1}{2}nd - m} \;\geq\; e^{-\log n}$$

(For m and $\frac{1}{2}nd - m$ tending to infinity the inequality follows easily from Stirling's formula, for the remaining m's observe first that it suffices to consider the "small" m's, then use $\binom{\frac{1}{2}nd}{m} \geq (\frac{nd}{2m})^m$ and the fact that $(1 - \frac{2m}{nd}) \geq e^{-2\log n}$ for m small enough.)

Hence, it suffices to show

$$\mathbf{P}\left[|\Delta^{(1)}(RG_p) - \mu_m^{(1)}| \geq \max\left\{d^4\sqrt{24\mu_m^{(1)} \cdot \log n}\,, 24d^5 \log n\right\}\right] \leq e^{-5d^2 \log n - \log n}/(nd).$$

This we will now do.

Observe that (by Brooks' theorem, for example) we can color the vertices of G with $d+1$ colors in such a way that no two vertices with the same color are adjacent.

We handle each color class separately. Let X_i denote the number of tuples (x, u, y) such that $x \neq y$, $\{x, u\}$ and $\{u, y\}$ belong to $E(G) \setminus E(RG_p)$ and u belongs to the ith color class. Let $X = \sum_{i=1}^{d+1} X_i$. Then $\mathbf{E}X$ is equal to $2\Delta^{(1)}(RG_p)$ and we therefore have to show that

$$\mathbf{P}[|X - \mathbf{E}X| \geq \max\{2d^4\sqrt{24\mu_m^{(1)} \cdot \log n}\,, 48d^5 \log n\}] \leq e^{-5d^2 \log n - \log n}/(nd).$$

Clearly, it suffices to show that for all $1 \leq i \leq d+1$

$$\mathbf{P}[|X_i - \mathbf{E}X_i| \geq \max\{d^3\sqrt{24\mathbf{E}X \cdot \log n}\,, 24d^4 \log n\}] \leq 2e^{-6d^2 \log n}.$$

Let n_i be the number of vertices in the ith color class. As X_i can be written as the sum of n_i independent variables with values in the interval $[0, d(d-1)]$ we may apply the Hoeffding bounds for the random variables $Y_i := X_i/d^2$ and $Y := X/d^2$.

First we handle those m's for which $\mathbf{E}Y \geq 24d^2 \log n$. Here we apply Corollary 3.2 with $\delta = d\sqrt{24\mathbf{E}Y \cdot \log n}/\mathbf{E}Y_i$:

$$\begin{aligned}\mathbf{P}[Y_i \geq \mathbf{E}Y_i + d\sqrt{24\mathbf{E}Y \cdot \log n}] &= \mathbf{P}[Y_i \geq (1+\delta)\mathbf{E}Y_i] \\ &\leq \begin{cases} e^{-\frac{1}{4}\cdot 24d^2 \mathbf{E}Y \log n/\mathbf{E}Y_i} \leq e^{-6d^2 \log n} & \text{if } \delta \leq \frac{4}{5} \\ e^{-\frac{1}{4}\cdot d\sqrt{24\mathbf{E}Y \cdot \log n}} \leq e^{-6d^2 \log n} & \text{otherwise} \end{cases}\end{aligned}$$

and

$$\begin{aligned}&\mathbf{P}[Y_i \leq \mathbf{E}Y_i - d\sqrt{24\mathbf{E}Y \cdot \log n}] = \mathbf{P}[Y_i \leq (1-\delta)\mathbf{E}Y_i] \\ &\leq \begin{cases} e^{-\frac{1}{2}\cdot 24d^2 \mathbf{E}Y \log n/\mathbf{E}Y_i} \leq e^{-12d^2 \log n} & \text{if } \mathbf{E}Y_i \geq d\sqrt{24\mathbf{E}Y \cdot \log n} \\ 0 & \text{otherwise.} \end{cases}\end{aligned}$$

For m's such that $\mathbf{E}Y \le 24d^2 \log n$ we just apply the first of the inequalities in Corollary 3.2 with $\delta = 24d^2 \log n/\mathbf{E}Y_i \ge 1$ to deduce

$$\begin{aligned}\mathbf{P}[|Y_i - \mathbf{E}Y_i| \ge 24d^2 \log n] &= \mathbf{P}[Y_i \ge \mathbf{E}Y_i + 24d^2 \log n] \\ &= \mathbf{P}[Y_i \ge (1+\delta)\mathbf{E}Y_i] \ \le \ e^{-6d^2 \log n}.\end{aligned}$$

□

Lemma 3.8.

$$|\mathcal{M}_2(G)| \le e^{-5d^2 \log n} \cdot |Paths(G)|.$$

Proof. Proceeding as in the proof of Lemma 3.7 we first observe that it suffices to show that

$$\mathbf{P}[|\Delta^{(2)}(RG_p) - \mathbf{E}\Delta^{(2)}_m| \ge \max\{d^5\sqrt{24\mathbf{E}\Delta^{(2)}_m \cdot \log n}, 24d^6 \log n\}] \le e^{-5d^2 \log n - 2\log n}/(nd).$$

Observe that we can color the edges of G with $2d(d-1)+1$ colors in such a way that any two edges of the same color are not connected by an edge in G, *i.e.* every color class is the edge set of an *induced matching* in G. To see this consider the graph where every edge of G corresponds to a vertex and every such vertex is connected to all vertices corresponding to edges which have to be colored differently. As the maximum degree of this graph is $2d(d-1)$ the claim follows from Brook's theorem.

With this fact in hand we can now again proceed similarly to the proof of Lemma 3.7. Let X_i denote the number of tuples (x,u,v,y) such that $\{x,u\}$ and $\{v,y\}$ belong to $E(G) \setminus E(RG_p)$, while $\{u,v\}$ belongs to the ith color class of G and to RG_p. Let $X = \sum_{i=1}^{2(d-1)^2+1} X_i$. Then $\mathbf{E}X$ is equal to $2\Delta^{(2)}(RG_p)$ and it therefore suffices to show that for all $1 \le i \le 2(d-1)^2+1$

$$\mathbf{P}[|X_i - \mathbf{E}X_i| \ge \max\{d^3\sqrt{24\mathbf{E}X \cdot \log n}\ , 24d^4 \log n\}] \le 2e^{-6d^2 \log n}.$$

By the choice of the coloring we again have that each X_i is the sum of independent variables with values in the interval $[0,(d-1)^2]$. An application of Corollary 3.2 similar to the one in the proof of Lemma 3.7 thus concludes the proof of Lemma 3.8. □

3.5. Proof of the Main Theorems. Our aim in this section is to verify the remaining hypotheses of Lemma 3.1, with "misbehaving" sets $\mathcal{M}_i(G)$ defined as in the previous section. We start with some auxiliary lemmas.

Lemma 3.9. *Let $\xi = \xi(n,d)$ be an arbitrary function such that $\xi \ge d^2+1$. For all paths $\mathcal{P} \in Paths(G)$:*

$$\prod_{m=\frac{1}{2}nd-\xi}^{\frac{1}{2}nd-1} \frac{\binom{nd-2m}{2} - \mu_m}{\binom{nd-2m}{2} - \Delta_m(\mathcal{P})} \le e^{d^2 \log(4\xi d^4)}.$$

Proof. Observe first that

$$\frac{\binom{nd-2m}{2}-\mu_m}{\binom{nd-2m}{2}-\Delta_m(\mathcal{P})} \le \binom{nd-2m}{2}-\mu_m \le \binom{2d^2}{2} \qquad \text{for all } \tfrac{1}{2}nd-d^2 \le m \le \tfrac{1}{2}nd-1.$$

Using Corollary 3.1 twice we furthermore observe that

$$\frac{\Delta_m(\mathcal{P})}{\binom{nd-2m}{2}-\Delta_m(\mathcal{P})} \le \frac{\Delta_m(\mathcal{P})}{\frac{1}{4}(nd-2m)^2} \le \frac{2d^2}{nd-2m} \qquad \text{for all } \tfrac{1}{2}nd-\xi \le m \le \tfrac{1}{2}nd-(d^2+1).$$

Hence,

$$\begin{aligned}
\prod_{m=\frac{1}{2}nd-\xi}^{\frac{1}{2}nd-1} \frac{\binom{nd-2m}{2}-\mu_m}{\binom{nd-2m}{2}-\Delta_m(\mathcal{P})} &\le \binom{2d^2}{2}^{d^2} \cdot \prod_{m=\frac{1}{2}nd-\xi}^{\frac{1}{2}nd-d^2-1} \left(1+\frac{\Delta_m(\mathcal{P})}{\binom{nd-2m}{2}-\Delta_m(\mathcal{P})}\right) \\
&\le (2d^4)^{d^2} \cdot \exp\left(\sum_{m=\frac{1}{2}nd-\xi}^{\frac{1}{2}nd-d^2-1} \frac{\Delta_m(\mathcal{P})}{\binom{nd-2m}{2}-\Delta_m(\mathcal{P})}\right) \\
&\le \exp(d^2\log(2d^4)) \cdot \exp\left(\sum_{i=d^2+1}^{\xi} \frac{d^2}{i}\right) \\
&\le \exp(d^2\log(2d^4) + d^2\log(\xi+1)).
\end{aligned}$$

□

Corollary 3.4. *Assume that $d = o(n^{1/3})$. Then*

$$\mathbf{P}_A[\mathcal{P}] \le e^{4d^2\log n} \cdot av(G) \qquad \textit{for all paths } \mathcal{P} \in Paths(G).$$

Proof. The expression in Lemma 3.9 for $\xi = \frac{1}{2}nd$ is

$$\frac{\mathbf{P}_A[\mathcal{P}]}{av(G)} \le \exp(d^2\log(2nd^5)).$$

□

Lemma 3.10. *All paths $\mathcal{P} \in \mathcal{A}v(G)$ satisfy*

$$\frac{|\Delta_m(\mathcal{P})-\mu_m|}{\binom{nd-2m}{2}-\Delta_m(\mathcal{P})} \le \frac{200d^{11/2}\sqrt{\log n}}{\sqrt{n}\cdot(nd-2m)} + \frac{200d^6\log n}{(nd-2m)^2} \qquad \textit{for all } m \le \tfrac{1}{2}nd-(d^2+1).$$

Proof. First we observe that Corollary 3.1 implies that

$$\binom{nd-2m}{2} - \Delta_m(\mathcal{P}) \ge \frac{1}{4}(nd-2m)^2 \qquad \text{for all } m \le \tfrac{1}{2}nd-(d^2+1).$$

Consider an arbitrary path $\mathcal{P} \in \mathcal{A}v(G)$. From the definition of $\mathcal{M}_1(G)$ and Lemma 3.4 we know

$$|\Delta_m^{(1)}(\mathcal{P}) - \mu_m^{(1)}| \quad \le \quad \begin{cases} 24d^5\log n & \text{if } \mu_m^{(1)} \le 24d^2\log n \\ 24d^4(nd-2m)\sqrt{\log n/n} & \text{otherwise.} \end{cases}$$

Similarly, the definition of $\mathcal{M}_2(G)$ and Lemma 3.5 imply

$$|\Delta_m^{(2)}(\mathcal{P}) - \mu_m^{(2)}| \quad \le \quad \begin{cases} 24d^6 \log n & \text{if } \mu_m^{(2)} \le 24d^2 \log n \\ 24d^5(nd-2m)\sqrt{m \log n}/n & \text{otherwise.} \end{cases}$$

Hence,

$$\begin{aligned}
\frac{|\Delta_m(\mathcal{P}) - \mu_m|}{\binom{nd-2m}{2} - \Delta_m(\mathcal{P})} \quad &\le \quad \frac{|\Delta_m^{(1)}(\mathcal{P}) - \mu_m^{(1)}| + |\Delta_m^{(2)}(\mathcal{P}) - \mu_m^{(2)}|}{\frac{1}{4}(nd-2m)^2} \\
&\le \quad 4 \cdot \left(\frac{(24+24)d^6 \log n}{(nd-2m)^2} + \frac{24d^4\sqrt{\log n}}{\sqrt{n}(nd-2m)} + \frac{24d^5\sqrt{m \log n}}{n(nd-2m)} \right) \\
&\le \quad \frac{192d^6 \log n}{(nd-2m)^2} + \frac{192d^{11/2}\sqrt{\log n}}{\sqrt{n} \cdot (nd-2m)}.
\end{aligned}$$

□

Corollary 3.5. *Assume that $d = o(n^{1/3})$ and let $\xi = \xi(n,d)$ be an arbitrary function such that $\xi \gg d^4\sqrt{\log n}$. All paths $\mathcal{P} \in \mathcal{A}v(G)$ satisfy*

$$\prod_{m=0}^{\frac{1}{2}nd-\xi} \frac{\binom{nd-2m}{2} - \mu_m}{\binom{nd-2m}{2} - \Delta_m(\mathcal{P})} = \exp(\mathcal{O}(d^6(\sqrt{\log n / nd} + \xi^{-1}) \log n)).$$

Proof. Let $\mathcal{P}$ be an arbitrary path in $\mathcal{A}v(G)$. Using Lemma 3.10 we deduce

$$\begin{aligned}
\prod_{m=0}^{\frac{1}{2}nd-\xi} \frac{\binom{nd-2m}{2} - \mu_m}{\binom{nd-2m}{2} - \Delta_m(\mathcal{P})} \quad &= \quad \prod_{m=0}^{\frac{1}{2}nd-\xi} \left(1 + \frac{\Delta_m(\mathcal{P}) - \mu_m}{\binom{nd-2m}{2} - \Delta_m(\mathcal{P})} \right) \\
&\le \quad \prod_{m=0}^{\frac{1}{2}nd-\xi} \left(1 + \frac{|\Delta_m(\mathcal{P}) - \mu_m|}{\binom{nd-2m}{2} - \Delta_m(\mathcal{P})} \right) \\
&\le \quad \exp\left(\sum_{m=0}^{\frac{1}{2}nd-\xi} \frac{|\Delta_m(\mathcal{P}) - \mu_m|}{\binom{nd-2m}{2} - \Delta_m(\mathcal{P})} \right) \\
&\le \quad \exp\left(\mathcal{O}\left(\sum_{i=\xi}^{\frac{1}{2}nd} \left(\frac{d^{11/2}\sqrt{\log n}}{\sqrt{n} \cdot i} + \frac{d^6 \log n}{i^2} \right) \right) \right) \\
&\le \quad \exp(\mathcal{O}(d^6(\sqrt{\log n/nd} + \xi^{-1}) \log n)).
\end{aligned}$$

For the proof of the lower bound observe first that the assumption on ξ together with Lemma 3.10 implies that for sufficiently large n all $m \le \frac{1}{2}nd - \xi$ satisfy

$$\frac{|\Delta_m(\mathcal{P}) - \mu_m|}{\binom{nd-2m}{2} - \Delta_m(\mathcal{P})} \le \frac{1}{2}.$$

Using the fact that $1 - x \ge e^{-2x}$ for all $x \le \frac{1}{2}$ we deduce that

$$1 + \frac{\Delta_m(\mathcal{P}) - \mu_m}{\binom{nd-2m}{2} - \Delta_m(\mathcal{P})} \ge 1 - \frac{|\Delta_m(\mathcal{P}) - \mu_m|}{\binom{nd-2m}{2} - \Delta_m(\mathcal{P})} \ge \exp\left(-\frac{2|\Delta_m(\mathcal{P}) - \mu_m|}{\binom{nd-2m}{2} - \Delta_m(\mathcal{P})} \right),$$

from which the desired inequality follows in the same way as above. □

Lemma 3.11. *Assume that $d = o(n^{1/3})$. Then for all $1 \le \xi \le \frac{1}{2}nd$:*

$$e^{-\frac{5d\xi}{n}} \le \prod_{m=\frac{1}{2}nd-\xi}^{\frac{1}{2}nd-1} \left(1 - \frac{\mu_m}{\binom{nd-2m}{2} - \mu_m}\right) \le \prod_{m=\frac{1}{2}nd-\xi}^{\frac{1}{2}nd-1} \left(1 - \frac{\mu_m}{\binom{nd-2m}{2}}\right) \le e^{-\frac{\xi}{2n}}.$$

Proof. From Corollary 3.3 we deduce that

$$\frac{1}{2n}\binom{nd-2m}{2} \le \mu_m \le \frac{2d}{n} \cdot \binom{nd-2m}{2} \qquad \text{for all } 0 \le m \le \tfrac{1}{2}nd - 1.$$

Using that $e^{-2x} \le 1 - x \le e^{-x}$ for all $x \le \frac{1}{2}$ (which is most easily verified by differentiation) we therefore conclude

$$\prod_{m=\frac{1}{2}nd-\xi}^{\frac{1}{2}nd-1} \left(1 - \frac{\mu_m}{\binom{nd-2m}{2}}\right) \le \prod_{m=\frac{1}{2}nd-\xi}^{\frac{1}{2}nd-1} \left(1 - \frac{1}{2n}\right) \le e^{-\frac{\xi}{2n}}$$

and

$$\prod_{m=\frac{1}{2}nd-\xi}^{\frac{1}{2}nd-1} \left(1 - \frac{\mu_m}{\binom{nd-2m}{2} - \mu_m}\right) \ge \prod_{m=\frac{1}{2}nd-\xi}^{\frac{1}{2}nd-1} \left(1 - \frac{\frac{2d}{n}}{1 - \frac{2d}{n}}\right) \ge e^{-\frac{4d\xi}{n-2d}} \ge e^{-\frac{5d\xi}{n}}.$$

□

Corollary 3.6. *Assume that $d = o(n^{1/3})$. Then*

$$\mathbf{P}_A[\mathcal{P}] \ge e^{-\mathcal{O}\left(d^2 \cdot \sqrt{(d^7(\log n)^3)/n}\right)} \cdot av(G)$$

for all paths $\mathcal{P} \in \mathcal{A}v(G)$.

Proof. Let $\mathcal{P}$ be an arbitrary path in $\mathcal{A}v(G)$ and let $\xi = \sqrt{nd^5 \log n}$. Observe that for $m \ge \frac{1}{2}nd - \xi$ we trivially have

$$1 + \frac{\Delta_m(\mathcal{P}) - \mu_m}{\binom{nd-2m}{2} - \Delta_m(\mathcal{P})} \ge \begin{cases} 1 & \text{if } \Delta_m(\mathcal{P}) \ge \mu_m \\ 1 - \frac{\mu_m}{\binom{nd-2m}{2} - \mu_m} & \text{otherwise.} \end{cases}$$

Thus, using Corollary 3.5 and 3.11 we deduce

$$\begin{aligned}
\frac{\mathbf{P}_A[\mathcal{P}]}{av(G)} &= \prod_{m=0}^{\frac{1}{2}nd-1} \left(1 + \frac{\Delta_m(\mathcal{P}) - \mu_m}{\binom{nd-2m}{2} - \Delta_m(\mathcal{P})}\right) \\
&\ge \left(\prod_{m=0}^{\frac{1}{2}nd-\xi} \frac{\binom{nd-2m}{2} - \mu_m}{\binom{nd-2m}{2} - \Delta_m(\mathcal{P})}\right) \cdot \left(\prod_{m=\frac{1}{2}nd-\xi}^{\frac{1}{2}nd-1} \left(1 - \frac{\mu_m}{\binom{nd-2m}{2} - \mu_m}\right)\right) \\
&\ge \exp\left(-\mathcal{O}\left(\sqrt{\frac{d^{11}(\log n)^3}{n}} + \frac{d^6 \log n}{\xi} + \frac{d\xi}{n}\right)\right)
\end{aligned}$$

$$\geq \exp\left(-\mathcal{O}\left(d^2 \cdot \sqrt{\frac{d^7(\log n)^3}{n}}\right)\right).$$

□

Lemma 3.12. *Assume that $d = O(n^{\frac{1}{28}})$ and $\xi \leq d^6(\log n)^2$. Then for a path $\mathcal{P}$ selected uniformly at random from $\mathcal{A}v(G)$, the expected value of*

$$\prod_{m=\frac{1}{2}nd-\xi}^{\frac{1}{2}nd-1} \frac{\binom{nd-2m}{2} - \mu_m}{\binom{nd-2m}{2} - \Delta_m(\mathcal{P})} \tag{12}$$

is asymptotic to 1 as $n \to \infty$.

Proof. Let p_1 denote this expected value. We compute with a path $\mathcal{P}$ selected uniformly at random from $Paths(G)$, and let p_0 denote the expected value of (12) for such $\mathcal{P}$. Later we deduce what we need about p_1 in the restricted space. Note that p_0 is determined by the last ξ edges in the ordering of the edges of G corresponding to $\mathcal{P}$, which form a random ordered ξ-subset $S = S(\mathcal{P})$ of the edges of G. Let $\mathcal{S}$ denote the probability space of these subsets, determined by all $\mathcal{P} \in Paths(G)$. Let $r(S)$ denote the number of edges in S which have distance at most 2 from a later edge in S (where by distance at most 2 we mean that one vertex of the new edge and one vertex of a later edge in S are at distance 2). Let R_r denote the set of S for which $r(S) = r$. Then for $S \in \mathcal{S}$,

$$\begin{aligned} \mathbf{P}(S \in R_r) &\leq \binom{\xi}{r}\left(\frac{2d^3\xi}{\frac{1}{2}nd - \xi}\right)^r \\ &= \left(\mathcal{O}\left(\frac{d^2\xi^2}{n}\right)\right)^r. \end{aligned} \tag{13}$$

Consider an arbitrary path $\mathcal{P}$ and assume that $S = S(\mathcal{P}) \in R_r$. We first aim at proving upper bound on $\Delta_m(\mathcal{P})$ for all $m \geq \frac{1}{2}nd - \xi$. In order to do so we imagine the sequential generation of S starting with the *last* edge. Let H_k denote the subgraph of G where the last k edges of S are removed from G. Observe that, trivially, $\Delta(H_0) = 0$. Now consider what happens if we change from H_{k-1} to H_k, that is, if we remove the kth edge. An important observation is that in fact $\Delta(H_k) = \Delta(H_{k-1})$ if the kth edge has distance at least 2 from all previously removed edges. That is, in this respect we only have to consider the edges which have distance at most 1 from some later edge. In order to obtain the type of bounds we need, we will, however, instead consider all edges which have distance at most 2 from a previously removed edge separately. Consider the ith such edge, say edge $\{u, v\}$, and assume it is the kth edge in total. We want an upper bound for $\Delta(H_k) - \Delta(H_{k-1})$. One easily checks that $\Delta^{(1)}(H_k) - \Delta^{(1)}(H_{k-1}) \leq i$ and we claim that $\Delta^{(2)}(H_k) - \Delta^{(2)}(H_{k-1}) \leq 4i$. To see this observe that by removing edge $\{u, v\}$ only the summands corresponding to edge $\{u, v\}$ itself and those of the adjacent edges can contribute. Edge $\{u, v\}$ contributes a negative term, so we don't have to worry about it. On the other hand, an edge adjacent to $\{u, v\}$, say $\{u, w\}$, adds to the difference *exactly* the degree of w in $G - H_{k-1}$. That is,

$$\Delta^{(2)}(H_k) - \Delta^{(2)}(H_{k-1}) \le \sum_{x \in N_{H_{k-1}}(u)} d_{G-H_{k-1}}(x) + \sum_{y \in N_{H_{k-1}}(v)} d_{G-H_{k-1}}(y).$$

Observe that any two edges in $G - H_{k-1}$ which contribute to the sum for u are at distance 2 from each other, and there are at most i such edges. Each such edge can count up to two times for u (once at each of its ends) and up to two times for w. This shows the desired bound of $4i$. Summarizing, we have for all $1 \le k \le \xi$

$$\Delta(H_k) \le \Delta(H_\xi) \le \sum_{i=1}^{r} 5i \le 5\binom{r+1}{2}$$

implying that $5\binom{r+1}{2}$ is also an upper bound on $\Delta_m(\mathcal{P})$ for $m \ge \frac{1}{2}nd - \xi$. Also note that $\Delta_m(\mathcal{P}) < \binom{2t}{2}$ for $nd - 2m = 2t$. Hence for $S(\mathcal{P}) \in R_r$,

$$\begin{aligned}
\prod_{m=\frac{1}{2}nd-\xi}^{\frac{1}{2}nd-1} \frac{\binom{nd-2m}{2}}{\binom{nd-2m}{2} - \Delta_m(\mathcal{P})} &\le \left(\prod_{t=1}^{t_0} \binom{2t}{2}\right) \left(\prod_{t=t_0+1}^{\xi} \frac{\binom{2t}{2}}{\binom{2t}{2} - 5\binom{r+1}{2}}\right) \\
&= \left(\mathcal{O}(r^2)\right)^{t_0} \exp\left(\sum_{t=t_0+1}^{\xi} \mathcal{O}\left(\frac{r^2}{t^2}\right)\right) \\
&= \left(\mathcal{O}(r^2)\right)^{t_0} \qquad (14)
\end{aligned}$$

provided t_0 is chosen such that $t_0 = \mathcal{O}(r)$ and $\binom{2t_0}{2} > 5(1+\epsilon)\binom{r+1}{2}$ for some $\epsilon > 0$. Recalling Lemma 3.11 that

$$e^{-\frac{2d\xi}{n-d}} \le \prod_{m=\frac{1}{2}nd-\xi}^{\frac{1}{2}nd-1} \frac{\binom{nd-2m}{2} - \mu_m}{\binom{nd-2m}{2}} \le e^{-\frac{\xi}{2n}}, \qquad (15)$$

we see that $p_0 \ge 1 - o(1)$, and by (13) and (14) with $t_0 = \frac{28}{25}r > \frac{\sqrt{5}}{2}r$,

$$p_0 \le 1 + \sum_{r=1}^{\xi} \left(\mathcal{O}\left(\frac{r^{56/25}d^2\xi^2}{n}\right)\right)^r = 1 + o(1) \qquad (16)$$

since $r \le \xi$.

This was all for $\mathcal{P}$ selected uniformly at random from $Paths(G)$; note that (15) similarly implies $p_1 \ge 1 - o(1)$. Furthermore, since (12) is always positive, we have

$$p_1 \le \frac{p_0}{1 - \mathbf{P}(\mathcal{M}_1(G) \cup \mathcal{M}_2(G))} \le 1 + o(1)$$

by Lemmas 3.7 and 3.8 and (16). □

Proof of Theorem 2.1. Using (5) and the definition (10) of $av(G)$ we obtain for a path $\mathcal{P}$ chosen uniformly at random from $\mathcal{A}v(G)$

$$\mathbf{E}(\mathbf{P}_A[\mathcal{P}]) = \frac{1}{\mathcal{A}v(G)} \cdot \sum_{\mathcal{P} \in \mathcal{A}v(G)} \mathbf{P}_A[\mathcal{P}]$$

$$= \frac{av(G)}{\mathcal{A}v(G)} \cdot \sum_{\mathcal{P} \in \mathcal{A}v(G)} \prod_{m=0}^{\frac{1}{2}nd-1} \frac{\binom{nd-2m}{2} - \mu_m}{\binom{nd-2m}{2} - \Delta_m(\mathcal{P})}.$$

From Corollary 3.5 and Lemma 3.12 using $\xi = d^6(\log n)^2$ we obtain (8). Hence from Lemmas 3.7 and 3.8 and Corollary 3.4 we deduce from Lemma 3.1 and Lemma 3.6

$$\begin{aligned} \mathbf{P}_A[G] &= (1+o(1)) \cdot av(G) \cdot |Paths(G)| \\ &= (1+o(1)) \cdot e^{\frac{1}{4}(d^2-1)} \cdot \mathbf{P}_U[G]. \end{aligned}$$

As we know from Corollary 1.1 that $\mathbf{P}_U[G] = (1+o(1)) \cdot e^{-\frac{1}{4}(d^2-1)}/|\mathcal{R}(n,d)|$, the theorem follows. Note that the little oh terms do not depend on G. $\square$

Proof of Theorems 2.2 and 2.3. Combining Lemmas 3.7 and 3.8 and Corollary 3.6 we deduce from Lemma 3.6 and Corollary 1.1 that for $d = o((n/(\log n)^3)^{\frac{1}{11}})$

$$\begin{aligned} \mathbf{P}_A[G] &\geq (1+o(1)) \cdot av(G) \cdot |Paths(G)| \\ &= (1+o(1)) \cdot e^{\frac{1}{4}(d^2-1)} \cdot \mathbf{P}_U[G] \\ &= (1+o(1)) \cdot \frac{1}{|\mathcal{R}(n,d)|}. \end{aligned}$$

As the sum $\sum \mathbf{P}_A[G]$ over all d-regular graphs G can be at most 1, we have that this sum is asymptotic to 1, and the claims of Theorems 2.2 and 2.3 follow. $\square$

3.6. Concluding Remarks. The following example shows that Theorem 2.1 is in some sense best possible. More precisely, this example shows that the probability distribution of two d-regular graphs G are not exactly equal.

Consider the following two 2-regular graphs on 6 vertices:

a) C_6 and

b) two disjoint copies of a C_3.

For each graph there are $6! = 720$ different paths. With the help of a small computer program one easily checks that

$$\mathbf{P}_A[C_6] = \tfrac{90455}{8775000} \qquad \text{while} \qquad \mathbf{P}_A[2\ C_3\text{'s}] = \tfrac{232544}{8775000}.$$

It is also possible to prove that infinitely many larger examples exist, by showing that the primes occurring in the denominators of the probabilities for some graphs will not occur in others. But it seems annoyingly difficult to show what we believe is true, that graphs usually occur with different probabilities.

We also do not have good examples of d-regular graphs which are considerably more (or less) likely than most of the d-regular graphs, which would limit the possible strengthenings of Theorem 2.1. A good candidate for this might be the graph which consists of the union of pairwise disjoint K_d's.

4. Performance of the Algorithm

Theorem 4.1. *For all $d = \mathcal{O}(n^{1/2})$ Step 2 of Algorithm 1 can be implemented so as to require expected time $\mathcal{O}(nd^2)$, and space $\mathcal{O}(nd)$.*

Proof. It can very easily be implemented in three phases. In the first phase, keep a list L of all the points, as an array with those in U at the front, and the other points in pairs afterwards, as well as another array I whose i'th entry is the position in L of the point i. Then two points i and j can be chosen randomly in U in constant time (assuming the number of digits in n is not a problem, and assuming a perfect constant-time random number generator). Moreover, they can be checked for suitability in time $\mathcal{O}(d)$, since all $d-1$ points in the same group can be found in L in time $\mathcal{O}(d)$ using I, and for each such point, if it is in a pair, its mate is next to it (on a known side) in L. This process is repeated until a suitable pair has been found. Then, if they are suitable, update L by swapping the chosen points with the last two listed in U, (which are $2m+1$ and $2m+2$ from the end of L if m pairs have already been added) and update I for the (up to) four points so moved. This takes constant time.

Hence, the running time of phase 1 is $\mathcal{O}(d)$ times the total number of points i and j checked for suitability throughout this phase. Phase 1 stops when the number of points in U first falls below $2d^2$.

Note that each point has at most $(d-1)^2$ other points in U with which it does *not* make a suitable pair. Hence, in phase 1, when there are $k > 2d^2$ points left in U, the expected number of trials of two points before a suitable pair is found is at most 2. The sum of this over $d^2 < k < nd$ is $\mathcal{O}(nd)$, giving the bound $\mathcal{O}(nd^2)$ on the total time to reach this stage.

On the other hand, when the number of points in U first falls below $2d^2$, phase 2 begins. Instead of choosing a pair at random, choose a random pair of groups (i.e. vertices of the graph G), say u and v, which do not yet have all points matched. If $\{u,v\}$ is already an edge of G, repeat this step. Then randomly choose a pair of points i and j in the groups corresponding to u and v respectively. If these points i and j are in U, they are a randomly chosen suitable pair; if not, repeat the choice of u, v, i, j again. Phase 2 lasts until the number of available groups drops below $2d$. The list of available groups can be maintained just like the list of available points, with negligible time required. Thus, choosing an available pair of groups takes constant time. The probability they form a non-edge of G is at least $1/2$ as there are at least $2d$ groups and each has an edge to at most $d-1$ others. Also, for each such u and v, the probability of finding a pair of points in U is at least $1/d^2$, and so for the $\mathcal{O}(d^2)$ pairs added this stage takes expected time $\mathcal{O}(d^4)$.

When the number of available groups falls below $2d$, phase 3 begins. Construct (in time $\mathcal{O}(d^2)$) the graph induced by all vertices of G of degree less than d, and form its complement H (on these same vertices). Then work as in phase 2, but instead of choosing u and v at random from the vertices of H, choose an edge $\{u,v\}$ of H uniformly at random and accept it with probability x_{uv}/d^2, where x_{uv} denotes the product of the number of suitable points in u times the number of suitable points in v. Given a list of edges of H, finding the next edge requires expected time at most $\mathcal{O}(d^2)$. Note that H can be updated in constant time each time an edge has been chosen, with the help of suitable data structures. The algorithm stops when no more suitable points exist, that is, H has no more edges. Hence phase 3 requires time $\mathcal{O}(d^4)$. This establishes the time requirements; the space requirements are several lists of the points ($\mathcal{O}(nd)$) and the space required for H ($\mathcal{O}(d^2)$). $\square$

The theorem gives the running time for one trial of Step 2: if the algorithm fails due to no regular graph being produced, it has to start again. By Theorem 2.3 the probability of failure of Step 2 in this sense is $o(1)$ for $d = o((n/(\log n)^3)^{\frac{1}{11}}$, and so the expected time for the algorithm to produce a regular graph is $\mathcal{O}(nd^2)\mathcal{O}(1)$ for such d. Moreover, experiments with n ranging from 50 to 400 and d ranging from $0.05n$ to $0.5n$ strongly suggest that the probability of succeeding on one trial is at least 0.3 for such n and d. (In fact, it appears roughly equal to $1-4d/(3n)$ up to $d = 0.75n$.) Thus, the algorithm seems usable, expecting a bounded number of repetitions of Step 2 per graph produced, for all d up to $n/2$ (though for such d we do not know much about the resulting probability distribution!). For larger d, one can of course generate the complementary graph. (Note that for d's larger than $\sqrt{n}$ the proof of Theorem 4.1 only gives a complexity of $\mathcal{O}(d^4)$, but this can be improved by the use of more clever data structures.) We stress that we have no proof that asymptotically for say $d \sim n/2$ the expected number of repetitions required is bounded. In addition, we have reasons to believe that for $d \gg n^{1/2}$ the resulting probability distribution will no longer be approximately uniform. Clarifying these possible extensions of Theorem 2.3 would be interesting.

References

[1] E. A. Bender and E. R. Canfield. The asymptotic number of labeled graphs with given degree sequences. *J. Combinatorial Theory, Series A*, 24:296–307, 1978.

[2] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, 1:311–316, 1980.

[3] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.*, 58:13–30, 1963.

[4] M.R. Jerrum, B.D. McKay, and A.J. Sinclair. When is a graphical sequence stable? In A. Frieze and T. Łuczak, editors, *Random Graphs Vol. 2*, pages 148–188. Wiley, 1992.

[5] M.R. Jerrum and A.J. Sinclair. Fast uniform generation of regular graphs. *Theoretical Computer Science*, 73:91–100, 1990.

[6] T. Łuczak and N.C. Wormald. Phase transition for random graph processes. In preparation.

[7] C. McDiarmid. On the method of bounded differences. In *Surveys in combinatorics, Lond. Math. Soc. Lect. Note Ser.* **141**, pages 148–188, 1989.

[8] B.D. McKay. Asymptotics for symmetric 0-1 matrices with prescribed row sums. *Ars Combinatorica*, 19A:15–25, 1985.

[9] B.D. McKay and N.C. Wormald. Uniform generation of random regular graphs of moderate degree. *Journal of Algorithms*, 11:52–67, 1990.

[10] B.D. McKay and N.C. Wormald. Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$. *Combinatorica*, 11:369–382, 1991.

[11] A. Ruciński and N.C. Wormald. Random graph processes with degree restrictions. *Combinatorics, Probability and Computing*, 1:169–180, 1992.

[12] G. Tinhofer. On the generation of random graphs with given properties and known distribution. *Appl. Comput. Sci., Ber. Prakt. Inf.*, 13:265–297, 1979.

[13] N.C. Wormald. Some problems in the enumeration of labelled graphs. Ph. D. thesis, University of Newcastle, 1978.

[14] N.C. Wormald. Generating random regular graphs. *Journal of Algorithms*, 5:247–280, 1984.