ON RANDOM REGULAR GRAPHS WITH NON-CONSTANT DEGREE

by

A. M. Frieze

Department of Mathematics
Carnegie Mellon University
Pittsburgh, PA 15213

Research Report No. 88-2

February 1988

On random regular graphs with non-constant degree

by

A.M. Frieze
Department of Mathematics
Carnegie Mellon University
Pittsburgh, PA 15213
USA

February 1988

Abstract

We study random r-regular labelled graphs with n vertices. For $r = O(n^{1/5-\epsilon})$ we give the asymptotic number of them and show that they are almost all r-connected and hamiltonian. Proofs are based on the analysis of a simple algorithm for finding "almost" random regular graphs.

§1. Introduction

This paper is mainly concerned with random r-regular labelled graphs with n vertices and $r = r(n) \to \infty$ with n. Bender and Canfield [1] and Bollobás [2] discuss models for the analysis of r regular graphs when r is constant. Bollobás's model is more transparent and in fact deals with $r(n) \leq (2 \log n)^{1/2}$. By and large however what is known about random regular graphs is restricted to constant r.

We will describe a method for analysing random regular graphs which is valid for $r(n) = O(n^{1/5-\epsilon})$ where for simplicity we take $\epsilon > 0$ to be an arbitrarily small constant.

Let REG (n,r) denote the set of r-regular graphs with vertex set $V_n = \{1,2,\ldots,n\}$. We first extend the range of applicability of the asymptotic formula for the number of labelled r-regular graphs given in [1] and [2].

As customary, we use the relation $A \sim B$ to denote the relation $A/B \to 1$ as $n \to \infty$.

Theorem 1

Assume $r = O(n^{1/5-\epsilon})$ for some fixed $\epsilon > 0$. Then

$$|REG(n,r)| \sim e^{-\lambda-\lambda^2} \frac{(rn)!}{(\frac{1}{2} rn)!\ 2^{\frac{1}{2} rn} (r!)^n},$$

where $\lambda = \frac{r-1}{2}$. □

We will deduce Theorem 1 as a special case of Theorem 2 below, which deals with general degree sequences (as in [1], [2]).

Suppose $\underline{d} = (d_1, d_2, \ldots, d_n)$ is a sequence of positive integers where $d_1 \geq d_2 \geq \ldots \geq d_n \geq 2$ and $\sum_{i=1}^{n} d_i = 2m$ is even. Let $\mathcal{G}(\underline{d})$ denote the set of graphs with vertex set V_n and vertex degrees $d_1, d_2, \ldots, d_n$.

Theorem 2

Let $\underline{d}$ be as above and assume $d_1^3/d_n^{1/2} = O(n^{1/2-\epsilon})$. Then

$$|\mathcal{G}(\underline{d})| \sim e^{-\lambda-\lambda^2} \frac{(2m)!}{m!2^m \prod_{i=1}^{n} d_i!}$$

where $\lambda = \frac{1}{2m} \sum_{i=1}^{n} \binom{d_i}{2}$. □

The method we use is based on the analysis of a constructive method (CHOOSE below). It seems likely that Theorems 1 and 2 could be proved more directly. However the advantage to our approach is that it will allow us to analyse random regular graphs of this size, whereas a direct approach might not. (Of course, a "direct" proof has not as yet been constructed.)

Let G_r denote a random graph sampled uniformly from REG (n,r). We will prove two properties of G_r extending what is known for constant r.

Theorem 3

Assume $r = O(n^{1/5-\epsilon})$ for some fixed $\epsilon > 0$. Then

(a) $\lim_{n\to\infty} \Pr(G_r$ is r-connected$) = 1$

(b) $\lim_{n\to\infty} \Pr(G_r$ is hamiltonian$) = 1$.

□

Neither result is surprising, they just seem difficult to establish. In particular, consider (b). It has already been established for various values of r_0 (10^{10} in Bollobás [3], 796 in Fenner and Frieze [4], 85 in Frieze [5]) that if r is constant and $r \geq r_0$ then almost every G_r is hamiltonian. Unfortunately one cannot obtain G_{r+1} by adding random edges to G_r and so none of these results imply (b) of the theorem, which seems obviously true.

Recently, Robinson and Wormald claim to have extended their approach in [7] to prove that $r_0 = 3$. At the time of writing only an abstract of the proposed paper is available.

One possibly useful outcome of this work is an efficient procedure for generating "nearly" random regular graphs.

(If say $r = 10$ and one wanted to use Bollobas's construction repeatedly until one found a simple graph one would expect to have to make approximately e^{25} trials).

The basis of our proofs is a simple algorithm CHOOSE which outputs a regular graph G_{CHOOSE} (in "configuration" form–see §2) with probability $1 - o(1)$ and satisfies the following:

let $\mathcal{A}$ be any graph property. Then under the assumptions of Theorem 1

$$\Pr(G_r \in \mathcal{A}) = \Pr(G_{CHOOSE} \in \mathcal{A}) + o(1).$$

We point out now that because we are dealing with asymptotic results all our inequalities are only required to hold for large enough n.

§2. Configurations

We are going to work with the Configuration Model of Bollobás [2]. We let $W = \{1,2,\ldots,2m\}$ and $W_i = \{\sum_{j=1}^{i-1} d_j + 1,\ldots,\sum_{j=1}^{i} d_j\}$ so that $W_1,W_2,\ldots,W_n$ is a partition of W into sets of size $d_1,d_2,\ldots,d_n$. For $k \in W$ we define $\psi(k) \in V_n$ so that $k \in W_{\psi(k)}$. We will let $\Delta = d_1$.

A configuration F is a partition of W into m pairs. Φ denotes the set of configurations. For $F = (\{x_i,y_i\},\ i = 1,2,\ldots,m) \in \Phi$ we let $\mu(F)$ be the multigraph with vertex set V_n and the m edges $\{(x_iy_i,\psi(x_i),\psi(y_i)) : i = 1,2,\ldots,m\}$. The notation $(xy,\psi(x),\psi(y))$ refers to an edge with label xy, $x < y$ and endpoints $\psi(x),\psi(y)$.

Notice that $F \to \mu(F)$ defines a bijection, because we can construct F from the edge labels of $\mu(F)$.

We consider Φ to be a probability space in which each $F \in \Phi$ is equally likely.

The multigraph $\mu(F)$ will in general contain parallel edges and loops. Let $\{x,y\} \in F$ be a <u>loop</u> if $\psi(x) = \psi(y)$ and let $\beta(F)$ denote the number of loops in F. $\{x,y\} \in F$, $x < y$ will be a <u>parallel pair</u> if there exists $\{x',y'\}$, $x' < y'$ where $\psi(x) = \psi(x') < \psi(y) = \psi(y')$. The parallel pair $\{x,y\}$ just considered is <u>redundant</u> if the $\{x',y'\} \in F$ just considered satisfies $x' < x$. Let $\alpha(F)$ denote the number of redundant pairs of F.

Thus the simple graph $\gamma(F)$ obtained from $\mu(F)$ by deleting loops and coalescing parallel edges has $m - (\alpha(F) + \beta(F))$ edges.

We let $\Phi_{a,b} = \{F \in \Phi : \alpha(F) = a,\ \beta(F) = b\}$ for $a,b \geq 0$ and let $F_{a,b}$ be a random configuration sampled uniformly from $\Phi_{a,b}$. We are of course most interested in $\Phi_{0,0}$, the case where $\mu(F) = \gamma(F)$. We know that each graph in $\mathcal{G}(\underline{d})$ is the image under μ of exactly $\prod_{i=1}^{n} d_i!$ configurations. Our task generally then is to estimate $|\Phi_{0,0}|$ and prove properties of almost every $F \in \Phi_{0,0}$.

§3. Sampling from $\Phi_{0,0}$.

Our problem is to describe a simple process of generating configurations which, with sufficiently high probability, produces a ("nearly") random member of $F_{0,0}$. (When Λ is constant one simply chooses a random $F \in \Phi$.)

We will now describe an algorithm which generates an $F \in \Phi_{0,0}$ with the required distribution, given the conditions of the theorem. To do this we define first for $a > 0$, $b \geq 0$ a bipartite graph $H_{a,b}$ with vertex partition $\Phi_{a,b}, \Phi_{a-1,b}$ and an edge FG, $F \in \Phi_{a,b}$, $G \in \Phi_{a-1,b}$ whenever the following condition holds:

(A) $\exists$ non-loops $e = \{u,v\}$, $f = \{x,y\} \in F$ such that

(i) e is redundant and f is not parallel in F

(ii) $G = (F \cup \{\{u,x\}, \{v,y\}\}) - \{e,f\}$

(see figure 1).

Condition A describes the set of neighbours of $F \in \Phi_{a,b}$. We should also describe the set of neighbours of $G \in \Phi_{a-1,b}$. Thus equivalently there is an edge FG if

(B) $\exists$ non-loops $e_1 = \{u,x\}$, $e' = \{x',y'\}$, $x' < y'$, $e_2 = \{v,y\} \in G$ such that

(i) $\psi(u) = \psi(x')$, $\psi(v) = \psi(y')$, $x' < u$,

(ii) $F = (G \cup \{\{u,v\}, \{x,y\}\}) - \{e_1,e_2\}$.

We must also define a similar bipartite graph $H'_{a,b}$, $a \geq 0$, $b > 0$ with bipartition $\Phi_{a,b}, \Phi_{a,b-1}$ and an edge FG, $F \in \Phi_{a,b}$, $G \in \Phi_{a,b-1}$ whenever

(A′) $\exists$ loop $e = \{u,v\}$, non-loop $f = \{x,y\} \in F$ such that

(i) $G = F \cup \{\{u,x\}, \{v,y\}\} - \{e,f\}$.

(see figure 2).

The equivalent conditions in terms of G are

(B′) $\exists$ non-loops $e_1 = \{u,x\}$, $e_2 = \{v,y\}$ such that

(i) $\psi(u) = \psi(v)$

(ii) $F = (G \cup \{\{u,v\}, \{x,y\}\}) - \{e_1,e_2\}$.

If $H \in \Phi_{a,b} \cup \Phi_{a-1,b}$ (resp. $H \in \Phi_{a,b} \cup \Phi_{a,b-1}$) then we let $N_{a,b}(H)$ (resp. $N'_{a,b}(H)$) denote its neighbour set in $H_{a,b}$ (resp. $H'_{a,b}$).

Consider now the following method of choosing $F \in \Phi_{0,0}$.

For a finite set S, RANSELECT $x \in S$ means select x randomly from S.

<u>Algorithm CHOOSE</u>

0. RANSELECT $F \in \Phi$;
   let $a = \alpha(F)$, $b = \beta(F)$
1. $F_0 := F$;
2. <u>for</u> $i = 1$ <u>to</u> a <u>do</u>
3. RANSELECT $F_i \in N_{a-i+1,b}(F_{i-1})$;
4. <u>for</u> $i = a+1$ <u>to</u> $a+b$ <u>do</u>
5. RANSELECT $F_i \in N'_{0,a+b-i+1}(F_{i-1})$;
6. <u>Output</u> $F_{CHOOSE} = F_{a+b}$.

Note that $F_{a+b} \in \Phi_{0,0}$. We do not intend to prove that F_{a+b} is equally likely to be any member of $F_{0,0}$. Indeed it is not. However, within the limits set by Theorem 1, it is "nearly so".

Theorem 4

Let $\mathcal{A} \subseteq \Phi$ be some property of configurations. Then with the same assumptions as Theorem 2

$$\Pr(F_{0,0} \in \mathcal{A}) = \Pr(F_{CHOOSE} \in \mathcal{A}) + o(1). \qquad \square$$

We devote the next section, the main body of the paper, to the proof of Theorem 4.

§4. Proof of Theorem 4.

Observe that in line 1 of CHOOSE F_0 is a random member of $\Phi_{a,b}$ but there is bias in the selection of $F_1 \in \Phi_{a-1,b}$. We modify CHOOSE by adding a 'rarely used' emergency procedure for reselecting F_i, so that F_i remains a random member of $\Phi_{a-i+1,b}$ throughout.

To motivate CHOOSEA below assume that F is chosen randomly from $\Phi_{a,b}$ and G is chosen randomly from $N_{a,b}(F)$ as in CHOOSE. If $\pi_{a,b}(G)$ is the probability that a particular G is chosen then a moments reflection will convince the reader that

$$\pi_{a,b}(G) = |\Phi_{a,b}|^{-1} \sum_{F \in N_{a,b}(G)} |N_{a,b}(F)|^{-1}.$$

Let $\Phi^+_{a-1,b} = \{G \in \Phi_{a-1,b}: \pi_{a,b}(G) > |\Phi_{a-1,b}|^{-1}\}$ be the set of graphs which are being chosen too often and $\Phi^-_{a-1,b} = \Phi_{a-1,b} - \Phi^+_{a-1,b}$. The modification that we make to CHOOSE in CHOOSEA is that, if in line 3, we find $F_i \in \Phi^+_{a-i,b}$, then we reject it, with a small probability, and re-choose from $\Phi^-_{a-i+1,b}$ in order to maintain uniformity of selection.

Algorithm CHOOSEA (modifications to CHOOSE)

Add lines 3a, 3b and lines 5a, 5b to CHOOSE

3a. if $F_i \in \Phi^+_{a-i,b}$ then with probability $\dfrac{\pi_{a-i+1,b}(F_i) - |\Phi_{a-i,b}|^{-1}}{\pi_{a-i+1,b}(F_i)}$ do

3b. $F_i := G \in \Phi^-_{a-i,b}$ with probability $\dfrac{|\Phi_{a-i,b}|^{-1} - \pi_{a-i+1,b}(G)}{\sum_{G \in \Phi^-_{a-i,b}} (|\Phi_{a-i,b}|^{-1} - \pi_{a-i+1,b}(G'))}$

lines 5a, 5b are similar, $\Phi_{0,b}$ replaces $\Phi_{a-i+1,b}$, $\Phi_{0,b-1}$ replaces $\Phi_{a-i,b}$ and $\pi_{0,b}$ replaces $\pi_{a-i+1,b}$.

Also we refer to the output of CHOOSEA as $F_{CHOOSEA}$.

Lemma 5

In Algorithm CHOOSEA, F_i is equally likely to be any member of $\Phi_{a-i,b}$, $i \leq a$ or $\Phi_{0,a+b-i}$, $i > a$.

Proof

Assume inductively that F_{i-1} is equally likely to be any member of $\Phi_{a-i+1,b}$ (we assume $i \leq a$, but only for notational purposes. The case $i > a$ is handled similarly). If $G \in \Phi^+_{a-i,b}$ then

$$\Pr(F_i = G) = \pi_{a-i+1,b}(G)\left(1 - \frac{\pi_{a-i+1,b}(G) - |\Phi_{a-i,b}|^{-1}}{\pi_{a-i+1,b}(G)}\right)$$

$$= |\Phi_{a-i,b}|^{-1}.$$

If $G \in \Phi^-_{a-i,b}$ then (where $\pi(G)$ denotes $\pi_{a-i+1,b}(G)$),

$$\Pr(F_i{=}G) = \pi(G) + \sum_{G'\in\Phi^+_{a-i,b}} (\pi(G') - |\Phi_{a-i,b}|^{-1}) \frac{|\Phi_{a-i,b}|^{-1} - \pi(G)}{\sum_{G''\in\Phi^-_{a-i,b}} (|\Phi_{a-i,b}|^{-1} - \pi(G''))}$$

$$= |\Phi_{a-i,b}|^{-1}$$

since

$$\sum_{G'\in\Phi^+_{a-i,b}} (\pi(G') - |\Phi_{a-i,b}|^{-1}) = \sum_{G''\in\Phi^-_{a-i,b}} (|\Phi_{a-i,b}|^{-1} - \pi(G'')).$$

□

Thus F_{CHOOSEA} is a random choice from $\Phi_{0,0}$.

Let now $\mathcal{E}^b$ denote the event {CHOOSEA executes line 3b or 5b}.

<u>Lemma 6</u>

Let $\mathcal{A} \subseteq \Phi$ be some property of configurations. Then

$$|\Pr(F_{0,0} \in \mathcal{A}) - \Pr(F_{\mathrm{CHOOSE}} \in \mathcal{A})| \le 2\Pr(\mathcal{E}^b).$$

Proof

We consider the following experiment: we run CHOOSE and CHOOSEA in parallel with the same random choices until the first execution of lines 3b or 5b. If $\mathcal{E}^b$ does not occur then they have the same output. Thus

$$\Pr(F_{0,0} \in \mathcal{A}) = \Pr(F_{CHOOSEA} \in \mathcal{A})$$

$$\leq \Pr(F_{CHOOSEA} \in \mathcal{A} \text{ and } F_{CHOOSE} = F_{CHOOSEA}) + \Pr(\mathcal{E}^b)$$

$$= \Pr(F_{CHOOSE} \in \mathcal{A} \text{ and } F_{CHOOSE} = F_{CHOOSEA}) + \Pr(\mathcal{E}^b)$$

$$\leq \Pr(F_{CHOOSE} \in \mathcal{A}) + \Pr(\mathcal{E}^b)$$

By using the same inequality for $\bar{\mathcal{A}}$ we obtain

$$\Pr(F_{0,0} \in \mathcal{A}) \geq \Pr(F_{CHOOSE} \in \mathcal{A}) - \Pr(\mathcal{E}^b)$$

and the lemma follows. □

Thus to prove Theorem 3, it is sufficient to show $\Pr(\mathcal{E}^b) = o(1)$.

In order to prove $\Pr(\mathcal{E}^b)$ is small we must examine the degree structure of the graphs $H_{a,b}$ and $H'_{a,b}$. For $F \in \Phi_{a,b} \cup \Phi_{a-1,b}$ let $d_{a,b}(F)$ denote its degree in $H_{a,b}$. Similarly, for $F \in \Phi_{a,b} \cup \Phi_{a,b-1}$ let $d'_{a,b}(F)$ denote its degree in $H'_{a,b}$.

For $k \geq 1$ let $\#C_k = \#C_k(F)$ be the number of cycles of length k in $\mu(F)$. ($k = 1$ counts loops, $k = 2$ counts pairs of parallel edges.)

Lemma 7

(a) $F \in \Phi_{a,b}$, $a > 0$, implies

$$2am - 2a(2a + b + 2\Delta(\Delta-2)) \leq d_{a,b}(F) \leq 2am$$

(b) $G \in \Phi_{a-1,b}$, $a > 0$, implies

$$\Sigma_G - (b+2a)(\Delta-1)^2 - 3\#C_3(G) - 4\#C_4(G) \leq d_{a,b}(G) \leq \Sigma_G$$

where

$$\Sigma_G = \sum_{\substack{e=\{x,y\}\in G \\ x<y}} (d_{\psi(x)} - t(x))(d_{\psi(y)} - 1)$$

and if $e = \{x,y\}$, $x < y$ then $t(x) = x - \sum_{i=1}^{\psi(x)-1} d_i$ (= the rank of x in $W_{\psi(x)}$).

(c) $F \in \Phi_{a,b}$, $b > 0$, implies

$$2b(m - b - (\Delta-2) - 2a) \leq d'_{a,b}(F) \leq 2b(m - b).$$

(d) $G \in \Phi_{a,b-1}$, $b > 0$, implies

$$\sum_{i=1}^{n} \binom{d_i}{2} - (b-1) - 3\#C_3 - 2a(\Delta-1) \leq d'_{a,b}(G) \leq \sum_{i=1}^{n} \binom{d_i}{2} - (b-1).$$

Proof

(a) Consider condition A. There are at most am pairs e,f satisfying (i) and each pair yields 2 edges of $H_{a,b}$. The upper bound follows. Now fix a redundant e. For the lower bound we must estimate the number of choices f so that G as defined in (ii) does not belong to $\Phi_{a-1,b}$. In the upper bound we ignored the fact that f was not be be a loop or parallel edge, which accounts for the $2a+b$ term. We need $|\psi(\{u,v,x,y\})| = 4$ to avoid creating loops, which removes at most $2(\Delta-2)$ choices for f. We must avoid creating new parallel edges which can be done by avoiding the existence of $\{p,q\} \in F$ such that $p \in \psi(e)$, $q \in \psi(f)$. This removes at most $2(\Delta-2)(\Delta-1)$ choices and gives us the lower bound.

(b) Consider condition B. For each non-loop $e' = \{x',y'\} \in G$, $x' < y'$ there are $d_{\psi(x')} - t(e')$ choices for u and $d_{\psi(y')}-1$ choices for v such that (i) holds and this justifies the upper bound. To get a lower bound we eliminate choices for which (α) $\exists$ $\{p,q\} \in G$ with $\psi(p) = \psi(x)$, $\psi(q) = \psi(y)$, creating extra parallel edges in F; there are at most $4\#C_4$ cases or, (β) $|\psi(\{u,v,x,y\})| \leq 3$, creating extra loops in F; there are at most $3\#C_3$ cases. Finally we must avoid destroying any loops or parallel edges of G; this accounts for at most $(b+2a)(\Delta-1)^2$ cases.

(c) Consider condition A'. There are at most $b(m - b)$ pairs e,f each giving 2 edges of $H'_{a,b}$, and this gives the upper bound. For a fixed loop e at most $\Delta-2$ choices of f create extra loops or parallel edges through $|\psi(\{u,x,y\})| = 2$ or the existence of $\{p,q\} \in F$ such that $\psi(p) = \psi(u)$, $\psi(q) = \psi(f)$. There are at most 2a choices of f which destroy a parallel edge. This gives the lower bound.

(d) Consider condition B'. There are at most $\sum_{i=1}^{n} \binom{d_i}{2}$ ways of choosing e_1,e_2 so that $\psi(u) = \psi(v)$. b-1 of these choices u,v correspond to loops of G and this gives the upper bound. There are at most $3\#C_3$ cases in which the construction yields a new parallel edge and at most $2a(\Delta-1)$ distinct cases in which a parallel edge is destroyed. □

Our next task is to give probability bounds for the quantities defined in the previous lemma.

Let $\sigma_k = \sum_{i=1}^{n} d_i^k$ for k integer.

Lemma 8

Let F be chosen randomly from Φ. Then

(a) $\Pr(\#C_k \geq \rho(\sigma_2/m)^k) \leq \rho^{-\rho(\sigma_2/m)^k/10}$ for $\rho_0 \leq \rho \leq m^{k+1}/k\sigma_2^k$, $1 \leq k \leq 4$.

where ρ_0 is a constant.

(b) $E(\Sigma_F) = \frac{1}{2m-1}\left(\left(\sum_{i=1}^{n}\binom{d_i}{2}\right)^2 - \sum_{i=1}^{n}\frac{d_i(d_i-1)^2}{4}\right)$

(c) $\Pr(|\Sigma_F - E(\Sigma_F)| \geq t) \leq 2e^{-t^2/16\sigma_5}$ for all $t \geq 0$.

Proof

(a) Let μ, t be positive integers.

$$\Pr(\#C_k \geq \mu) \leq E((\#C_k)_t/(\mu)_t)$$

$$\leq \Big(\sum_{\substack{S \subseteq V_n \\ |S|=k}} \prod_{j \in S} \binom{d_j}{2} \frac{(k-1)!\, 2^{k-1}}{(2m-2kt)^k}\Big)^t/(\mu)_t$$

$$\leq \frac{((k-1)!)^t}{2^{(k+1)t}(m-kt)^{kt}(\mu)_t}\Big(\sum_{\substack{S \subseteq V_n \\ |S|=k}} \prod_{j \in S} d_j^2\Big)^t$$

$$\leq \frac{((k-1)!)^t \sigma_2^{kt}}{2^{(k+1)t}(m-kt)^{kt}(\mu)_t}$$

$$\leq \left(\frac{(k-1)!\ \sigma_2^k}{2^{k+1} m^k \mu}\right)^t \left(1 - \frac{kt}{m}\right)^{-kt} \frac{\mu^t}{(\mu)_t} .$$

Let now $\mu = \lceil \rho(\sigma_2/m)^k \rceil$ and $t = \lfloor \mu/9 \rfloor$. Then

$$\left(1 - \frac{kt}{m}\right) \geq \frac{8}{9}$$

and

$$\frac{\mu^t}{\mu_{(t)}} \leq \left(\frac{\mu}{\mu - t}\right)^t \leq (8/7)^t \text{ , for sufficiently large } \mu$$

and (a) follows.

(b) Suppose we construct F by choosing a random pair of elements e_1 from W, then a random pair of elements e_2 from $W - \{e_1\}$ and so on. Viewed in this light

$$E(\Sigma_F) = m\ E(\zeta(x,y))$$

where x is chosen uniformly from $W-\{rn\}$ and y is chosen uniformly from $W - \{1,2,\ldots,x\}$ and $\zeta(x,y) = (d_{\psi(x)} - t(x))(d_{\psi(y)} - 1)$.

Now

$$E(\zeta(x,y)) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E(\zeta(x,y)\,|\,\psi(x) = i,\ \psi(y) = j)\ \frac{d_i d_j}{m(2m-1)}$$

$$+ \sum_{i=1}^{n} E(\zeta(x,y)\,|\,\psi(x) = \psi(y) = i)\ \frac{d_i(d_i-1)}{2m(2m-1)}$$

$$= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{d_i-1}{2}(d_j-1)\,\frac{d_i d_j}{m(2m-1)} + \sum_{i=1}^{n} \frac{(d_i-1)^2}{2}\,\frac{d_i(d_i-1)}{2m(2m-1)}$$

$$= \frac{1}{m(2m-1)}\left(2\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \binom{d_i}{2}\binom{d_j}{2} + \sum_{i=1}^{n} \binom{d_i}{2}^2 - \sum_{i=1}^{n} \frac{d_i(d_i-1)^2}{4}\right)$$

and the result follows.

We prove (c) by using a martingale inequality. Let $X_1, X_2, \ldots, X_m$ be random variables, and for each $i = 1, \ldots, m$ let $\underline{X}^{(i)}$ denote $(X_1, X_2, \ldots, X_i)$. Suppose that the random variable M is determined by $\underline{X}^{(m)}$. For each $i = 1,2,\ldots,m$. Let

(4) $$\delta_i = \sup\ |E(M|\underline{X}^{(i-1)}) - E(M|\underline{X}^{(i)})| \qquad i = 1,2,\ldots,m.$$

Here $E(M|\underline{X}^{(0)})$ means just $E(M)$. The following inequality is a special case of a martingale inequality due to Azuma.

(5) $$\Pr(|M - E(M)| \geq u) \leq 2e^{-u^2/2\sum_{i=1}^{m}\delta_i^2}$$

for any $u \geq 0$.

To apply these ideas for $F \in \Phi$ we let $X_i = X_i(F) = \{p_i, q_i\}$, $p_i > q_i$, $i = 1,2,\ldots,m$ denote the pairs of F sorted into lexicographically decreasing order. We prove that when $M = \Sigma_F$ in (4)

(6) $$\delta_i \le 4d^2_{\psi(i)} \qquad i = 1,2,\ldots,m.$$

We can then apply (5) and $\sum_{i=1}^{m} d^4_{\psi(i)} \le \sigma_5$ to obtain our result. Fix i and $\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_{i-1}$ and let $\hat{\Phi} = \{F \in \Phi : X_j = \hat{X}_j,\ j = 1,2,\ldots,i-1\}$. Let $Y = W - \bigcup_{j=1}^{i-1} \hat{X}_j$ and $x = \min(Y)$. For $y \in Y - \{x\}$ let $\hat{\Phi}_y = \{F \in \hat{\Phi} : X_i = \{x,y\}\}$ and observe that these sets partition $\hat{\Phi}$. Now

(7) $$E(\Sigma_F | \underline{X}^{(i-1)}) = \frac{1}{2m-2i+1} \sum_{y \in Y-\{x\}} E(\Sigma_F | F \in \hat{\Phi}_y).$$

We show that for $y, y' \in Y$

(8) $$|E(\Sigma_F |\ F \in \hat{\Phi}_y) - E(\Sigma_F |\ F \in \hat{\Phi}_{y'})| \le 4d^2_{\psi(x)}$$

and then (6) follows from (7), (8), $x \ge i$ and

$$\delta_i = \max_{y \in Y-\{x\}} (|E(\Sigma_F | \underline{X}^{(i-1)}) - E(\Sigma_F | F \in \hat{\Phi}_y)|).$$

To prove (8) we consider the following mapping $f_{y,y'} : \hat{\Phi}_y \to \hat{\Phi}_{y'}$ defined as follows: suppose $F \in \hat{\Phi}_y$ contains the pair $\{x', y'\}$ containing y', then

$$f_{y,y'}(F) = (F - \{\{x,y\}, \{x',y'\}\}) \cup \{\{x,y'\}, \{x',y\}\} \in \hat{\Phi}_{y'}.$$

These functions are all bijections, $f^{-1}_{y,y'} = f_{y',y}$ and furthermore $F \in \hat{\Phi}_y$, $y' \neq y$ implies

$$|\Sigma_F - \Sigma_{f_{y,y'}(F)}| \le 4d^2_{\psi(x)}$$

and (c) follows.

($4d^2_{\psi(x)}$ is a crude but obvious upper bound here.) □

We now apply this lemma to obtain information about the graphs $H_{a,b}$, $H'_{a,b}$. Let now $K = \{(a,b): 0 \le a \le \alpha_0, 0 \le b \le \beta_0\}$ where $\alpha_0 = \sigma_2^2 \log n/m^2 < 4\Lambda^2 \log n$ and $\beta_0 = \sigma_2 \log n/m < 2\Lambda \log n$ ($\sigma_2 \le 2m\Lambda$).

Lemma 9

(a) If F is chosen randomly from Φ then

$$\Pr((\alpha(F), \beta(F)) \notin K) \le (\log n)^{-(\sigma_2/m)^2 \log n/10} + (\log n)^{-(\sigma_2/m)\log n/10}.$$

(b) $\Pr_{a,b}(|d_{a,b}(G) - 2m\lambda^2)| \ge 6\, \sigma_2\sigma_5^{1/2} \log n/m) \le e^{-(\sigma_2 \log n/m)^2/2}$

for $G \in \Phi_{a-1,b}$, $(a,b) \in K$, $a > 0$.

(c) $|\Phi_{a-1,b}|/|\Phi_{a,b}| = \frac{a}{\lambda^2}(1 + \epsilon_{a,b})$

where $|\epsilon_{a,b}| \le 100 \dfrac{\sigma_5^{1/2} \log n}{\sigma_2}$ for $(a,b) \in K$, $a > 0$.

(d) $|\Phi_{a,b-1}|/|\Phi_{a,b}| = \frac{b}{\lambda}(1 + \epsilon'_{a,b})$

where $|\epsilon_{a,b}| \le \sigma_5^{1/2} \log n/\sigma_2$, for $(a,b) \in K$, $b > 0$.

Proof

(a) This follows from Lemma 8(a) with $k = 1,2$ and $\rho = \log n$. Note that this implies

$$(9) \qquad |\Phi_{\bar{a},\bar{b}}|/|\Phi| \geq 1/(2\alpha_0\beta_0) \quad \text{for some } (\bar{a},\bar{b}) \in K.$$

Before proceeding to (b) we obtain some approximations to the ratios estimated in (c), (d). Now

$$(10a) \qquad \sum_{F\in\Phi_{a,b}} d_{a,b}(F) = \sum_{G\in\Phi_{a-1,b}} d_{a,b}(G) \quad \text{for } a > 0,$$

and

$$(10b) \qquad \sum_{F\in\Phi_{a,b}} d'_{a,b}(F) = \sum_{G\in\Phi_{a,b-1}} d'_{a,b}(G) \quad \text{for } b > 0.$$

Now (10a) and Lemma 7 imply

$$(11a) \qquad \frac{|\Phi_{a-1,b}|}{|\Phi_{a,b}|} \geq \frac{2am - 2a(2a+b + 2\Delta(\Delta-2))}{E_{a-1,b}(\Sigma_G)}$$

(where $E_{a,b}$ denotes expectation for random variables over $\Phi_{a,b}$)

$$\geq \frac{2am\left(1 - \frac{9\Delta^2 \log n}{m}\right)}{\frac{1}{2}(\Delta-1)^2\,\Delta n}$$

(the denominator here is a crude upper bound for $\max\{\Sigma_G\colon\ G \in \Phi_{a-1,b}\}$)

$$\geq \Delta^{-3}.$$

Similarly

(11b) $$\frac{|\Phi_{a-1,b}|}{|\Phi_{a,b}|} \le \frac{2am}{E_{a-1,b}(\Sigma_G - 3\#C_3(G) - 4\#C_4(G)) - 2a(\Delta-1)^2 - b\Delta(\Delta-1)},$$

(11c) $$\frac{|\Phi_{a,b-1}|}{|\Phi_{a,b}|} \ge \frac{2b(m - b - (\Delta-2) - 2a)}{\sum_{i=1}^{n} \binom{d_i}{2} - (b-1)}$$

$$\ge \frac{2bm}{\sum_{i=1}^{n} \binom{d_i}{2}} \left(1 - \frac{9\Delta^2 \log n}{n}\right),$$

$$\ge \frac{b}{\Delta},$$

and

(11d) $$\frac{|\Phi_{a,b-1}|}{|\Phi_{a,b}|} \le \frac{2b(m - b)}{\sum_{i=1}^{n} \binom{d_i}{2} - (b-1) - 3E_{a,b-1}(\#C_3(G)) - 2a(\Delta-1)}.$$

(b), (c), (d).

Let $(\bar{a},\bar{b})$ be as in (9).

<u>Case 1</u>: $0 \le a \le \bar{a}$, $b = \bar{b}$. Proof of (b) and (c).

It follows from (9) and (11a) that

(12) $$\frac{|\Phi_{a,\bar{b}}|}{|\Phi|} \ge \Delta^{-3\alpha_0/2\alpha_0\beta_0} \ge e^{-(\log n\, \sigma_2/m)^2}, = \theta \text{ say},$$

for $0 \le a \le \bar{a}$.

But then

(13) $$\Pr_{a,\bar{b}}(|\Sigma_F - E(\Sigma_F)| \geq 5\, \sigma_2\sigma_5^{1/2} \log n/m)$$

$$\leq \Pr(|\Sigma_F - E(\Sigma_F)| \geq 5\, \sigma_2\sigma_5^{1/2} \log n/m)/\theta$$

$$\leq e^{-9(\sigma_2 \log n/m)^2/16}$$

on using Lemma 8 (c).

Hence, since $d_{a,\bar{b}}(G) \leq \Sigma_G$ for $G \in \Phi_{a-1,\bar{b}}$,

(14) $$\Pr_{a-1,\bar{b}}(d_{a,\bar{b}}(G) - E(\Sigma_F) \geq 5\, \sigma_2\sigma_5^{1/2} \log n/m) \leq e^{-9(\sigma_2 \log n/m)^2 16}$$

for $0 < a \leq \bar{a}$.

For the subsequent analysis we need two inequalities:

(15a) $$m^{k+1}/\sigma_2^k \geq n^{1-\frac{k}{5}}$$

and

(15b) $$\frac{m^3\sigma_5^{1/2}}{\sigma_2^3} \geq n^{3/20}\ .$$

<u>Proof of (15a)</u>

Use $m \geq 2n$ and $\sigma_2 \leq 2m\Delta$ and $\Delta < n^{1/5}$.

<u>Proof of (15b)</u>

Observe first that $\sigma_5 \geq \sigma_2^{5/2}/n^{3/2}$ and so

$$\frac{m^3\sigma_5^{1/2}}{\sigma_2^3} \geq \frac{m^3}{n^{3/4}\sigma_2^{7/4}}$$

$$\geq \frac{m^3}{4n^{3/4}m^{7/4}\Delta^{7/4}} .$$

Now use $m \geq 2n$ and $\Delta \leq n^{1/5-\epsilon}$.

Now

(16) $$\Pr_{a,\bar{b}}(\#C_4(F) \geq 10^{-2}\sigma_2\sigma_5^{1/2}\log n/m)$$

$$\leq \Pr_{a,\bar{b}}(\#C_4(F) \geq n^{3/20}(\sigma_2/m)^4) \qquad \text{by (15b)}$$

$$\leq \Pr(\#C_4(F) \geq n^{3/20}(\sigma_2/m)^4/\theta$$

$$\leq n^{-\frac{3}{20}\cdot\frac{1}{10}\cdot n^{3/20}\cdot(\sigma_2/m)^4/\theta}$$

(on using Lemma 8(a) and (15a))

$$\leq e^{-(\sigma_2\log n/m)^2},$$

for $0 \leq a \leq \bar{a}$.

Similarly,

(17) $$\Pr_{a,\bar{b}}(\#C_3(F) \geq 10^{-2}\sigma_2\sigma_5^{1/2}\log n/m) \leq e^{-(\sigma_2\log n/m)^2} \quad \text{for } 0 \leq a \leq \bar{a}.$$

Thus, on using the lower bound for $d_{a,\bar{b}}(G)$ in Lemma 7(b) (and

$\sigma_5 \geq \sigma_2^{5/2}/n^{3/2}$), we find

$$\text{(18)} \qquad Pr_{a-1,\bar{b}}(d_{a,\bar{b}}(G)-E(\Sigma_F) \leq -5.9\sigma_2\sigma_5^{1/2}\log n/m)$$

$$\leq e^{-(9(\sigma_2\log n/m)^2/16} + 2e^{-(\sigma_2 \log n/m)^2}$$

for $0 < a \leq \bar{a}$.

Now observe that

$$\text{(19)} \qquad |E(\Sigma_F) - 2m\lambda^2| = O(\sigma_3/m) = o(\sigma_2\sigma_5^{1/2}/m)$$

Using this in (18) completes the proof of (b) for Case 1.

We now show that for $0 \leq a \leq \bar{a}$,

$$\text{(20)} \qquad |E_{a,\bar{b}}(\Sigma_F) - E(\Sigma_F)| \leq 7\sigma_2\sigma_5^{1/2}\log n/m$$

and

$$\text{(21a)} \qquad E_{a,\bar{b}}(\#C_3(F)) \leq \sigma_2\sigma_5^{1/2}\log n/99m.$$

$$\text{(21b)} \qquad E_{a,\bar{b}}(\#C_4(F)) \leq \sigma_2\sigma_5^{1/2}\log n/99m.$$

Proof of (20)

It follows from (13) that

$$E_{a,\bar{b}}(\Sigma_F) \geq (E(\Sigma_F) - 6\sigma_2\sigma_5^{1/2}\log n/m)(1 - e^{-9(\sigma_2 \log n/m)^2/16})$$

$$\geq E(\Sigma_F) - 7\sigma_2\sigma_5^{1/2}\log n/m$$

and

$$E_{a,\bar{b}}(\Sigma_F) \leq E(\Sigma_F) + 6\sigma_2\sigma_5^{1/2}\log n/m + \tfrac{1}{2}(\Delta-1)^2\Delta n\ e^{-9(\sigma_2 \log n/m)^2/16}$$

$$\leq E(\Sigma_F) + 7\sigma_2\sigma_5^{1/2}\log n/m$$

and (20) follows.

Proof of (21b)

It follows from (16) that

$$E_{a,\bar{b}}(\#C_4) \leq 10^{-2}\sigma_2\sigma_5^{1/2}\log n/m + \binom{2m}{8}\Pr_{a,\bar{b}}(\#C_4 \geq 10^{-2}\sigma_2\sigma_5^{1/2}\log n/m)$$

$$\leq 10^{-2}\sigma_2\sigma_5^{1/2}\log n/m + \binom{2m}{8}e^{-(\sigma_2 \log n/m)^2}$$

$$\leq \sigma_2\sigma_5^{1/2}\log n/99m.$$

The proof of (21a) is almost identical.

Now from (11a) and (20) we have

(22) $$\frac{|\Phi_{a-1,b}|}{|\Phi_{a,b}|} \geq \frac{2am\ (1 - \frac{9\Lambda^2 \log n}{n})}{E(\Sigma_F) + 7\sigma_2\sigma_5^{1/2}\log n/m} .$$

But now we can use (19) and

(23) $$2m\lambda^2 \geq \sigma_2^2/8m$$

in (22) to give us the required lower bound. The upper bound follows in the same way from (11b), (19), (20), (21) and (23).

<u>Case 2</u>: $\bar{a} < a \leq \alpha_0$, $b = \bar{b}$. Proof of (b) and (c).

We inductively show that, for $\bar{a} \leq a < \alpha_0$,

(24) $$\frac{|\Phi_{a,\bar{b}}|}{|\Phi|} \geq \frac{1}{2}\Lambda^{-3}\lambda^{2(a-\bar{a})}\frac{\bar{a}!}{a!}\left(1 - 100\frac{\sigma_5^{1/2}\log n}{\sigma_2}\right)^{a-\bar{a}} \geq \theta.$$

Now for $a = \bar{a}$ this follows from (8) and so assume it is true for $\bar{a},\bar{a}+1,\ldots,a-1$ for some $a > \bar{a}$. Note that this implies (20) and (21) (with a replaced by a-1). Using these inequalities in (10b) yields (24).

This then implies (20) and (21), (with a now and not a-1), which yields (b). Finally (20) and (21) and (10) give (c).

<u>Case 3</u>: $0 \leq a \leq \alpha_0$, $0 \leq b < \bar{b}$. Proof of (b) and (d)

Using (11c) and (24) we obtain

$$(25) \qquad \frac{|\Phi_{a,b}|}{|\Phi|} \geq \Delta^{-\beta_0}\theta \qquad 0 \leq b < \bar{b}.$$

This enables us to show (see 21a) $E_{a,b-1}(\#C_3(G)) \leq \sigma_2\sigma_5^{1/2}\log n/99m$ in (11d) and then (d) follows from (11c) and (11d). The proof of (b) follows as in previous cases.

Case 4: $0 \leq a \leq \alpha_0$, $\bar{b} < b \leq \beta_0$. Proof of (b) and (d).

This time we inductively show that for $\bar{b} \leq b \leq \beta_0$

$$(26) \qquad \frac{|\Phi_{a,b}|}{|\Phi|} \geq \theta\,\lambda^{b-\bar{b}}\,\frac{\bar{b}!}{b!}\left(1 - \frac{\sigma_5^{1/2}\log n}{\sigma_2}\right)^{b-\bar{b}}$$

$$\geq \theta e^{-\sigma_5^{1/2}(\log n)^2/m}.$$

Now for $b = \bar{b}$ this follows from (20) and so assume it is true for $\bar{b},\bar{b}+1,\ldots,b-1$ for some $b > \bar{b}$. Note that this enables us to show $E_{a,b-1}(\#C_3(G)) \leq \sigma_2\sigma_5^{1/2}\log n/99m$ in (10d) and then (26) follows from (10d). But then (10c) and (10d) now imply (d) and the proof of (b) follows as in previous cases. □

We can now estimate the probability that $\mathcal{E}^b$ occurs.

Lemma 10

Let $\mathcal{E}^b$ be as defined prior to Lemma 4. Then, with the assumptions of Theorem 2,

$$\Pr(\mathcal{E}^b) = o(1).$$

Proof

Let K as defined prior to Lemma 8. Then, by Lemma 8(a),

$$\text{(27)} \qquad \Pr(\mathcal{E}^b) \le \sum_{(a,b)\in K} \Pr(\mathcal{E}^b | F_0 \in \Phi_{a,b}) \Pr(F_0 \in \Phi_{a,b}) + o(1).$$

So we can restrict our attention to estimating

$$\Pr(\mathcal{E}^b | F_0 \in \Phi_{a,b}) \quad \text{where} \quad (a,b) \in K.$$

Let now $\mathcal{E}_i^b$ denote the occurrence of $\mathcal{E}^b$ on the i^{th} execution of the main loop of CHOOSEA. Then

$$\Pr(\mathcal{E}^b | F_0 \in \Phi_{a,b}) \le \sum_{i=1}^{a+b} \Pr(\mathcal{E}_i^b | F_0 \in \Phi_{a,b}).$$

Case 1: $1 \le i \le a$.

Let ξ^+ denote $\max\{0,\xi\}$ when $\xi \in \Re$. Then

$$\Pr(\mathcal{E}_i^b | F_0 \in \Phi_{a,b}) = \sum_{G\in\Phi^+_{a-1,b}} (\pi_{a-i+1,b}(G) - |\Phi_{a-i,b}|^{-1})^+$$

$$\le \Pr_{a-i,b}(d_{a-i+1,b}(G) \ge 2m\lambda^2 + 6\sigma_2\sigma_5^{1/2}\log n/m)$$

$$+ \sum_{G\in\Phi^+_{a-i+1,b}} (|\Phi_{a-i+1,b}|^{-1}(2m\lambda^2 + 6\sigma_2\sigma_5^{1/2}\log n/m)/(2(a-i+1)m(1 - \frac{3\Lambda^2\log n}{n}))$$

$$- |\Phi_{a-i,b}|^{-1})^+$$

(on using $\pi_{a-i+1,b}(G) \leq |\Phi_{a-i+1,b}|^{-1} d_{a-i+1,b}(G)/(2(a-i+1)m(1 - \frac{3\Delta^2 \log n}{m})))$

$$\leq e^{-(\sigma_2 \log n/m)^2/2} + |\Phi_{a-i,b}|^{-1} \sum_{G \in \Phi^+_{a-i+1,b}} \left(\frac{|\Phi_{a-i,b}|}{|\Phi_{a-i+1,b}|} \frac{\lambda^2}{(a-i+1)}(1 + \frac{200 \log n \sigma_5^{1/2}}{\sigma_2}) - 1\right)^+$$

$$\leq e^{-(\sigma_2 \log n/m)^2/2} + |\Phi_{a-i,b}|^{-1} \sum_{G \in \Phi^+_{a-i+1,b}} (1 + 301 \log n\ \sigma_5^{1/2}/\sigma_2 - 1)^+$$

$$\leq 302 \log n\ \sigma_5^{1/2}/\sigma_2.$$

<u>Case 2</u>: $a < i \leq a+b$ (let $b' = a+b-i$)

$$\Pr(\mathcal{E}_i^b | F_0 \in \Phi_{a,b}) \leq$$

$$\sum_{G \in \Phi^+_{0,b'}} \left(|\Phi_{0,b'+1}|^{-1} \left(\sum_{i=1}^{n} \binom{d_i}{2} - b'+1\right)(2b'(m - b' - \Delta+2))^{-1} - |\Phi_{0,b'}|^{-1}\right)^+$$

$$\leq 2 \log n\ \sigma_5^{1/2}/\sigma_2.$$

after making approximations as in Case 1.

Thus

$$\Pr(\mathcal{E}^b) = o(1) + O((\log n\ \sigma_5^{1/2}/\sigma_2)(\sigma_2^2 \log n/m^2))$$

$$= o(1) + O((\log n)^2\ \sigma_5^{1/2} \sigma_2/m^2)$$

Now let $\hat{d}_i = d_i/d_n$, $i = 1,2,\ldots,n$, $\hat{m} = m/d_1$ and $\hat{\sigma}_k = \sigma_k/d_1$ for $k \geq 2$. Then

$$\sigma_5^{1/2}\sigma_2/m^2 = d_n^{5/2}\,\hat{\sigma}_5^{1/2}\,\hat{\sigma}_2/\hat{m}^2$$

$$\leq d_n^{5/2}(2\hat{m})^{1/2}\,\hat{d}_1^2\,2\hat{m}\,\hat{d}_1/\hat{m}^2$$

$$\leq 4\,d_1^3/(nd_n)^{1/2}$$

$$= O(n^{-\epsilon})$$

and the result follows. □

Theorem 4 now follows from Lemma 6 and Lemma 10.

§5. Proof of Theorem 2

It follows from Lemma 7(a) that

$$|\Phi| \sim \sum_{(a,b)\in K} |\Phi_{a,b}|.$$

Furthermore $(a,b) \in K$ implies

$$\frac{|\Phi_{a,b}|}{|\Phi_{0,0}|} = \frac{\lambda^{2a+b}}{a!b!} \prod_{i=0}^{a-1} (1 + \epsilon_{a-i,b}) \prod_{j=0}^{b-1} (1 + \epsilon'_{0,b-j})$$

$$= \frac{\lambda^{2a+b}}{a!b!} (1 + \theta_{a,b})$$

where $|\theta_{a,b}| \leq n^{-\epsilon/2}$.

Hence

$$|\Phi| \sim |\Phi_{0,0}| \sum_{a=0}^{\alpha_0} \sum_{b=0}^{\beta_0} \frac{\lambda^{2a+b}}{a!b!}$$

$$\sim |\Phi_{0,0}|\, e^{\lambda+\lambda^2} \qquad \text{since } \alpha_0 >> \lambda^2 \text{ and } \beta_0 >> \lambda.$$

The result now follows from

$$|\Phi| = \frac{(2m)!}{m!2^m} \text{ and } |\Phi_{0,0}| = (\prod_{i=1}^{n} d_i!)\,|\mathscr{G}(\underline{d})|.$$

□

Observe that the following (technical) strengthening of Theorem 2 is possible. It is used in the proof of Theorem 3b.

<u>Theorem 2*</u>

The condition $d_1^3/d_n^{1/2} = 0(n^{1/2-\epsilon})$ in Theorem 2 can be weakened to

$$\sigma_5^{1/2}\sigma_2/m^2 \leq An^{-\epsilon}$$

for some absolute constant $A > 0$.

The conclusion can then be expressed

$$|(e^{\lambda+\lambda^2}|\Phi_{0,0}|/|\Phi|) - 1| \leq n^{-\epsilon/2} \qquad \text{for large } n.$$

Proof

Lemma 9 was proved with the stronger assumption, which was then shown to imply the weaker (less complex) assumption at the end of the proof of Lemma 10. □

§6. Proof of Theorem 3

We proceed as in the proofs for r constant, but make minor changes. The only technical difficulty lies in the fact that our construction of F_{CHOOSE} requires us to delete edges from the initially chosen F_0. However we delete few edges and so a given edge is unlikely to be deleted.

Lemma 11

Consider Algorithm CHOOSE. Assume $F_0 \in K$ and $u \in F_0$ is not a loop or redundant edge. Then, independently of previous choices,

$$\Pr(u \notin F_i \mid u \in F_{i-1}) \leq \frac{3}{rn}$$

for $i = 1,2,\ldots,a+b$.

Proof

Assume first that $i \leq a$. Consider condition A of §3 and assume edge $F_{i-1}F_i$ of $H_{a-i+1,b}$ comes from some fixed e in (i). Then there are at least $\frac{1}{2}rn - (2a + b + 2r(r-2))$ equally likely choices for f (see the proof of Lemma 7a). This implies the result for $i \leq a$. For $i > a$ use the proof of Lemma 7c. □

Corollary 12

Assume $S \subseteq F_0$. Then $\Pr(F_{CHOOSE} \cap S = \phi) \leq (\frac{4r \log n}{n})^{|S|}$.

Proof

Let $S = \{u_1, u_2, \ldots, u_s\}$. If $S \cap F_{CHOOSE} = \phi$ then for some $1 \leq i_1 < i_2 < \ldots < i_s \leq \alpha_0 + \beta_0 \leq r^2 \log n + r \log n$ we find that an element of S is deleted at iteration $i_1, i_2, \ldots, i_s$ of Algorithm CHOOSE. Then, by Lemma 11,

$$\Pr(S \cap F_{CHOOSE} = \phi) \leq s! \binom{\lfloor r^2 \log n \rfloor + \lfloor r \log n \rfloor}{s} (\frac{3}{rn})^s$$

$$\leq (\frac{4r \log n}{n})^s \qquad \square$$

Proof of Theorem 3a

Since the result is known for r constant, we shall assume $r \geq 100$. If F_{CHOOSE} is not r-connected then there exists a set $R \subseteq V_n$, $|R| = r-1$ and sets A,B, $2 \leq |A| \leq |B|$, partitioning $V_n - R$ such that every edge joining A,B in F_0 is missing in F_{CHOOSE}. We refer to this event as $\mathcal{E}_{R,A,B}$. Let $\mathcal{E}'_A$ be the event F_0 contains $2k$ or more pairs contained in $W_A = \bigcup_{i \in A} W_i$. Now fix A, $2 \leq |A| = k \leq r(\log n)^2$. Then

$$\Pr(\bigcup_R \mathcal{E}_{R,A,B}) \leq \Pr(\mathcal{E}'_A) + \sum_R \Pr(\mathcal{E}_{R,A,B} | \bar{\mathcal{E}}'_A)$$

$$\leq \binom{rk}{2k}(\frac{k}{n})^{2k} + \binom{n}{r-1} \sum_{t=0}^{(r-2)k} \binom{(r-2)k}{t} (\frac{r-1}{n})^t (\frac{4r \log n}{n})^{(r-2)k-t}$$

$$\leq (\frac{rek}{2n})^{2k} + n^{r-1} 2^{(r-2)k} (\frac{4r \log n}{n})^{(r-2)k}.$$

Hence

$$\Pr\Big(\bigcup_{|A|\le r(\log n)^2}\bigcup_{R}\mathcal{E}_{R,A,B}\Big) \le \sum_{k=2}^{r(\log n)^2}\binom{n}{k}\Big(\Big(\frac{rek}{2n}\Big)^{2k} + n^{r-1}\Big(\frac{8r\ \log n}{n}\Big)^{(r-2)k}\Big)$$

$$\le \sum_{k=2}^{r(\log n)^2}\Big(\Big(\frac{r^3e^3k}{4n}\Big)^{k} + n^{k+r-.8(r-2)k}\Big)$$

$$= o(1).$$

For $|A| = k > r(\log n)^2$ we observe that if $\mathcal{E}_{R,A,B}$ occurs then there must be at most $r^2\log n + r\ \log n + r(r-1) \le rk/10$ edges between A and $V_n - A$ in $\mu(F_0)$. But the probability that this occurs for some A, $r(\log n)^2 \le |A| \le \frac{1}{2}n$ is at most

$$\sum_{k=r(\log n)^2}^{\frac{1}{2}n}\binom{n}{k}\binom{rk}{rk/10}\Big(\frac{k}{n}\Big)^{9rk/10}$$

$$\le \sum_{k=r(\log n)^2}^{\frac{1}{2}n}\Big(\frac{ne}{k}\Big)^{k}(10e)^{rk/10}\Big(\frac{k}{n}\Big)^{9rk/10}$$

$$= \sum_{k=r(\log n)^2}^{\frac{1}{2}n}\Big(\Big(\frac{k}{n}\Big)^{\frac{9}{10}r-1} e\ (10e)^{r/10}\Big)^{k}$$

$$\le \sum_{k\ge r(\log n)^2}\Big(\Big(\frac{1}{2}\Big)^{9r/10} 2e\ (10e)^{r/10}\Big)^{k}$$

$$= o(1) \qquad \text{as } r \ge 100.$$

Thus all possibilities for $|A|$ are covered and the part (a) of the theorem is proved. □

Proof of Theorem 3b

Since the case r constant (and large) is already established we shall assume that $r \to \infty$ with n.

We modify the argument from Fenner and Frieze [4].

We need to show that

$$\Pr(\phi(F_{CHOOSE}) \text{ is hamiltonian}) = 1 - o(1).$$

We consider the following blue-green colouring of $F \in \Phi$: let $W_1' \subseteq W_1, \dots W_n' \subseteq W_r$ be fixed subsets of size $\lceil r/2 \rceil$ and $W' = \bigcup_{i=1}^{n} W_i'$. Now for $i = 1,2,\dots,n$ choose $w_i \in W_i - W_i'$ and let $W'' = W' \cup \{w_1',\dots,w_n'\}$.

We will use σ to denote an arbitrary choice of $w_1, w_2, \dots, w_n$.

Let $F_b = F_b(\sigma) = \{e \in F\colon e \cap W'' \neq \phi\}$ and $F_g = F - F_b$.

A graph G with vertex set V_n is said to have the neighbourhood property if $S \subseteq V_n$, $|S| \leq \frac{1}{4} n$ implies $|N_G(S)| \geq 2|S|$ where

$$N_G(S) = \{w \in V_n - S\colon \exists\, v \in S \ \text{ s.t. } \{v,w\} \in E(G)\}.$$

If $F \in \Phi$ let $\phi'(F)$ be the graph $(V_n, \{\phi(e)\colon e \in F \text{ and } e \cap W' \neq \phi\}$.

Lemma 13

$\Pr(\phi'(F_{CHOOSE}) \text{ has the neighbourhood property}) = 1 - o(1)$.

Proof

Fix $k \geq r/6$ and let $A = \{1,2,\ldots,k\}$, $B = \{k+1,\ldots,3k-1\}$ and $C = \{3k+1,\ldots,n\}$. Then by considering the "first" $rk/4$ pairs involving $A \cap W'$ we see that

$$\Pr(F'_{CHOOSE} \text{ has no A–C pairs}) \leq \left(\frac{3k}{n} + \frac{4r\log n}{n}\right)^{kr/4}.$$

Hence

$$\Pr(\phi(F'_{CHOOSE}) \text{ does not have the neighbourhood property})$$

$$\leq \sum_{k=r/6}^{n/4} \binom{n}{k}\binom{n}{2k}\left(\frac{3k}{n} + \frac{4r\log n}{n}\right)^{rk/4}$$

$$\leq \sum_{k=r/6}^{n/4} u_k^k$$

where $u_k = \frac{n^3e^3}{4k^3}\left(\frac{3k}{n} + \frac{4r\log n}{n}\right)^{r/4}$.

For $k \leq 32r\log n$, $u_k \leq n^3e^3n^{-r/5}$ $(r = O(n^{1/5-\epsilon}))$ and for $k > 32r\log n$, $u_k \leq (n^3e^3/k^3)(25k/8n)^{r/4}$ and it is easy to see that

$$\sum_{k=r/6}^{n/4} u_k^k = o(1).$$

□

Lemma 14

$$\Pr(|F_{CHOOSE} - F'_{CHOOSE}| \geq \frac{rn}{10}) = 1 - o(1).$$

<u>Proof</u>

It is sufficient to show that

$$\Pr(|F_0 - F_0'| \geq \frac{rn}{9}) = 1 - o(1).$$

let $Z = |F_0 - F_0'|$. It is easy to see that $E(Z) \sim \frac{rn}{8}$. Next consider (4) and (5) with $M = Z$. Using the same sort of argument as in Lemma 8(c) we see that $\delta_i \leq 1$ and the lemma follows from (5). □

Now define

$$\Psi_0 = \{F \in \Phi_{0,0}: \begin{array}{ll} \text{(i)} & \phi(F) \text{ is connected,} \\ \text{(ii)} & \phi'(F) \text{ has the neighbourhood property,} \\ \text{(iii)} & |F - F'| \geq rn/10\} \end{array}$$

We deduce from Theorems 3a, 4 and Lemmas 13 and 14 that $|\Psi_0|/|\Phi_{0,0}| = 1 - o(1)$. Let $\Psi_1 = \{F \in \Psi_0: \Phi(F)$ is not hamiltonian$\}$. We must show

(28) $$\frac{|\Psi_1|}{|\Phi_{0,0}|} = o(1).$$

For $F \in F_{0,0}$ we let

$$a(F,\sigma) = 1 \text{ if } \begin{array}{ll} \text{(i)} & \phi(F_b) \text{ has the neighbourhood property,} \\ \text{(ii)} & \text{longest paths in } \phi(F_b) \text{ are longest in } \phi(F), \\ \text{(iii)} & \text{no } e \in F_g \text{ is such that } \psi(e) \text{ joins the endpoints} \\ & \text{of a longest path of } \phi(F_b) \\ \text{(iv)} & F \in \Psi_1 \end{array}$$

$$= 0 \qquad \text{otherwise.}$$

((ii) above is not actually necessary but helps to verify (iii)). Observe that

(29) $$a(F,\sigma) = 1 \quad \text{implies} \quad |F_g| \geq rn/11.$$

We observe next (as in [4]) that

$$F \in \Psi_1 \quad \text{implies} \quad \sum_{\sigma} a(F,\sigma) \geq 1$$

and so

$$|\Psi_1| \leq \sum_{F \in \Phi_{0,0}} \sum_{\sigma} a(F,\sigma).$$

For a fixed values $\bar{\sigma}$ of σ and $\bar{F}$ of F_b let

$$X_0(\bar{\sigma},\bar{F}) = |\{F_g\colon F = F_g \cup \bar{F} \ \text{ satisfies } \ F_b(\bar{\sigma}) = \bar{F}\}|$$

and

$$X_1(\bar{\sigma},\bar{F}) = |\{F_g\colon F = F_g \cup \bar{F} \ \text{ satisfies } \ F_b(\bar{\sigma}) = \bar{F} \ \text{ and } \ a(F,\bar{\sigma}) = 1\}|.$$

We show that there exists a constant $0 < \gamma < 1$ such that

(30) $$|X_1(\bar{\sigma},\bar{F})| \leq \gamma^{rn}|X_0(\bar{\sigma},\bar{F})| \quad \text{for all } \ \bar{\sigma},\bar{F}.$$

It will then follow that

$$|\Psi_1| \le \sum_{F}\sum_{\sigma} a(F,\sigma)$$

$$= \sum_{\bar{F}}\sum_{\bar{\sigma}} |X_1(\bar{\sigma},\bar{F})|$$

$$\le \gamma^{rn} \sum_{\bar{F}}\sum_{\bar{\sigma}} |X_0(\bar{\sigma},\bar{F})|$$

$$= \gamma^{rn}|\Phi_{0,0}|\ \lfloor r/2 \rfloor^n$$

and (28) follows.

To prove (30) we observe that if $n_1(\bar{\sigma},\bar{F}) > 0$ then using (i) and (iii) of the definition of $a(F,\sigma)$ and Posa's Theorem [6] we see that

(31) $\exists\ S = \{v_1 v_2,\dots,v_k\} \subseteq V_n$ and sets $S_1,S_2,\dots,S_k$ $\subseteq V_n$ of size $k \ge \frac{1}{4} n$ such that F_g has no pair $\{x,y\}$ with $\Psi(x) = v_i$ and $\Psi(y) \in S_i$ for some $i = 1,2,\dots,k$.

Suppose that for some such fixed $(\bar{\sigma},\bar{F})$ we find $\phi(\bar{F})$ is a graph with degrees $d_1',d_2',\dots,d_n'$. Then $X_0(\bar{\sigma},\bar{F})$ is (essentially) the set of configurations with degree sequence $\tilde{d}_1 = r - d_1',\ \tilde{d}_2 = r - d_2',\dots,\tilde{d}_n = r - d_n'$ and X_1 is the subset of these configurations which contain no edge as described in (31). Observe that (29) implies $\sum_{i=1}^{n} \tilde{d}_1 \ge \frac{rn}{11}$. Let $\tilde{\Phi}$ be the set of configurations defined using $\underline{\tilde{d}}$ and let $\tilde{\Phi}_{0,0}$ be those without loops or parallel edges. We need to show that

(32) $$\Pr_{0,0}(\tilde{F} \text{ contains no pair as defined in (31)}) \le \gamma^{rn}$$

for some $\gamma < 1$.

But it is simple to show that

(33) $$\Pr(\tilde{F} \text{ contains no pair as defined in (31)}) \le \gamma_1^{rn}$$

for some $\gamma_1 < 1$.

Also by Theorem 2^*

(34) $$\frac{|\tilde{\Phi}_{0,0}|}{|\tilde{\Phi}|} \ge \frac{1}{2} e^{-(\tilde{\lambda}+\tilde{\lambda}^2)}$$

where $\tilde{\lambda} = O(r)$. (We use $\sum_{i=1}^{n} \tilde{d}_i \ge rn/11$ at this point.)

(33) and (34) imply (32) which implies (30) and hence the theorem. □

References

[1] E.A. Bender and E.R. Canfield, "The asymptotic number of labelled graphs with given degree sequences". *Journal of Combinatorial Theory (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] B. Bollobás, "Almost all regular graphs are hamiltonian", *European Journal on Combinatorics* **4** (1983) 97-106.

[4] T.I. Fenner and A.M. Frieze, "Hamilton cycles in random regular graphs", *Journal of Combinatorial Theory (B)* **37** (1984) 103-112.

[5] A.M. Frieze, "Finding hamilton cycles in sparse random graphs", to appear in *Journal of Combinatorial Theory (B)*.

[6] L. Pósa, "Hamiltonian circuits in random graphs", *Discrete Mathematics* **14** (1976) 359-364.

[7] R.W. Robinson and N.C. Wormald, "Existence of long cycles in random cubic graphs", *Progress in Enumeration and Design, Proceedings of Waterloo Conference on Combinatorics* (1984) 251-270.