ON SUBGRAPH SIZES IN RANDOM GRAPHS

Neil Calkin[*]and Alan Frieze[†]

Department of Mathematics,

Carnegie Mellon University,

Pittsburgh, U.S.A.

and

Brendan D. McKay

Department of Computer Science,

Australian National University

Canberra, Australia

September 4, 2008

Abstract

For a graph G with m edges let its Range of Subgraph Sizes (RSS)

$$\rho(G) = \{t : G \text{ contains a vertex-induced subgraph with } t \text{ edges}\}.$$

G has a *full* RSS if $\rho(G) = \{0, 1, \ldots, m\}$. We establish the threshold for a random graph to have a full RSS and give tight bounds on the likely RSS of a dense random graph.

1

1 INTRODUCTION

In this paper we discuss the *Range of Subgraph Sizes* (RSS) of a random graph. For a graph G with $e(G) = m$ edges we define

$$\rho(G) = \{t : G \text{ contains a vertex-induced subgraph with } t \text{ edges}\}.$$

G has a *full* RSS if $\rho(G) = \{0, 1, \ldots, m\}$.

The RSS of a graph is a graph parameter of intrinsic interest. To the best of our knowledge it has not been seriously studied before, mainly because of the difficulty of proving anything interesting about it. On the other hand, for random graphs we can answer two natural questions with high precision.

First of all G can only have a full RSS if it has minimum degree $\delta(G) \leq 1$, otherwise no vertex-induced subgraph can have $m - 1$ edges. Our first result is that the threshold for the random graph $G_{n,m}$ to have a full RSS is the same as that for minimum degree 2.

Theorem 1 *Let $m = \frac{n}{2}(\log n + \log \log n + c_n)$ and let c be a constant.*

(All logarithms in this paper are natural unless indicated otherwise.)

$$\lim_{n \to \infty} \mathbf{Pr}(G_{n,m} \text{ has a full RSS}) = \begin{cases} 1 & \text{if } c_n \to -\infty \\ 1 - e^{e^{-c}} & \text{if } c_n \to c \\ 0 & \text{if } c_n \to \infty \end{cases}$$

$$= \lim_{n \to \infty} \mathbf{Pr}(\delta(G_{n,m}) \leq 1).$$

$\square$

It will be convenient to prove a slightly stronger result about the *graph process*

$\Gamma_i = (\{1, 2, \ldots, n\}, E_i)$ where $E_0 = \emptyset$ and E_{i+1} is obtained from E_i by adding a random edge. Let $m^* = \min\{i : \delta(\Gamma_i) \geq 2\}$. Theorem 1 follows from

Theorem 2 $\Gamma_0, \Gamma_1, \ldots \Gamma_{m^*-1}$ **a.s.** *all have full RSS's.*

$\square$

An event occurs **a.s.** (almost surely) if it occurs with probability 1-o(1) as $n \to \infty$. This is non-standard for probability theorists but has become accepted in the study of random graphs. An event occurs **q.s.** (quite surely) [2] if the probability of occurrence is $1 - o(n^{-A})$ for any constant $A > 0$.

Our second result concerns the random graph $G_{n,p}$ when p is constant. We show that in a sense this graph **a.s.** has a nearly complete RSS of subgraph sizes.

Theorem 3 *There are constants $A = A(p), A' = A'(p)$ such that* **a.s.**

(a)
$$\rho(G_{n,p}) \supseteq [0, e(G_{n,p}) - \frac{An^{3/2}}{\sqrt{\log n}}].$$

(b)
$$\rho(G_{n,p}) \not\supseteq [0, e(G_{n,p}) - \frac{A'n^{3/2}}{\sqrt{\log n}}].$$

$\square$

These questions were first raised in discussions between Brendan McKay and Paul Erdős.

Notation

(For real values $a < b$ the notation $[a, b] = \{\lfloor a \rfloor, \lfloor a \rfloor + 1, \ldots, \lceil b \rceil\}$ will be used throughout the paper. Also $[n] = [1, n]$.)

For a graph $G = (V, E)$, $\Delta(G)$ is its maximum degree. If $v \in V$ then $d(v, G)$ is the degree of v. For $S \subseteq V$, $d(v, S)$ is the number of neighbours of v in S. (The specific graph G will be

understood from the context.)

$Bin(n, p)$ denotes a Binomial random variable and $N(\mu, \rho^2)$ denotes a Normal random variable. The Chernoff bound refers to

$$\mathbf{Pr}(|Bin(n, p) - np| \geq \epsilon np) \leq 2e^{-\epsilon^2 np/3}$$

for $0 \leq \epsilon \leq 1$.

2 PROOF OF THEOREM 1

In the following all probabilistic statements hold *simultaneously* for $i \in [m^* - 1]$. We remark that $\Delta(\Gamma_{m^*}) \leq 3 \log n$ **a.s.** .

Case 1: $i \in [\frac{1}{5} n \log n]$

One can easily show that **a.s.** in this case Γ_i has more than $\Delta(\Gamma_i)$ isolated edges. We order the vertices $v_1, v_2, \ldots v_n$ so that the vertices of the isolated edges come last. Let H_j be the subgraph of Γ_i induced by $v_1, v_2, \ldots v_j$ and let $k_j = e(H_j)$. We show that we are always able to find a subgraph with t edges, for all $t \in [k_j + 1, k_{j+1} - 1]$. In this case we simply add $t - k_j$ isolated edges to H_j. (We apply the same approach throughout this proof.)

Case 2: $i \in [\frac{1}{5} n \log n, \frac{2}{5} n \log n]$

4

We note that if $A_t = \{v : d(v, \Gamma_i) = t\}$ and $A_1' = \{$ vertices on isolated edges$\}$ then **a.s.** $|A_1 \setminus A_1'| \geq n^{1/6}$ and the neighbours B of $A_1 \setminus A_1'$ induce an acyclic subgraph.

We now order the vertices $v_1, v_2, \ldots v_n$ so that the vertices in B come first and the vertices in $A_1 \setminus A_1', A_1'$ come last. We again consider the sequence of subgraphs H_j. Now trees clearly have full spectra and this deals with the case $j \leq |B|$. For $j < n - |A_1|$ we can always add $t - k_j$ vertices from $A_1 \setminus A_1'$ to H_j. For $j \geq n - |A_1|$ we have $k_{j+1} \leq k_j + 1$ and there is nothing to do.

Case 3: $i \in [\frac{2}{5} n \log n, m^* - 1]$

(We note that $|m^* - \frac{n}{2}(\log n + \log \log n)| \leq \log \log \log n$ **a.s.**.)

We now note that **a.s.**

$$|A_2| \geq \frac{\log n}{\log \log n},$$

$$|A_3| \geq \frac{(\log n)^2}{\log \log n},$$

and the neighbours B of $A = A_1 \cup A_2 \cup A_3$ form an independent set.

We now order the vertices $v_1, v_2, \ldots v_n$ so that B is first in the order and A_3, A_2, A_1 finish the order. Now $k_j = 0$ for $j \leq |B|$. For $j < n - |A|$ we have $k_{j+1} - k_j \leq \Delta \leq 3 \log n$. So to get a graph of size $k_j + t < k_{j+1}$ we simply add $\lfloor t/3 \rfloor$ vertices from A_3, and if necessary a vertex from A_1 or A_2 depending on $t \bmod 3$. For $j > n - |A|$ we add vertices from A_2 or A_1 as necessary to H_j.

This completes the proof of Theorem 2 and also of Theorem 1. $\qquad \square$

3   PROOF OF THEOREM 3

In this section ℓ denotes the required subgraph size. We break the proof of part (a) into 3 sub-cases depending on the size of ℓ.

Case 1: $0 \leq \ell \leq \ell_0 = (\log_\rho n)^2/3$ where $\rho = (\min\{p, 1-p\}))^{-1}$.

This is particularly straightforward as G **a.s.** contains a copy of any subgraph H with at most $k_0 = \lfloor (1-\epsilon) \log_\rho n \rfloor$ vertices, where $\epsilon > 0$ is fixed. This result is not best possible and each such H can be chosen "greedily" one vertex at a time.

Indeed, if H has k vertices we choose vertices $1 = v_1 < v_2 < \ldots v_k$ which induce a copy of H in G. After having chosen $v_1, v_2, ..., v_t$ the probability that a particular vertex $v > v_t$ cannot be added is at most $1 - \rho^{-t}$. Hence

$$\mathbf{Pr}(v_{t+1} - v_t \geq n^{1-\epsilon/2}) \leq (1 - \rho^{-t})^{n^{1-\epsilon/2}} \leq \exp(-n^{\epsilon/2}),$$

and the probability that H cannot be constructed is at most t times this. Hence the probability that there is an H that cannot be constructed in this way is at most

$$\sum_{t=2}^{\ell_0} 2^{\binom{t}{2}} t \exp(-n^{\epsilon/2}) = o(1).$$

Case 2: $\ell_0 \leq \ell \leq \ell_1 = p\binom{n}{2} - n^{3/2+\epsilon}$ where $\epsilon > 0$ is arbitrarily small (actually the reader can easily verify that $\epsilon = \frac{1}{4}$ suffices).

Let

$$\omega = c_0(\ell^{1/2}/\log \ell)^{1/2}$$

where $c_0 = 5/\sqrt{\epsilon p^{3/2}(1-p)}$ and let

$$k = \lfloor (2\ell/p)^{1/2} - \omega \rfloor.$$

Here k is small enough to ensure that the number of edges in $K = [k]$ is less than ℓ, but is sufficiently close that our modification process will work.

Note that

$$\omega \approx c_0(p/8)^{1/4}(k/\log k)^{1/2}$$

and

$$k \leq n - n^{\frac{1}{2}+\epsilon}. \tag{1}$$

Define s to be the unique positive integer such that

$$e(K) + p\left(\binom{s}{2} + sk\right) = \ell + t \qquad 0 \leq t \leq p(k+s).$$

This is well defined **q.s.** since $e(K)$ is distributed as $Bin\left(\binom{k}{2}, p\right)$ and

$$\mathbf{E}(e(K)) = p\binom{k}{2} = \ell - (2\ell p)^{1/2}\omega + O(\ell^{1/2}).$$

Thus, using the Chernoff bound,

$$\mathbf{Pr}((\ell p)^{1/2}\omega \leq \ell - e(K) \leq 2(\ell p)^{1/2}\omega) \geq 1 - e^{-c_2\ell/\log \ell} \tag{2}$$

for some $c_2 > 0$.

Note that $\sum_{\ell \geq \ell_0} e^{-c_2\ell/\log \ell} = o(1)$ and so this probability bound is good for all $\ell \geq \ell_0$.

If $e(K) = \ell - a\omega(\ell p)^{1/2}$ for some $1 \leq a \leq 2$ then s is well defined and

$$s \approx \frac{a\omega}{\sqrt{2}} \geq \sqrt{\frac{25k}{2^{5/2}\epsilon p(1-p)\log k}}. \tag{3}$$

Our aim now is to find vertices $v_1, v_2, ..., v_s$ so that $K_s = K \cup \{v_1, v_2, ..., v_s\}$ satisfies $e(K_s) = \ell$.

7

Define $t_i, i = 1, 2, ..., s$ by

$$e(K_i) + p\left(\binom{s-i}{2} + (s-i)(k+i)\right) = \ell + t_i.$$

We add $v_1, v_2, ..., v_s$ in three phases. In Phase 1 we reduce t_i to $O(\sqrt{k \log k})$. At the end of Phase 2 we have $-1 < t_i < 1$ and in Phase 3 we keep t_i in this range and so we have the integer $t_s = 0$ and $e(K_s) = \ell$ as required. For this purpose it will be convenient to divide $V \setminus K$ into sets W_1, W_2, W_3 each of size at least $\lfloor |V \setminus K|/3 \rfloor$.

Phase 1

The "K-degrees" $d(v, K)$ for $v \notin K$ are independently distributed as $Bin(k, p) \approx N(kp, kp(1-p))$. Hence for $v \notin K$

$$\mathbf{Pr}\left(d(v, K) \leq kp - \sqrt{\alpha p(1-p)k \log k}\right) \approx \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2}$$

where $x = \sqrt{\alpha \log k}$.

Hence, where

$$I_\epsilon = \left[kp - \sqrt{\epsilon p(1-p)k(\log k)/2}, \; kp - \sqrt{\epsilon p(1-p)(k \log k)/3}\right]$$

we find

$$p_\epsilon = \mathbf{Pr}(d(v, K) \in I_\epsilon) \geq \frac{1}{k^{\epsilon/2}} \text{ for } v \notin K.$$

Hence if $W_{1,\epsilon} = \{v \in W_1 : d(v, K) \in I_\epsilon\}$, then by (1)

$$\mathbf{E}(|W_{1,\epsilon}|) \geq \frac{1}{3} n^{\frac{1}{2}+\frac{\epsilon}{2}}$$

and so, using the Chernoff bound on the tails of $Bin(|W_1|, p_\epsilon)$ we see that

$$|W_{1,\epsilon}| \geq \frac{1}{4} n^{\frac{1}{2}+\frac{\epsilon}{2}} \quad \mathbf{q.s.} \; . \tag{4}$$

8

Now

$$\begin{aligned} t_{i+1} - t_i &= e(K_{i+1}) - e(K_i) - (k+i)p \\ &= d(v_{i+1}, K_i) - (k+i)p. \end{aligned} \qquad (5)$$

So if $v_{i+1} \in W_{1,\epsilon}$ (and $i \le s$)then

$$\begin{aligned} t_{i+1} - t_i &\le d(v_{i+1}, K) + i - (k+i)p \\ &\le -\frac{1}{2}\sqrt{\epsilon p(1-p)k(\log k)/3} \end{aligned} \qquad (6)$$

and

$$t_{i+1} - t_i \ge -\sqrt{\epsilon p(1-p)k \log k}. \qquad (7)$$

In Phase 1 we only use vertices from $W_{1,\epsilon}$. Note that (3) and (4) imply $s < |W_{1,\epsilon}|$ **q.s.** and so we have sufficient vertices to carry out this phase. It follows from (6) and (7) that after adding at most

$$\frac{k+s}{\sqrt{\epsilon p(1-p)k(\log k)/3}} \approx \sqrt{\frac{3k}{\epsilon p(1-p)\log k}} \qquad (8)$$

vertices we find

$$0 \le t_i \le \sqrt{\epsilon p(1-p)k \log k}. \qquad (9)$$

Note that the RHS of (8) is less than s (see (3)) and so (9) eventually occurs. As soon as this happens Phase 1 ends.

Phase 2

Suppose Phase 1 adds r_1 vertices. If $-1 < t_{r_1} < 1$ we go on to Phase 3, otherwise choose $a_{r_1+1}, a_{r_1+2}, ..., a_{r_2} \in [\lceil \sqrt{k} \rceil], 1 \le r_2 - r_1 \le \sqrt{2\epsilon p(1-p)\log k}$ satisfying

$$a_{r_1+1} + a_{r_1+2} + ... + a_{r_2} = \lfloor t_{r_1} \rfloor + \lceil f_{r_1+1} + f_{r_1+2} + ... + f_{r_2} \rceil,$$

9

where
$$f_j = \lceil (k+j-1)p \rceil - (k+j-1)p \qquad r_1 < j \leq r_2.$$

We divide W_2 into $r_2 - r_1$ sets $W_{2,j}, r_1 < j \leq r_2$ of almost equal size. Thus

$$|W_{2,j}| \geq n^{\frac{1}{2}+\epsilon-o(1)}, \ r_1 < j \leq r_2 \tag{10}$$

The aim now is to choose $v_j \in W_{2,j}, r_1 < j \leq r_2$,such that

$$d(v_j, K_{j-1}) = \lceil (k+j-1)p \rceil - a_j. \tag{11}$$

It follows from (5) that $-1 < t_{r_2} < 1$. Indeed

$$
\begin{aligned}
t_{r_2} - t_{r_1} &= \sum_{j=r_1+1}^{r2} (d(v_j, K_{j-1}) - (k+j-1)p) \\
&= \sum_{j=r_1+1}^{r_2} (f_j - a_j) \\
&= -\lfloor t_{r_1} \rfloor + (f_{r_1+1} + \ldots + f_{r_2}) - \lceil f_{r_1+1} + \ldots + f_{r_2} \rceil.
\end{aligned}
$$

Now when we examine $W_{2,j}$ we find that the degrees $d(v, K_{j-1})$ are independently distributed as $Bin(k+j-1, p)$. We observe that

$$\mathbf{Pr}(Bin(n,p) = np+h) \approx \frac{1}{\sqrt{2\pi pqn}} \exp\left(-\frac{h^2}{2pqn}\right) \tag{12}$$

for $h = o(n^{2/3})$ - see for example Bollobás [1]. Since $a_j \in [[\lceil \sqrt{k} \rceil]]$ we see immediately that

$$\mathbf{Pr}(d(v, K_{j-1}) = \lceil (k+j-1)p \rceil - a_j) \geq \frac{c_1}{\sqrt{n}} \tag{13}$$

for some $c_1 = c_1(p)$. It follows now from (10), (13) and the Chernoff bound that **q.s.** there are at least $n^{\epsilon-o(1)}$ choices for v_j. Thus Phase 2 can be completed **a.s.**.

Phase 3

If $r_2 = s$ then we have $t_s = 0$ because $-1 < t_s < 1$ and t_s is an integer. On the other hand suppose that $r = s - r_2 > 0$. We know from the analysis of Phase 2 that at the start of Phase 3 **q.s.** there are at least $n_1 = n^{\epsilon - o(1)}$ vertices in W_3 which have $d(v, K_{r_2}) = \lfloor (k + r_2)p \rfloor + u$, for each integer $u, |u| \le 2\sqrt{k}$.

Let $\mathcal{E}$ denote the event that for some $c_2 = c_2(p)$, throughout Phase 3 there are at least $c_2 n_1$ vertices v in $W_3 \setminus K_j$ with $d(v, K_j) = \lfloor (k + j)p \rfloor + u$, for each integer $u, |u| \le \sqrt{k}$ and for each $r_2 \le j \le s$.

Assume for the moment that $\mathcal{E}$ occurs.

Suppose first that $s - r_2$ is odd. If $t_{r_2} > 0$ let $v_{r_2 + 1}$ be any vertex v in W_3 for which $d(v, K_{r_2}) = \lfloor (k + r_2)p \rfloor$. Then

$$t_{r_2 + 1} = t_{r_2} + \lfloor (k + r_2)p \rfloor - (k + r_2)p.$$

From $0 < t_{r_2} < 1$ we deduce that $-1 < t_{r_2 + 1} < 1$. If $t_{r_2} \le 0$ we choose v with $d(v, K_{r_2}) = \lceil (k + r_2)p \rceil$.

We maintain $-1 < t_j < 1$ throughout Phase 3.

As t_s is an integer this will be sufficient to prove that $t_s = 0$. Define j_0 by $j_0 = r_2$ if $s - r_2$ is even and $j_0 = r_2 + 1$ otherwise.

Hence we can assume that $s - j_0 > 0$ and $s - j_0$ is even. We choose an integer u at random from $[\lfloor \sqrt{k} \rfloor]$ and then $v_{j+1} \in W_3 \setminus K_j$ such that

$$d(v_{j+1}, K_j) = \lfloor (k + j)p \rfloor + u \tag{14}$$

and then $v_{j+2} \in W_3 \setminus K_{j+1}$ such that

$$d(v_{j+2}, K_{j+1}) = \lfloor (k + j + 1)p \rfloor - u + \delta \tag{15}$$

11

where $\delta \in \{0, 1, 2\}$.

Then

$$t_{j+2} = t_j + (\lfloor (k+j)p \rfloor + u - (k+j)p) + (\lfloor (k+j+1)p \rfloor - u - (k+j+1)p) + \delta.$$

δ is chosen so that $-1 < t_{j+2} < 1$. To complete the proof we must justify the assumption that $\mathcal{E}$ occurs.

For $|u| \leq 2\sqrt{k}$ let S_u be any n_1-subset of $\{v \in W_3 : d(v, K_{r_2}) = \lfloor (k+r_2)p \rfloor + u\}$. Fix $\hat{u}$ and consider $|S_{\hat{u}} \cap K_j|$ for $r_2 \leq j \leq s$. If $v \in S_{\hat{u}} \setminus K_{j-1}$ and $d(v, K_{j-1}) = (k+j-1)p + \tilde{u}, \tilde{u} \geq 0$ then

$$\begin{aligned}
\mathbf{Pr}(v = v_{j+1}) &\leq& \mathbf{Pr}(u = \tilde{u}) \\
&\leq& \frac{1}{\sqrt{k}}
\end{aligned}$$

independent of K_{j-1}. Similary, if $\tilde{u} \leq 0$ then crudely,

$$\mathbf{Pr}(v = v_{j+1}) \leq \frac{6e}{\sqrt{k}}.$$

Hence

$$\begin{aligned}
\mathbf{Pr}(|S_{\hat{u}} \cap K_j| \geq \frac{n_1}{2}) &\leq& \binom{n_1}{n_1/2} j^{n_1/2} \left(\frac{3}{\sqrt{k}}\right)^{n_1/2} \\
&\leq& \left(\frac{12s}{\sqrt{k}}\right)^{n_1/2}.
\end{aligned}$$

Explanation: $\binom{n_1}{n_1/2}$ counts the possibilities for $S_{\hat{u}} \cap K_j$ and $j^{n_1/2}$ bounds the number of orderings of the ways a fixed set can be placed into K_j. Since $s = o(\sqrt{k})$ we deduce that $|S_u \setminus K_j| \geq n_1/2$ **q.s.** for all $|u| \leq \sqrt{k}$.

We can now estimate $\mathbf{Pr}(\bar{\mathcal{E}})$. A key fact to observe is that when adding v_{j+1} we do not look at the edges from v_{j+1} to $W_3 \setminus K_{j+1}$ i.e. their occurrences

remain independent. Fix $j = r_2 + a$ and $|u| \leq \sqrt{k}$. Let

$$X = \bigcup_{\hat{u}=u-\lfloor ap+\sqrt{a}\rfloor}^{u-\lfloor ap-\sqrt{a}\rfloor} \{v \in S_{\hat{u}} \setminus K_j\}.$$

Arguing as in (13) we see that if $v \in X$ then

$$\mathbf{Pr}(d(v, K_j) = (k+j)p + u) \geq \frac{c_3}{\sqrt{a}}$$

for some $c_3 = c_3(p)$. The degrees $d(v, K_j)$ are independent of each other and the events $\{|S_u \setminus K_j| \geq n_1/2\}$ since they depend on different edges. Since $|X| \geq n_1\sqrt{a}$ **q.s.** we see from the Chernoff bound that

$$|\{v \in X : d(v, K_j) = (k+j)p + u\}| \geq c_3 n_1/2 \quad \textbf{q.s..}$$

Thus $\mathcal{E}$ occurs **q.s.** on taking $c_2 = c_3/2$. This ends the analysis of Case 2.

Case 3: $\ell_1 < \ell \leq e(G) - \frac{An^{3/2}}{\sqrt{\log n}}$

where $A = 10\sqrt{1/(\epsilon p(1-p))}$.

Let $m = p\binom{n}{2} - \ell$. Note that **q.s.** $|e(G) - p\binom{n}{2}| \leq n \log n$ so that

$$\frac{An^{3/2}}{\sqrt{\log n}} - n \log n \leq m \leq n^{3/2+\epsilon} + n \log n.$$

Let $\hat{s}$ be defined by

$$m = p\left(\binom{\hat{s}}{2} + \hat{s}(n - \hat{s})\right) + \sigma$$

where $0 \leq \sigma < n$.

Clearly $\hat{s} \approx \frac{m}{np} \geq \frac{A}{2p}\sqrt{\frac{n}{\log n}}$. Next let $s_0 = \hat{s} - \lceil B\sqrt{n/\log n}\rceil$ where $B = A/3p$ and $W = [s_0]$ and define s by

$$e(G - W) - p\left(\binom{s}{2} + s(n - s_0 - s)\right) = \ell + t_0$$

13

where $0 \leq t_0 < n - s - s_0$.

Now **q.s.**

$$|e(G) - e(G - W) - p\left(\binom{s_0}{2} + s_0(n - s_0)\right)| \leq \sqrt{s_0 n} \log n$$

and so

$$
\begin{aligned}
p\left(\binom{s}{2} + s(n - s - s_0)\right) &= e(G - W) - \ell - t_0 \\
&= \left(e(G) - p\binom{n}{2}\right) + m + (e(G - W) - e(G)) - t_0 \\
&\leq m + n \log n - p\left(\binom{s_0}{2} + s_0(n - s_0)\right) + \sqrt{s_0 n} \log n \\
&= (1 + o(1))(\hat{s} - s_0)np.
\end{aligned}
$$

Consequently

$$s \approx B\sqrt{\frac{n}{\log n}} \qquad \textbf{q.s.}.$$

Our aim now is to find s vertices $v_1, v_2, ..., v_s$ in $G \setminus W$ such that, where $W_i = W \cup \{v_1, v_2, ..., v_i\}$ and $K_i = [n] \setminus W_i$

$$e(K_s) = \ell. \tag{16}$$

To this end we will define t_i by

$$e(K_i) - p\left(\binom{s - i}{2} + (s - i)(n - s_0 - s)\right) = \ell + t_i$$

and aim to achieve $t_s = 0$.

Corresponding to (5) we have

$$t_{i+1} - t_i = -d(v_{i+1}, K_i) + p(n - s_0 - i). \tag{17}$$

14

Out strategy now is to examine the vertices in $K = [n] \backslash W$ in order $1, 2, ..., n - s_0$ until we have found s suitable vertices. Note that the edges of K are unconditioned at present (this explains why we take $m = p\binom{n}{2} - \ell$ and not the more natural $e(G) - \ell$.) We proceed in three phases as for Case 2. The analysis however is slightly different.

Phase 1

Let now

$$I_\epsilon = [(n - s_0 - 1)p + \sqrt{\epsilon p(1-p)n(\log n)/3}, (n - s_0 - 1)p + \sqrt{\epsilon p(1-p)(n \log n)/2}].$$

As previously stated, we examine the vertices of K in order, accepting a vertex v as the next v_j if $d(v, K) \in I_\epsilon$. We stop after r_1 vertices have been accepted where r_1 is now the smallest index such that $0 \leq t_{r_1} \leq \sqrt{\epsilon p(1-p)n \log n}$.

Note that $d(v, K) \geq d(v, K_i) \geq d(v, K) - i$. It follows that for $i \geq 0$

$$-\sqrt{\epsilon p(1-p)n(\log n)/2} - ip \leq t_{i+1} - t_i \leq -\sqrt{\epsilon p(1-p)n(\log n)/3} + i(1-p).$$

Hence this phase ends after deleting at most $\sqrt{4n/(\epsilon p(1-p)\log n)}$ $(< s/2)$ vertices, provided that we can find a suitable number with degrees in I_ϵ.

We show next that **q.s.** there are at least $n^{1/2}$ such vertices in $[n_2]$ where $n_2 = \lfloor 2n^{1/2 + \epsilon/2} \rfloor$.

The proof of this involves the important observation that having examined vertices $1, 2, ..., k - 1$ the edges between vertices $i, j \geq k$ are still unconditioned. Let us observe next that **q.s.**

$$d(k, [k-1]) \in [kp - \omega \sqrt{k \log n}, kp + \omega \sqrt{k \log n}] \tag{18}$$

15

for any $\omega = \omega(n) \to \infty$.

This is because $d(k, [k-1])$ is a binomial random variable. So suppose now that $k \leq n_2$. Assume $d(k, [k-1]) = kp + \delta$ where $|\delta| \leq \omega\sqrt{k \log n} \leq n^{1/4+\epsilon}$. Then $d(v, K) \in I_\epsilon$ iff $d(k, [k+1, n-s_0]) \in J$ where J is the interval

$$[(n-s_0-k)p-\delta+\sqrt{\epsilon p(1-p)n(\log n)/3}, (n-s_0-k)p-\delta+\sqrt{\epsilon p(1-p)n(\log n)/2}]$$

and since $\delta = o(\sqrt{n \log n})$, the probability of this happening is at least $n^{-\epsilon/2}$ and this event is independent of previous similar events, assuming only that (18) holds. Thus the number of vertices of the required degree dominates $Bin(n_2, n^{-\epsilon/2})$ and there are enough **q.s.**.

We will refer to this as "the forward looking argument".

Phase 2

The aim of Phase 2 is as before to reduce t_i so that at its end $-1 < t_i < 1$. If t_r satisfies this then we go straight to Phase 3. Otherwise let $b_{r_1+1}, ..., b_{r_2}$ be a sequence of positive integers satisfying

$$1 \leq b_j \leq \lceil \sqrt{n} \rceil, \quad r_1 < j \leq r_2 \leq \sqrt{2\epsilon p(1-p) \log n}$$

and

$$b_{r_1+1} + ... + b_{r_2} = \lfloor t_{r_1} \rfloor - \lceil f_{r_1+1} + ... + f_{r_2} \rceil$$

where

$$f_j = \lceil p(n - s_0 - j) \rceil - p(n - s_0 - j).$$

We continue our examination of the remaining vertices and at the j'th step of Phase 2 we look for a v_j where

$$d(v_j, K_{j-1}) = \lceil p(n - s_0 - (r_1 + j - 1)) \rceil + b_j.$$

16

If we can find $v_1, v_2, ..., v_{r_2}$ then Phase 2 ends with $-1 < t_{r_2} < 1$ as before.

Using a forward looking argument we see that independently, the next vertex has probability at least $c_4/\sqrt{n}$ of being selected as a v_j. Thus **q.s.** at most n_2 vertices need be examined before Phase 2 comes to an end.

Phase 3

We start Phase 3 with $-1 < t_{r_2} < 1$ and **q.s.** we have only examined at most $2n_2$ vertices. We must now see precisely how we delete $s - r_2$ more vertices and obtain $t_s = 0$.

If $s - r_2$ is odd we proceed similarly to Case 2 but using the forward looking argument.

In general, in analogy to (14), (13), we search forward until we have found a pair v_{j+1}, v_{j+2} such that

$$d(v_{j+1}, K_j) = \lceil p(n - s_0 - j) \rceil + u$$

$$d(v_{j+2}, K_{j+1}) = \lceil p(n - s_0 - j - 1) \rceil - u + \delta$$

where $u \in [\sqrt{n}]$ and $\delta = 0, 1,$ or 2.

The proof of this case is complete if we show that **q.s.** we can find pairs after examining at most $2\sqrt{n}$ vertices per pair.

First of all suppose we have examined $\sqrt{n}$ vertices in our search. Using the forward looking argument we see that **q.s.** we will have found at least $c_5\sqrt{n}$ distinct values for $\alpha_r = d(v, K_j) - \lceil p(n - s_0 - j) \rceil$ in $[-\lfloor\sqrt{n}\rfloor, \lfloor\sqrt{n}\rfloor]$ (the probability that α_r is in this range is bounded below by a constant and the probability that any particular value occurs twice is $O(1/\sqrt{n})$.) At this stage the probability that the next vertex can be coupled with one of these $c_5\sqrt{n}$

17

is bounded by a constant, again by a forward looking argument. Thus **q.s.** $O(\log n)$ further examinations are needed to find a pair.

There is one final point to consider before claiming that the proof for Case 3 is done with. The forward looking argument is valid at vertex k only if

$$\omega\sqrt{k\log n} = o(\sqrt{n\log n}). \tag{19}$$

Now $s = O(\sqrt{n/\log n})$ and the above analysis shows that **q.s.** we only examine $k \leq 2s\sqrt{n} + 2n_2$ and (19) holds provided $\omega = o(\sqrt{\log n})$.

This completes the proof of part (a) of the theorem.

To prove part (b) we let

$$m_k = e(G) - e(G_{n-k})$$

where G_{n-k} is the subgraph of G induced by $[n-k]$. Then we let

$\bar{m}_k = $ the maximum of $e(G) - e(H)$ over all induced subgraphs H with $n-k$ vertices,

$\underline{m}_k = $ the minimum of $e(G) - e(H)$ over all induced subgraphs H with $n-k$ vertices.

Now m_k is distributed as $Bin\left(\binom{k}{2} + k(n-k)), p\right)$ which has mean $\mu_k \approx knp$ for $k = o(n)$. Hence on using the Chernoff bound with $k = o(n)$ and $\epsilon np/(3\mu_k) \approx 1/(3k)$ we see that

$$\mathbf{Pr}(\bar{m}_k \geq \mu_k + \frac{1}{3}np) \leq 2\binom{n}{k}\exp\left\{-\frac{1}{3}\left(\frac{1}{4k}\right)^2 knp\right\} \tag{20}$$

and

$$\mathbf{Pr}(\underline{m}_{k+1} \leq \mu_{k+1} - \frac{1}{3}np) \leq 2\binom{n}{k}\exp\left\{-\frac{1}{3}\left(\frac{1}{4k}\right)^2 knp\right\}. \tag{21}$$

Now if $k = \sqrt{np/(50 \log n)}$ then the RHS's of (20) and (21) both tend to zero super-polynomially fast. Thus **q.s.** G contains no induced subgraph containing ℓ edges when $\ell \in J = [e(G) - \mu_{k+1} + \frac{1}{3}np, e(G) - \mu_k - \frac{1}{3}np]$. Now $|J| \geq \frac{1}{4}np$ and part (b) follows. $\qquad \square$

References

[1] B. Bollobás, *Random Graphs*, Academic Press, London, 1985.

[2] D.E.Knuth, R.Mothwani and B.Pittel, *Stable husbands* Random Structures and Algorithms 1, (1990) 1-14.