Generating Random Regular Graphs

Jeong Han Kim
Microsoft Research
One Microsoft Way
Redmond, WA 98052
jehkim@microsoft.com

Van H. Vu
Department of Mathematics
UCSD
La Jolla, CA 92093
vanvu@ucsd.edu

ABSTRACT

Random regular graphs play a central role in combinatorics and theoretical computer science. In this paper, we analyze a simple algorithm introduced by Steger and Wormald [9] and prove that it produces an asymptotically uniform random regular graph in a polynomial time. Precisely, for fixed d and n with $d = O(n^{1/3-\varepsilon})$, it is shown that the algorithm generates an asymptotically uniform random d-regular graph on n vertices in time $O(nd^2)$. This confirms a conjecture of Wormald. The key ingredient in the proof is a recently developed concentration inequality by the second author.

Besides being perhaps the only algorithm which works for relatively large d in practical time, our result also has a significant theoretical value, as it can be used to derive many properties of uniform random regular graphs.

Categories and Subject Descriptors

F.2.2 [**ANALYSIS OF ALGORITHMS AND PROBLEM COMPLEXITY**]: Nonnumerical Algorithms and Problems; G.2.2 [**DISCRETE MATHEMATICS**]: Graph Theory

General Terms

Algorithms

Keywords

Random Regular Graph, Configuration Model, Concentration Inequality

1. INTRODUCTION

One of the most important and interesting models for random graphs is the model of random regular graphs. The study of this model has been pursued by a large number of researchers through decades. Random regular graphs have been playing a crucial role in theoretical computer science, especially in the theory of expanders. For instance,it is now known that a uniform random d-regular graph has asymptotically the best possible expansion rate, namely, the second eigenvalue of it is almost surely $(2 + o(1))\sqrt{d}$, for any constant d (see [3]). For more information, the reader is referred to a recent survey by Wormald [11], which contains lost of results, questions, and more than one hundred references. Here and later, by graphs we always mean simple graphs, that is, without loops and parallel edges.

Fix a pair of positive integers $1 \leq d < n$ and let $S(n, d)$ be the set of all simple d-regular graphs on a set of n vertices. The uniform random d-regular graph $G_{n,d}$ is obtained by sampling with respect to the uniform distribution from $S(n, d)$. The probability of each simple d-regular graph is thus $1/|S(n, d)|$. Since nd is twice the number of edges in $G_{n,d}$, we always assume nd is even.

Despite an extensive study, a very fundamental, and perhaps practically the most important, question concerning random regular graphs has not yet been answered in a satisfactory way:

Question. *How do we generate a uniform random regular graph?*

Beside its obvious practical importance, the above question is also critical from the theoretical point of view, as it is difficult to study properties of random regular graphs (or any random objects) without knowing how to generate them.

By definition, the most straightforward method is making a list of all (simple) regular graphs and then choosing one uniformly at random from the list. This, unfortunately, could never be done in practice, as the number of regular graphs is huge (see Section 3 for an asymptotic estimate).

A better approach is to follow the configuration model proposed by Bollobás [1], and Bender and Canfield [2]. In order to generate a uniform random d-regular graph on n vertices, we consider a family of n sets of size d. Each set may be regarded as a set of d copies of each vertex. All together there are nd points, or nd copies of the vertices. Draw a uniform random perfect matching on these copies and connect two vertices i and j if the matching contains an edge consisting of copies of i and j. It is easy to see that the resulting multigraph is d regular and the distribution, conditioned on the graph being simple, is uniform. It is known that if d is a constant then the probability of being simple is bounded by a positive constant, uniformly in n. This, however, is no longer true if d is large, for instance $\log n$ or

STOC'03, June 9–11, 2003, San Diego, California, USA.

n^a, so that most of the time one gets a non-simple graph. More precisely, it has been shown [11] that the probability is about $\exp(-\frac{d^2}{4})$, which tends to 0 extremely fast in d. Even for a modest parameter such as $d = 14$, one, in expectation, has to try $e^{49} > 10^{20}$ times until he or she obtains a simple graph. This clearly rules out the possibility of using this approach in practice unless d is very small. From the theoretical point of view, the configuration model is not very useful if $d \gg \log n$, as one has to beat the failure probability $\exp(-\frac{d^2}{4})$. For more discussion about this point, we refer to Wormald's survey [11].

For many practical and theoretical purposes, it is usually sufficient to generate random d-regular graphs which are asymptotically uniform. After all, most statements we want to prove about a random model are of asymptotic nature. It has turned out that allowing asymptotically uniform distributions makes the problem more feasible.

A natural way to generate an asymptotically uniform random regular graph is to use the Markov chain technique. This was done by Jerrum and Sinclair [4], using their algorithm for generating random perfect matchings in dense graphs. On the other hand, although the algorithm is proved to be in **P**, its running time is $\Omega(n^{10})$, which is only theoretically polynomial. Moreover, it seems to us that it is very hard to prove properties of random regular graphs based on this algorithm.

More recently, a faster algorithm was analyzed by Steger and Wormald [9], following an earlier work of Rucinski and Wormald [8]. This algorithm is a natural refinement of the configuration model algorithm of Bollobás, and Bender and Canfield. Again let us consider a set V of nd points which is the union of n disjoint sets of size d. We want to construct a perfect matching which does not give rise to loops and parallel edges. We achieve this by building the matching one edge at a time. First of all, we never pick an edge which creates a loop. Namely, we never pick an edge with both ends in the same set. Assume that a bunch of edges are picked. In the next step, we only pick an edge which does not create a parallel edge, namely, we ignore all the edges connecting two sets I and J after some $i \in I$ and $j \in J$ are connected for the first time. When we pick an edge, we pick it with respect to the uniform distribution over the set of all available edges. At the end, we obtain a random perfect matching , which provides a simple, d-regular random graph. The implementation of the algorithm is straightforward and its running time is only $O(nd^2)$, which is sub-quadratic for $d \ll n^{1/2}$. The only matter here is the distribution of this random graph.

Steger and Wormald [9] proved that if $d = o(n^{1/28})$, then the distribution in question is asymptotically uniform. Wormald [11] conjectured that the distribution is still asymptotically uniform for any d significantly less than $n^{1/3}$.

The first purpose of this paper is to prove this conjecture. This way, we obtain a very fast method to generate asymptotically uniform random regular graphs with degree up to $n^{1/3-\varepsilon}$, for any positive constant ε. Our result also has a considerable theoretical value, as we can rely on the algorithm to prove many highly non-trivial properties of uniform random regular graphs. In fact, it was shown, using a weaker version of this result, that if $\log n \ll d \leq n^{1/3-\varepsilon}$, then the random regular graph $G_{n,d}$ behaves very much like the (non-regular) Erdős-Rényi's random graph $G(n,p)$ of the same density [6]. Among others, it follows immediately that in this range of d, $G_{n,d}$ almost surely contains a hamiltonian cycle, since the Erdős-Rényi graph with density significantly larger than $\log n/n$ has this property. The interested reader is referred to [6] for a precise statement.

The other purpose of the paper is to introduce, via the proof, a recently developed concentration inequality by the second author [10] , which can be applied for functions with large Lipchitz coefficients, a situation where standard tools such as Azuma's and Talagrand's are ineffective. This concentration result plays a critical role in the proof.

In the whole paper, Pr denotes probability and $\mathbb{E}$ denotes expectation. 1_X is the characteristic function of the event X: $1_X = 1$ if X occurs and $1_X = 0$ otherwise. The asymptotic notation is used under the assumption that n tends to infinity. All logarithms have the natural base.

2. THE ALGORITHM AND THE RESULT

Let $N(d,n)$ be the number of d-regular graphs on n vertices. McKay and Wormald [7] have shown that for $d = o(n^{1/2})$

$$N(d,n) = \frac{(nd)!}{(\frac{1}{2}nd)!2^{nd/2}(d!)^n} \exp\Big(\frac{1-d^2}{4} - \frac{d^3}{12n} + O(\frac{d^2}{n})\Big)$$

Consider the set $S(n,d)$ of all d-regular graphs on n points equipped with the uniform distribution. The (uniform) probability of each graph in $S(n,d)$ is $p_u = 1/N(d,n)$.

Let us now describe our algorithm:

Algorithm A.
(I) Start with a set $\mathbf{U}$ of nd points (nd even) partitioned into n group of size d.
(II) Repeat the following until no suitable pair can be found. Choose two random points i and j in $\mathbf{U}$ and if ij is suitable, pair i with j and delete them from $\mathbf{U}$.
(III) Create a graph G with an edge from r to s if and only if there is a pair containing points in the r^{th} and s^{th} groups. If G is regular, output it, otherwise return to step (I).

Here and later, a pair (i,j), where i is a point in group I and j is a point in group J, is *suitable* if $I \neq J$ and no other pair $(i',j'), i' \in I, j' \in J$ has been chosen.

Throughout the paper, $Pr_A(G)$ denotes the probability that the output of A is a particular graph G. In [9], Steger and Wormald analyzed the above algorithm and proved the following theorem.

THEOREM 2.1. *If $d = o(n^{1/28})$, then for every simple d-regular graph G on n vertices*

$$Pr_A(G) = (1 + o(1))p_u.$$

Our goal is to extend the above result for larger d. We shall prove

THEOREM 2.2. *For any positive constant $\varepsilon < 1/3$ the following holds. For any simple $d \leq n^{1/3-\varepsilon}$ and any d-regular graph G on n vertices*

$$Pr_A(G) = (1 + o(1))p_u.$$

The rest of the paper is devoted to the proof of Theorem 2.2. The next two sections describe the main ideas of the proof. In particular, we shall state a series of properties which imply Theorem 2.2 (see Section 4). The essential technical ingredient, the new large deviation inequality, will be presented in Section 5. The remaining sections of the paper are devoted to the proofs of the properties stated in Section 4.

3. PRELIMINARIES

It is well-known (and easy to check) that for each d-regular graph G, there are exactly $(d!)^n$ different simple perfect matchings on $\mathbf{U}$ which give rise to G. (A perfect matching is simple if it gives rise to a simple regular graph.) So to prove Theorem 2.2, it suffices to show that if d satisfies the assumption of the theorem, then for any simple perfect matching M

$$Pr_A(M) = (1+o(1))\frac{(\frac{1}{2}nd)!2^{nd/2}}{(nd)!}\exp\Big(\frac{d^2-1}{4}\Big). \qquad (1)$$

Consider an ordering $\mathcal{M} = x_1, \ldots, x_{nd/2}$ where x_m are the edges of M. Assume that the first m edges of $\mathcal{M}$ are obtained and let $G_m(\mathcal{M})$ be the graph formed by the projection of these edges. Thus $G_m(\mathcal{M})$ is a subgraph of G. To count the number of suitable edges, notice that there are $\binom{nd-2m}{2}$ ways to form an edge. However, an edge is suitable if and only if it does not join two vertices from the same group or two vertices come from two already adjacent groups. The number of edges of the first type is $\Delta_m^{[1]}(\mathcal{M}) = \sum_u \binom{d-d_u}{2}$ and the number of unsuitable edges of the second type is $\Delta_m^{[2]}(\mathcal{M}) = \sum_{u\sim_{G_m(\mathcal{M})}v}(d-d_u)(d-d_v)$, where d_u is the degree of u in $G_m(\mathcal{M})$. Set $\Delta_m(\mathcal{M}) = \Delta_m^{[1]}(\mathcal{M}) + \Delta_m^{[2]}(\mathcal{M})$, the number of suitable edges is $\binom{nd-2m}{2} - \Delta_m(\mathcal{M})$. It follows that

$$Pr_A(\mathcal{M}) = \prod_{m=0}^{nd/2-1}\frac{1}{\binom{nd-2m}{2}-\Delta_m(\mathcal{M})}. \qquad (2)$$

Set $\delta_m^{[1]} = \frac{1}{2}(nd-2m)^2\frac{(d-1)}{nd}$ and $\delta_m^{[2]} = (nd-2m)^2\frac{m(d-1)^2}{n^2d^2}$. It will be useful to think of $\delta_m^{[i]}$ as a sort of expectation of $\Delta_m^{[i]}$ with respect to a random choice of $\mathcal{M}$. Define $\delta_m = \delta_m^{[1]} + \delta_m^{[2]}$. A routine, but somewhat tedious, calculation shows that for $d = o(n^{1/3})$ (a similar calculation is presented in [9] so we omit the details)

$$\prod_{m=0}^{nd/2-1}\frac{\binom{nd-2m}{2}}{\binom{nd-2m}{2}-\delta_m} = \exp\Big(\frac{d^2-1}{4}+o(1)\Big). \qquad (3)$$

Also notice that

$$\prod_{m=0}^{nd/2-1}\binom{nd-2m}{2} = \frac{(nd)!}{2^{nd/2}}. \qquad (4)$$

Define

$$f(\mathcal{M}) = \prod_{m=0}^{nd/2-1}\frac{\binom{nd-2m}{2}-\delta_m}{\binom{nd-2m}{2}-\Delta_m(\mathcal{M})}.$$

Using (2, 3, 4), it is clear that to prove (1) it suffices to show

$$\sum_{\mathcal{M}\in\mathcal{S}(M)} f(\mathcal{M}) = (1+o(1))(nd/2)!, \qquad (5)$$

where $\mathcal{S}(M)$ denotes the set of all orderings of the edges of M. Notice that $\mathcal{S}(M)$ has exactly $(nd/2)!$ elements, so (5) is equivalent to saying that the expectation of $f(\mathcal{M})$ with respect to the uniform distribution on $\mathcal{S}(M)$ is $1+o(1)$

$$\mathbb{E}\Big(f(\mathcal{M})\Big) = 1+o(1). \qquad (6)$$

To prove (6), we shall consider the upper bound and the lower bound separately

$$\mathbb{E}\Big(f(\mathcal{M})\Big) \le 1+o(1). \qquad (7)$$

$$\mathbb{E}\Big(f(\mathcal{M})\Big) \ge 1-o(1). \qquad (8)$$

The proof of (7) is rather difficult and occupies most of the rest of the paper, including Sections 3,4 and 5. The proof of (8) follows easily from a lemma developed for the proof of (7) and is presented in Section 7.

The general idea for proving (7) is the following: we first partition $\mathcal{S}(M)$ into many classes according to the order of magnitude of $f(\mathcal{M})$. Next, we upper bound the measures of these classes with respect to the uniform distribution on $\mathcal{S}(M)$. There will be one main class which contributes $1-o(1)$; all other classes together contribute $o(1)$.

The partition is non-trivial and bounding the measures is not at all easy. We need very strong tools and the key one is a recent concentration result, proved in [10]. This result provides a very tight control on the deviation tail $\Delta_m(\mathcal{M})-\delta_m$, where $\mathcal{M}$ is sampled randomly from $\Omega(M)$. Consequently, we obtain sufficiently good bound on the measures of the classes in the partition.

4. PARTITION

Let $\omega = \log^{0.01} n$. Set $\lambda_0 = \omega\log n$ and $\lambda_i = 2\lambda_{i-1}$ for all $i = 1, 2, \ldots, L$, where L is the smallest integer such that $\lambda_L \ge cd^2\log n$, where $\bar{c}$ is a sufficiently large constant. For each $0 \le m \le nd/2$, we will define a function T_m so that the sequence $\{T_m(\lambda_0) < T_m(\lambda_1) \le \cdots \le T_m(\lambda_L)\}$ satisfies certain properties. We first present the properties and then define T_m.

Let $\varepsilon \le 1/3$ be an arbitrarily small positive constant and let $\mathcal{S}^*(M)$ be the set of all orderings $\mathcal{M}$ with $\Delta_m(\mathcal{M}) \le (1-\varepsilon/2)\binom{nd-2m}{2}$ for all $m = 1, 2, ..., nd/2$. In order to prove (6), we will show that

$$\mathbb{E}\Big(f(\mathcal{M})1_{\mathcal{S}^*(M)}\Big) = 1-o(1). \qquad (9)$$

and

$$\mathbb{E}\Big(f(\mathcal{M})1_{\mathcal{S}(M)\setminus\mathcal{S}^*(M)}\Big) = o(1). \qquad (10)$$

Here and later $1_{\mathcal{S}^*(M)}$ is the characteristic function of the event $\mathcal{M}\in\mathcal{S}^*(M)$: $1_{\mathcal{S}^*(M)} = 1$ if $\mathcal{M}\in\mathcal{S}^*(M)$ and $1_{\mathcal{S}^*(M)} = 0$ otherwise. Similarly, $1_{\mathcal{S}(M)\setminus\mathcal{S}^*(M)}$ is the the characteristic function of the event $\mathcal{M}\in\mathcal{S}(M)\setminus\mathcal{S}^*(M)$.

To bound the contribution of $\mathcal{S}^*(M)$, we partition $\mathcal{S}^*(M)$ as follows. For $i = 1, 2, \ldots, L$, let A_i be the set of all orderings $\mathcal{M}$ satisfying

$$\Delta_m(\mathcal{M}) - \delta_m < T_m(\lambda_i) \text{ for all } m,$$

and A_∞ be the set of $\mathcal{M}$ such that there is some m such that $\Delta_m(\mathcal{M}) - \delta_m \geq T_m(\Delta_L)$. Clearly,

$$\mathcal{S}^*(M) = A_0 \cup A_\infty \cup \Big(\cup_{i=1}^{L} A_i \setminus A_{i-1} \Big).$$

We further partition A_0: For each $j = 0, 1, 2, \ldots, K$, where K is the smallest integer satisfying $2^K \geq \lambda_0 \log n + 1$, let B_j be the set of all ordering $\mathcal{M} \in A_0$ with $\Delta_m(\mathcal{M}) < 2^j$ for all $m \geq (nd - \omega\lambda_0)/2$. Let $C = B_0$. Then $\mathcal{M} \in C$ yields $\Delta_m(\mathcal{M}) = 0$ for all $m \geq (nd - \omega\lambda_0)/2$ since $\Delta_m(\mathcal{M})$ is a nonnegative integer. The desired properties of T_m are as follows.

Property 4.0 *For all m,*

$$T_m(\lambda_0) \leq \lambda_0 \log n \quad \forall m \geq (nd - \omega\lambda_0)/2.$$

This property implies that

$$\Delta_m(\mathcal{M}) \leq T_m(\lambda_0) + \delta_m \leq \lambda_0 \log n + \delta_m,$$

for $\mathcal{M} \in A_0$ and $m \geq (nd - \omega\lambda_0)/2$. Since $\delta_m = o(1)$ if $m \geq (nd - \omega\lambda_0)/2$, we have $A_0 = \Big(\cup_{j=1}^{K} B_j \Big) \cup C$ and hence

$$\mathcal{S}^*(M) = A_\infty \cup \Big(\cup_{i=1}^{L} A_i \setminus A_{i-1} \Big) \cup \Big(\cup_{j=1}^{K} B_j \setminus B_{j-1} \Big) \cup C. \quad (11)$$

For (9), let us consider

$$\mathbb{E}\Big(f(\mathcal{M}) 1_{\mathcal{S}^*(M)} \Big) = \mathbb{E}\Big(f(\mathcal{M}) 1_{A_\infty} \Big) + \sum_{i=1}^{L} \mathbb{E}\Big(f(\mathcal{M}) 1_{A_i \setminus A_{j-1}} \Big) + \sum_{j=1}^{K} \mathbb{E}\Big(f(\mathcal{M}) 1_{B_j \setminus B_{j-1}} \Big) + \mathbb{E}\Big(f(\mathcal{M}) 1_C \Big).$$

We will show that $\mathbb{E}\Big(f(\mathcal{M}) 1_C \Big) \leq 1 + o(1)$ and the other terms are $o(1)$. These estimates are direct consequences of the following properties:

PROPERTY 4.1. *For all $\mathcal{M} \in C$, $f(\mathcal{M}) \leq 1 + o(1)$.*

PROPERTY 4.2. *(a) $Pr(A_i \setminus A_{i-1}) \leq \exp(-\Omega(\lambda_i))$, for all $1 \leq i \leq L$*
(b) $f(\mathcal{M}) \leq \exp(o(\lambda_i))$ for $\mathcal{M} \in A_i \setminus A_{i-1}$.

PROPERTY 4.3. *(a) $Pr(B_j \setminus B_{j-1}) \leq \exp(-\Omega(2^{j/2} \log n))$ for all $1 \leq j \leq K$*
(b) $f(\mathcal{M}) \leq \exp(O(2^{3j/4}))$ for $\mathcal{M} \in B_j \setminus B_{j-1}$.

PROPERTY 4.4. *(a) $Pr(A_\infty) \leq \exp(-5d^2 \log n)$,*
(b) $f(\mathcal{M}) \leq \exp(4d^2 \log n)$ for all $\mathcal{M} \in A_\infty$.

Now we define the critical parameters $T_m(a)$. The heart of the proof is to show that the $T_m(a)$, as defined, satisfy the above properties.

DEFINITION 4.5. *First set $q_m = (nd - 2m)/nd$ and $p_m = 1 - q_m$ for all $m = 0, 1, \ldots, nd/2 - 1$. Next, let*

$$\begin{aligned} \alpha_m(a) &= c\sqrt{a(nd^2q_m^2 + a^2)(dq_m + a)} \\ \beta_m(a) &= c\sqrt{a(nd^3q_m^2 + a^2)(d^2q_m + a)} \\ \gamma_m(a) &= c\sqrt{a(nd^3q_m^3 + a^3)(d^2q_m^2 + a^2)} \\ \nu_m &= 8nd^3q_m^3, \end{aligned}$$

where c is a sufficiently large constant. Define

$$T_m(a) = \begin{cases} \min\Big(\alpha_m(a) + \beta_m(a) + \gamma_m(a), \alpha_m(a) + \beta_m(a) + \nu_m\Big) & \\ \qquad\qquad \text{if } nd - 2m \geq \omega a & \\ a^2/\omega \qquad \text{otherwise} & \end{cases}$$

Property 4.0 follows from the definition. The function T_m arises naturally from the large deviation consideration in the next section. Technically speaking, T_m has been set so that we can conveniently derive the large deviation parts, or parts (a), of Properties 4.2-4.4 from a general concentration result. On the other hand, it turns out that this definition of T_m does satisfy the second parts of the properties (see Section 5).

The proofs of the properties will be presented in the next two sections. A routine calculation shows that these properties together imply that $\mathbb{E}(f(\mathcal{M})1_{\mathcal{S}^*}) \leq 1 + o(1)$. In Section 8, we prove (10), showing that $\mathbb{E}\Big(f(\mathcal{M}) 1_{\mathcal{S}(M) \setminus \mathcal{S}^*(M)} \Big) = o(1)$. These two expectation estimates together yield (7). The proof of (8) is presented in Section 7. This proof follows relatively easily from certain estimates in Section 6.

5. LARGE DEVIATION

In this section, we prove the large deviation part of Properties 4.2-4.4. The most difficult proof is that of Property 4.2 and the vital tool for this proof is the following concentration result, proved by the second author in [10]. This result refines an earlier result by the authors in [5].

Consider independent random variables $t_1, \ldots, t_n$ with arbitrary distributions on the interval $[0, 1]$. For a polynomial $Y = Y(t_1, \ldots, t_n)$ of degree k and a multi-set A of size at most k, $\partial_A Y$ denotes the partial derivative of Y with respect to A. For instance, if $Y = t_1^2 t_2^2 t_3 + t_4^5$ and $A_1 = \{1, 2\}$, $A_2 = \{1, 1, 3\}$, then $\partial_{A_1}(F) = 4t_1 t_2 t_3$ and $\partial_{A_2}(F) = 2t_2^2$, respectively. If the set A is empty then $\partial_A Y = Y$. For all $0 \leq j \leq k$, let

$$\mathbb{E}_j(Y) = \max_{|A| \geq j} \mathbb{E}(\partial_A(Y))$$

be the maximum expectation of a partial derivative of order at least j of Y.

We define two parameters c_k and 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 5.1. *Let Y be a polynomial of degree k with positive coefficients at most 1. For any collection of positive numbers $\mathcal{E}_0 > \mathcal{E}_1 > \cdots > \mathcal{E}_k = 1$ and λ satisfying*
- *$\mathcal{E}_j \geq \mathbb{E}_j(Y)$*
- *$\mathcal{E}_j/\mathcal{E}_{j+1} \geq \lambda + 4j \log n$, $0 \leq j \leq k - 1$,*

the following holds

$$Pr\Big(|Y - \mathbb{E}(Y)| \geq c_k \sqrt{\lambda \mathcal{E}_0 \mathcal{E}_1} \Big) \leq d_k e^{-\lambda/4}.$$

In our present application t_i are i.i.d binary random variables and each monomial in the interested polynomials is a product of few different t_i's. While partial derivatives are convenient to use symbolically, the reader could also interpret the quantities $\mathbb{E}_j(Y)$ in a combinatorial way as follows. For $j \geq 1$, $\mathbb{E}_j(Y)$ is the maximum average effect of a group of at least j atom variables. In other words, changing the value of any group of at least j variables would change Y, in expectation with respect to the random variables outside the group, by at most $\mathbb{E}_j(Y)$.

The crucial advantage of Theorem 5.1 is that the average effects are usually much less than the maximum effect (or maximum martingale difference) one needs to consider when apply an Azuma type inequality. This advantage thus enables us to often derive a tighter concentration result, compared to what one has from Azuma's. (It would be useful for the reader to try to use Azuma's inequality instead of Theorem 5.1 and see what he or she obtain.) Theorem 5.1 and many variants are discussed in details in [10], which also contains a variety of applications.

Since a sum of independent random variables can be seen as a polynomial of degree one, one can also view Theorem 5.1 as an extension of Chernoff's bound.

5.1 Proof of Property 4.2, part (a)

We need to show

$$Pr(A_i \setminus A_{i-1}) \leq \exp(-\Omega(\lambda_i)).$$

Since $\lambda_i = 2\lambda_{i-1}$, $T_m(\lambda_i) \leq 8T_m(\lambda_{i-1})$ by definition. Given the definition of the set A_i and the fact that $\lambda_i \gg \log n$, it suffices to prove the following two lemmas

LEMMA 5.2. *For any m such that $nd - 2m \geq \omega\lambda_i$*

$$Pr\Big(\Delta_m(\mathcal{M}) - \delta_m \geq \frac{1}{8}\min\Big(\alpha_m(\lambda_i) + \beta_m(\lambda_i) + \gamma_m(\lambda_i), \alpha_m(\lambda_i) + \beta_m(\lambda_i) + \nu_m\Big)\Big) \leq \exp(-\Omega(\lambda_i)).$$

LEMMA 5.3. *For any m such that $nd - 2m < \omega\lambda_i$*

$$Pr\Big(\Delta_m(\mathcal{M}) \geq \lambda_i^2/4\omega\Big) \leq \exp(-\Omega(\lambda_i)). \quad (12)$$

Part (a) of Property 4.4 also follows instantly from Lemma 5.2 by adjusting the constant c in the definition of λ_L.

Proof of Lemma 5.2 It is more convenient to consider the following model: G_{p_m} is the random subgraph of G obtained by keeping each edge of G with probability $p_m = 2m/nd$, independently. A standard calculation shows that with probability at least $1/nd$, G_{p_m} has exactly m edges. Since $\lambda_i \gg \log nd$, to prove the lemma, it suffices to show that for each m

$$Pr_{G_{p_m}}\Big(\Delta_m(\mathcal{M}) - \delta_m \geq \tfrac{1}{8}\min\Big(\alpha_m(\lambda_i) + \beta_m(\lambda_i) + \gamma_m(\lambda_i), \alpha_m(\lambda_i) + \beta_m(\lambda_i) + \nu_m\Big)\Big) \leq \exp(-\Omega(\lambda_i)). \quad (13)$$

Till the end of this proof, we fix m and i and ignore the subindices i and m in all other relevant quantities; Pr means $Pr_{G_{p_m}}$.

Notice that δ has been defined (on purpose) to be exactly the expectation of $\Delta(G_p)$ (and $\delta^{[1]}, \delta^{[2]}$ the expectations of $\Delta^{[1]}(G_p), \Delta^{[2]}(G_p)$, respectively). Therefore, (13) is a corollary of the following fact

FACT 5.4. *We have*

$$Pr\Big(\Delta^{[1]}(G_p) - \mathbb{E}(\Delta^{[1]}(G_p)) \geq \alpha/8\Big) \leq \exp(-\Omega(\lambda)) \quad (14)$$

$$Pr\Big(\Delta^{[2]}(G_p) - \mathbb{E}\big(\Delta^{[2]}(G_p)\big) \geq \min(\frac{\beta+\gamma}{8}, \frac{\beta+\nu}{8})\Big) \leq \exp(-\Omega(\lambda)).$$

For each edge of G, define a random variable t_e as follows: $t_e = 1$ if e is not chosen in G_p and 0 otherwise. Obviously, the t_e's are i.i.d. binary random variables with mean q. We have

$$\Delta^{[1]} = \sum_u \binom{(d-d_u)}{2} = \frac{1}{2}\sum_u \sum_{\substack{e \ni u, f \ni u \\ e \neq f}} t_e t_f.$$

First, we have $\mathbb{E}(\Delta^{[1]}) \leq \frac{1}{2}nd^2q^2$; there are n choices for u, each u provides at most $\binom{d}{2} < d^2/2$ pairs e, f and each product $t_e t_f$ has expectation q^2. Next, for each t_e, $\partial_{t_e}\Delta^{[1]} = \sum_{f: f\cap e \neq \emptyset, f \neq e} t_f$; there are at most $2(d-1) < 2d$ terms in the sum, so the expectation of a partial derivative of order 1 is less than $2dq$. Finally, any partial derivative of order 2 of $\Delta^{[1]}$ is 0 or 1. So

$$\mathbb{E}_0(\Delta^{[1]}) \leq \max(\frac{1}{2}nd^2q^2, 2dq, 1), \quad \mathbb{E}_1(\Delta_1) \leq \max(2dq, 1)$$

and $\mathbb{E}_2(\Delta^{[1]}) = 1$. Now set $\mathcal{E}_0 = nd^2q^2 + 2\lambda^2$, $\mathcal{E}_1 = 2dq + \lambda$ and $\mathcal{E}_2 = 1$. Since $\lambda \gg \log nd$, it is easy to show that the conditions of Theorem 5.1 are satisfied. On the other hand, setting the constant c in Definition 4.5 large enough, one can guarantee that $c_2\sqrt{\lambda\mathcal{E}_0\mathcal{E}_1} \leq \alpha/8$. Theorem 5.1 yields

$$\begin{aligned} &Pr\Big(|\Delta^{[1]}(G_p) - \mathbb{E}(\Delta^{[1]}(G_p))| \geq \alpha/8\Big) \\ &\leq Pr\Big(|\Delta^{[1]}(G_p) - \mathbb{E}(\Delta^{[1]}(G_p))| \geq c_2\sqrt{\lambda\mathcal{E}_0\mathcal{E}_1}\Big) \\ &\leq \exp(-\Omega(\lambda)), \end{aligned}$$

completing the proof of (14).

The proof of the second part is similar. We shall present the details in the Appendix. □

Remark. Instead of $\beta/8$ and $\gamma/8$ we can write $\beta/4$ and γ, respectively. However, this makes no difference.

Proof of Lemma 5.3. In this lemma, $nd - 2m$ is small so that G_p is very dense. Consider its complement G_q. Let $N_0(u) = N(u) \cup \{u\}$. Then

$$\Delta(G_p) \leq \sum_u d_{G_q}(u) \sum_{v \in N_0(u)} d_{G_q}(v).$$

If $\Delta(G_p) \geq \lambda^2/4\omega$ then one of the following should hold

$$G_q \text{ has more than } \frac{1}{4}\omega^2\lambda \text{ edges} \quad (15)$$

$$\text{For some } u, \sum_{v \in N_0(u)} d_{G_q}(v) \geq \lambda/\omega^3. \quad (16)$$

The probability that (15) holds is at most

$$\binom{nd/2}{\frac{1}{4}\omega^2\lambda} q^{\frac{1}{4}\omega^2\lambda} \leq \Big(\frac{2end}{\omega^2\lambda}q\Big)^{\frac{1}{4}\omega^2\lambda}. \quad (17)$$

Recall that $q \le \omega\lambda/nd$; it follows that the right hand side in (17) is at most

$$\exp(-\frac{1}{4}\omega^2\lambda) = \exp(-\Omega(\lambda)) \tag{18}$$

For (16), notice that there are at most d^2 edges in G that contain at least one vertex in $N_0(u)$, and each edge contributes at most two in the sum $\sum_{v\in N_0(u)} d_{G_q}(v)$. In particular,

$$\begin{aligned} Pr(\sum_{v\in N_0(u)} d_{G_q}(v) \ge \lambda/\omega^3) &\le \binom{d^2}{\lambda/2\omega^3} q^{\lambda/2\omega^3} \\ &\le n\Big(\frac{ed^2}{\lambda/2\omega^3}q\Big)^{\lambda/2\omega^3} \\ &\le n\Big(\frac{2ed\omega^4}{n}\Big)^{\lambda/2\omega^3} \\ &\le \exp(-\Omega(\frac{\lambda}{\omega^3}\log n)) \\ &= \exp(-\Omega(\lambda)), \end{aligned} \tag{19}$$

as we have chosen $\omega \ll \log^{1/3} n$. □

5.2 Proof of Properties 4.3 and 4.4, parts (a)

This proof is more or less identical to the proof of Lemma 5.3. Again we omit the sub-index m, but let us remark that we are interested in only those m where $nd - 2m \le \omega\lambda_0$. Since $\omega\lambda_0 = \omega^2 \log n \ll \log^2 n$, there are only $o(\log^2 n)$ such indices. Therefore, it is enough to prove the bound for each individual m.

Similarly to (15) and (16), if $\Delta(G_p) \ge 2^{j-1}$ then one of the following should hold

$$G_q \text{ has more than } 2^{j/2}/2 \text{ edges}$$

$$\text{For some } u, \sum_{v\in N_0(u)} d_{G_q}(v) \ge 2^{j/2}.$$

The rest of the proof can be easily worked out along the lines of (17, 18, 19). □

By repeating the proof of Property 4.2 for the special case $i = L$, we have that

$$Pr(A_\infty) \le \exp(-\Omega(\lambda_L)) = \exp(-\Omega(\bar{c}d^2 \log n)).$$

where $\bar{c}$ is the constant in the definition of λ_L. The hidden constant in Ω does not depend on $\bar{c}$, so by setting $\bar{c}$ sufficiently large, we can have $\Omega(cd^2 \log n)) \ge 5d^2 \log n$ and

$$Pr(A_\infty) \le \exp(-5d^2 \log n),$$

as claimed. □

Remark. In fact, by increasing $\bar{c}$ we can have that

$$Pr(A_\infty) \le \exp(-ad^2 \log n),$$

for any constant a. This fact will be important in the proof of (10).

6. BOUNDS

In this section, we prove the second half of Properties 4.1, 4.2, 4.3 and 4.4. Consider

$$f(\mathcal{M}) = \prod_{m=0}^{nd/2-1} \Big(1 + \frac{\Delta_m(\mathcal{M}) - \delta_m}{\binom{nd-2m}{2} - \Delta_m(\mathcal{M})}\Big).$$

Since $\binom{nd-2m}{2} \ge (1-\varepsilon/2)^{-1}\Delta_m(\mathcal{M})$ for $\mathcal{M} \in S^*(M)$,

$$f(\mathcal{M}) \le \prod_{m=0}^{nd/2-1} \Big(1 + \frac{(3/\varepsilon)\max(\Delta_m(\mathcal{M}) - \delta_m, 0)}{(nd-2m)^2}\Big). \tag{20}$$

6.1 Property 4.2, part (b)

It is sufficient to show that for any $\mathcal{M} \in A_i$

$$\sum_{m=0}^{nd/2-1} \frac{\max(\Delta_m(\mathcal{M}) - \delta_m, 0)}{(nd-2m)^2} = o(\lambda_i)$$

To this end, we omit the sub-index i and use the short hand $g(m)$ for $\frac{\max(\Delta_m(\mathcal{M})-\delta_m,0)}{(nd-2m)^2}$. Since $g(m) = O(1)$ for any m, $\sum_{nd-2m=2}^{\lambda/\omega^{1/2}} g(m) = o(\lambda)$, so we only have to consider the sub-sum starting from $nd - 2m = \lambda/\omega^{1/2}$. Notice that the numerator in $g(m)$ is at most $T_m(\lambda)$, so due to the definition of $T_m(\lambda)$ we have

$$\begin{aligned} \sum_{nd-2m=\lambda/\omega^{1/2}}^{nd} g(m) \le \sum_{nd-2m=\lambda/\omega^{1/2}}^{\omega\lambda} \frac{\lambda^2}{\omega} + \\ \sum_{nd-2m=\omega\lambda}^{\omega\lambda_i^2} \Big(\alpha_m(\lambda) + \beta_m(\lambda) + \nu_m\Big) + \\ \sum_{nd-2m=\omega\lambda^2}^{nd/2} \Big(\alpha_m(\lambda) + \beta_m(\lambda) + \gamma_m(\lambda)\Big). \end{aligned}$$

To bound the three sums on the right hand side, we require a series of elementary estimates, presented below. The need for these estimates will be justified in the rest of the proof. In the following we use the fact that $q_m = (nd-2m)/nd$

$$\begin{aligned} \sum_{nd-2m=2}^{nd} \frac{(\lambda nd^3 q_m^3)^{1/2}}{(nd-2m)^2} &= \frac{\lambda^{1/2}}{n} \sum_{nd-2m=2}^{nd} \frac{1}{\sqrt{nd-2m}} \\ &= O\Big(\frac{\lambda^{1/2}}{n}\int_{x=2}^{nd/2} \frac{1}{\sqrt{x}}\partial x\Big) O\Big(\frac{\lambda^{1/2}}{n}\sqrt{nd}\Big). \end{aligned}$$

$$\begin{aligned} \sum_{nd-2m=2}^{nd} \frac{\lambda(nd^2 q_m^2)^{1/2}}{(nd-2m)^2} &= O\Big(\frac{\lambda}{n^{1/2}} \sum_{nd-2m=2}^{nd} \frac{1}{nd-2m}\Big) \\ &= O\Big(\frac{\lambda}{n^{1/2}}\log nd\Big) \end{aligned} \tag{21}$$

$$\begin{aligned} \sum_{nd-2m=2}^{nd} \frac{(\lambda^3 dq_m)^{1/2}}{(nd-2m)^2} &= O\Big(\frac{\lambda^{3/2}}{n^{1/2}} \sum_{nd-2m=2}^{nd} \frac{1}{(nd-2m)^{3/2}}\Big) \\ &= O\Big(\frac{\lambda^{3/2}}{n^{1/2}}\Big) \end{aligned} \tag{22}$$

$$\sum_{nd-2m=\omega\lambda}^{nd} \frac{\lambda^2}{(nd-2m)^2} \le \lambda^2 \int_{x=\omega\lambda}^{nd} x^{-2} = o(\lambda). \tag{23}$$

(21, 21, 22) imply the next three estimates, respectively.

$$\sum_{nd-2m=2}^{nd} \frac{(\lambda nd^5 q_m^3)^{1/2}}{(nd-2m)^2} = O\Big(\lambda^{1/2}\sqrt{\frac{d^3}{n}}\Big) \tag{24}$$

$$\sum_{nd-2m=2}^{nd} \frac{\lambda(nd^3q_m^2)^{1/2}}{(nd-2m)^2} = O\Big(\lambda \frac{d^{1/2}\log nd}{n^{1/2}}\Big) \qquad (25)$$

$$\sum_{nd-2m=2}^{nd} \frac{(\lambda^3 d^2 q_m)^{1/2}}{(nd-2m)^2} = O\Big(\frac{\lambda^{3/2}d^{1/2}}{n^{1/2}}\Big) \qquad (26)$$

Furthermore,

$$\begin{aligned}\sum_{nd-2m=2}^{nd} \frac{(\lambda^3 nd^3q^3)^{1/2}}{(nd-2m)^2} &= \frac{\lambda^{3/2}}{n}\sum_{nd-2m=2}^{nd}\frac{1}{(nd-2m)^{1/2}} \\ &= O\Big(\lambda^{3/2}\sqrt{\frac{d}{n}}\Big) \qquad (27)\end{aligned}$$

$$\begin{aligned}\sum_{nd-2m=2}^{nd} \frac{\lambda^2 dq}{(nd-2m)^2} &= O\Big(\frac{\lambda^2}{n}\sum\frac{1}{nd-2m}\Big) \\ &= O\Big(\frac{\lambda^2\log n}{n}\Big) \qquad (28)\end{aligned}$$

$$\begin{aligned}\sum_{nd-2m=w\lambda^2}^{nd} \frac{\lambda^3}{(nd-2m)^2} &= O\Big(\lambda^3\int_{w\lambda^2}^{\infty} x^{-2}\partial x\Big) \\ &= O\Big(\frac{\lambda^3}{\omega\lambda^2}\Big) = o(\lambda) \qquad (29)\end{aligned}$$

$$\sum_{nd-2m=2}^{w\lambda^2} \frac{nd^3q^3}{(nd-2m)^2} = \sum\frac{(nd-2m)}{n^2} = O\Big(\frac{\omega^2\lambda^4}{n^2}\Big) \qquad (30)$$

Important remark. Notice $\lambda \le \lambda_L = O(d^2\log n)$, so for $d = o(n^{1/3}/\log^{1/2} n)$ the right most formula in (21)-(30) is $o(\lambda)$. In fact, only in (30) do we need this bound on d. The stronger assumption $d \le n^{1/3-\varepsilon}$ is required only for an estimate in Section 8.

Due to the basic inequality $\sqrt{A+B} \le \sqrt{A}+\sqrt{B}$ for any two non-negative numbers A and B, we have

$$\alpha_m(\lambda) = O\Big((\lambda nd^3q_m^3)^{1/2} + \lambda(nd^2q_m^2)^{1/2} + (\lambda^3 dq_m)^{1/2} + \lambda^2\Big)$$
$$\beta_m(\lambda) = O\Big((\lambda nd^5q_m^3)^{1/2} + \lambda(nd^3q_m^2)^{1/2} + \lambda(\lambda d^2 q_m)^{1/2} + \lambda^2\Big)$$
$$\gamma_m(\lambda) = O\Big((\lambda(nd^5q_m^5)^{1/2} + (\lambda^3 nd^3q_m^3)^{1/2} + (\lambda^4 d^2 q_m^2)^{1/2} + \lambda^3\Big).$$

It follows from (21, 21, 22, 23) that

$$\sum_{nd-2m=\omega\lambda}^{nd/2} \alpha_m = o(\lambda). \qquad (31)$$

Similarly, it follows from (23, 24, 25, 26) that

$$\sum_{nd-2m=\omega\lambda}^{nd/2} \beta_m = o(\lambda). \qquad (32)$$

Notice that in (31, 32), the sum start from $\omega\lambda$. We need the bound $nd-2m \ge w\lambda^2$ only for γ (see (29)). Using (27, 28, 29), we have that

$$\sum_{nd-2m=\omega\lambda^2}^{nd/2} \gamma_m = o(\lambda). \qquad (33)$$

(31, 32, 33) shows that last sum is $o(\lambda)$. To handle the second sum, given (31) and (32), we need only show

$$\sum_{nd-2m=\omega\lambda}^{\omega\lambda^2} \frac{\nu_m}{(nd-2m)^2} = o(\lambda).$$

This follows trivially from (30) and the fact that we can set ω so that $\omega^2\lambda^3 = o(n^2)$.

Finally, by the definition of γ_m and (30), the first sum is also $o(\lambda)$, completing the proof. □

6.2 Property 4.1

By the definition of the set C, $\sum_{nd-2m=2}^{\omega\lambda_0} g(m) = 0$. It remains to show that if $\Delta_m(\mathcal{M}) - \delta_m \le T_m(\lambda_0)$ for all m such that $nd-2m \ge \omega\lambda_0$ then

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

Clearly it is sufficient to prove that

$$\sum_{nd-2m=\omega\lambda_0}^{nd/2} \frac{T_m(\lambda_0)}{(nd-2m)^2} = o(1). \qquad (34)$$

The proof of this is similar to the proof in the previous subsection, with a slight modification. Instead of (29) and (30), we shall use

$$\begin{aligned}\sum_{nd-2m=w\lambda_0^3}^{nd} \frac{\lambda_0^3}{(nd-2m)^2} &= O\Big(\lambda_0^3\int_{w\lambda^3}^{\infty} x^{-2}\partial x\Big) \\ &= O\Big(\frac{\lambda_0^3}{\omega\lambda_0^3}\Big) = o(1)\end{aligned}$$

$$\sum_{nd-2m=2}^{w\lambda_0^3} \frac{nd^3q^3}{(nd-2m)^2} = \sum\frac{(nd-2m)}{n^2} = O\Big(\frac{\omega^2\lambda_0^6}{n^2}\Big) = o(1),$$

respectively. In all other estimates, the right most formula is $o(1)$ for $\lambda = \lambda_0$. We invite the reader to work out the rest of the proof. □

6.3 Properties 4.3 and 4.4 , parts (b)

By the definition of the set B_j,

$$\sum_{nd-2m=2}^{\omega\lambda_0} g(m) \le \sum_{nd-2m=2}^{\omega\lambda_0} \frac{2^j}{(nd-2m)^2} \le 2^j,$$

completing the proof. □

Part (b) of Property 4.4 was proved in [9].

7. PROOF FOR THE LOWER BOUND

In this section we prove (8). Due to (12, 42, 43)

$$Pr\Big(|\Delta_m - \mu_m| \ge \alpha_m(\lambda_0)+\beta_m(\lambda_0)+\gamma_m(\lambda_0)\Big) = o(1), \quad (35)$$

for all m such that $nd-2m \geq \omega\lambda_0$. Consider an ordering $\mathcal{M}$ where $|\Delta_m - \mu_m| \leq \alpha_m(\lambda_0)+\beta_m(\lambda_0)+\gamma_m(\lambda_0)$ for these m. We have

$$f(\mathcal{M}) \geq \prod_{nd-2m=\omega\lambda_0^3}^{nd} \Big(1-(3/\varepsilon)\frac{\alpha_m(\lambda_0)+\beta_m(\lambda_0)+\gamma_m(\lambda_0)}{(nd-2m)^2}\Big) \times \prod_{nd-2m=2}^{nd-2m=\omega\lambda_0^3} \Big(1-(3/\varepsilon)\frac{\delta_m}{(nd-2m)^2}\Big) \qquad (36)$$

Part of the proof of (34) shows that $\sum_{nd-2m=\omega\lambda_0^3}^{nd} \alpha_m(\lambda_0)+\beta_m(\lambda_0)+\gamma_m(\lambda_0) = o(1)$. This implies that the first product is $1-o(1)$. The second product is also $1-o(1)$ since $\delta_m = O\Big((nd-2m)^2(\frac{1}{n}+\frac{m}{n^2})\Big)$ and $\omega\lambda_0 = o(\log^2 n)$. These together with (35) yield that

$$\mathbb{E}(f(\mathcal{M})) \geq 1-o(1), \qquad (37)$$

completing the proof. □

Remark. The proofs in this and the previous section yield the following corollary, which will be useful in the next section.

COROLLARY 7.1. *By setting the constant $\bar{c}$ in the definition of λ_L sufficiently large, we have*

$$\mathbb{E}\Big(\exp\Big(\frac{10}{\varepsilon^2}\sum_{m=0}^{nd/2-1}\frac{\max(\Delta_m-\delta_m(\mathcal{M}),0)}{(nd-2m)^2}\Big)\Big) = 1+o(1). \qquad (38)$$

In fact, $\frac{10}{\varepsilon^2}$ can be replaced by any constant.

To prove (38), it suffices to notice the following:

- In the proofs of the (b) part of the Property 4.2, we actually bound

  $$\exp(\sum_{m=0}^{nd/2-1}\frac{\max(\Delta_m-\delta_m(\mathcal{M}),0)}{(nd-2m)^2})$$

  instead of $f(\mathcal{M})$. We first upper bound $f(\mathcal{M})$ by $\prod_{m=0}^{nd/2-1}\Big(1+\frac{(3/\varepsilon)\max(\Delta_m(\mathcal{M})-\delta_m,0)}{(nd-2m)^2}\Big)$ (see (20)). Next, we upper bound this product by

  $$\exp\Big((3/\varepsilon)\sum_{m=0}^{nd/2-1}\frac{\max(\Delta_m-\delta_m(\mathcal{M}),0)}{(nd-2m)^2}\Big).$$

  Then we actually prove that the sum in the exponent is $o(\lambda_i)$. This, together with the (a) part of Property 4.2, imply that the contribution of the sets $A_i \backslash A_{i-1}$'s in the expectation in (38) is $o(1)$. The extra factor $10/\varepsilon^2$ does not really matter since $10/\varepsilon^2 \times o(\lambda_i)$ is still $o(\lambda_i)$.

- Similarly, one can show that the contribution from the sets $B_i \backslash B_{i-1}$'s is also $o(1)$ and the contribution from C is $1+o(1)$. Here the factor $10/\varepsilon^2$ is swallowed by the extra $\log n$ term in the (a) part of Property 4.3.

- To bound the contribution from A_∞, it is not enough to use the stated bounds in Property 4.4, because of the extra factor $10/\varepsilon^2$. To handle this extra factor, we invoke the remark at the end of subsection 5.3. This remark allows us to replace the bound $\exp(-5d^2\log n)$ in part (a) of Property 4.4 by $\exp(-\frac{50}{\varepsilon^2}d^2\log n)$. Following the proof of the (b) part of this property we have that

  $$\exp\Big(\frac{10}{\varepsilon^2}\sum_{m=0}^{nd/2-1}\frac{\max(\Delta_m-\delta_m(\mathcal{M}),0)}{(nd-2m)^2}\Big) \leq \exp(\frac{40}{\varepsilon^2}d^2\log n).$$

  Thus, the contribution of A_∞ is at most $\exp(-\frac{50}{\varepsilon^2}d^2\log n + \frac{40}{\varepsilon^2}d^2\log n) = o(1)$.

8. PROOF OF (10)

Let us recall the definition of $\mathcal{S}^*(M)$: $\mathcal{S}^*(M)$ is the set of all orderings $\mathcal{M}$ with $\Delta_m(\mathcal{M}) \leq (1-\varepsilon/2)\binom{nd-2m}{2}$ for all $m = 1,2,...,nd/2$, where ε is an arbitrarily small positive constant. We assume that $d \leq n^{1/3-\varepsilon}$. In this section, we deal with the set $\mathcal{S}(M) \setminus \mathcal{S}^*(M)$ which contains those $\mathcal{M}$ where the condition $\Delta_m(\mathcal{M}) \leq (1-\varepsilon/2)\binom{nd-2m}{2}$ is violated for some m.

It is known ([9] p.383) and not hard to prove that

$$\Delta_m(\mathcal{M}) \leq d^2(nd-2m)/2.$$

In particular, if $nd-2m \geq (1-\varepsilon/2)^{-1}d^2+1$, then $\Delta_m(\mathcal{M}) \leq (1-\varepsilon/2)\binom{nd-2m}{2}$. Thus, the condition $\Delta_m(\mathcal{M}) \leq (1-\varepsilon/2)\binom{nd-2m}{2}$ could be violated only when m is relatively close to $nd/2$, namely, $m \leq nd/2-(1-\varepsilon/2)^{-1}d^2$. Let $\mathcal{S}_i(M)$, $i = 1,...,(1-\varepsilon/2)^{-1}d^2$, be the set of all orderings $\mathcal{M}$ of M such that

$$\Delta_m(\mathcal{M}) \leq (1-\varepsilon/2)\binom{nd-2m}{2} \quad \forall\ m < nd/2-i$$

and

$$\Delta_m(\mathcal{M}) > (1-\varepsilon/2)\binom{nd-2m}{2} \quad \text{for } m = nd/2-i.$$

Since $\sum_{i=1}^{\infty} n^{-\varepsilon i} = o(1)$, in order to prove (10), it suffices to show that

$$\mathbb{E}(f(\mathcal{M})1_{S_i}) \leq (1+o(1))n^{-\varepsilon i} \text{ for all } i = 1,...,(1-\varepsilon/2)^{-1}d^2.$$

Notice that $\binom{nd-2m}{2} - \Delta_m(\mathcal{M})$ is at least $nd/2-m$, for $nd/2-m$ edges in M remain to be selected. This yields

$$\frac{\binom{nd-2m}{2}}{\binom{nd-2m}{2}-\Delta_m(\mathcal{M})} \leq nd-2m$$

and hence

$$\prod_{m=nd/2-i}^{nd/2-1}\frac{\binom{nd-2m}{2}}{\binom{nd-2m}{2}-\Delta_m(\mathcal{M})} \leq 2^i i! \leq 2i^i,$$

where the last inequality $2^i i! \leq 2i^i$ can be verified by induction. Moreover, the fact that $\Delta_m(\mathcal{M}) \leq (1-\varepsilon/2)\binom{nd-2m}{2}$ for all $m < nd/2-i$ yields

$$\prod_{m=0}^{nd/2-i-1}\frac{\binom{nd-2m}{2}-\delta_m}{\binom{nd-2m}{2}-\Delta_m(\mathcal{M})} = \prod_{m=0}^{nd/2-i-1}\Big(1+\frac{\Delta_m(\mathcal{M})-\delta_m}{\binom{nd-2m}{2}-\Delta_m(\mathcal{M})}\Big) \leq \exp\Big(\frac{5}{\varepsilon}\sum_{m=0}^{nd/2-i-1}\frac{\max(\Delta_m(\mathcal{M})-\delta_m,0)}{(nd-2m)^2}\Big).$$

Therefore,

$$f(\mathcal{M})1_{\mathcal{S}_i} = 1_{\mathcal{S}_i}\prod_{m=0}^{nd/2-1}\frac{\binom{nd-2m}{2}-\delta_m}{\binom{nd-2m}{2}-\Delta_m(\mathcal{M})} \le 2i^i 1_{\mathcal{S}_i}\exp\Big(\frac{5}{\varepsilon}\sum_{m=0}^{nd/2-i-1}\frac{\max(\Delta_m(\mathcal{M})-\delta_m,0)}{(nd-2m)^2}\Big),$$

and Hölder's inequality gives

$$\mathbb{E}(f(\mathcal{M})1_{\mathcal{S}_i}) \le 2i^i\mathbb{E}\Big(1_{\mathcal{S}_i}\exp\Big(\frac{5}{\varepsilon}\sum_{m=0}^{nd/2-i-1}\frac{\max(\Delta_m(\mathcal{M})-\delta_m,0)}{(nd-2m)^2}\Big)\Big) \le 2i^i\mathbb{E}(1_{\mathcal{S}_i})^{1-\varepsilon/2}\mathbb{E}\Big(\exp\Big(\frac{10}{\varepsilon^2}\sum_{m=0}^{nd/2-i-1}\frac{\max(\Delta_m(\mathcal{M})-\delta_m,0)}{(nd-2m)^2}\Big)\Big)^{\varepsilon/2}$$

Since (38) gives

$$\mathbb{E}\Big(\exp\Big(\frac{10}{\varepsilon^2}\sum_{m=0}^{nd/2-i-1}\frac{\max(\Delta_m(\mathcal{M})-\delta_m,0)}{(nd-2m)^2}\Big)\Big)^{\varepsilon/2} = 1+o(1),$$

we only need to show that

$$2i^i Pr(\mathcal{S}_i)^{1-\varepsilon/2} \le (1+o(1))n^{-\varepsilon i}.$$

Let $m = nd/2 - i$ and $\Gamma(u) = \Gamma_{\mathcal{M},m}(u)$ be the set of all neighbors of u (excluding u) in the graph generated by first m edges of $\mathcal{M}$. Since

$$\Delta_m(\mathcal{M}) = \frac{1}{2}\sum_u (d-d_u)\sum_{v\in\Gamma(u)\cup\{u\}}(d-d_v-1_{\{u=v\}})$$

and

$$\binom{nd/2-m}{2} = \frac{1}{2}\sum_u(d-d_u)\sum_v(d-d_v-1_{\{u=v\}}),$$

$\Delta_m(\mathcal{M}) > (1-\varepsilon/2)\binom{nd/2-m}{2}$ implies that there is a vertex u with $d-d_u>0$ such that

$$\sum_{v\in\Gamma(u)\cup\{u\}}(d-d_v-1_{\{u=v\}}) \ge (1-\varepsilon/2)\sum_v(d-d_v-1_{\{u=v\}}),$$

or equivalently

$$\sum_{v\notin\Gamma(u)\cup\{u\}}(d-d_v) \le \frac{\varepsilon}{2}\sum_v(d-d_v-1_{\{u=v\}}) = \frac{\varepsilon}{2}(nd-2m-1) \le \varepsilon i. \quad (39)$$

Notice that $d_u = |\Gamma(u)|$ and $d-d_u>0$ means $|\Gamma(u)|<d$. Moreover, any of the last i edges of $\mathcal{M}$ that is not entirley in $\Gamma(u)$ contributes at least 1 in the first sum of (39). Hence the number of such edges is at most εi, or all but at most εi of the last i edges of $\mathcal{M}$ are entirely in $\Gamma(u)$. Let $j = d-|\Gamma(u)|$ and l be the number of edges entirely in $\Gamma(u)$. Then it is required that $j\ge 1$ and $l\ge(1-\varepsilon)i$. The probability that $d-d_u>0$ and (39) without the two middle steps occur for a (fixed) vertex u is upper bounded by

$$\sum_{j\ge1}\sum_{l\ge(1-\varepsilon)i}\frac{\binom{d}{j}\binom{\binom{d-j}{2}}{l}\binom{nd/2-d-\binom{d-j}{2}}{i-j-l}}{\binom{nd/2}{i}},$$

and hence

$$Pr(\mathcal{S}_i) \le Pr\Big(\exists\text{ such a }u\Big) \le n\sum_{j\ge1}\sum_{l\ge(1-\varepsilon)i}\frac{\binom{d}{j}\binom{\binom{d-j}{2}}{l}\binom{nd/2-d-\binom{d-j}{2}}{i-j-l}}{\binom{nd/2}{i}}.$$

Using

$$\binom{d}{j}\le\frac{d^j}{j!},\quad \binom{\binom{d-j}{2}}{l}\le\frac{(d^2/2)^l}{l!},$$

$$\binom{nd/2-d-\binom{d-j}{2}}{i-j-l}\le\frac{(nd/2)^{i-j-l}}{(i-j-l)!}$$

and, as $i = O(d^2) = o((nd)^{1/2})$,

$$\binom{nd/2}{i} = (1+o(1))\frac{(nd/2)^i}{i!},$$

we have that

$$Pr(\mathcal{S}_i) \le (1+o(1))n\sum_{j\ge1}\sum_{l\ge(1-\varepsilon)i}\binom{i}{j\ \ l}\Big(\frac{2d}{nd}\Big)^j\Big(\frac{d^2}{nd}\Big)^l = (1+o(1))n\sum_{j\ge1}\sum_{l\ge(1-\varepsilon)i}\binom{i}{j\ \ l}\Big(\frac{2}{n}\Big)^j\Big(\frac{d}{n}\Big)^l,$$

where $\binom{i}{j\ \ l} = \frac{i!}{j!l!(i-j-l)!}$. Since the summand decreases at least geometrically in the ratio $O(id/n) = o(1)$ as j or l increases, the last double sum asymptotically equals the (first) term with $j=1$ and $l=(1-\varepsilon)i$. So

$$Pr(\mathcal{S}_i) \le (2+o(1))i\binom{i-1}{(1-\varepsilon)i}\Big(\frac{d}{n}\Big)^{(1-\varepsilon)i}.$$

Since $\varepsilon\le1/3$, we have

$$\binom{i-1}{(1-\varepsilon)i}\Big(\frac{d}{n}\Big)^{(1-\varepsilon)i} \le 2^{i-1}\Big(\frac{d}{n}\Big)^{(1-\varepsilon)i} \le \Big(\frac{2^{3/2}d}{n}\Big)^{(1-\varepsilon)i},$$

which together with $i\le(1-\varepsilon/2)^{-1}d^2$ and $d\le n^{1/3-\varepsilon}$ implies that

$$2i^i Pr(\mathcal{S}_i)^{1-\varepsilon/2} \le (4+o(1))i\Big(\frac{d^2}{1-\varepsilon/2}\Big(\frac{2^{3/2}d}{n}\Big)^{(1-3\varepsilon/2)}\Big)^i \le (4+o(1))i\Big(\frac{2^{3/2}}{1-\varepsilon/2}n^{1-3\varepsilon-(1-3\varepsilon/2)}\Big)^i \le (1+o(1))n^{-\varepsilon i},$$

completing the proof. □

9. REMARKS AND OPEN QUESTIONS

To start this section, we would like to make some remark concerning the sharpness of our analysis. The assumption $d\ll n^{1/3}$ was used at two places. The first is in the proof of (30) (see the remark in Section 6), where we need to assume that $d = o(n^{1/3}/\log^{1/2}n)$. The second place is the end of the proof of (10) (see the last paragraph of Section 8), where we need to assume that $d\le n^{1/3-\varepsilon}$. Our feeling is that the arguments in Section 8 are somewhat more flexible, so there might be a chance to improve the calculation here. On the

other hand, improving the analysis in Section 6 seems like a bigger challenge. Thus, while we think that our current analysis might be tightened to yield a slightly better bound, say $d \leq n^{1/3}/\text{polylog}n$, we do not see how one can obtain $d \leq n^{1/3+\varepsilon}$.

The next, and natural issue is whether one could really expect the bound $d \leq n^{1/3+\varepsilon}$ from by the proposed algorithm (Algorithm A). Wormald [11] conjectures that $n^{1/3}$ may be the right threshold for this algorithm. Except Jerrum-Sinclair result (mentioned in the introduction) there is no other result, as far as we know, about generating random regular graphs with degree larger than $n^{1/3}$.

In many cases (in both theory and practice), one would be satisfied with a random sample whose probability is comparable to the uniform probability. In other words, the distribution is not uniform, but has a constant torsion factor. Technically, we want $Pr_A(G)/p_u = \Theta(1)$ for any d-regular graph G. In fact, it is already useful to have $Pr_A(G)/p_u = \Theta(1)$ for most d-regular graphs G. The formal question we would like to pose is:

Question. *For which d $Pr_A(G)/p_u = \Theta(1)$ holds for all d-regular graph G on n vertices, with a possible exception of a set of measure $o(1)$.*

Finally, the random graphs created by Algorithm A, regardless their distribution, form a quite reasonable model for random regular graphs. The study of this model has been suggested by several researchers [8, 11]. In a recent paper [6], the present authors proved that random graphs created by Algorithm A behave very much like Erdős-Rényi graphs $(G(n,p)$ of the same density, for all $\log n \ll d \leq n-1$. By comparing with Erdős-Rényi graphs, one can easily compute several parameters (such as the chromatic number) of the random graphs in the new model.

10. APPENDIX

Here we present the proof of the second statement of Fact 5.4. To prove this statement we need to verify the following two inequalities

$$Pr\Big(\Delta^{[2]}(G_p) - \mathbb{E}(\Delta^{[2]}(G_p)) \geq \frac{\beta+\gamma}{8}\Big) \leq \exp(-\Omega(\lambda)). \quad (40)$$

and

$$Pr\Big(\Delta^{[2]}(G_p) - \mathbb{E}(\Delta^{[2]}(G_p)) \geq \frac{\beta+\nu}{8}\Big) \leq \exp(-\Omega(\lambda)). \quad (41)$$

Consider the set X of all triples (e,g,f) where e,g,f are edges of G and they (in this order) form a path of length 3 in G. A short consideration shows that $\Delta^{[2]} = \sum_{u \nsim_{G_p} v}(d-d_u)(d-d_v)$ can be expressed as

$$\sum_{(e,g,f)\in X} t_e t_f (1-t_g) = Y_1 - Y_2,$$

where $Y_1 = \sum_{(e,g,f)\in X} t_e t_f$ and $Y_2 = \sum_{(e,g,f)\in X} t_e t_f t_g$. A routine argument (similar to the one presented for $\Delta^{[1]}$) shows that $\mathbb{E}_0(Y_1) \leq \max(nd^3q^2, 2d^2q, 1)$ and $\mathbb{E}_1(Y_1) \leq \max(2d^2q, 1)$. Set $\mathcal{E}_0 = nd^3q^2 + 2\lambda^2$, $\mathcal{E}_1 = 2d^2q + \lambda$ and verify that they satisfy the conditions of Theorem 5.1. By adjusting the constant c in Definition 4.5, we can assume that $c_2\sqrt{\lambda\mathcal{E}_0\mathcal{E}_1} \leq \beta/8$ (where c_2 is the constant in Theorem 5.1). Theorem 5.1 yields

$$\begin{aligned} Pr\Big(|Y_1 - \mathbb{E}(Y_1)| \geq \beta/8\Big) &\leq Pr\Big(|Y_1 - \mathbb{E}(Y_1)| \geq c_2\sqrt{\lambda\mathcal{E}_0\mathcal{E}_1}\Big) \\ &\leq \exp(-\Omega(\lambda)). \end{aligned} \quad (42)$$

For Y_2, one can show that $\mathbb{E}_0(Y_2) \leq \max(nd^3q^3, 2d^2q^2, dq, 1)$, $\mathbb{E}_1(Y_2) \leq \max(2d^2q^2, dq, 1)$ and $\mathbb{E}_2(Y_2) \leq \max(dq, 1)$. We define $\mathcal{E}_0 = nd^3q^3 + 3\lambda^3$, $\mathcal{E}_1 = 2d^2q^2 + 2\lambda^2$ and $\mathcal{E}_2 = dq + \lambda$ and verify the conditions in Theorem 5.1. Again by adjusting the constant c in Definition 4.5, we can assume that $c_3\sqrt{\lambda\mathcal{E}_0\mathcal{E}_1} \leq \gamma/8$ (where c_3 is the constant in Theorem 5.1). Similarly to (42), we have

$$\begin{aligned} Pr\Big(|Y_2 - \mathbb{E}(Y_2)| \geq \gamma/8\Big) &\leq Pr\Big(|Y_2 - \mathbb{E}(Y_2)| \geq c_3\sqrt{\lambda\mathcal{E}_0\mathcal{E}_1}\Big) \\ &\leq \exp(-\Omega(\lambda)). \end{aligned} \quad (43)$$

(42) and (43) imply (40); (41) follows from (42) and the following simple observation

$$\begin{aligned} \Delta^{[2]} - \mathbb{E}(\Delta^{[2]}) &\leq |Y_1 - \mathbb{E}(Y_1)| + \mathbb{E}(Y_2) \\ &\leq |Y_1 - \mathbb{E}(Y_1)| + nd^3q^3 \\ &= |Y_1 - \mathbb{E}(Y_1)| + \nu/8 \end{aligned}$$

Our proof of Lemma 5.3 is thus completed.

11. REFERENCES

[1] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.* 1 (1980), no. 4, 311–316.

[2] E. Bender and R. Canfield, The asymptotic number of labeled graphs with given degree sequences, *J. Combinatorial Theory Ser.* A 24 (1978), no. 3, 296–307.

[3] J. Friedman, A proof of Alon's second eigenvalue conjecture, *preprint.*

[4] M. Jerrum and A. Sinclair, Fast uniform generation of regular graphs, *Theoret. Comput. Sci.* 73 (1990), no. 1, 91–100.

[5] J. H. Kim and V. H. Vu, Concentration of multivariate polynomials and its applications, *Combinatorica* 20 (2000), no. 3, 417–434.

[6] J. H. Kim and V. H. Vu, Sandwitching random graphs, *submitted.*

[7] B. McKay and N. Wormald, Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$, *Combinatorica* 11 (1991), no. 4, 369–382.

[8] A. Rucinski and N. Wormald, Random graph processes with degree restrictions, *Combin. Probab. Comput.* 1 (1992), no. 2, 169–180.

[9] A. Steger and N. Wormald, Generating random regular graphs quickly. (English. English summary) Random graphs and combinatorial structures (Oberwolfach, 1997), *Combin. Probab. Comput.* 8 (1999), no. 4, 377–396.

[10] V. H. Vu, Concentration of non-Lipschitz functions and applications. Probabilistic methods in combinatorial optimization, *Random Structures Algorithms* 20 (2002), no. 3, 262–316.

[11] N.Wormald, Models of random regular graphs. Surveys in combinatorics, 1999 (Canterbury), 239–298, London Math. Soc. Lecture Note Ser., 267, Cambridge Univ. Press, Cambridge, 1999.