How many random edges make a dense graph Hamiltonian?

Tom Bohman* Alan Frieze† Ryan Martin‡

Department of Mathematical Sciences,
Carnegie Mellon University
Pittsburgh PA 15213.

February 19, 2002

Abstract

This paper investigates the number of random edges required to add to an arbitrary dense graph in order to make the resulting graph Hamiltonian with high probability. Adding $\Theta(n)$ random edges is both necessary and sufficient to ensure this for all such dense graphs. If, however, the original graph contains no large independent set, then many fewer random edges are required.

1 Introduction

In the classical model of a random graph (Erdős and Rényi [3]) we add random edges to an empty graph, all at once or one at a time and then ask for the probability that certain structures occur. At the present time, this model and its variants, have generated a vast number of research papers and at least two excellent books, Bollobás [1] and Janson, Łuczak and Ruciński [5]. It is also of interest to consider random graphs generated in other ways. For example there is a well established theory of considering random subgraphs of special graphs, such as the n-cube. In this paper we take a slightly different line. We start with a graph H chosen *arbitrarily* from some class of graphs and then consider adding a random set of edges R. We then ask if the random graph $G = H + R$ has a certain property. This for example would model graphs which were basically deterministically produced, but for which there is some uncertainty about the complete structure. In any case, we feel that there is the opportunity here for asking interesting and natural questions.

As an example we consider the following scenario: Let $0 < d < 1$ be a fixed positive constant. We let $\mathcal{G}(n, d)$ denote the set of graphs with vertex set $[n]$ which have minimum degree $\delta \geq dn$. We choose H arbitrarily from $\mathcal{G}(n, d)$ and add a random

*Supported in part by a grant from the NSA
†Supported in part by NSF grant CCR-9818411.
‡Supported in part by NSF VIGRE grant DMS-9819950.

set of m edges R to create the random graph G. We prove two theorems about the number of edges needed to have G Hamiltonian **whp**. Since $d \geq 1/2$ implies that H itself is Hamiltonian (Dirac's Theorem), this could be considered to be a probabilistic generalisation of this theorem to the case where $d < 1/2$. It also means that we will assume $d \leq 1/2$ from here on.

Theorem 1. *Suppose $0 < d < 1/2$ is constant, $H \in \mathcal{G}(n,d)$ and let $\theta = \log d^{-1} \geq .69$. Let $G = H + R$ where $|R| = m$ is chosen randomly from $\overline{E} = n^{(2)} \setminus E(H)$.*

(a) *If $m \geq 100\theta n$ then G is Hamiltonian* **whp**.

(b) *For $d \leq 1/10$ there exist graphs $H \in \mathcal{G}(n,d)$ such that if $m < \theta n/3$ then* **whp**[1] *G is not Hamiltonian.*

So it seems that we have to add $\Theta(n)$ random edges in order to make G Hamiltonian **whp**. Since a random member of $\mathcal{G}(n,d)$ is already likely to be Hamiltonian, this is a little disappointing. Why should we need so many edges in the worst-case? It turns out that this is due to the existence of a large independent set. Let $\alpha = \alpha(H)$ be the independence number of H.

Theorem 2. *Suppose $H \in \mathcal{G}(n,d)$ and $1 \leq \alpha < d^2 n/2$ and so $d > n^{-1/2}$ (d need not be constant in this theorem). Let $G = H + R$ where $|R| = m$ is chosen randomly from $\overline{E}$. If*

$$\frac{md^3}{\log d^{-1}} \to \infty \tag{1}$$

then G is Hamiltonian **whp**.

Note that if d is constant then Theorem 2 implies that $m \to \infty$ is sufficient. Theorem 1 is proven in the next sections and Theorem 2 is proven in Section 3.

2 The worst-case

We will follow a well-trodden route, using Posá's lemma [6] and the colouring argument of Fenner and Frieze [4].

Observe first that if we randomly delete edges with probability 3/4 then **whp** we will have a graph H' with

$$\delta(H') \geq \frac{dn}{5} \text{ and } \Delta(H') \leq \frac{n}{3}. \tag{2}$$

Since our property is monotone, let us assume that H itself satisfies (2). We will advise the reader later on when we make use of this technical assumption.

[1] A sequence of events $\mathcal{E}_n$ is said to occur "with high probability" (**whp**) if $\lim_{n\to\infty} \mathbf{Pr}(\mathcal{E}_n) = 1$

We will assume from now on that m is exactly $\lceil 100\theta n \rceil$.

We first show that

Lemma 1. *G is connected* **whp**.

Proof Let $N = \binom{n}{2}$. If $u, v \in [n]$ then either they are at distance one or two in H or

$$\mathbf{Pr}(dist_G(u,v) > 3) \leq \left(1 - \frac{m}{N}\right)^{d^2n^2/25} \leq e^{-4\theta d^2 n}.$$

Hence,

$$\mathbf{Pr}(diam(G) > 3) \leq n^2 e^{-4\theta d^2 n} = o(1).$$

□

Given a longest path P in a graph Γ with end-vertices x_0, y and an edge yv where v is an internal vertex of P, we obtain a new longest path $P' = x_0..vy..w$ where w is the neighbour of v on P between v and y. We say that P' is obtained from P by a *rotation* with x_0 *fixed.*

Let $END_\Gamma(x_0, P)$ be the set of end-vertices of longest paths of Γ which can be obtained from P by a sequence of rotations keeping x_0 as a fixed end-vertex. For each $y \in END_\Gamma(x_0, P)$ let $END_\Gamma(y, P)$ be the set of end-vertices of longest paths of Γ which can be obtained from P by a sequence of rotations keeping y as a fixed end-vertex. Let $END_\Gamma(P) = \{x_0\} \cup END_\Gamma(x_0, P)$. Note that if Γ is connected and non-Hamiltonian then there is no edge (x, y) where $x \in END_\Gamma(P)$ and $y \in END_\Gamma(x, P)$.

It follows from Posá [6] that

$$|N_\Gamma(END_\Gamma(P))| < 2|END_\Gamma(P)|, \tag{3}$$

where for a graph Γ and a set $S \subseteq V(\Gamma)$

$$N_\Gamma(S) = \{w \notin S : \exists v \in S \text{ such that } vw \in E(\Gamma)\}.$$

Lemma 2. Whp

$$|N_G(S)| \geq 3|S| \tag{4}$$

for all $S \subseteq [n]$, $|S| \leq n/5$.

Proof Clearly, $|N_H(S)| \geq 3|S|$ for all $S \subseteq [n]$, $|S| \leq dn/20$. So,

$$\begin{aligned}
\mathbf{Pr}(\exists |S| \leq n/5 : \ |N_G(S)| < 3|S|) &\leq \sum_{k=dn/20}^{n/5} \binom{n}{k}\binom{n}{3k}\left(1 - \frac{m}{N}\right)^{k(n-4k)} \\
&\leq \sum_{k=dn/20}^{n/5} \left(\frac{n^4e^4}{27k^4} e^{-40\theta}\right)^k \\
&= o(1).
\end{aligned}$$

$\square$

Let G_R be the graph induced by the edges R. Let $d_R(v)$ be the number of edges in R which are incident with v and let $\Delta_R = \max_v d_R(v)$. We show next that

$$\Delta_R \leq \ln n \qquad \textbf{whp}. \tag{5}$$

Let $\lambda = \ln n$. Then, using (2),

$$\begin{aligned}
\mathbf{Pr}(\Delta_R \geq \lambda) &\leq n\binom{m}{\lambda}\left(\frac{n}{n^2/6-m}\right)^{\lambda} \\
&\leq n\left(\frac{601e\theta}{\lambda}\right)^{\lambda} \\
&= o(1).
\end{aligned}$$

Let

$$\mathcal{R} = \{R \subseteq [n]^{(2)} \setminus E(H) : |R| = m, G \text{ is connected and satisfies } (4),(5)\}.$$

Now let $\omega = \log n$. We say $X \subseteq R$ is *deletable* if

D1 $|X| = \omega$.

D2 X is a matching.

D3 X avoids some longest path P of G.

D4 There does not exist $(x,y) \in X$ such that $x \in END_{x_0,G-X}(P)$ and $y \in END_{G-X}(y,P)$, where x_0 is an endpoint of P.

Finally let

$$\alpha(R,X) = \begin{cases} 1 & G \text{ is non-Hamiltonian}, R \in \mathcal{R} \text{ and } X \text{ is deletable.} \\ 0 & \text{Otherwise.} \end{cases}$$

We will prove the following two inequalities: Let $N_1 = |\overline{E}|$.

$R \in \mathcal{R}$ and G is non-Hamiltonian implies

$$\sum_{X \subseteq R} \alpha(R,X) \geq \left(1 - \frac{1}{99\theta}\right)^{\omega}\binom{m}{\omega}. \tag{6}$$

$R_1 \subseteq \overline{E}$, $|R_1| = m - \omega$ implies

$$\sum_{X \cap R_1 = \emptyset} \alpha(R_1 \cup X, X) \leq \binom{N_1 - m + \omega}{\omega}\left(\frac{49}{50}\right)^{\omega}. \tag{7}$$

Before verifying (6), (7) we see how Theorem 1 follows. Let $B_{\mathcal{R}}$ be the number of choices of $R \in \mathcal{R}$ for which G is not Hamiltonian. Then (6) implies

$$B_{\mathcal{R}} \leq \left(1 - \frac{1}{99\theta}\right)^{-\omega} \binom{m}{\omega}^{-1} \sum_{R \in \mathcal{R}} \sum_{X \subseteq R} \alpha(R, X).$$

But (7) implies

$$\sum_{R \in \mathcal{R}} \sum_{X \subseteq R} \alpha(R, X) \leq \binom{N_1}{m - \omega} \binom{N_1 - m + \omega}{\omega} \left(\frac{49}{50}\right)^{\omega}.$$

So,

$$\begin{aligned} B_{\mathcal{R}} &\leq \left(\frac{49}{50} \cdot \left(1 - \frac{1}{99\theta}\right)^{-1}\right)^{\omega} \binom{N_1}{m - \omega} \binom{N_1 - m + \omega}{\omega} \binom{m}{\omega}^{-1} \\ &= \left(\frac{49}{50} \cdot \left(1 - \frac{1}{99\theta}\right)^{-1}\right)^{\omega} \binom{N_1}{m}. \end{aligned}$$

Thus,

$$\mathbf{Pr}(G \text{ is Hamiltonian}) = o(1) + \frac{B_{\mathcal{R}}}{\binom{N_1}{m}} = o(1).$$

Proof of (6)

Fix $R \in \mathcal{R}$, G non-Hamiltonian and let P be some longest path of G. If we choose X to satisfy D1,D2,D3 then D4 is automatically satisfied since G is connected. Thus the number of choices for X satisfying $\alpha(R, X) = 1$ is at least

$$(m - n)(m - n - 2\ln n) \cdots (m - n - 2(\omega - 1)\ln n)/\omega! \geq \left(1 - \frac{1}{99\theta}\right)^{\omega} \binom{m}{\omega}.$$

$\square$

Proof of (7)

Fix $R_1 \subseteq \overline{E}$, $|R_1| = m - \omega$. If there exists X such that $\alpha(R_1 \cup X, X) = 1$ then (4) and X being a matching implies that $\Gamma = H + R_1$ satisfies $|N_\Gamma(S)| \geq 2|S|$ for all $S \subseteq [n], |S| \leq n/5$. It follows from (3) that to choose X such that D4 holds, we have to avoid choosing any of a set of at least $\binom{n/5}{2}$ edges. This implies (7) and completes the proof of (a).

Remark 1. *The calculations above go through quite happily for $\delta(H) \geq n^{3/4}$, say. For this degree bound the number of additional edges required in the worst-case is $\Omega(n \log n)$. But now we realise that it only requires $\frac{1}{2} n \log n$ edges starting with the empty graph and there is no point in considering smaller values of d, unless we can improve the constant factor 100.*

(b) Let $m = cn$ for some constant c and let H be the complete bipartite graph $K_{A,B}$ where $|A| = dn$ and $|B| = (1-d)n$. Let I be the set of vertices of B which are not incident with an edge in R. If $|I|>|A|$ then G is not Hamiltonian. Instead of choosing m random edges for R, we choose each possible edge independently with probability $p=\frac{2m}{(d^2+(1-d)^2)n^2}$. (We can use monotonicity, see for example Bollobás II.1 to justify this simplification). Then

$$\mathbf{E}(|I|)=(1-d)n(1-p)^{(1-d)n-1}\sim(1-d)\exp\left\{-\frac{2(1-d)m}{(d^2+(1-d)^2)n}\right\}n.$$

We can use the Chebychef inequality to show that $|I|$ is concentrated around its mean and so G will be non-Hamiltonian **whp** if c satisfies

$$c<\frac{1}{2(1-d)}(d^2+(1-d)^2)\ln(d^{-1}-1).$$

This verifies (b). □

3 Small independence number

Proof of Theorem 2
We will first show that we can decompose H into a few large cycles.

Lemma 3. *Suppose that G has minimum degree dn where $d\leq 1/2$ and that $\alpha(G)<d^2n/2$. Let $k_0=\lfloor\frac{2}{d}\rfloor$. Then the vertices of G can be partitioned into $\leq k_0$ vertex disjoint cycles.*

Proof Let C_1 be the largest cycle in H. $|C_1|\geq dn+1$ and we now show that the graph $H\setminus C_1$ has minimum degree $\geq dn-\alpha$.

To see this, let $C_1=v_1,\dots,v_c,v_{c+1}=v_1$. Let $w\in V(H\setminus C_1)$. Because C_1 is maximum sized, no such w is adjacent to both v_i and v_{i+1}. Also, if $w\sim v_i$ and $w\sim v_j$ with $i<j$ and $v_{i-1}\sim v_{j-1}$, then

$$w,v_j,\dots,v_c,v_1,\dots,v_{i-1},v_{j-1},\dots,v_i,w$$

is a larger cycle. So the predecessors of $N(w)$ in C_1 must form an independent set and $|N(w)\cap C_1|\leq\alpha$. Similar arguments are to be found in [2].

We can repeat the above argument to create disjoint cycles $C_1,\dots,C_r$ where $|C_1|\geq|C_2|\geq\cdots\geq|C_r|$ and C_j is a maximum sized cycle in the graph $H_{j-1}=H\setminus(C_1\cup\cdots\cup C_{j-1})$ for $j=1,2,\dots,r$. Now H_k has minimum degree at least $dn-k\alpha$ and at most $n-dn-1-(dn-\alpha+1)-\cdots-(dn-(k-1)\alpha+1)=n-k(dn+1-(k-1)\alpha/2)$ vertices. Since $d^2n>2\alpha$, H_{k_0}, if it existed, would have minimum degree at least 2 and a negative number of vertices. □

It will simplify the analysis if the edges of R are chosen from $\overline{E}$ by including each $e\in\overline{E}$ independently with probability $\frac{m}{|\overline{E}|}$. Because Hamiltonicity is a monotone property, showing that G is Hamiltonian **whp** in this model implies the theorem.

It further simplifies things if we consider $R = R_1 \cup R_2 \cup \cdots \cup R_r$ where each edge set R_i is independently chosen by including $e \in \overline{E}$ with probability p, where $1-(1-p)^r = \frac{m}{|\overline{E}|}$. Each R_i will be used to either extend a path or close a cycle and will only be used for one such attempt. In this way each such attempt is independent of the previous. To this end let $G_t = H \cup \bigcup_{s=1}^{t} R_t$ for $t = 0, 1, \ldots, r$. Thus $G_0 = H$ and $G_r = G$.

Let $e = \{x, y\}$ be an edge of C_r and let Q be the path $C_r - e$. In the procedure below we will have a current path P with endpoints x, y together with a collection of vertex disjoint cycles $A_1, A_2, \ldots, A_s$ which cover V. Initially $P = Q$, $s = r-1$ and $A_i = C_i$, $i = 1, 2, \ldots, r-1$. Also, we will have constructed G_{t-1}, so that initially $t = 1$.

Now consider the set $Z = END_{G_{t-1}}(x, P)$ created from rotations with x as a fixed endpoint, as in Section 2. We identify the following possibilities:

Case 1: There exists $z_1 \in Z$, $z_2 \notin P$ such that $f = (z_1, z_2)$ is an edge of H.
Let Q be the corresponding path with endpoints x, z_1 which goes through $V(P)$. Now suppose that $z_2 \in C_i$ and let $f' = (z_2, z_3)$ be an edge of C_i incident with z_2. Now replace P by the path Q, f, Q' where $Q' = C_i - f$. This construction reduces the number of cycles by one.

Case 2: $|V(P)| \leq n/2$ and $z \in Z$ implies that $N_{G_{t-1}}(z) \subseteq V(P)$.
It follows from (3) that $|Z| \geq dn/3$. Now add the next set R_t of random edges. Since $|V(P)| \leq n/2$, the probability that no edge in R_t joins $z_1 \in Z$ to $z_2 \in V \setminus V(P)$ is at most $(1-p)^{(dn/3)(n/2)}$. If there is no such edge, we fail, otherwise we can use (z_1, z_2) to proceed as in Case 1. We also replace t by $t+1$.

Case 3: $|V(P)| > n/2$ and $z \in Z$ implies that $N_{G_{t-1}}(z) \subseteq V(P)$.
Now we close P to a cycle. For each $z \in Z$ let $A_z = END_{G_{t-1}}(z, Q_z)$ where Q_z is as defined in Case 1. Each A_z is of size at least $dn/3$. Add in the next set R_t of random edges. The probability that R_t contains no edge of the form (z, z') where $z \in Z$ and $z' \in A_z$ is at most $(1-p)^{d^2n^2/10}$. If there is no such edge, we fail. Otherwise, we have constructed a cycle C through the set $V(P)$ in the graph G_t. If C is Hamiltonian we stop. Otherwise, we choose a remaining cycle C', distinct from C and replace P by $C' - e$ where e is any edge of C'. Now $|V(P)| < n/2$ and we can proceed to Case 1 or Case 2.

After at most r executions of each of the above three cases, we either fail or produce a Hamilton cycle. The probability of failure is bounded by

$$\begin{aligned}
k_0((1-p)^{(dn/3)(n/2)} + (1-p)^{d^2n^2/10}) &\leq 2d^{-1}\left(\left(1-\frac{m}{\overline{\overline{E}}}\right)^{\frac{dn^2}{6r}} + \left(1-\frac{m}{\overline{\overline{E}}}\right)^{\frac{d^2n^2}{10r}}\right) \\
&\leq 4d^{-1}e^{-md^3/10} \\
&= o(1)
\end{aligned}$$

provided (1) holds. $\square$

An observation: We do not actually need the condition that $\alpha(H) \leq d^2n/2$ to complete this proof. The weaker condition that $d^2n/2$ bounds the independence number of the neighborhood of each vertex is enough.

Acknowledgement Wethank the referees for a careful reading which found several small errors.

References

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

[2] V. Chvátal and P. Erdős, A note on Hamiltonian circuits, *Discrete Math* 2 (1972) 111-113.

[3] P. Erdős and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* 5 (1960) 17-61.

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

[5] S. Janson, T. Łuczak, and A. Ruciński, Random Graphs, Wiley - Interscience Series, New York, 2000.

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