ON THE NUMBER OF HAMILTON CYCLES IN A RANDOM GRAPH

by

C. Cooper
Department of Computing, Mathematics and Statistics
Polytechnic of North London
Holloway Road
London N78DB
England

and

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

Research Report No. 88-27

August 1988

On the number of hamilton cycles in a random graph

by

C. Cooper

Department of Computing, Mathematics and Statistics
Polytechnic of North London
Holloway Road
London N78DB
England

and

A.M. Frieze

Department of Mathematics
Carnegie Mellon University
Pittsburgh, PA 15213
USA

August 1988

Abstract

Let a random graph G be constructed by adding random edges one by one, starting with n isolated vertices. We show that with probability going to one as n goes to infinity, when G first has minimum degree two, it has at least $(\log n)^{(1-\epsilon)n}$ distinct hamilton cycles for any fixed $\epsilon > 0$.

§1. Introduction

Let $V_n = \{1,2,\ldots,n\}$ and consider the random graph process (Bollobás [3]) $G_0, G_1, \ldots, G_\nu$, $\nu = \binom{n}{2}$ where $G_m = (V_n, E_m)$, $E_0 = \phi$ and E_{m+1} is obtained from E_m by adding an edge e_{m+1} chosen randomly from $[n]^{(2)} - E_m$. Now let

$$m^* = \min\{m: \delta(G_m) \geq 2\}.$$

Bollobás [2] (see also Ajtai, Komlos and Szemerédi [1]) showed that

$$\lim_{n\to\infty} \Pr(G_{m^*} \text{ is hamiltonian}) = 1$$

which was claimed but not proved by Kòmlós and Szemerédi [7] when they first established the exact threshold for the existence of hamilton cycles in a random graph.

Knowing that G_{m^*} usually has at least one hamilton cycle raises the question of how many distinct hamilton cycles does it usually contain. We prove

Theorem

If $\epsilon > 0$ is fixed then

$$\lim_{n\to\infty} \Pr(G_{m^*} \text{ has at least } (\log n)^{(1-\epsilon)n} \text{ distinct hamilton cycles}) = 1.$$

□

Thus at m^* the number of hamilton cycles jumps dramatically from 0 to at least $(\log n)^{n-o(n)}$. On the other hand the expected number of hamilton cycles at this point is $n!p^n = (\log n)^n e^{-n+o(n)}$ and so the theorem gives the right order of magnitude for the number of hamilton cycles in G_{m^*}.

§2. Notation and preliminaries

We say that almost every (a.e.) graph process satisfies a certain property if this property holds with probability tending to 1 as n tends to ∞. Let $m_1 = \lfloor \frac{1}{2} n(\log n + \log\log n - \log\log\log n) \rfloor$ and $m_2 = \lceil \frac{1}{2} n(\log n + \log\log n - \log\log\log n) \rceil$. It follows from Erdös and Renyi [4] that $m_1 \le m^* \le m_2$ in a.e. graph process.

In what follows our inequalities need only be true for large enough n. It is always useful to bear in mind the relationship between G_m and G_p, $p = m/\nu$, $\nu = \binom{n}{2}$, the random graph in which each possible edge appears independently with probability p. Let E_p denote the edge set of G_p.

The properties we need are (see [2]): suppose $\mathcal{A}$ is some property of graphs then

(2.1a) $\quad \Pr(G_m \in \mathcal{A}) \le 3\sqrt{n} \log n \Pr(G_p \in \mathcal{A}) \qquad m_1 \le m \le m_2$

(2.1b) $\quad$ a.e. $G_p \in \mathcal{A}$ and $\mathcal{A}$ is monotone implies a.e. $G_m \in \mathcal{A}$.

(2.1c) $\quad$ a.e. $G_p \in \mathcal{A}$ implies there exists m', $m - \sqrt{n} \log n \le m' \le m$ such that a.e. $G_{m'} \in \mathcal{A}$.

Now let $\epsilon > 0$ be fixed and small from now on and $V_n^+ = V_n - V_{n_\epsilon}$ where $n_\epsilon = \lfloor (1-\epsilon)n/2 \rfloor$.

$$L_m = \{ v \in V_n : d_m(v) \leq \log n/10\}$$

where $d_m(v)$ is the degree of v in G_m and

$$L_m^+ = \{v \in V_n : d_m^+(v) \leq \log n/10\}$$

where $d_m^+(v)$ is the number of neighbours of v in V_n^+.

For $S \subseteq V_n$ let

$$N_m(S) = \{w \in V_n - S : \exists\, v \in S \text{ such that } vw \in E_m\}$$

and let $N_p(S)$ be defined similarly.

For $S, T \subseteq V_n$, $S \cap T = \phi$, $e_m(S,T) = |\{vw \in E_m : v \in S, w \in T\}|$.

Let $NL = L_m \cup L_m^+ \cup (N_m (L_m \cup L_m^+) \cap V_{n_\epsilon})$.

We now describe the basic properties of G_m, $m_1 \leq m \leq m_2$ which are needed for the paper.

<u>Lemma 2.1</u>

Almost every graph process is such that <u>simultaneously</u> for all $m_1 \leq m \leq m_2$, G_m satisfies

(2.2a) $\Delta(G_m) \leq 3 \log n$. (maximum degree)

(2.2b) $|L_m| \leq n^{2/5}$, $|L_m^+| \leq n^{4/5}$.

(2.2c) No pair of vertices $v,w \in L_m$ are within distance 4 of each other.

(2.2d) No pair of vertices $v, w \in V_n$ have 3 or more common neighbours

(2.2e) $T \subseteq V_n$, $|T| \leq \frac{n}{(\log n)^2}$ implies that T contains at most $3|T|$ edges.

(2.2f) * * $S \subseteq V_n L_m L_m$, $|S| < ^$ implies $\backslash ^(S) \backslash \geq ^ |s|$.

(2.2g) ^ $S \subseteq V_{n-L}$; , $|S|_ < ^$ implies INJS) fl V^L > ^ |S|

(2.2h) $S, T \subseteq V_n$, $S \cap T = *$, $|s| = |T| = [\frac{n}{(\log\log n)}]$ implies $e_m(S,T) \geq \frac{n \log n}{2(\log\log n)^6}$.

(2.2i) V_n^+ contains at least $\frac{1}{3}$ n logn edges.

<u>Proof</u> (Outline: details of similar results can be found in [2])

Let $p_1 = m_1/N$, $p_2 = m_2/N$.

<u>Proof of (2.2a)</u>

$$Pr(A(G_{p_2}) > 3 \log n) \leq n \sum_{k>3\log n} (V^oU-\overset{k}{P}p)^{11"1"1'} = o(1).$$

Hence (2.1b) implies $Pr(A(G_{m_2}) > 3 \log n) = o(1)$ and then the result follows from $A(G_m) \leq A(G_{m_2})$.

Proof of (2.2b)

$$E(|L_{p_1}|) = n \sum_{k\le\frac{1}{10}\log n} \binom{n-1}{k} p_1^k(1-p_1)^{n-1-k}$$

$$= O(n^{0.34}).$$

Now use the Markov inequality and proceed as in the proof of (2.2a). The proof of the upper bound for $|L_m^+|$ is similar.

Proof of (2.2c)

$$\Pr((2.2c) \text{ fails in } G_{p_1}) \le n^5 p_1^4 \Big(\sum_{k\le\frac{1}{10}\log n} \binom{n-1}{k} p_1^k(1-p_1)^{n-1-k}\Big)^2$$

$$= o(1)$$

Let now m' be as in (2.1c). then

$$\Pr((2.2c) \text{ fails from some } G_m,\ m' \le m \le m_2 \mid (2.2a) - (2.2c) \text{ holds in } G_{m'})$$

$$\le \Pr(\exists\ e = uv \in E_{m_2} - E_{m'} \text{ such that } \mathrm{dist}(u,L_{m'}),\ \mathrm{dist}(v,L_{m'}) \le 3 \text{ in } G_{m'} \mid$$
$$(2.2a) - (2.2c) \text{ holds in } G_{m'}))$$

$$= O(n \log\log\log n (n^{2/5}(\log n)^3)^2/\nu) \qquad [\nu = \tbinom{n}{2}$$

$$= o(1).$$

Proof of (2.2d)

$$\Pr(G_p \text{ has 2 vertices with 3 or more common neighbours}) \le \binom{n}{2}\binom{n-2}{3}p_2^6$$

$$\le (\log n)^6/n.$$

We can now use (2.1b) to 'extend' this to G_{m_2}. But if (2.2f) holds for G_{m_2}, it must also hold for $m \le m_2$.

Proof of (2.2e)

Fix m and $p = \frac{m}{\nu}$. Then

$$\Pr((2.2e) \text{ fails in } G_p) \le \sum_{k=8}^{n/(\log n)^2} \binom{n}{k} \begin{bmatrix} \binom{k}{2} \\ 3k+1 \end{bmatrix} p^{3k+1}$$

$$= O(n^{-16}).$$

Hence, by (2.1a),

$$\Pr(\exists m,\ m_1 \le m \le m_2 \text{ such that (2.2e) fails in } G_m) = o(1).$$

Proof of (2.2f)

Now if (2.2e) holds then this on its own implies

$$|N_m(S)| \ge \frac{\log n}{60}|S| \quad \text{for } S \subseteq V_n - L_m,\ |S| \le \frac{n}{(\log n)^4}.$$

For larger S, we drop the condition $S \cap L_m = \phi$.

Suppose $S \subseteq V_n$, $|S| \leq \frac{n}{\log n}$. If $v \in V_n - S$ then $\Pr(v \in N_p(S)) = 1 - (1-p)^{|S|} \geq \frac{|S|p}{2}$. Hence

$$\Pr(\exists\, S \subseteq V_n : \frac{n}{(\log n)^4} \leq S \leq \frac{n}{\log n} \text{ and } |N_p(S)| \leq \frac{\log n}{60}|S|)$$

$$\leq \sum_{s=\frac{n}{(\log n)^4}}^{\frac{n}{\log n}} \binom{n}{s} \Pr(B(n-s, \frac{sp}{2}) \leq \frac{s \log n}{60})$$

$$\leq \sum_{s \geq \frac{n}{(\log n)^4}} (\frac{ne}{s})^s e^{-\alpha nps} \qquad \text{for some constant } \alpha > 0$$

$$= o(n^{-2}).$$

Proof of (2.2g)

Similar to that of (2.2f).

Proof of (2.2h)

Let $s = \lceil \frac{n}{(\log\log n)^3} \rceil$. Now $e_p(S,T)$ is distributed as the binomial random variable $B(s^2,p)$. But

$$\Pr(B(s^2,p) \leq \frac{1}{2}s^2p) \leq e^{-\frac{1}{8}s^2p}.$$

Hence

$$\Pr((2.2h) \text{ fails in } G_p) \le \binom{n}{s}^2 e^{-\frac{1}{8}s^2p}$$

$$= o(n^{-2})$$

and the result follows in the usual manner.

Proof of (2.2i)

The number of edges of G_p which are contained in V_n^+ dominates $B(\frac{1}{8}n^2,p)$. □

Now let $\mathcal{G}_m = \{G_m: (2.2) \text{ holds and } \delta(G_m) \ge 2\}$.

§3. Proof of the theorem

We now describe a way of choosing a large set $\mathcal{H}$ of subgraphs of $G_m \in \mathcal{G}_m$, most of which are hamiltonian and such that if C,C' are hamilton cycles of distinct $H,H' \in \mathcal{H}$ then $C \ne C'$.

Let $A_m = V_{n_\epsilon} - NL$, $B_m = V_n^+ - NL$ and for $v \in A_m$ let

$W(v) = \{vw \in E_m: w \in B_m\}$.

Let $L_0 = \lceil \log n/10 \rceil$ and r be a prime satisfying $(\log\log n)^2 \le r \le 2(\log\log n)^2$, let $k = \lfloor \log_r L_0 \rfloor$ and $L = r^k$. We treat $\{1,2,\dots,L\}$ as the points of the k-dimensional vector space over the field with r elements, GF_r. This space has $K = r^{k-1}(r^k-1)/(r-1)$ lines. Let the point sets for these lines be the r-subsets $X_1,X_2,\dots,X_K$ of L. The only property of these sets used is $|X_i \cap X_j| \le 1$ for $i \ne j$.

For each $v \in A_m$ we choose a random L-subset $W'(v) \subseteq W(v)$ plus a random ordering $w_1,w_2,\dots,w_L$ (of $W'(v)$). We then define r-subsets $W(v,k) \subseteq W'(v)$, $k = 1,2,\dots,K$ by letting $W(v,k) = \{w_{i_1},w_{i_2},\dots,w_{i_r}\}$ when

$X_k = \{i_1, i_2, \dots, i_r\}$.

Now let $\Phi = \{f: A_m \to \{1,2,\dots,K\}\}$. For each $f \in \Phi$ we will define a subgraph H_f of G_m as follows: delete from G_m all edges incident with A_m other than $\bigcup_{v \in A_m} W(v,f(v))$. Let now $\mathcal{H} = \{H_f: f \in \Phi\}$. Observe

$$(3.1) \qquad |\Phi| \geq K^{(n_\epsilon - n^{4/5})}$$

$$= (\log n)^{(1-\epsilon-o(1))n}$$

(3.2) If C_f, C_g are hamilton cycles of H_f, H_g, $f \neq g$ then $C_f \neq C_g$.

For if $f(v) \neq g(v)$ then C_f uses 2 edges of $W(v,f(v))$ and C_g can use at most one edge of $W(v,f(v))$.

Now let $Z_m = |\{f \in \Phi: H_f \text{ is not hamiltonian}\}|$. We prove

$$(3.3) \qquad E(Z_m | G \in \mathcal{G}_m) \leq |\Phi|/n^3$$

and so

$$\Pr(Z_m \geq \frac{|\Phi|}{n} | G \in \mathcal{G}_m) = O(n^{-2}).$$

Thus

$$(3.4) \qquad \Pr(G_m \text{ has fewer than } (1 - \tfrac{1}{n}) (\log n)^{(1-\epsilon-o(1))n} \text{ hamilton cycles } | G_m \in \mathcal{G}_m) = O(n^{-2}).$$

The theorem follows immediately from (3.4).

We must now show that most H_f are hamiltonian.

Consider now a fixed $f \in \Phi$. To prove (3.3) we show

$$(3.5) \qquad Pr(H_f \text{ is not hamiltonian} | G \in \mathcal{G}_m) = O(n^{-3}).$$

First of all consider the distribution of the edges in the sets $W(v,f(v))$.

Lemma 3.1

Conditional on the sub-graph induced by $V_n - A_m$, the sets $W(v,f(v))$ are an independent collection of random r-subsets of B_m.

Proof

Consider a fixed G_m, $v \in A_m$ and $W(v) = N_m(v) \cap B_m$. (We cannot assume $G_m \in \mathcal{G}_m$ here.) Replacing $W(v)$ by another subset of B_m of the same size does not change A_m or NL. We use here the fact that $w \in B_m$ has at least $\log n/10$ neighbours in V_n^+ and so changing the neighbours of $v \in A_m$ cannot place w in NL. It follows that the sets $W(v)$ are independent random subsets and the lemma follows as the $W(v,f(v))$ are random subsets of these.

□

Let now $X \subseteq E_m$ and $H_{f,X} = H_f - X$. We say that X is deletable if

$$(3.6a) \qquad |X^+| = n \quad \text{where} \quad X^+ = \{e \in X : e \subseteq V_n^+\} ,$$

$$(3.6b) \qquad |X \cap W(v,f(v))| = 3 \quad \text{for} \quad v \in A_m,$$

(3.6c) X is not incident with any vertex in

$$\hat{L}_m = \{v \in V_n : d_m(v) < 10 + 2EL + 2 \frac{\log n}{\log\log n}\}$$

(3.6d) If $v \in B_m$ and $d^+(v) = [\log n/10] + k$ then v is incident with at most $k-1$ edges in X.

(3.6e) No $v \in B_m$ is incident with $\frac{2}{7}\frac{\log n}{\log\log n}$ or more edges in X^+.

(3.6f) $M(H_f) = X(H_{f,X})$ where X denotes the length of the longest path in the appropriate graph.

Observe that a calculation similar to that given for (2.2b) shows that $|\hat{L}_m| \le n^{2/5}$ in a.e. G_m. We now incorporate this condition into the definition of $\mathcal{S}_m$.

Our next lemma deals with the number of neighbours of subsets of A_m.

For $S \subseteq V_n$ and subgraph H of G_m let $N_H(S) = \{w \notin S: vw \in E(H)$ for some $v \in S\}$.

<u>Lemma 3.2</u>

The following hold with probability $1 - o(n^{-})$. Here let $H = H_f$.

(i) $S \subseteq A_m$, $1 \le |S| \le \frac{n}{10}$ implies $|N_H(S)| \ge 80\,|S|$,

(ii) $S \subseteq A$, $T \subseteq B$, $|S| = |T| = \lceil \frac{n}{\log\log n} \rceil$ implies that H contains at least $n \log\log n$ edges joining S and T.

(iii) $T \subseteq B_m$, $|T| \le \frac{n}{\log n}$ implies $|N_H(T) \cap A| < 3r\,|T|$.

Proof

(i)

We first consider $|S| \leq n/3r$ and show $|N_H(S)| \geq r|S|/2$ with the required probability.

$$\Pr(\exists S:\ |S| \leq n/3r \text{ and } |N_H(S)| \leq r|S|/2) \leq \sum_{s=1}^{\frac{n}{3r}} \binom{n_\epsilon}{s} \binom{n-n_\epsilon}{rs/2} \left[\frac{\binom{rs/2}{r}}{\binom{n-n_\epsilon}{r}}\right]^s$$

$$\leq \sum_{s=1}^{\frac{n}{3r}} \left(\frac{n_\epsilon e}{s} \left(\frac{2(n-n_\epsilon)e}{rs}\right)^{r/2} \left(\frac{rs}{2(n-n_\epsilon)}\right)^r\right)^s$$

$$\leq \sum_{s=1}^{\frac{n}{3r}} \left(\frac{ne}{s} \left(\frac{ers}{2(n-n_\epsilon)}\right)^{r/2}\right)^s$$

$$= o(n^{-3}).$$

Suppose now $n/3r < |S| \leq n/600$. Let $S' \subseteq S$ be of size $\lfloor n/3r \rfloor$. Then

$$|N_H(S)| \geq |N_H(S')|$$

$$\geq r\lfloor n/3r \rfloor/2$$

$$\geq n/7$$

$$\geq 80\ |S|.$$

(ii)

Consider the selection of the sets $W(v,f(v))$ for $v \in S$. This involves rs ($s = |S|$) choices of elements in B_m and each choice always has probability at least $\frac{s - r+1}{n - n_\epsilon}$ of being in T. Thus the number of choices, and hence edges in question, stochastically dominates the binomial $B(rs, \frac{s - r+1}{n - n_\epsilon})$. Hence

$$\Pr((iii)\text{ fails}) \le \binom{n}{s}^2 \Pr(B(rs, \frac{s - r+1}{n - n_\epsilon}) \le n \log\log n)$$

and the result follows from the Chernoff bound (see for example [3]) for the tails of the binomial since $E(B(rs, \frac{s - r+1}{n - n_\epsilon})) \approx \frac{2rs^2}{n(1+\epsilon)} \ge \frac{2n \log\log n}{1+\epsilon}$.

(iii)

Fix $T \subseteq B_m$, $\frac{n}{r \log n} \le |T| = t \le \frac{n}{6r}$ and $S \subseteq A_m$, $|S| = 3r|T|$. Now if $\hat{n} = |B_m|$ then

$$\Pr(W(v,f(v)) \cap T \ne \phi \text{ for all } v \in S) = \left(1 - \frac{\binom{\hat{n}-t}{r}}{\binom{\hat{n}}{r}}\right)^{3rt}$$

$$\le \left(1 - \left(1 - \frac{t}{\hat{n}-r}\right)^r\right)^{3rt}$$

$$\le \left(\frac{2rt}{n}\right)^{3rt}.$$

Hence

$$\Pr((\text{iii})\text{ fails}) \leq \sum_{t=n/(r\log n)}^{n/6r} \binom{\hat{n}}{t} \begin{bmatrix} \frac{1}{2}n \\ 3rt \end{bmatrix} \left(\frac{2rt}{n}\right)^{3rt}$$

$$\leq \sum_{t=n/r\log n}^{n/6r} \left(\frac{ne}{t}\right)^{t} \left(\frac{e}{3}\right)^{3rt}$$

$$= o(n^{-3}).$$ □

Let $\mathcal{E}_f$ be the event denoting the occurrence of the conditions in the above lemma.

Lemma 3.3

Suppose $G_m \in \mathcal{G}_m$, $f \in \Phi$, $\mathcal{E}_f$ occurs, X is deletable and $H = H_{f,X}$. Then

(i) $S \subseteq V_n$, $|N_H(S)| < 2|S|$ implies

(a) $|S| \geq \frac{n}{600}$

(b) $|(S \cup N_H(S)) \cap (B_m)| \geq \frac{n}{2} + \frac{\epsilon n}{3}$.

(ii) H is connected.

Proof

Suppose $S \subseteq V_n$. Let $S_0 = S \cap L_m$, $S_1 = S \cap (L_m^+ - L_m)$, $S_2 = S \cap A_m$ and $S_3 = S - (S_0 \cup S_1 \cup S_2)$.

Assume first that $|S_3| \leq \frac{n}{\log n}$ and $|S_2| \leq \frac{n}{600}$.

Case 1: $|S_2| \leq |S_1 \cup S_3|$.

(a) $|S - S_2| < 2|NL|$.

Let S^* be the larger and $\hat{S}$ the smaller of S_1, S_3. Then

$$|N_H(S)| \geq |N_m(S_0)| + |N_m(S^*)| - \frac{2\ \mathrm{logn}}{\mathrm{loglogn}} |S^*| - |S_2 \cup \hat{S}|$$

$$- |N_m(S^*) \cap (S_0 \cup N_m(S_0))|$$

$$\geq 2|S_0| + (\frac{\mathrm{logn}}{60} - \frac{2\ \mathrm{logn}}{\mathrm{loglogn}})|S^*| - 3|S^*| - |S^*|$$

$$\geq 2|S|,$$

(after using (2.2c), (2.2f), (2.2g) and (3.6e) to obtain the second inequality).

(b) $|S - S_2| \geq 2|NL|$.

$$|N_H(S)| \geq |N_H(S_3)| - |NL \cup S_2|$$

$$\geq (\frac{\mathrm{logn}}{60} - \frac{2\ \mathrm{logn}}{\mathrm{loglogn}})\ |S_3| - |NL| - |S_2|$$

$$\geq 2|S|,$$

(using $S_0 \cup S_1 \subseteq NL$ and $|S_2| \leq |S_3| + |NL|$).

<u>Case 2</u>: $|S_2| > |S_1 \cup S_3|$.

$$|N_H(S)| \geq 80|S_2| - 3|S_2| + 2|S_0| - |S_1 \cup S_3| \geq 2|S|.$$

Suppose now that $|S_2| \leq \frac{n}{600}$ and $\frac{n}{\mathrm{logn}} \leq |S_3| \leq \frac{n}{600}$. Choose $S_3' \subseteq S_3$ of size $\lfloor \frac{n}{\mathrm{logn}} \rfloor$ and let $S' = (S-S_3) \cup S_3'$. Then

$$|N_H(S)| \geq |N_H(S')| - |S_3 - S_3'|$$

$$\geq 2|S_0| + 22|S_2| + \frac{\log n}{200}(|S_1| + |S_3'|) - |S_3 - S_3'|$$

$$\geq 2|S_0| + 22|S_2| + \frac{\log n}{200}|S_1| + \frac{n}{200} - \frac{\log n}{200} - |S_3| + \left\lfloor\frac{n}{\log n}\right\rfloor$$

$$\geq 2|S|.$$

We have thus proved (i), part (a).

For part (b), we know, from part (a), that $|S| \geq \frac{n}{600}$ and hence $|S_2 \cup S_3| \geq \frac{n}{700}$.

Assume first that $|S_3| \geq \frac{n}{1400}$. Suppose $|(S_3 \cup N_H(S_3)) \cap B_m| < \frac{1}{2}n + \frac{\epsilon n}{3}$. Then there exists $T \subseteq B_m$ of size at least $\frac{\epsilon n}{7}$ such that $N_H(S_3) \cap T = \phi$. Now it follows from (2.2h) that G_m contains at least $\frac{n \log n}{2(\log\log n)^6}$ edges joining S_3 and T. But X contains at most n edges joining S_3 and T and so $N_H(S_3) \cap T \neq \phi$ – contradiction.

Assume next that $|S_2| \geq \frac{n}{1400}$. The proof here is similar to that above, but relying on Lemma 3.2(ii) in place of (2.2h), and the fact that X contains only 3 edges incident with each $v \in A_m$.

(ii)

Suppose H is not connected and there exists $S \subseteq V_n$, $|S| \leq \frac{1}{2}n$ such that there are no S to $V_n - S$ edges in H. Now $|(V_n-S) \cap (B_m)| \geq \frac{\epsilon n}{3}$ and (i) implies $|S| \geq \frac{n}{600}$. We obtain a contradiction using (2.2h) or Lemma 3.2(ii) as in (i)(b). □

Suppose now that H_f is not hamiltonian and X is deletable. Let $P = (x_0, x_1, \ldots, x_\lambda)$ be a longest path of both H_f and $H = H_{f,X}$. If

$x_i x_\lambda \in E(H_f)$, $i \neq 0$, then the associated <u>rotation</u> with x_0 <u>fixed</u> and <u>broken edge</u> $x_i x_{i+1}$ yields a new longest path $\rho(P,x_0,x_i) = (x_0,x_1,\ldots,x_i,x_\lambda,x_{\lambda-1},\ldots,x_{i+1})$.

Let $END(P,x_0)$ denote the set of other endpoints of longest paths which are obtainable in H from P by a sequence of rotations, with x_0 fixed, and starting from P.

We will restrict our allowable rotations to those where the broken edge is an edge of the starting path P. We further restrict ourselves so that if P' is obtained from P by a sequence of rotations through paths $P = P_0,P_1,\ldots,P_k = P'$ then the paths $P_1,P_2,\ldots,P_k$ have distinct endpoints, other than x_0.

Suppose that the paths produced in the construction of $END(P,x_0)$ are $\mathcal{P} = \{P^0,P^1,P^2,\ldots\}$ where $P^0 = P$ and P^{i+1} is obtained from some P^j, $j \leq i$, by a single rotation.

Let $END = END(P,x_0) \cup \{x_0\}$ and for each $x \in END$ let P_x denote the first path (in the above ordering) with endpoint x (so that $P_{x_0} = P$). For $x \neq x_0$ let $END(x) = END(P_x,x)$. Now a simple modification of the argument of Posa [6] shows that

$$|N_H(END(x))| < 2|END(x)|.$$

(Indeed, all we have to show is that if $v \in N_H(END)$ with neighbours w_1,w_2 on P then $\{w_1,w_2\} \cap END \neq \phi$. Suppose $w' \in END$ and $vw' \in E(H)$. Consider the neighbours w_1',w_2' of v on $P_{w'}$. If $\{w_1',w_2'\} = \{w_1,w_2\}$ then some allowable rotation from $P_{w'}$ shows one of w_1,w_2 is in END. If say $w_1 \notin \{w_1',w_2'\}$ then the sequence of rotations that created $P_{w'}$ deleted the edge vw_1 and so $w_1 \in END$.)

We deduce from Lemma 3.3 that

(3.7a) $|END(x)| > ^$ for $x \in END$

(3.7b) $|END| \geq ggy$

(3.7c) Each P_x, $x \in END$, contains at least $\frac{2}{*3} zr$ en edges with both endpoints in B_m.

To see (3.7c) let n_i, i = 0,1,2 denote the number of edges of P_x with i vertices in B_m. Then $i^ - n_Q$ 1 (|V(P_x) fl BJ - |V(P_x) fl (V_n U NL)|) - 1.

Since P_x is a longest path, it must contain N_H(END(x)). But then Lemma 3.3 implies |(END(x) U N_{fl}(END(x))) D (B_m) |_> | n + ^- and so $n_2 - n_Q \geq$ | n + $\frac{en}{3}$ - (|n - ^- + o(n)) - 1 and (3.7c) follows.

Given (3.7) we consider two possibilities.

Case l: there exists $x \in END$ such that |END(x) fl B_m | I $\frac{S}{1200}$.

Case 2: |END(x) 0 B I < *rr^r* for all $x \in END$.

Case 1 is easier to deal with and is considered first. Without loss of generality assume |END fl B_m| _> TSTTPT i.e. $x = x_n$ suffices above. Observe that because H_i is connected,

(3.8) $x \in END$, $y \in END(x)$ implies $xy \in E(H_f)$.

(We use the "colouring" argument of Fenner and Frieze [5] to show this is unlikely when a large number of $x \in B_m$. Since A_m contains no edges in H_{r_i}

(3.8) does not help so much in Case 2 and we are in a similar situation to that encountered in the case of random bipartite graphs, Frieze [6]).

Suppose now that given $G_m \in \mathcal{G}_m$, we randomly pick $X \subseteq E_m$ satisfying (3.6a), (3.6b). We consider two events:

$\mathcal{E}_1 = \mathcal{E}_f \cap \{G_m \in \mathcal{G}_m,\ H_f$ is not hamiltonian, Case 1 occurs$\}$

$\mathcal{E}_2 = \mathcal{E}_1 \cap \{X$ is deletable$\}$.

We show

(3.9a) $$\Pr(\mathcal{E}_2|\mathcal{E}_1) \geq \frac{1}{2}(1 - \frac{2}{r})^{n_\epsilon} (1 - \frac{20}{\log n})^n$$

(3.9b) $$\Pr(\mathcal{E}_2) \leq c_1^n \quad \text{for some constant } 0 \leq c_1 < 1.$$

We can then deduce

(3.10) $$\Pr(\mathcal{E}_1) \leq (c_1 + o(1))^n.$$

<u>Proof of (3.9a)</u>

Fix $G \in \mathcal{G}_m$ and the choices $W(v,f(v))$ for $v \in A_m$. Fix some longest path P of H_f. Consider first the edges of X that meet A_m. Each $W(v,f(v))$ contains at most 2 edges of P. This accounts for the term $(1 - \frac{2}{r})^{n_\epsilon}$. Now consider the remaining n edges of X. Now to avoid P and the edges incident with NL, X must avoid at most $n + o(n)$ edges, given (2.2a), (2.2b). Using this and (2.2i) we obtain $(1 - \frac{20}{\log n})^n$ as a lower bound for the probability of avoiding these edges. Given that these edges are not selected, the probability that (3.6d) or (3.6e) fails is $o(1)$, which accounts for the $\frac{1}{2}$.

Proof of (3.9b)

Consider fixed graphs $\hat{G}$, $\hat{H}$. We show

$$(3.11) \qquad \Pr(\mathcal{E}_2 | G_m - X = \hat{G},\ H_{f,X} = \hat{H}) \le c_1^n$$

and (3.9b) follows.

Observe that $G_m - X$, $H_{f,X}$ together determine A_m by $v \in A_m$ iff $v \le n_\epsilon$ and it loses edges in $H_{f,X}$. NL is then determined by $v \in NL$ iff $v \notin A_m$ and $d^+(v) \le \frac{\log n}{10}$ or $v \in V_{n_\epsilon}$ and v is the neighbour of such a vertex.

If $\Pr(\mathcal{E}_2 | G_m - X = \hat{G},\ H_{f,X} = \hat{H}) > 0$ then there exists X such that $\mathcal{E}_2$ occurs for $\hat{G} + X$, $\hat{H} + X$. Hence we may assume that (3.7) holds where END, END(x), $x \in$ END are <u>determined by $\hat{H}$ only</u> (and are independent of X). We may also assume Case 1 occurs in $\hat{H}$.

Furthermore the edges in X are required to conform to (3.8). Thus let $\hat{\mathcal{E}}_2$ denote the event $\{x \in \text{END},\ y \in \text{END}(x) \text{ implies } xy \notin X\}$. Then

$$(3.12) \quad \Pr(\mathcal{E}_2 | G_m - X = \hat{G},\ H_{f,X} = \hat{H}) \le \Pr(\hat{\mathcal{E}}_2 | G_m - X = \hat{G},\ H_{f,X} = \hat{H},\ (3.6c), (3.6d)).$$

(For (3.12) use $\Pr(A|BC) \ge \Pr(AB|C)$ for events A,B,C).

Let us now consider the distribution of X given $G_m - X$, $H_{f,X}$ and (3.6c), (3.6d). Let $X = X^+ \cup (\bigcup_{v \in A_m} Y_v)$, where for $v \in A_m$, $Y_v = \{vw \in X\}$. We claim that

(3.13a) X^+ is a random n-subset of $B_m^{(2)} - E(\hat{G})$,

(3.13b) For $v \in A_m$, Y_v is a random 3-subset of $\{vw \notin E(\hat{G}): w \in B_m\}$ and these subsets are independent of each other.

(3.13a) follows from the fact that given (3.6c), (3.6d) holds for one X, the addition (and subsequent deletion) of any n-subset of $B_m^{(2)} - E(\hat{G})$ does not affect $H_{f,X}$ and (3.6c), (3.6d) will still hold. (3.13b) follows from Lemma 3.1 and its proof.

Now for $w \in END \cap B_m$ let $\beta(w) = |END(w) \cap B_m|$. The following 2 subcases cover all possibilities:

Case 1a: $|\{w: \beta(w) > \frac{n}{1200}\}| \geq \frac{n}{2400}$

Case 1b: $|\{w: \beta(w) < \frac{n}{1200}\}| \geq \frac{n}{2400}$.

It follows from (3.13a) that, where $\nu^+ = \binom{n-n_\epsilon}{2}$ and $\hat{m} \leq m$,

$$\Pr(\hat{\mathcal{E}}_2|\text{Case 1a}) \leq \binom{\nu^+ - \hat{m} - 3n^2/(2(2400)^2)}{n} \Big/ \binom{\nu^+ - \hat{m}}{n}$$

$$\leq (\frac{95999}{96000})^n.$$

It follows from (3.13b) that

$$\Pr(\hat{\mathcal{E}}_2|\text{Case 1b}) \leq (1 - \frac{3}{2400})^{n/1200}.$$

We have thus confirmed (3.9b).

Let us now consider Case 2. Let $\mathcal{E}_1$ be as before, except that Case 2 replaces Case 1 and let $\mathcal{E}_2$ now be defined with respect to the new $\mathcal{E}_1$. (3.9a) continues to hold. We prove

(3.9b') $\Pr(\mathcal{E}_2 | G_m - X = \hat{G},\ H_{f,X} = \hat{H}) \le c_2^n$ for some constant $0 < c_2 < \ = c_2(\epsilon) < 1$

which combined with (3.9a) yields

$$\Pr(\mathcal{E}_1) \le (c_2 + o(1))^n. \tag{3.10'}$$

From (3.10) and (3.10') and the fact that $\Pr(\mathcal{E}_f | G \in \mathcal{G}_m) = 1 - o(n^{-3})$ we obtain (3.3) and the theorem.

We observe that (3.13) continues to hold. We can assume that $\hat{H}$ contains a longest path P with endpoints x_0, x_1 and $\frac{n}{1200}$ vertices $\mathrm{END} \subseteq A_m$ and for each $x \in \mathrm{END}$ there is a set of $\frac{n}{600}$ paths $\mathcal{P}_x$ with distinct endpoints (END(x)). These will have been constructed from a path P_x by rotations as in the discussion prior to (3.7).

We now consider in more detail the construction of $\mathrm{END}(P,x_0)$. Let $T = T(x_0)$ denote the tree with vertex set $\mathrm{END}(P,x_0)$, rooted at x_1 and with an edge directed from x to y if P_y is obtained by a single rotation from P. Let $\mathcal{T}$ be the set of possible trees that can be so constructed.

Consider the following condition:

$\mathcal{A}$: there exists $T \in \mathcal{T}$ such that T contains a subtree T', rooted at x_1, which has (i) $|V(T') \cap A_m| \ge \frac{n}{1200}$ and (ii) $|V(T') \cap B_m| \le \frac{n}{4800r}$.

Suppose now that $\mathcal{A}$ holds. For each $v \in \mathrm{END}' = V(T') \cap A_m$ let $\phi(v)$ denote the neighbour of v on P_v.

Lemma 3.4

If $\mathscr{A}$ holds then $|\phi(\text{END}')| \geq \frac{n}{9600}$.

Proof

We show first

$$(3.14) \qquad y \in \phi(\text{END}') - V(T') \text{ implies } |\phi^{-1}(y)| \leq 2.$$

We do this by showing that if $y = \phi(x)$ then xy is an edge of P. This is clearly true if $x = x_1$. If $x \neq x_1$ then y is adjacent to x on P_x. If xy is not an edge of P then y is an ancestor of x in T', a contradiction, as $y \notin V(T')$.

Now (3.14) implies that

$$(3.15) \qquad |\phi(\text{END}')| \geq \frac{1}{2}|\text{END}' - \phi^{-1}(\phi(\text{END}') \cap V(T'))|.$$

But since $\phi^{-1}(\phi(\text{END}') \cap V(T')) \subseteq N_{\hat{H}}(B_m) \cap A_m$ we see from Lemma 3.3 and $\mathscr{A}$(ii) that

$$|\phi^{-1}(\phi(\text{END}') \cap V(T'))| \leq \frac{n}{4800r} \cdot 3r$$

and the lemma follows from this and (3.15). □

It is important to note that any path obtained from P_x, $x \in \text{END}'$ by a sequence of rotations with x fixed has $\phi(x)$ as x's neighbour.

Suppose now that $\mathscr{A}$ does not hold. We will obtain a contradiction. Let $T \in \mathscr{T}$. Since $|V(T) \cap A_m| \geq \frac{n}{1200}$ we must have $|V(T) \cap B_m| > \frac{n}{4800r}$. Then T

contains a subtree $\hat{T}$ with $|V(\hat{T}) \cap B_m| = \frac{n}{?}$ and since ** does not hold $|V(\hat{T}) \cap A_m| < \frac{n}{?}$. Let $S = V(\hat{T}) \cap B_m$. It follows from (2.2h) that $|N_H(S) \cap B_m| \geq \frac{n}{?}$. Now if $v \in S$, $w \in N_H(S) \cap B_m$ and $vw \in E(\hat{H})$ then we can legitimately construct $\rho(P_v, x_0, w)$ unless the associated broken edge $ww' \in E(P)$. But this latter condition rules out at most $2|V(T)|$ rotations: (2 for each added edge of each P_v, $v \in V(T)$). The same w' can be produced at most twice in this way. Thus there exists $T \in J$ which contains a subtree which is obtained from $\hat{T}$ by adding at least $\frac{1}{2}\{ ? - 2(\frac{n}{1200} + |N_L|)\} \geq \frac{n}{?}$ leaves. Since $\mathcal{A}$ does not occur, at least $\frac{n}{?} - \frac{n}{1?00} > \frac{n}{?}$ of these new leaves are in B_m. But this means Case 1 holds, a contradiction.

Applying this argument for each $x \in END$ i.e. constructing a tree $T(x)$ of paths starting with P_x, we deduce, from Lemma 3.4 that the following is true:

Lemma 3.5

In H there are $\frac{n}{9600}$ vertices $y_1, y_2, \ldots$ in $END \cap A_m$ and a set of $\frac{n}{9600}$ vertices $z_1, z_2, \ldots$ in B_m such that for each i there are $\frac{n}{1200}$ longest paths with one endpoint y_i, z_i adjacent to y_i on each path and the other endpoints of each set of $\frac{n}{1200}$ paths are distinct members of A_m. □

Let Y_i, $i = 1,2,\ldots,\frac{n}{9600}$ denote the set of other endpoints of the paths with one fixed endpoint y_i.

We can now confirm (3.9b'). We must add random edges, as in (3.13), and show that with high probability these extra edges make the resulting graph hamiltonian or have a longer path than H.

We consider the edges in (3.13b) to be added randomly in 3 waves $X_1, X_2,$

$K_j \cup X^+$ where $|X_j| = l^\wedge l = l^\wedge l = |A_m|$ and each $v \in A_m$ is incident with one edge of each X_t, $t = 1,2,3$.

Adding X_1

For $y \in Y = \bigcup_i Y_i$ let $\delta(y) = |\{i: y \in Y_i\}|$. Clearly $|Y'| \geq \frac{n}{30}$ where $Y' = \{y \in Y: \delta(y) \geq \frac{n}{8(1200)^2}\}$.

If $y \in Y'$ then independently of other members of Y'

$$Pr(\text{for some } i, X_1 \text{ contains an edge } yz_i \text{ where } y \in Y_i) \geq \frac{1}{4(1200)^2}.$$

Hence there exist constants $0 < f_1, \eta_1 < 1$ such that

$$Pr(\mathcal{S}_3) \geq 1 - \eta_1$$

where

$$\mathcal{S}_3 = \{X_1 \text{ contains } f_1 n \text{ edges of the form } z_i y, y \in Y_i\}.$$

Assume now that $\mathcal{S}_3$ occurs.

We now have $f_1 n$ cycles $C_1, C_2, \ldots$ say, plus an edge joining y_i to C_i. Applying (3.7c) we see that each C_i contains a set of vertices K_i, $|K_i| \geq \frac{\varepsilon}{3} \varepsilon n$, where $v \in K_i$ implies v lies on an edge of C_i with both endpoints in B_m.

Adding X_2

Now, independently, for each i, $Pr(X_2$ contains an edge $y_i u$ where

$u \in K_i) \geq \epsilon$. By considering these cycles one by one, we see that there exist constants $0 < \xi_2 = \xi_2(\epsilon)$, $\eta_2 = \eta_2(\epsilon) < 1$ such that

$$\Pr(\mathcal{E}_4 | \mathcal{E}_3) > 1 - \eta_2^n$$

where

$\mathcal{E}_4 = \{X_2$ contains $\xi_2 n$ edges of the form $y_i u_i$, $u_i \in K_i$ and the B_m neighbours $v_1, v_2, \ldots$ of $u_1, u_2, \ldots$ on $C_1, C_2, \ldots$ are distinct$\}$.

Now each time X_2 contains an edge $y_i u_i$, $u_i \in K_i$, we can obtain a longest path of $\hat{H} + (X_1 \cup X_2)$ with one endpoint y_i and the other endpoint in B_m by using the edges $(C_i \cup \{y_i u_i\}) - \{u_i v_i\}$.

Assume that $\mathcal{E}_4$ occurs.

<u>Adding $X_3 \cup X^+$</u>

We now have $\xi_3 n$ longest paths $Q_1, Q_2, \ldots$ of $\hat{H} + (X_1 \cup X_2)$, each with a distinct endpoint $v_i \in B_m$. We are now essentially in a Case 1 situation. Take each Q_i and using v_i as a fixed endpoint generate $\geq \frac{n}{600}$ longest paths by rotations. Now throw in $X_3 \cup X^+$. The probability that we fail to close one of these paths is exponentially small. (3.9b′) follows and we are done.

<u>References</u>

[1] M. Ajtai, J. Komlós and E. Szemerédi, "First occurrence of hamilton cycles in random graphs", *Annals of Discrete Mathematics* **27** (1985) 173-178.

[2] B. Bollobás, "The evolution of sparse graphs", in *Graph Theory and Combinatorics, Proceedings of Cambridge Combinatorial Conference in honour of Paul Erdös* (B. Bollobás Ed.), Academic Press (1984) 35-57.

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

[4] P. Erdös and A. Rényi, "On the strength of connectedness of a random graph", *Acta Math. Acad. Sci. Hungar.* **12** (1961) 261-267.

[5] T.I. Fenner and A.M. Frieze, "On the existence of hamilton cycles in a class of random graphs", *Discrete Mathematics* **45** (1983) 301-305.

[6] A.M. Frieze, "Limit distribution for the existence of hamilton cycles in a random bipartite graph", *European Journal on Combinatorics* **6** (1985) 327-334.

[7] J. Komlós and E. Szemerédi, "Limit distribution for the existence of hamilton cycles in random graphs", *Discrete Mathematics* **43** (1983) 55-63.