Random regular graphs of non-constant degree: independence and chromatic number

Colin Cooper[*]   Alan Frieze[†]   Bruce Reed[‡]   Oliver Riordan[§]

November 15, 2001

Abstract

Let $r(= r(n)) \to \infty$ with $3 \le r \le n^{1-\eta}$ for an arbitrarily small constant $\eta > 0$, and let G_r denote a graph chosen uniformly at random from the set of r-regular graphs with vertex set $\{1, 2, \dots, n\}$. We prove that with probability tending to 1 as $n \to \infty$, G_r has the following properties: the independence number of G_r is asymptotically $\frac{2n \log r}{r}$ and the chromatic number of G_r is asymptotically $\frac{r}{2 \log r}$.

1   Introduction

The properties of random r-regular graphs have received much attention. For a comprehensive discussion of this topic, see the recent survey by Wormald [15] or Chapter 9 of the book *Random Graphs* by Janson, Łuczak and Ruciński [8].

A major obstacle in the development of the subject has been a lack of suitable techniques for modelling simple random graphs over the entire range, $0 \le r \le n - 1$, of possible values of r. The classical method for generating uniformly distributed simple r-regular graphs is by rejection sampling using the configuration model of Bollobás [2]. The configuration model is a probabilistic interpretation of a counting formula of Bender and Canfield [1]. The method is most easily applied when r is constant or

---

[*]Department of Mathematical and Computing Sciences, Goldsmiths College, University of London, New Cross, London SE14 6NW. Research supported by the STORM Research Centre, UNL. E-mail: `c.cooper@gold.ac.uk`.

[†]Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh PA15213, U.S.A., Supported in part by NSF grant CCR-9818411. E-mail: `alan@random.math.cmu.edu`.

[‡]Equipe Combinatoire, CNRS, Univ. de Paris VI, 4 Place Jussieu, Paris 75005, France. E-mail: `reed@ecp6.jussieu.fr`.

[§]Trinity College, Cambridge CB2 1TQ. E-mail: `omr10@dpmms.cam.ac.uk`.

grows slowly with n, the number of vertices, as n tends to infinity. The formative paper [2] on this topic considered the case where $r = O((\log n)^{1/2})$. McKay [12] and McKay and Wormald [13, 14] subsequently gave alternative approaches which are useful for $r = o(n^{1/2})$ or $r = \Omega(n)$.

Let G_r denote a graph chosen uniformly at random from the set $\mathcal{G}_r$ of simple r-regular graphs with vertex set $V = \{1, 2, \dots, n\}$. We consider the independence and chromatic number of G_r for the case where $r \to \infty$ as $n \to \infty$, but $r \le n^{1-\eta}$ for some $\eta > 0$. These are also studied in a recent paper by Krivelevich, Sudakov, Vu and Wormald [9], for the case where $r(n) \ge \sqrt{n} \log n$. Our paper complements [9] both in both in the range of r studied and in the techniques applied. Throughout the paper we say that a sequence of events $\mathcal{E}_n$ holds *with high probability* (**whp**) if $\mathbf{Pr}(\mathcal{E}_n) \to 1$ as $n \to \infty$.

Theorem 1 *Let* ϵ, η *be positive constants, then for any* $n^{1/4} \le r \le n^{1-\eta}$ **whp** *the independence number* α *of* G_r *satisfies*

$$\left| \alpha(G_r) - \frac{2n}{r}(\log r - \log \log r + 1 - \log 2) \right| \le \frac{\epsilon n}{r}. \tag{1}$$

Our proof of Theorem 1 is easily adapted to prove:

Theorem 2 *If* $n^{1/4} \le r \le n^{1-\eta}$ *then* **whp** *the chromatic number* χ *of* G_r *satisfies*

$$\chi(G_r) = \frac{r}{2 \log r} \left(1 + O\left(\frac{\log \log r}{\log r} \right) \right).$$

Frieze and Luczak [6] showed that for any fixed $\epsilon, \eta > 0$ there exists r_ϵ such that if $r_\epsilon \le r \le n^{1/3-\eta}$ then **whp** (1) is true, and that if $r \le n^{1/3-\eta}$ and r is sufficiently large then **whp**

$$\left| \chi(G_r) - \frac{r}{2 \log r} \right| \le \frac{16 r \log \log r}{(\log r)^2}. \tag{2}$$

The paper [9] also gives asymptotically tight estimates for $\alpha(G_r)$ and $\chi(G_r)$ when $n^{6/7+\eta} \le r \le 0.9n$, $\eta > 0$ constant. By proving the theorems above, we have closed the gap in the middle range of r.

We will use *edge switching*: [12], [13, 14], [4], [10], [9] and [3]. We also make use of a partitioning technique which allows us to build up simple r-regular graphs as the union of smaller less dense graphs.

2 Generating graphs with a fixed degree sequence.

Let $\mathbf{d} = (d_1, d_2, \dots, d_n)$, and let $|\mathbf{d}| = (d_1 + d_2 + \cdots + d_n)/2$. Let $\mathcal{G}_{\mathbf{d}}$ be the set of simple graphs G with vertex set $V = [n]$ and degree sequence $\mathbf{d}$ (and hence $|\mathbf{d}|$ edges).

Let W be any totally ordered set of size $2|\mathbf{d}|$ and $\Omega = \Omega(W)$ be the set of all $(2|\mathbf{d}|)!/(|\mathbf{d}|!2^{|\mathbf{d}|})$ partitions of W into $|\mathbf{d}|$ 2-element sets. An element of Ω is a *configuration*. The constituent 2-element sets of a configuration F are referred to as the *edges* of F.

Let $W_1, W_2, ..., W_n$ be the natural ordered partition $P_\mathbf{d}$ of W into sets of size $|W_i| = d_i$, where $(\max W_i) < \min W_{i+1}$ for $i < n$.

Let $\Omega = \Omega_\mathbf{d} = \Omega(W)$ with the understanding that the underlying set W is partitioned into $P_\mathbf{d}$. The *degree sequence* of an element F of $\Omega_\mathbf{d}$ is $\mathbf{d}$. Define $\phi = \phi_{P_\mathbf{d}} : W \to [n]$ by $\phi(w) = i$ if $w \in W_i$. Let $\gamma(F)$ denote the *configuration multigraph* with vertex set $[n]$ and edge multiset $E_F = \{\{\phi(x), \phi(y)\} : \{x, y\} \in F\}$.

Definition: Let $\Omega_\mathbf{d}^*$ denote those configurations F for which $\gamma(F)$ is simple relative to $P_\mathbf{d}$.

Remark 1 Note that each member of $\mathcal{G}_\mathbf{d}$ is the image under γ of precisely $\prod_{i=1}^n d_i!$ members of $\Omega_\mathbf{d}^*$. Thus sampling uniformly from $\Omega_\mathbf{d}^*$ is in some sense equivalent to sampling uniformly from $\mathcal{G}_\mathbf{d}$.

If $d_i = r$, $(1 \le i \le n)$ we will say the configuration F is r-regular. The probability $|\Omega^*|/|\Omega|$ that the underlying r-regular multigraph $\gamma(F)$ of such a configuration F is simple is $\exp(-\Theta(r^2))$. For $r = o(n^{1/2})$ this follows from [13, 14] and for larger values of r from Lemma 2 below. This result allows us to prove many results directly via configurations and then condition the probability estimates for simple graphs.

The following two lemmas are proved in [3]:

Lemma 1 *Let* $\Delta = \max_{i \in [n]} d_i$. *Suppose that* $\Delta \le n/1000$ *and that* $\mathbf{d}$ *satisfies* $\min_{i \in [n]} d_i \ge \Delta/4$. *Given* $a, b \in [n]$, *if* G *is sampled uniformly at random (u.a.r.) from* $\mathcal{G}_\mathbf{d}$, *then*
$$\mathbf{Pr}(\{a, b\} \in E(G)) \le \frac{20\Delta}{n}.$$

Lemma 2 *Suppose* $100 \le r \le n/1000$. *Let* $d_j = r$, $1 \le j \le n$. *If* F *is chosen u.a.r. from* Ω *then for* n *sufficiently large,*
$$\mathbf{Pr}(F \in \Omega^*) \ge e^{-2r^2}.$$

Remark 2 We also need to consider random bipartite graphs with a given degree sequence and a similar configuration model is available for this purpose. We just consider a multigraph obtained from random pairings F between two equal sized disjoint sets $W, \tilde{W}$ equipped with partitions $W_1, W_2, \ldots, W_n$ and $\tilde{W}_1, \tilde{W}_2, \ldots, \tilde{W}_{\tilde{n}}$. A pair (x, y), $x \in W_i, y \in \tilde{W}_j$ giving rise to the edge (i, j).

At this point we describe two algorithms for obtaining a u.a.r. configuration. We describe bipartite versions since we covered the non-bipartite versions in [3]. In any case the necessary changes for this case should be clear.

Algorithm CONSTRUCT
begin
    $F_0 := \emptyset$; $R_0 := W$, $\tilde{R}_0 := \tilde{W}$
    For $i = 1$ **to** $|\mathbf{d}|$ **do**
    begin
        Choose $u_i \in R_{i-1}$ *arbitrarily*
        Choose v_i *uniformly at random* from $\tilde{R}_{i-1}$
        $F_i := F_{i-1} \cup \{\{u_i, v_i\}\}$; $R_i := R_{i-1} \setminus \{u_i\}, \tilde{R}_i := \tilde{R}_{i-1} \setminus \{v_i\}$
    end
    Output $F := F_{|\mathbf{d}|}$.
end

Algorithm GENERATE
begin
    $F_0 := \emptyset$
    Let $\boldsymbol{\sigma}_1 = (x_1, x_2, \ldots, x_{|\mathbf{d}|-1}, x_{|\mathbf{d}|})$ be an ordering of W
    Let $\boldsymbol{\sigma}_2 = (y_1, y_2, \ldots, y_{|\mathbf{d}|-1}, y_{|\mathbf{d}|})$ be an ordering of $\tilde{W}$
    For $i = 1$ **to** $|\mathbf{d}|$ **do**
    begin

$$F_i := \begin{cases} F_{i-1} \cup \{\{x_i, y_i\}\} & \left(\text{With probability } \tfrac{1}{i}\right) \ \mathbf{A} \\[2mm] F_{i-1} \cup \{(x_i, \eta_1\}, \{\xi_1, y_i\}\} - \{\xi_1, \eta_1\} & \left(\text{With probability } \tfrac{i-1}{i}\right) \ \mathbf{B} \end{cases}$$

        Here $\{\xi_1, \eta_1\}$ is chosen u.a.r. from F_{i-1} (with $\xi_1 \in W$).
    end
    Output $F := F_{|\mathbf{d}|}$
end

Remark 3 Neither of the algorithms generating $F_{|\mathbf{d}|}$ use any information about the partition $P_{\mathbf{d}}$ associated with the configuration. After iteration i of GENERATE, F_i is a u.a.r. element of Ω_i, the set of configurations on $\{x_1, \ldots, x_i\}$, $\{y_1, \ldots, y_i\}$. We can, if we wish, construct an $F_I \in \Omega_I$ in some other way and and then switch to GENERATE, starting from step $I + 1$. Instead of initialising the orderings $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$ used in algorithm GENERATE with W, $\tilde{W}$ we initialize $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$ with R_I, $\tilde{R}_I$, the remaining unmatched points. If F_I is distributed as a u.a.r. element of Ω_I then the F obtained in this way will be uniformly chosen from Ω.

3 The independence number

3.1 The lower bound

To prove the lower bound in Theorem 1 we will follow the basic strategy of [6]. We start with the following result of Frieze [5]. Let $c_\epsilon n \le m \le n^2/\log^2 n$ and let $d = 2m/n$, then

$$\left| \alpha(G_{n,m}) - \frac{2n}{d}(\log d - \log\log d + 1 - \log 2) \right| \le \frac{\epsilon n}{d} \qquad (3)$$

with probability

$$1 - \exp\left\{ -\Omega\left(\frac{\epsilon^2 n}{2d(\log d)^2} \right) \right\}. \qquad (4)$$

We will choose $m \sim rn/2$ and define a random multi-graph as follows: let $(v_1, v_2, \ldots, v_{2m})$ be chosen u.a.r. from $[n]^{2m}$ and let H_m have vertex set $[n]$ and edge set $\{\{v_{2i-1}, v_{2i}\} : i = 1, 2, \ldots, m\}$.

While [5] proves the independence number result (3) for the standard model $G_{n,m}$, we can deduce essentially the same result for H_m by removing loops and repeated edges to obtain $G_{n,m'}$ with $m' \sim m$. The paper [6] starts with H_m and then transforms it into $F \in \Omega_{\mathbf{r}}$ ($\mathbf{r} = (r, r, \ldots, r)$) and then into G_r without changing the independence number by much. This needed $r \le n^{1/3-\eta}$ so that the transition from $F \in \Omega_{\mathbf{r}}$ to G_r could be done easily.

In this paper, because the degree, r, is larger, we introduce a decomposition technique (Section 3.1.1) which enables us to apply the results and techniques of [5, 6] to larger values of r.

3.1.1 A decomposition of G_r

Let $s = r^{1-\eta/10}$ where η is as in Theorem 1 and is sufficiently small. Let $\nu = n/s$ and let $V_1, V_2, \ldots, V_s$ be a random partition of $[n]$ into sets of size ν. (We assert that we can afford to ignore the niceties of rounding. In reality $s = \lfloor r^{1-\eta/10} \rfloor$ and $|V_i| = \lfloor \nu \rfloor$ or $\lceil \nu \rceil$).

Let $\Gamma_{i,i} = G_r[V_i]$ be the subgraph of G_r induced by V_i and for $i \ne j$ let $\Gamma_{i,j}$ be the bipartite subgraph of G_r with vertex partition V_i, V_j and all G_r-edges joining V_i to V_j. Let $\mathbf{d}_{i,j}$ denote the degree sequence of $\Gamma_{i,j}$. We observe that if i, j and $\mathbf{d}_{i,j}$ are given then $\Gamma_{i,j}$ is a random graph or bipartite graph with this degree sequence and that, furthermore, the $\Gamma_{i,j}$ are conditionally independent once the V_i and $\mathbf{d}_{i,j}$ are given. This

is because any two graphs on V_i, V_j with degree sequence $\mathbf{d}_{i,j}$ have precisely the same set of extensions to an r-regular graph on V.

The degree $d_{i,j}(v)$ of $v \in V_i$ in $\Gamma_{i,j}$ is sharply concentrated around its mean. Indeed the randomness of the partition and Theorem 2 of Hoeffding [7] (sampling without replacement) gives
$$\mathbf{Pr}(|d_{i,j}(v) - \rho| \geq \kappa\rho) \leq 2e^{-\kappa^2 r/(4s)}$$
where $\rho = r/s$ and we have replaced the usual 3 by 4 to account for some rounding.

Putting $\kappa = (s/r)^{1/2} \log n$ we see that **whp**
$$|d_{i,j}(v) - \rho| < \rho^{1/2} \log n \quad \text{for all } i, j, v \in V_i. \tag{5}$$

To generate G_r given degree sequences satisfying (5), it is enough to independently generate $\Gamma_{i,j}$, where the $\Gamma_{i,j}$ are random graphs on $V_i \cup V_j$ with degree sequence $\mathbf{d}_{i,j}$.

We can therefore analyse G_r by focusing on a typical set of degree sequences $\{\mathbf{d}_{i,j},\ i,j = 1, ..., s\}$ and then *independently* generating the $\Gamma_{i,j}$. One can thus analyse G_r as the union of an independently chosen collection of random graphs. Each $\Gamma_{i,j}$ will have ν or 2ν vertices and maximum degree $\sim \rho = r^{\eta/10} \leq \nu^{1/10}$ when η is small. This is small enough that simple switching analysis will be practical. We expect this model to be useful for proving many properties of G_r.

For the rest of this section we fix degree sequences $\mathbf{d}_{i,j}$ satisfying (5).

3.1.2 From H_m to $F_{i,j}$

The transition from H_m to G_r will be in two stages. The first stage is to obtain independent u.a.r. configurations with degree sequences $\mathbf{d}_{i,j}$.

We define m by
$$m = \frac{rn}{2}\left(1 - 3\rho^{-1/2} \log n\right). \tag{6}$$

We fix $V_1, V_2, \ldots, V_s$ and a set of degree sequences $\mathbf{d}_{i,j}$ satisfying (5). We then generate H_m by choosing a sequence $\boldsymbol{\tau} = (v_1, v_2, \ldots, v_{2m})$ u.a.r. from $[n]^{2m}$. For each $1 \leq i \leq j \leq s$ we let $X^*_{i,j}$ denote the set of pairs $\{\{x_{2t-1}, x_{2t}\} : x_{2t-1} \in V_i, x_{2t} \in V_j$ or vice-versa$\}$. The multigraph on $V_i \cup V_j$ defined by these pairs is denoted by $H^*_{i,j}$. The degree $d^*_{i,j}(v)$ of $v \in V_i$ in $H^*_{i,j}$ is distributed as a binomial $B(m, 2(sn)^{-1})$, $i \neq j$ or as the sum of m independent copies of a certain $\{0, 1, 2\}$-valued random variable if $i = j$. Given its degree sequence $\mathbf{d}^*_{i,j}$, $H^*_{i,j}$ will be distributed as a u.a.r. configuration multigraph with this degree sequence. Also, given the $\mathbf{d}^*_{i,j}$, the $H^*_{i,j}$ are independent. Formally, if $i = j$, let $W^{(i,*)} = \bigcup_{v \in V_i} W^{(i,*)}_v$ where $|W^{(i,*)}_v| = \mathbf{d}^*_{i,i}(v)$. We associate each $X^*_{i,j}$ with $\prod_{v \in V_i} |W^{(i,*)}_v|!$ distinct pairings of $W^{(i,*)}$. A similar argument holds for $i \neq j$ and sets $W^{(i,j,*)}_v, v \in V_i$ and $\tilde{W}^{(i,j,*)}_v, v \in V_j$.

6

By standard calculations we see that **whp**

$$\rho - 4\rho^{1/2}\log n \le d_{i,j}^*(v) \le \rho - 2\rho^{1/2}\log n \qquad \forall i,j,v \in V_i. \tag{7}$$

At this point we remind the reader that from (3) and the fact (6) that m is sufficiently close to $rn/2$, with probability given in (4) the simple graph underlying H_m has an independent set I where

$$\left| |I| - \frac{2n}{r}(\log r - \log\log r + 1 - \log 2) \right| \le \frac{2\epsilon n}{r}. \tag{8}$$

We now wish to transform each $H_{i,j}^*$ into a random configuration multigraph with degree sequence $\mathbf{d}_{i,j}$ in place of $\mathbf{d}_{i,j}^*$. For $i \ne j$ we imagine that $H_{i,j}^*$ has been obtained through a random pairing $F_{i,j}^*$ of $W^{(i,j,*)} \subseteq W^{(i,j)}$ with $\tilde{W}^{(i,j,*)} \subseteq \tilde{W}^{(i,j)}$. We now apply algorithm GENERATE initialized with $F_{i,j}^*$ and with $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2$ as random orderings $\boldsymbol{\sigma}_1^{(i,j)}, \boldsymbol{\sigma}_2^{(i,j)}$ of $W^{(i,j)} \setminus W^{(i,j,*)}$ and $\tilde{W}^{(i,j)} \setminus \tilde{W}^{(i,j,*)}$. The configuration so constructed is denoted by $F_{i,j}$, and is a u.a.r. configuration from $\Omega_{\mathbf{d}_{i,j}}$. A similar construction is applied when $i = j$. The union of the multigraphs $\gamma(F_{i,j})$ is an r-regular multigraph and in the next section we bound the effect of GENERATE on the independent set I.

3.1.3   Edges created inside I

In going from the $F_{i,j}^*$'s to the $F_{i,j}$'s algorithm GENERATE will probably add some edges with both endpoints in I. Fix some $i \ne j$, and a configuration $F_{i,j}^*$ whose degree sequence satisfies (7). Let $\boldsymbol{\sigma}_1^{(i,j)}, \boldsymbol{\sigma}_2^{(i,j)} = (x_1, x_2, \dots, x_l), (y_1, y_2, \dots, y_l)$ be defined as above by randomly ordering the remaining configuration points. Note that $l = \Theta(\nu\rho^{1/2}\log n)$.

Fix $u \in V_i$ and $v \in V_j$. What is the probability that the edge uv is added when we run GENERATE? For steps of type A we have a bound

$$O\left(l \times \frac{1}{\nu\rho} \times \frac{\rho^{1/2}\log n}{l} \cdot \frac{\rho^{1/2}\log n}{l}\right) = O\left(\frac{\log n}{\nu^2\rho^{1/2}}\right).$$

Explanation: The first term is the number of steps, the second term is the probability that we execute step A as opposed to step B. From the degree bounds (5) and (7), there are $O(\rho^{1/2}\log n)$ x's with $\phi(x) = u$. Using only this and the randomness of the ordering we obtain the third term as a bound on the probability that $\phi(x_k) = u$ and $\phi(y_k) = v$ for a given k.

For steps of type B the corresponding bound is

$$O\left(2 \times l \times \frac{\rho^{1/2}\log n}{l} \times \frac{\rho}{\rho\nu}\right) = O\left(\frac{\rho^{1/2}\log n}{\nu}\right).$$

Explanation: The second term is the number of steps. The third term bounds the probability that $\phi(x_k) = u$ and the last term the probability that the edge (ξ_1, η_1) has $\phi(\eta_1) = v$. The factor 2 allows for $\phi(y_k) = v$, $\phi(\xi_1) = u$.

A similar argument holds when $i = j$ (even if $u = v$), so (assuming (7)) the probability that for fixed $u, v \in V$ the edge uv is added when we run GENERATE on the relevant $F_{i,j}^*$ is $O(\nu^{-1}\rho^{1/2}\log n)$. The expected number of edges added to I is thus

$$O\left(\frac{\rho^{1/2}\log n}{\nu}|I|^2\right) = o(n/r),$$

so **whp** the r-regular multigraph $M_r = \bigcup_{i,j}\gamma(F_{i,j})$ contains an independent set $\hat{I}$ where

$$\left||\hat{I}| - \frac{2n}{r}(\log r - \log\log r + 1 - \log 2)\right| \leq \frac{3\epsilon n}{r}. \tag{9}$$

3.1.4   Bad loops and multiple edges

Returning for the moment to H_m and the associated $F_{i,j}^*$, for each i, j let $G_{i,j}$ be the simple graph obtained by merging multiple edges and deleting loops of $\gamma(F_{i,j}^*)$ and let $\Gamma_r = \bigcup_{i,j} G_{i,j}$. The choice of I can be assumed to depend only on Γ_r. Suppose that for each i, j there are $a_{i,j}$ loops and $b_{i,j}$ parallel (non-loop) edges in $\gamma(F_{i,j}^*)$.

Remark 4 From the definition of H_m, given $G_{i,i}$ and $a_{i,i}$, the $a_{i,i}$ loops in $\gamma(F_{i,i}^*) = H_m[V_i]$ are independently chosen uniformly from the $|V_i| = \nu$ possibilities, with replacement. Similarly, if we fix the multiplicities of the parallel edges in $\gamma(F_{i,j}^*)$ then the edges with multiplicity greater than one are chosen uniformly at random from the $\binom{\nu}{2}$ possibilities (without replacement).

Call a loop or parallel edge a *bad edge* if it contains a member of I. Now **whp** for each i, j we have $a_{i,j} + b_{i,j} = O(\rho^2\log\nu)$ (using (7) and Lemma 5 from the Appendix) and so using Remark 4 the expected number of bad edges in $\gamma(F_{i,j}^*)$ meeting I in V_i, say, is

$$O\left(\rho^2\log\nu \times \frac{|V_i \cap I|}{\nu}\right).$$

Using the calculation of the previous section, the expected number of bad loops added going from $F_{i,i}^*$ to $F_{i,i}$ is $O(\nu^{-1}\rho^{1/2}\log n|V_i \cap I|)$. Also, as $F_{i,j}^*$ has at most $\rho|V_i \cap I|$ edges with one end in $V_i \cap I$, the expected number of edges added parallel to such edges is

$$O\left(\frac{\rho^{3/2}\log n}{\nu}|V_i \cap I|\right).$$

Using an argument similar to that in the previous section one can also bound the probability that a certain edge is added twice. Putting these bounds together we see that:

the expected number of bad edges in $\gamma(F_{i,j})$ meeting $V_i \cap I$ is

$$O\left(\rho^2 \log \nu \times \frac{|V_i \cap I|}{\nu}\right) \qquad \forall i,j. \quad (10)$$

From the next section we can forget H_m and the $F_{i,j}^*$. We have now shown that if we take fixed $\mathbf{d}_{i,j}$ satisfying (5) and construct independent u.a.r. configurations $F_{i,j} \in \Omega_{\mathbf{d}_{i,j}}$ then **whp** (9) holds, and (10) holds. There is one more fact we shall need about $\hat{I}$: the partition of the vertex set can be taken to be independent of the choice of I. Thus given i and $|I|$ the expectation of $\binom{|V_i \cap I|}{2}$ is

$$\binom{|I|}{2}\frac{|V_i|}{n}\frac{|V_i|-1}{n} \leq |I|^2/s^2,$$

so as $\hat{I} \subset I$, **whp**

$$\sum_i \binom{|V_i \cap \hat{I}|}{2} = O(|\hat{I}|^2/s). \quad (11)$$

3.1.5 Simplification: from $F_{i,j}$ to $\Gamma_{i,j}$

The multigraphs $\gamma(F_{i,j})$ have the right degree sequences $\mathbf{d}_{i,j}$, but we now need to transform them into random simple graphs with these degree sequences.

We focus on graphs, the construction for bipartite graphs is similar. We show how to *simplify* a random configuration F from $\Omega_{\mathbf{d}}$ where

$$\mathbf{d} = (d_1, d_2, \dots, d_\nu) \text{ and } \rho/2 \leq \min_i d_i \leq \max_i d_i \leq 2\rho.$$

Recalling that $n^{1/4} \leq r \leq n^{1-\eta}$, let

$$\epsilon_1 = \frac{\rho^5}{\nu^{1/2}} \leq n^{-\eta/10}.$$

An edge $\{w_1, w_2\} \in F$ is a *loop* if $\phi(w_1) = \phi(w_2)$. An edge $\{w_1, w_2\} \in F$, $w_1 < w_2$ is *redundant* if F contains an edge $\{w_1', w_2'\}$, $w_1' < w_2'$ with $\phi(w_i') = \phi(w_i), i = 1, 2$ such that $w_1' < w_1$, recalling that ϕ is increasing. It is convenient to ignore multiple loops at the same vertex when computing the number of redundant edges. Let $\Omega_{a,b}$ be the set of configurations in $\Omega_{\mathbf{d}}$ which have a loops and b redundant edges. As an intermediate stage we will generate configurations in a set $\Omega_{0,0}^{good} \subset \Omega_{0,0}$ defined in the Appendix.

9

Since whether a configuration F is good (in $\Omega_{0,0}^{good}$) depends not just on the graph $\gamma(F)$, we must consider the set

$$\mathcal{G}_{\mathbf{d}}^{good} = \{G \in \mathcal{G}_{\mathbf{d}} :$$
$$|\{F \in \Omega_{0,0}^{good} : \gamma(F) = G\}| \geq (1 - \epsilon_1)|\{F \in \Omega_{0,0} : \gamma(F) = G\}|\}, \quad (12)$$

noting that the latter count depends only on $\mathbf{d}$.

Algorithm SIMPLIFY

1. Start with a u.a.r. configuration $F^* \in \Omega_{\mathbf{d}}$.

2. Suppose F^* has a loops and b redundant edges.

3. If $a > \rho \log \nu$ or $b > \rho^2 \log \nu$, output $G^S = \perp$ — construed as **failure**.

4. Suppose the redundant edges of F^* are $e_i, i = 1, 2, \dots, b$ and the loops of F^* are $e_i, i = b+1, b+2, \dots, a+b$. Here $e_i = \{x_i, y_i\}$, $x_i < y_i$ for $i = 1, 2, \dots, a+b$ and $x_i < x_{i+1}$ for $i = 1, 2, \dots, b-1$ and $i = b+1, b+2, \dots, a+b-1$.

5. Let Σ denote the set of sequences $(\alpha_1, \beta_1, \dots, \alpha_{a+b}, \beta_{a+b})$ such that

$$f_i = \{\alpha_i, \beta_i\} \in F^* \setminus \{e_1, \dots, e_{a+b}\}, \, i = 1, 2, \dots a + b,$$

$$dist(e_i, f_j) \geq 2 \text{ and } dist(f_i, f_j) \geq 1, \text{ for } 1 \leq i \neq j \leq a + b, \quad (13)$$

   and if

$$F^{**} = F^* - e_1 - f_1 - \cdots - e_{a+b} - f_{a+b} +$$
$$+ \{x_1, \alpha_1\} + \{y_1, \beta_1\} + \cdots + \{x_{a+b}, \alpha_{a+b}\} + \{y_{a+b}, \beta_{a+b}\} \quad (14)$$

   then
$$F^{**} \in \Omega_{0,0}.$$

   Here $dist$ denotes minimum distance in $\gamma(F^*)$ between vertices of the given edges.

   Choose a random member of Σ and carry out the corresponding switching $\sigma : F^* \to F^{**}$ defined by (14).

6. For $F \in \Omega_{0,0}$ define $\pi_{F,a,b}$ by letting $\frac{\pi_{F,a,b}}{|\Omega_{0,0}|}$ be the probability that $F^{**} = F$ at this point, conditional on a, b.
   As shown in the Appendix, for $F \in \Omega_{0,0}^{good}$ we have $\pi_{F,a,b} \in [1 - \epsilon_1, 1 + \epsilon_1]$. If $F^{**} \notin \Omega_{0,0}^{good}$ then Output $G^S = \perp$.

7. Let $G^{**} = \gamma(F^{**})$. If $G^{**} \notin \mathcal{G}_{\mathbf{d}}^{good}$ then Output $G^S = \perp$. Let $\frac{\tilde{\pi}_{G,a,b}}{|\mathcal{G}_{\mathbf{d}}|}$ be the probability that we reach this point with $G^{**} = G$, given a, b. As shown in the Appendix, $\tilde{\pi}_{G^{**},a,b} \in [1 - 2\epsilon_1, 1 + \epsilon_1]$.

$$\text{Output} \begin{cases} G^S = G^{**} & \text{Probability } \frac{1-2\epsilon_1}{\tilde{\pi}_{G^{**},a,b}} \\ G^S = \perp & \text{Probability } \frac{\tilde{\pi}_{G^{**},a,b}+2\epsilon_1-1}{\tilde{\pi}_{G^{**},a,b}} \end{cases}$$

The exchange of edges in Step 5 is called a *switching* (or set of switchings).

The properties of SIMPLIFY we need are given in the following Theorem. Here $\mathbb{P}$ refers to the probability space used to make the decisions in SIMPLIFY (i.e., to choose F^*, to choose σ, and then to make the choice in step 7).

Theorem 3

(a) $\mathbb{P}(G^S = \perp) \leq 4\epsilon_1$.

(b) $|\mathcal{G}_{\mathbf{d}}^{good}| \geq (1 - \nu^{-\frac{1}{2}\rho^2 \log \nu})|\mathcal{G}_{\mathbf{d}}|$.

(c) $\mathbb{P}(G^{**} \in \mathcal{G}_{\mathbf{d}}^{good}) \geq 1 - \nu^{-\frac{1}{3}\rho \log \log \nu}$.

(d) *If $G \in \mathcal{G}_{\mathbf{d}}^{good}$ then*
$$\mathbb{P}(G^S = G \mid G^S \neq \perp) = |\mathcal{G}_{\mathbf{d}}^{good}|^{-1}.$$

$\square$

The proof of this theorem is much as in previous papers [12, 13, 14, 4] but we give it in an appendix for completeness.

Let us apply this theorem to finish the proof of the lower bound in Theorem 1. We observe first that when a u.a.r. r-regular graph G_r is partitioned, the parts $\Gamma_{i,j}$ satisfy

$$\mathbf{Pr}(\exists i, j : \Gamma_{i,j} \notin \mathcal{G}_{\mathbf{d}_{i,j}}^{good}) \leq o(1) + s^2 \nu^{-\frac{1}{2}\rho^2 \log \nu} = o(1),$$

where the first $o(1)$ is the probability that (5) does not hold, and the second term is from Theorem 3(b). Thus if we first generate degree sequences $\mathbf{d}_{i,j}$ satisfying (5) according to an appropriate distribution, and then generate independent u.a.r. elements $\Gamma_{i,j}$ of $\mathcal{G}_{\mathbf{d}_{i,j}}^{good}$, the graph $\tilde{G}_r = \bigcup_{i,j} \Gamma_{i,j}$ is an almost uniform r-regular graph. 'Almost' means that the same events hold with probability $1 - o(1)$ in either model.

So, starting with the $F_{i,j}$'s we use SIMPLIFY to generate random members $\Gamma_{i,j}$ of $\mathcal{G}_{\mathbf{d}_{i,j}}^{good}$. One problem with SIMPLIFY is that we cannot guarantee that **whp** there is no i, j for

which the algorithm returns $\perp$. For those i,j which do return $\perp$ we generate $\Gamma_{i,j}$ by direct sampling, i.e., we choose it at random from $\mathcal{G}^{good}_{\mathbf{d}_{i,j}}$.

We next estimate the expected number of edges that SIMPLIFY introduces into $\hat{I}$. We create an edge contained in $\hat{I}$ only when we delete a bad edge e_k and $\{\alpha_k, \beta_k\} \cap \hat{I} \neq \emptyset$. Let $\beta_{i,j}$ denote the expected number of bad edges in $\gamma(F_{i,j})$ meeting $V_i \cap \hat{I}$. Since in step 5 of simplify only a small proportion of the possible sequences $(\alpha_1, \ldots, \beta_{a+b})$ are excluded, for each bad $e_k = \{x_k, y_k\}$ with $x_k \in V_i \cap \hat{I}$, conditional on everything relevant the probability that $\alpha_k \in V_j \cap \hat{I}$ is at most $(1 + o(1))|V_j \cap \hat{I}|/\nu$. Also, as $\phi(\alpha_k) = \phi(x_k)$ is ruled out, the probability that $\alpha_k \in V_i \cap \hat{I}$ is at most $(1 + o(1))(|V_i \cap \hat{I}| - 1)/\nu$. Hence the expected number of edges created in $\hat{I}$ is bounded by

$$
O\left(\sum_i \sum_j \beta_{i,j} \frac{|V_j \cap \hat{I}| + |V_i \cap \hat{I}| - 1}{\nu}\right)
$$

$$
= O\left(\sum_i \sum_j \rho^2 \log \nu \times \frac{|V_i \cap \hat{I}|}{\nu} \times \left(\frac{|V_j \cap \hat{I}| + |V_i \cap \hat{I}| - 1}{\nu}\right)\right)
$$

$$
= O\left(\rho^2 \log \nu \frac{|\hat{I}|^2}{\nu^2}\right)
$$

$$
= O((\log n)^3).
$$

Here we used (10) for the expected number of bad edges, and then (11).

The term $O((\log n)^3)$ is $o(n/r)$ as required.

Finally, we consider the edges introduced into $\hat{I}$ in the cases where SIMPLIFY produces $\perp$. It follows from Theorem 3(c) that with probability at least $1 - o(n^{-2})$, for every i, j, the execution of SIMPLIFY produces $G^{**} \in \mathcal{G}^{good}_{\mathbf{d}}$. Then conditional on this occurrence, the iterations that output $\perp$ are determined by the random choices in Step 7, which are independent of $\hat{I}$. Thus the number of edges introduced into $\hat{I}$ by failing iterations has expectation bounded by

$$
O(|\hat{I}|^2 \times \epsilon_1 \times 20\rho/\nu) = O\left(|\hat{I}| \frac{\rho^{9/2} \log r}{\sqrt{n/r}}\right) = o(n/r) \tag{15}
$$

where we have used (9) and the final factor $\frac{20\rho}{\nu}$ is from Lemma 1.

Thus the total number of edges introduced into I by our process has expectation $o(n/r)$ and this becomes a high probability bound using the Markov inequality. This completes our discussion of the lower bound in Theorem 1.

3.2 Upper bound

We now consider the upper bound of Theorem 1. This is usually a straightforward application of the first moment method. Here the model makes it more difficult. The proof we give is identical to the first part of the proof of Theorem 2.2 of [9] except that we give enough details to show that the conclusion (1) holds.

Lemma 3 *Fix $A \subseteq V, |A| = a \geq 2$. Let $\mathcal{C}_k \subseteq \mathcal{G}_r$ be the graphs for which A contains k edges. Then*

$$\frac{|\mathcal{C}_k|}{|\mathcal{C}_{k-1}|} = \frac{1}{k} \binom{a}{2} \frac{r}{n} \left(1 + O\left(\frac{1}{a} + \frac{1}{r} + \frac{k}{ar} + \frac{k}{a^2} + \frac{a}{n} + \frac{r}{n} \right) \right). \tag{16}$$

Proof   Fix an r-regular graph G and suppose A contains k edges. For $v \in A$, let d_v denote the number of neighbours of v in A. Let ϕ be given by

$$\phi = \sum_{i \neq j \in A} (r - d_i)(r - d_j) = ((r - d_1) + \cdots + (r - d_a))^2 - \sum_{i \in A} (r - d_i)^2.$$

The function $\sum (r - d_i)^2$ is minimized, subject to $\sum d_i = 2k$, at $d_i = 2k/a$. The maximum is at $d_i = r$, $i = 1, ..., 2k/r$, $d_i = 0$, $i = 2k/r + 1, ..., a$ (with suitable interpolation). Thus

$$(ar - 2k)^2 - r^2(a - 2k/r) \leq \phi \leq (ar - 2k)^2 - a(r - 2k/a)^2$$

and so after some simplification (note that $k \leq ar/2$) we see that

$$\phi = 2 \binom{a}{2} r^2 \left(1 + O\left(\frac{1}{a} + \frac{k}{ar} \right) \right).$$

Denote by ρ the number of (unordered) pairs of edges ux, vy of G_r between A and $V \setminus A$ which satisfy the properties

P0: $u, v \in A$ and $x, y \notin A$.

P1: $u \neq v$ and the edge $uv \notin A$,

P2: $x \neq y$.

Thus ρ is given by

$$\rho = \tfrac{1}{2}\phi - \psi - \eta,$$

where ψ is the sum of $(r - d_u)(r - d_v)$ over the k edges uv of A, so $\psi \leq kr^2$, and η is the overcounting due to coinciding pairs ux, vx, so $\eta \leq a^2 r$.

13

Hence,

$$\rho = \binom{a}{2} r^2 \left(1 + O\left(\frac{1}{a} + \frac{1}{r} + \frac{k}{ar} + \frac{k}{a^2} \right) \right)$$

Let B be a bipartite graph with bipartition $(\mathcal{C}_{k-1}, \mathcal{C}_k)$ and an edge from $G \in \mathcal{C}_{k-1}$ to $G' \in \mathcal{C}_k$ if a switching can be made, as described below. Thus $|\mathcal{C}_{k-1}| d_L = |\mathcal{C}_k| d_R$, where d_L (resp. d_R) is the average degree of the left (resp. right) bipartition.

What constraints must we place on the choice of edges to use in switching? We assume below that $u, v \in A$, $x, y \notin A$ and u, v, w, x, y, z are distinct.

Switch up $(G \to G')$: $ux, vy, wz \to uv, wx, yz$.

We require $wz \notin A$, excluding a total of $k - 1$ edges. To ensure simplicity of G', we require $wx, yz \notin G$. On choosing wz the vertices w, z must not be adjacent to $x, y \in N(u, v)$, excluding a total of at most $2r^2$ edges. Thus

$$d_L = 2\rho(nr/2 - k - O(r^2)) = \rho nr \left(1 + O\left(\frac{k}{rn} + \frac{r}{n} \right) \right).$$

Switch down $(G' \to G)$: $uv, wx, yz \to ux, vy, wz$.

To avoid the possibility that $wz \in A$, we avoid edges from A to $V \setminus A$ in our choices of xw, yz, a total of $ar - k$ edges. To ensure G is simple, when choosing xw, yz we require that w and z are not adjacent in G' and $x, y \notin N(u, v)$, ruling out a total of $O(r^2)$ choices for each edge. Thus

$$d_R = 4k(nr/2 - (ar - k) - O(r^2))^2 = kn^2 r^2 \left(1 + O\left(\frac{a}{n} + \frac{r}{n} \right) \right).$$

Hence

$$|\mathcal{C}_k| = |\mathcal{C}_{k-1}| \frac{1}{k} \binom{a}{2} \frac{r}{n} \left(1 + O\left(\frac{1}{a} + \frac{1}{r} + \frac{k}{ar} + \frac{k}{a^2} + \frac{a}{n} + \frac{r}{n} \right) \right).$$

$\square$

Now let $a = \frac{2n}{r}(\log r - \log \log r + 1 - \log 2 + \epsilon)$. Applying the above lemma we see that if $k = \left\lfloor \binom{a}{2} \frac{r}{n} \right\rfloor$ then

$$\frac{|\mathcal{C}_0|}{|\mathcal{C}_k|} = k! \left(\frac{n}{r} \binom{a}{2}^{-1} \right)^k \left(1 + O\left(\frac{\log r}{r} + \frac{r}{n} \right) \right)^k.$$

So the probability that G_r contains an independent set of size a is at most

$$
\binom{n}{a}\frac{|\mathcal{C}_0|}{|\mathcal{C}_k|} \;\leq\; \left(\frac{ne}{a}\right)^a \left(\frac{kn}{r(e+o(1))\binom{a}{2}}\right)^k \left(1 + O\left(\frac{\log r}{r} + \frac{r}{n}\right)\right)^k
$$

$$
= \left(\frac{ne}{a}\exp\left\{-\frac{(a-1)r}{2n}\left(1 + O\left(\frac{\log r}{r} + \frac{r}{n}\right)\right)\right\}\right)^a
$$

$$
\leq\; e^{-\epsilon a/2}
$$

$$
=\; o(1).
$$

This completes the proof of Theorem 1. $\qquad\square$

4 The chromatic number

The lower bound on the chromatic number in Theorem 2 follows from the upper bound on the independence number in Theorem 1. We use the same strategy of transforming H_m to G_r as in the previous section. It follows from Łuczak [11] that **whp**

$$
\chi(H_m) = \frac{r}{2\log r}\left(1 + O\left(\frac{\log\log r}{\log r}\right)\right).
$$

We start with a minimum proper colouring of H_m. Applying the analysis of the previous section to each individual colour class we see that **whp**, in going from H_m to G_r the number of edges which are improperly coloured is

$$
O\left(\frac{\rho^{1/2}\log n\,|I|^2}{\nu} \times \frac{r}{\log r}\right) = O\left(\frac{n(\log n)^2}{r^{\eta/20}}\right).
$$

(The main term is the number of edges added during GENERATE, see the analysis of Section 3.1.3. The terms from SIMPLIFY and from the case where SIMPLIFY fails are smaller.)

Lemma 4 *Fix* $C_1 > 0$ *constant. Then* **whp** *every* $A \subseteq V, a = |A| \leq a_0 = \frac{C_1 n(\log n)^2}{r^{\eta/20}}$, *contains at most* ℓa G_r-*edges, where* $\ell = r^{1-\eta/40}$.

Proof It follows from Lemma 3 that

$$
\mathbf{Pr}(A \text{ contains } \geq \ell a \text{ edges}) \leq |\mathcal{C}_0|^{-1}\sum_{k\geq \ell a}|\mathcal{C}_{\ell a}| \leq 2\frac{|\mathcal{C}_{\ell a}|}{|\mathcal{C}_0|} \leq \frac{1}{(\ell a)!}\binom{a}{2}^{\ell a}\left(\frac{r}{n}\right)^{\ell a}C_2^{\ell a}
$$

15

for some constant $C_2 > 0$. (Here we use $k \leq \min\left\{ar, \binom{a}{2}\right\}$ in the error term of (16).) Hence

$$
\begin{aligned}
\mathbf{Pr}(\exists A) &\leq \sum_{a=\ell}^{a_0} \binom{n}{a} \left(\frac{C_2 e}{\ell a} \binom{a}{2} \frac{r}{n}\right)^{\ell a} \\
&\leq \sum_{a=\ell}^{a_0} \left(\left(\frac{a}{n}\right)^{\ell-1} \left(\frac{C_3 r}{\ell}\right)^{\ell}\right)^{a} \qquad C_3 \leq C_2 e \\
&\leq \sum_{a=\ell}^{a_0} \left(\left(\frac{C_1 (\log n)^2}{r^{\eta/20}}\right)^{\ell-1} (C_3 r^{\eta/40})^{\ell}\right)^{a} \\
&= o(1).
\end{aligned}
$$

$\square$

It follows from this lemma that **whp** the vertices incident with the improperly coloured edges induce a subgraph H of G_r such that every subgraph of H has a vertex of degree at most $2r^{1-\eta/40}$. Consequently, H can be re-coloured using at most $2r^{1-\eta/40} + 1$ new colours, which is negligible and completes the proof of Theorem 2. $\square$

References

[1] E.A.Bender and E.R.Canfield, *The asymptotic number of labelled graphs with given degree sequences*, Journal of Combinatorial Theory, Series A 24 (1978) 296-307.

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

[3] C.D. Cooper, A.M. Frieze and B.A. Reed, *Random regular graphs of non-constant degree: connectivity and Hamiltonicity*, to appear.

[4] A.M.Frieze, *Random regular graphs of non-constant degree*, Technical Report, Department of Mathematical Sciences, Carnegie Mellon University (1988).

[5] A.M.Frieze, *On the independence number of a random graph*, Discrete Mathematics 81 (1990) 171-176.

[6] A.M.Frieze and T.Luczak, *On the independence and chromatic numbers of random regular graphs*, Journal of Combinatorial Theory Series B 54 (1992) 123-132.

[7] W.Höffding, *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association 58, (1963) 13-30.

[8] S.Janson, T.Łuczak and A.Ruciński, *Random Graphs*, John Wiley and Sons, 2000.

[9] M.Krivelevich, B.Sudakov, V.Vu and N.C.Wormald, *Random regular graphs of high degree*, to appear in Random Structures and Algorithms.

[10] L.Kučera, *Canonical labelling of regular graphs in linear expected time*, Proceedings of the 28th Annual IEEE Symposium on Foundations of Computer Science (1987) 271-279.

[11] T.Łuczak, *The chromatic number of random graphs*, Combinatorica 11 (1991) 45-54.

[12] B.D.Mckay, *Asymptotics for symmetric 0-1 matrices with prescribed row sums*, Ars Combinatoria 19A (1985) 15-25.

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

[14] B.D.Mckay and N.C.Wormald, *Asymptotic enumeration by degree sequence of graphs of high degree*, European Journal of Combinatorics 11 (1990) 565-580.

[15] N.C.Wormald, *Models of random regular graphs*, Surveys in Combinatorics, London Mathematical Society Lecture Note Series 267, Cambridge University Press, Cambridge, 1999 (J.D.Lamb and D.A.Preece, Eds.), Proceedings of the 1999 British Combinatorial Conference, Cambridge University Press, 239-298.

APPENDIX:
Proof of Theorem 3

We will give the proof for graphs, the proof for bipartite graphs is similar and slightly simpler in that we do not have to worry about loops or triangles.

Recall that F (called F^* in SIMPLIFY) is chosen randomly from $\Omega_{\mathbf{d}}$ where

$$\mathbf{d} = (d_1, d_2, \dots, d_\nu) \text{ and } \rho/2 \leq \min_i d_i \leq \max_i d_i \leq 2\rho \ll \nu$$

and $D = |\mathbf{d}| = \frac{1}{2} \sum_i d_i$.

Lemma 5
$$\mathbb{P}(F \text{ has } \geq \rho \log \nu \text{ loops }) \leq \nu^{-\frac{1}{2}\rho \log \log \nu}.$$

$$\mathbb{P}(F \text{ has } \geq \rho^2 \log \nu \text{ redundant edges }) \leq \nu^{-\frac{1}{2}\rho^2 \log \log \nu}.$$

Proof   Let $k_1 = \rho \log \nu$. Then

$$\mathbb{P}(F \text{ has } \geq k_1 \text{ loops}) \leq \sum_{x_1 + \dots + x_\nu = k_1} \prod_{i=1}^{\nu} \binom{d_i}{2}^{x_i} \left(\frac{1}{2D - 2k_1} \right)^{k_1}$$

$$\leq \binom{\nu + k_1 - 1}{k_1} \binom{2\rho}{2}^{k_1} \left(\frac{1}{2D - 2k_1} \right)^{k_1} \leq \left(\frac{3\nu}{k_1} \cdot \frac{2\rho^2}{\rho\nu/3} \right)^{k_1} \leq \left(\frac{18}{\log \nu} \right)^{k_1}.$$

Let $k_2 = \rho^2 \log \nu$. Then

$$\mathbb{P}(F \text{ has } \geq k_2 \text{ redundant edges})$$

$$\leq \sum_{\xi_2 + 2\xi_3 + \dots + k_2 \xi_{k_2+1} = k_2} \prod_{i=2}^{k_2+1} \left(\frac{1}{\xi_i!} \binom{\nu}{2}^{\xi_i} (2\rho)^{2i\xi_i} \left(\frac{1}{2D - 4k_2} \right)^{i\xi_i} \right)$$

$$\leq \sum \left(\frac{72e\rho^2}{\xi_2} \right)^{\xi_2} \prod_{i=3}^{k_2+1} (6\rho\nu)^{\xi_i} \left(\frac{12\rho}{\nu} \right)^{(i-1)\xi_i}$$

$$\leq \sum \left(\frac{72e\rho^2}{\xi_2} \right)^{\xi_2} \prod_{i=3}^{k_2+1} \left(\frac{12\sqrt{6}\rho^{3/2}}{\nu^{1/2}} \right)^{(i-1)\xi_i}$$

$$\leq 2^{k_2} \max_{\xi_2} \left(\frac{72e\rho^2}{\xi_2} \right)^{\xi_2} \left(\frac{12\sqrt{6}\rho^{3/2}}{\nu^{1/2}} \right)^{k_2 - \xi_2} = 2^{k_2} \left(\frac{72e\rho^2}{k_2} \right)^{k_2} = \left(\frac{144e}{\log \nu} \right)^{k_2}.$$

For the first inequality above, let the k_2 redundant edges arise from ξ_i edges of multiplicity i, $i = 2, \dots, k_2 + 1$. Now choose the corresponding end vertices in at most $\prod_{i=2}^{k_2+1} \frac{1}{\xi_i!} \binom{\nu}{2}^{\xi_i}$ ways and the configuration points in at most $\prod_{i=2}^{k_2+1} (2\rho)^{2i\xi_i}$ ways. Finally

18

multiply by $\prod_{i=2}^{k_2+1} \left(\frac{1}{2D-4k_2} \right)^{i\xi_i}$ to estimate the probability of this occurrence. For the subsequent inequalities note that $\sum_i (i-1)\xi_i = k_2$, and that in the sums displayed above there are at most 2^{k_2} terms. $\qquad\qquad\square$

This verifies that the chance of failure in Step 3 is less than ϵ_1.

Fix a, b for the remainder of the Appendix and let E be the set $\{(F^*, \sigma, F^{**}) : F^* \in \Omega_{a,b},\ F^{**} \in \Omega_{0,0}$ and $\sigma : F^* \to F^{**}$ is a switching allowed by SIMPLIFY$\}$. For $F_1 \in \Omega_{a,b}$, let $d_L(F_1) = |\{(F, \sigma, F') \in E : F = F_1\}|$ and for $F_2 \in \Omega_{0,0}$, let $d_R(F_2) = |\{(F, \sigma, F') \in E : F' = F_2\}|$. We note that where $\pi_{F,a,b}$ is as defined in Algorithm SIMPLIFY,

$$\frac{\pi_{F,a,b}}{|\Omega_{0,0}|} = |\Omega_{a,b}|^{-1} \sum_{\substack{F^* \in \Omega_{a,b} \\ (F^*, \sigma, F) \in E}} d_L(F^*)^{-1}. \qquad (17)$$

Lemma 6 *(i)*
$$(2D - 32(a+b)\rho^2)^{a+b} \le d_L(F) \le (2D)^{a+b}.$$

(ii)

$$(D_2 - 3C_3 - 128(a+b)\rho^6)^a \times (\Gamma_1 - 3C_3 - 2C_4 - 128(a+b)\rho^6)^b$$
$$\le a!\, b!\, d_R(F') \le D_2^a \Gamma_1^b$$

where C_i is the number of i-cycles in F',

$$D_2 = \sum_{i=1}^{\nu} \binom{d_i}{2} \quad and \quad \Gamma_1 = \Gamma_1(F') = \sum_{\substack{\{x,y\} \in F' \\ x < y}} (d_{\phi(x)} - \mathrm{rk}(x))(d_{\phi(y)} - 1),$$

and $\mathrm{rk}(x)$ is the rank of x in the set $W_{\phi(x)}$.

Proof   (i) The upper bound is obvious. For the lower bound, we observe that the conditions (13) together with $dist(e_i, f_i) \ge 2$ guarantee that $F^{**} \in \Omega_{0,0}$. $16(a+b)\rho^2$ is a crude upper bound on the number of edges excluded by these conditions.

(ii) We start with the upper bound. Consider a simple F' and the possibilities for σ, F such that $(F, \sigma, F') \in E$. In going from F to F' we simultaneously perform a loop-deleting switchings and b redundant-edge-deleting switchings. Although their vertex sets may overlap, the individual switchings do not interact in the sense that no element of W is involved in more than one. Given F', the triple (F, σ, F') is determined by specifying a switching τ consisting of a loop-creating switchings and b redundant-edge-creating switchings. (There are other conditions on τ relevant only for the lower bound.) When we delete a loop we create a path of length 2 through the corresponding vertex. So to create a loop $e = \{a_1, a_2\}$ we need a path of length 2 in $\gamma(F')$. So we take 2 pairs $\{a_j, b_j\}$, $j = 1, 2$ where $\phi(a_1) = \phi(a_2)$ and replace them by $\{a_1, a_2\}, \{b_1, b_2\}$. There are at most D_2 choices here for each loop and so at most $\binom{D_2}{a}$ choices for a loops.

19

For redundant edges note that the conditions on σ ensure that going from F to F' the edge e'_i with respect to which e_i $(i \le b)$ is redundant is also an edge of F'. So in the reverse direction we must choose switchings creating edges redundant with respect to edges of F'. To create a redundant edge $\{a_1, a_2\}$ we must take 2 pairs $\{a_j, b_j\}$, $j = 1, 2$ and replace them by $\{a_1, a_2\}, \{b_1, b_2\}$. Here F' must have another pair $\{x, y\}$, $x < y$ such that $\phi(a_1) = \phi(x), \phi(a_2) = \phi(y)$ and $a_1 > x$. The summand in the definition of Γ_1 bounds the number of choices for $\{a_i, b_i\}$, $i = 1, 2$ for a given $\{x, y\} \in F'$, so for each switching there are at most Γ_1 choices, and for b switchings at most $\binom{\Gamma_1}{b}$. This proves the upper bound.

For the lower bound we choose the individual switchings τ_i making up τ so that the loops to be added and the redundant edges to be created are all at distance at least 4 from each other **in** $\gamma(\mathbf{F'})$. Making the choices in some order this excludes (crudely) at most $128(a + b)\rho^6$ possibilities at each stage. This ensures that the vertex sets V_i of the individual switchings are disjoint, and also that for $i \ne j$ the configuration F', and hence F, contains no $V_i - V_j$ edge. Letting σ be the inverse of τ the conditions (13) are thus met. It only remains to ensure that the individual switchings τ_i create no extra loops or redundant edges. (The condition that F^{**} be simple in step 5 of SIMPLIFY is automatically met as we start from a simple F'.) First consider the loops. When we insert loop $\{a_1, a_2\}$ we need to avoid the case where the edge $\{b_1, b_2\}$ is parallel to some edge $\{x, y\} \in F'$. Here F' contains a triangle $(\{x, y\}, \{a_1, b_1\}, \{b_2, a_2\})$. There are $3C_3$ paths of length 2 in triangles. In the case of redundant edges when inserting the redundant edge $\{a_1, a_2\}$ we must make sure the other edge $\{b_1, b_2\}$ added is not parallel to some $\{u, v\} \in F'$. In this case F' contains a 4-cycle $(\{u, v\}, \{b_1, a_1\}, \{x, y\}, \{a_2, b_2\})$. There is also the case where $\{b_1, b_2\}$ is a loop. Here F' contains a 3-cycle $(\{x, y\}, \{a_1, b_1\}, \{b_2, a_2\})$. $\qquad \square$

Now let Pr, Pr_0 denote the uniform measure on $\Omega_{\mathbf{d}}, \Omega_{0,0}$ respectively. Let $\mathbf{E}$ denote expectation with respect to Pr.

Lemma 7

(i) *For $k = 3, 4$,*

$$Pr_0(C_k \ge (2\rho)^k + t) \le \exp\left\{-\frac{t^2}{2^9 \nu \rho^7} + O(\rho^2)\right\}.$$

(ii)

$$\mathbf{E}(\Gamma_1) \ge \frac{1}{40}\rho^3 \nu.$$

(iii)

$$Pr_0(|\Gamma_1 - \mathbf{E}(\Gamma_1)| \ge t) \le 2\exp\left\{-\frac{t^2}{128\nu\rho^5} + O(\rho^2)\right\}.$$

Proof   (i) We use a martingale argument on configurations F in $\Omega_{\mathbf{d}}$. We imagine that we produce F using CONSTRUCT. At stage t we choose u_t as the minimum of R_{t-1}. Consider fixing the first i pairs and denote them by $Y_1, Y_2, \ldots, Y_i$ and let w_a be the minimum of R_i. We compare $\mathbf{E}(Z \mid Y_1, Y_2, \ldots, Y_i, \{w_a, w_b\})$ and $\mathbf{E}(Z \mid Y_1, Y_2, \ldots, Y_i, \{w_a, w_x\})$ for arbitrary w_b, w_x. We use the following mapping between the conditional spaces: if w_x is paired with w_y in the first then in the second we pair w_b and w_y. Thus if θ_Z bounds the change in Z when pairs $\{\alpha, \beta\}, \{\gamma, \delta\}$ are replaced by $\{\alpha, \gamma\}, \{\beta, \delta\}$ we get, after applying Azuma-Hoeffding,

$$Pr(|Z - \mathbf{E}(Z)| \geq t) \leq 2\exp\{-\tfrac{t^2}{2D\theta_Z^2}\}.$$

We inflate the RHS by $e^{O(\rho^2)}$ in order to replace Pr by Pr_0, using a near-regular version of Lemma 2.

For $Z = C_k$ we have $\mathbf{E}(Z) \leq (2\rho)^k$ and $\theta_Z \leq 2(2\rho)^{k-1}$.

(ii) A random member $\{x, y\}$, $x < y$ of a random $F \in \Omega_{\mathbf{d}}$ is a random unordered pair of elements of W. Thus $\mathbf{E}(d_{\phi(x)} - \mathrm{rk}(x)) \geq \tfrac{1}{2}\mathbf{E}(d_{\phi(x)} - 1) \geq \tfrac{1}{2}(\tfrac{1}{2}\rho - 1)$. Hence

$$\mathbf{E}(\Gamma_1) \geq (\tfrac{1}{2}\rho - 1)\mathbf{E}\Big(\sum_{\substack{\{x,y\} \in F \\ x < y}} (d_{\phi(x)} - \mathrm{rk}(x)))\Big) \geq (\tfrac{1}{2}\rho - 1) \cdot \tfrac{1}{4}\rho\nu \cdot \tfrac{1}{2}(\tfrac{1}{2}\rho - 1).$$

For (iii) we simply observe that $\theta_{\Gamma_1} \leq 2(2\rho)^2$.   □

Now let

$$\Omega_{0,0}^{bad} = \{F \in \Omega_{0,0} : \ (a) \ C_3 + C_4 > \rho^5 \nu^{1/2} \log \nu \ \text{ or } (b) \ |\Gamma_1 - \mathbf{E}(\Gamma_1)| > \rho^4 \nu^{1/2} \log \nu\}$$

and

$$\Omega_{0,0}^{good} = \Omega_{0,0} \setminus \Omega_{0,0}^{bad}.$$

It follows from Lemma 7 that

$$\frac{|\Omega_{0,0}^{bad}|}{|\Omega_{0,0}|} \leq \nu^{-\rho^2 \log \nu}. \tag{18}$$

We are now ready to analyse the behaviour of SIMPLIFY.

Lemma 8 *For all $a \leq \rho \log \nu$, $b \leq \rho^2 \log \nu$,*

(i) $\mathbb{P}(F^{**} \in \Omega_{0,0}^{good} \mid a, b) = 1 - O(\nu^{-\frac{1}{2}\rho^2 \log \nu})$

(ii) *for all $F \in \Omega_{0,0}^{good}$ we have $\pi_{F,a,b} \in [1 - \epsilon_1, 1 + \epsilon_1]$.*

21

Proof

It follows from (17) and Lemmas 6, 7 that if $F \in \Omega_{0,0}^{good}$ and $(a,b) \in [\rho \log \nu] \times [\rho^2 \log \nu]$ then

$$\frac{\pi_{F,a,b}}{|\Omega_{0,0}|} = (1 + \theta_F)|\Omega_{a,b}|^{-1}(2D)^{-a-b}D_2^a \mathbf{E}(\Gamma_1)^b/(a!b!) \tag{19}$$

where $|\theta_F| \leq \epsilon_1/3$ and so

$$\mathbb{P}(F^{**} \in \Omega_{0,0}^{good} \mid a, b) = (1 + \theta')\frac{|\Omega_{0,0}^{good}|}{|\Omega_{a,b}|}(2D)^{-a-b}D_2^a \mathbf{E}(\Gamma_1)^b/(a!b!)$$

where $|\theta'| \leq \epsilon_1/3$ too.

Furthermore, since $\Gamma_1 \leq 4\rho^3 \nu$,

$$\mathbb{P}(F^{**} \in \Omega_{0,0}^{bad} \mid a, b) \leq 2\frac{|\Omega_{0,0}^{bad}|}{|\Omega_{a,b}|}(2D)^{-a-b}D_2^a (4\rho^3 \nu)^b/(a!b!).$$

Therefore, since $\mathbf{E}(\Gamma_1) \geq \frac{1}{40}\rho^3 \nu$,

$$\frac{\mathbb{P}(F^{**} \in \Omega_{0,0}^{bad} \mid a, b)}{\mathbb{P}(F^{**} \in \Omega_{0,0}^{good} \mid a, b)} \leq 3 \times 160^b \frac{|\Omega_{0,0}^{bad}|}{|\Omega_{0,0}^{good}|} = O(160^b \nu^{-\rho^2 \log \nu}) = O(\nu^{-\frac{1}{2}\rho^2 \log \nu}).$$

This verifies (i). To obtain (ii) we choose $\tilde{F} \in \Omega_{0,0}^{good}$ and write

$$1 - O(\nu^{-\frac{1}{2}\rho^2 \log \nu}) = \sum_{F \in \Omega_{0,0}^{good}} \frac{\pi_{F,a,b}}{|\Omega_{0,0}|} \geq \frac{1 - \epsilon_1/3}{1 + \epsilon_1/3}\frac{|\Omega_{0,0}^{good}|}{|\Omega_{0,0}|}\pi_{\tilde{F},a,b},$$

using (19). Thus

$$\pi_{\tilde{F},a,b} \leq \frac{1 + \epsilon_1/3}{1 - \epsilon_1/3}(1 + O(\nu^{-\frac{1}{2}\rho^2 \log \nu})) \leq 1 + \epsilon_1.$$

A lower bound follows in the same way. $\qquad \square$

We now turn to $\mathcal{G}_{\mathbf{d}}^{good}$. Let $\mathcal{G}_{\mathbf{d}}^{bad} = \mathcal{G}_{\mathbf{d}} \setminus \mathcal{G}_{\mathbf{d}}^{good}$. From the definition (12) of $\mathcal{G}_{\mathbf{d}}^{good}$ we see that

$$\frac{|\Omega_{0,0}^{bad}|}{|\Omega_{0,0}|} \geq \epsilon_1 \frac{|\mathcal{G}_{\mathbf{d}}^{bad}|}{|\mathcal{G}_{\mathbf{d}}|}.$$

Thus from (18),

$$\frac{|\mathcal{G}_{\mathbf{d}}^{bad}|}{|\mathcal{G}_{\mathbf{d}}|} \leq \epsilon_1^{-1}\nu^{-\rho^2 \log \nu} \leq \nu^{-\frac{1}{2}\rho^2 \log \nu}. \tag{20}$$

Lemma 9 *For all $a \le \rho \log \nu$, $b \le \rho^2 \log \nu$,*

(i) $\mathbb{P}(G^{**} \in \mathcal{G}_{\mathbf{d}}^{good} \mid a, b) = 1 - O(\nu^{-\frac{1}{2}\rho^2 \log \nu})$

(ii) *for all $G \in \mathcal{G}_{\mathbf{d}}^{good}$ we have $\tilde{\pi}_{G,a,b} \in [1 - 2\epsilon_1, 1 + \epsilon_1]$.*

Proof   (i) From Lemma 8(i), given a, b the probability of not reaching step 7 of SIMPLIFY is small enough. From Lemma 8(ii) $\mathbb{P}(G^{**} \in \mathcal{G}_{\mathbf{d}}^{bad}) \le (1 + \epsilon_1)|\mathcal{G}_{\mathbf{d}}^{bad}|/|\mathcal{G}_{\mathbf{d}}|$.

(ii) is immediate from Lemma 8(ii) and (12). $\square$

Theorem 3 follows by combining the results from above:

(a) From Lemma 5, and Lemma 9(i) and (ii).

(b) From (20).

(c) From Lemma 5 and Lemma 9(i).

(d) Each $G \in \mathcal{G}_{\mathbf{d}}^{good}$ has the same probability $\frac{1-2\epsilon_1}{|\mathcal{G}_{\mathbf{d}}^{good}|}$ of being output.

                                                                              $\square$