Finding Hamilton Cycles in Sparse Random Graphs

A. M. FRIEZE

Department of Computer Science and Statistics, Queen Mary College,
London E1 4NS, England

Communicated by the Managing Editors

Received January 28, 1986

We describe a polynomial time (\(O(n \log n)\)) algorithm which has a high probability of finding hamilton cycles in two classes of random graphs which have constant average degree: the m-out model and the random regular graph model.

We also show how the algorithm can be used to find a large cycle in a sparse random graph. © 1986 Academic Press, Inc.

INTRODUCTION

The past few years have seen some important progress with respect to the problem of the existence of hamilton cycles in random graphs. The paper of Komlós and Szemerédi [15] gave the exact threshold for the existence of hamilton cycles in the random graph \(G(n, p)\), tightening the result of Pósa [16]. Bollobás, Frenner and Frieze [7] described a polynomial time algorithm HAM for finding hamilton cycles which gives a constructive proof of the result in [15]. This improved the previous results of Aigner and Valliant [2] and Shamir [18].

In [13] we studied random travelling salesman problems and gave modifications to HAM which enabled us to prove that it has a high probability of success on graphs with considerably fewer edges than needed for the threshold [15] provided the minimum degree is high enough. In this paper we continue this development for three classes of random graph with constant average degree.

We first consider a variation on the class of graphs studied by Frenner and Frieze [9]. Let \(v \in V_r = \{1, 2, \ldots, n\}\) independently make \(m\) random choices \(c(i) \in V_r, i = 1, 2, \ldots, m\). These choices are not necessarily distinct. This is done independently for each \(v \in V_r\). Then \(E(n, m)\) is the multigraph \((V_r, E(n, m))\), where \(E(n, m) = \{(v, c(i), i) : v \in V_r, 1 \leq i \leq m\}\) and
There is an $O(n^2 \log n)$ time algorithm HAM1 which satisfies

$$\lim_{n \to \infty} \Pr(\text{HAM1 finds a hamilton cycle in } G(n, 10)) = 1.$$ 

We next consider random regular graphs. Here we let $R(n, r)$ denote a random regular graph chosen uniformly from the set of graphs on $V_r$ which are regular of degree $r$. Bollobás [5] and Fenner and Frieze [10] independently gave proofs that there is a constant $r_0$ such that for any constant $r \geq r_0$

$$\lim_{n \to \infty} \Pr(R(n, r) \text{ is hamiltonian}) = 1.$$ 

The smaller value of $r_0$ was 796 of [10]. This paper improves this, but more importantly gives a polynomial time constructive proof. More specifically we have

$$\lim_{n \to \infty} \Pr(\text{HAM2 finds a hamilton cycle in } R(n, r)) = 1$$

for any constant $r \geq 85$.

It is reasonable to conjecture that $r_0=3$, especially as Richmond, Robinson and Wormald [17] have proved the corresponding result for the bipartite case.

Our final result concerns the random graph $G_{c,p}$, $p = c/n$, $c$ constant, which has vertex set $V_n$ and in which each of the $c$ possible edges independently has probability $p$ of being included and $1-p$ of being excluded. Several papers [1, 11, 4, 6, 12] have been concerned with the length of the longest path or cycle in $G_{c,p}$. The strongest result, for large $c$, is given in [12]. For a graph $G$ let $\ell(G)$ be the length of the longest cycle in $G$. In [12] we show that

$$\lim_{n \to \infty} \Pr(\ell(G_{c,p}) = 1 - ce^{-c} + c(1 + o(1))n) = 1,$$

where $\lim_{n \to \infty} o(c) = 0$.

The proof in [12] was again existential. Our final result is
THEOREM 1.3. There is an $O(n \log n)$ time algorithm CYCLEFIND which satisfies

$$\lim_{n \to \infty} Pr(\text{CYCLEFIND constructs a cycle of length } n) = 1,$$

(1 - ce^{-[(1 + o(1))n]} = 1).

(The result is still valid if $e = o(n) \to \infty$.)

Note. We give some notation that is used throughout the paper.

A graph $G$ has vertex set $V = V(G)$ and edge set $E = E(G)$. It has minimum degree $\delta(G)$ and $v \in V$ has degree $d_G(v)$. If $S \subseteq V$ then $G[S] = (S, E_S)$, where $E_S = \{ e \in E : e \subseteq S \}$. Also $V(S, G) = \{ w \in V - S : \exists v \in S \text{ such that } (v, w) \in E \}$.

An event $E_n$ will be said to occur almost surely (a.s.) if $\lim_{n \to \infty} Pr(E_n) = 1$.

ALGORITHM HAM

The following idea has been used by many authors: given a path $P = (v_1, v_2, \ldots, v_k)$ plus an edge $e = (v_x, v_y)$, where $1 \leq i \leq k - 2$, we can create another of length $k - 1$ by deleting edge $(v_i, v_{i+1})$ and adding $e$.

Thus let

$$\text{ROTA}(P, e) = (v_1, v_2, \ldots, v_{i-1}, v_y, v_{i+1}, \ldots, v_k)$$

and $\text{NEW}(P, e) = (v_x, v_y)$. $v_i$ is called the fixed endpoint, $v_k$ is called the rotated endpoint and $e$ is called the rotation edge of the rotation.

The algorithm we describe proceeds by a number of stages. At the beginning of the $k$th stage we have a path $P_k$ of length $k$, with endpoints $v_0, v_k$. We try to extend $P_k$ from $v_k$. If we fail but $(v_n, v_t) \in E(H)$ then assuming connectivity we can find another path of length $k - 1$. Failing this, we do a sequence of rotations with $v_0$ as a fixed endpoint, which creates new paths that we can try to extend or close. We apply the same construction to all these paths and so on until we have succeeded in obtaining a path of length $k - 1$ or we have run out of paths to rotate. We then take this set of paths and treat each of them like $P_k$ but use $v_k$ as the first rotated endpoint.

We construct our sequence of paths in a "depth-first" manner. Suppose the "current" path is $Q$. One end $w$ will be kept fixed. Suppose its other end $x$ has neighbours $x_1, x_2, \ldots, x_j$, where $x_i \in Q$. We replace $Q$ by $\text{ROTA}(Q, (w, x_i))$ and continue with this "new" $Q$ before considering $x_j$ and the "old" $Q$, which will be done after backtracking.

The above procedures are all perfectly natural. We now come to a somewhat unnatural procedure. It is included because without it we cannot make our proofs work. We would like someday to avoid this trick but at
HAMilton Cycles in Random Graphs

present we cannot. Our algorithm HAM assumes the partition of the edges
$E(H)$ of the input graph $H$ into 2 sets $E_+$ and $E_-$. The edges in $E_+$ can
only be used to close cycles. We define $H_+ = (V(H), E_+)$

We now give a formal description:

HAM’s input is a connected graph $H$ with $\delta(H) \geq 2$ plus a partition of its
edges into $E_+$ and $E_-$. We also assume that the input includes specific
orderings of the vertex adjacency lists, i.e., for each $v \in V(H)$ HAM is given
a total ordering of the set $N(v, H)$ of neighbours of $v$ in $H$. In this context
$\min_f(X)$ for $X \subseteq N(v, H)$, is the first vertex of $X$ appearing in the adjacency
lists for $v$.

Algorithm HAM
begin
let $P_1$ be the degenerate path consisting of $v_1 = \min_f(V(H))$ alone
$k := 0$;
LO:
begin [stage $k$ begins here]
longerpathfound := false;
L1:
let $P_k$ have endpoints $w_0, w_1$ where $w_0 \notin P_{k-1}$ and $w_1 \notin P_{k-1}$; [of course
when $k = 1, w_0 = v_1 = w_1$,
storepaths := true; [When this variable is true the new paths generated
by rotations during SEARCH are stored for later.
END$_w$ := $\{w_1\}$ [We keep track of the endpoints of some paths.
SEARCH$(P_w, w_0, P_k)$; [Do rotations with $w_0$ as fixed endpoint
if longerpathfound then [i.e. if SEARCH has found a path longer than
$P_k$
begin
$k := k + 1$; $P_k := P_k$ goto LO
end else
storepaths := false; [No need to store paths now
for $w \in END_w$ do
[SEARCH$(P_w, w_0, P_k)$ constructs a set of paths
${\lfloor P(w_0, w); w \in END_w \rfloor}$ where $P(w_0, w)$
has endpoints $w_0$ and $w$.
begin
SEARCH$(P(w_0, w), w, P_k)$;
if longerpathfound then goto L1
end;
terminate unsuccessfully [successful termination with a hamilton cycle
occurs in SEARCH end;
end.
procedure SEARCH(Q, u, P);
begin
let v be the endpoint of Q other than u;
DFS(v, *)
[* is used as a marker here.]
end;
procedure DFS(v, y); [vertex y is such that \text{ROTATE}(Q, \{v, y\})
(reverses the rotation made immediately prior to this call of DFS.)
begin
L2:
let X := \{x \not\in Q : \{v, x\} \in E_u\};
if X \neq \emptyset then
begin
x := \min(X); p := Q + \{v, x\}; longerpathfound := true [extension
end else
L3:
if \{(u, e) \in E(H)\} then
begin [cycle extension
let C be the cycle Q + \{u, v\};
if C is a hamilton cycle then terminate HAM successfully
else
begin starting from u, let a be the first vertex along Q which is adjacent in
H_u to some vertex not in C;
let B := \{x \in C : \{a, x\} \in E_u\};
b := \min(B); let a_i and a_j be the neighbours of a on C where a_i < a_j;
P := P + \{a, b\} - \{a, a_i\}; longerpathfound := true
end
end
end;
else
begin
let X := \{x_1, x_2, ..., x_p\};
for i = 1 to p do
if \{(x_i, x_{i+1}) \in E_u\}, and not longerpathfound and e has not been used
previously as a rotation edge in the current execution of SEARCH
then
begin
Q := \text{ROTATE}(Q, \{v, x_i\}); e' := \text{NEW}(Q, \{v, x_i\});
if storepath and e' \neq END_2 then
begin
END_2 := END_2 \cup \{e'\}; P(e, e') := Q
end;
DFS(NEW(Q, \{v, x_i\}), x_i)
end;
end;
end;
If \( y \neq e \) then \( Q := \text{ROTATE}(Q, \{v, y\}) \) [backtrack to the parent path].

**Running Time of HAM**

The running time of HAM is dominated by a factor dependent on the number of rotations. Using the idea of \([2]\) we may do each rotation in \(O(\log n)\) time. Each execution of SEARCH requires \(\leq |E|\) rotations, each of which stages requires \(\leq n\) executions of SEARCH giving \(O(n^2 \log n)\) time overall as in our examples \(|E| = O(n)\) a.s.

For the remainder of this section we consider those executions of HAM that terminate unsuccessfully. In particular suppose that HAM terminates unsuccessfully in stage \(k\). Let

\[
\text{END}(H) = END_k \cup \{v_k\}
\]

and

\[
\text{END}(H, w) = \{x : x \text{ is an endpoint of a path created during the execution of SEARCH}(P(v_n, w), w, P_k)\}
\]

for \(w \in \text{END}(H)\).

The following lemma is clear.

**Lemma 2.1.** If HAM terminates unsuccessfully then \(w \in \text{END}(H)\), \(e \in \text{END}(H, w)\) implies that \((v, w) \notin E(H)\).

A set \(X \subseteq E\) is deleteable if no \(e \in X\) is used to close a cycle at statement L3 during the execution of HAM on \(H\). The following lemma is also clear.

**Lemma 2.2.** If HAM terminates unsuccessfully, \(X\) is deleteable, \(H_k = (V_k, E_k, X)\) and the adjacency lists of \(H_k\) conform with those of \(H\) then HAM terminates unsuccessfully on \(H_k\) in stage \(k\). Furthermore \(\text{END}(H_k) = \text{END}(H)\) and \(\text{END}(H_k, w) = \text{END}(H, w)\) for \(w \in \text{END}(H)\).

We will need to show that \(|\text{END}(H)|\) is a.s. large. This will always be shown to follow from

**Lemma 2.3.** If HAM terminates unsuccessfully then

\[
|\{w \in \text{END}(H, w), H, w \} | \leq 2|\text{END}(H)|. \tag{2.1a}
\]

\[
|\{w \in \text{END}(H, w), H, w \} | \leq 2|\text{END}(H, w)|. \tag{2.1b}
\]
Proof. We modify the argument of Pósa [16]. We prove (2.1a); an almost identical argument will prove (2.1b). To prove (2.1a) we show that

\[ x \in N(\text{END}(H), H') \text{ implies } y \in \text{END}(H) \text{ such that } (x,y) \text{ is an edge of } P_k. \]  

Suppose \( x \in N(\text{END}(H), H'), z \in \text{END}(H), e=(x,z) \in E_k - P_k \), and neither of the neighbours \( u, v \) of \( x \) on \( P_k \) is in \( \text{END}(H) \). \( x \neq u, v \) since stage \( k \) terminated unsuccessfully. Eventually HAM creates a path \( P \) with \( z \) as an endpoint and \( x \) will be considered for rotation. It will not have been used before as \( x \notin \text{END}(H) \) and \( P \) will contain both edges \((u, v)\), \((x, u)\) because when an edge is deleted by a rotation one of the vertices is placed in \( \text{END}(H) \). Thus \( e \) will be used to rotate and at least one of \( u, v \) is in \( \text{END}(H) \)--contradiction. \[ \]

**M-OUT**

This section is devoted to the proof of Theorem 1.1. Let \( m \) be a fixed integer; \( m \) will be 9 for the main result. We construct an edge coloured \( D(n, m+1) \) as the union of \( D(n, m) \), with red edges plus an independent \( D(n, 1) \) with blue edges.

The input to HAM is \( H = D(n, m+1) \) with (1) adjacency lists in random order (note that the adjacency list for \( x \) can contain two copies of a vertex \( w \) if \( x \) chooses \( w \) and \( w \) chooses \( x \)) and (2) \( E = E(m, n) \) (red edges of \( H \)) and \( E = E(n, 1) \) (blue edges of \( H \)).

We first note

**LEMMA 3.1.** \( D(n, m) \) is a.s. connected for \( m \geq 2 \).

The next lemma shows how we aim to prove Theorem 1.1.

**LEMMA 3.1.** Suppose that the following are true a.s. for \( n = n(m), \beta = \beta(m), 0 < \beta < \alpha < 1 \):

\[ S \subseteq V_n, |S| < \alpha n \]  \( \Rightarrow \) \( |N(D, S)| \geq 2|S| \) \hspace{1cm} (3.1a)

where \( D = D(n, m) \).

HAM applied to \( D(n, m) \) makes fewer than \( \beta n \) cycle extensions. \hspace{1cm} (3.1b)

\[ (1 - \beta) > (1 - x)^{\alpha n(1 - \beta)} - \beta. \] \hspace{1cm} (3.1c)

Then HAM a.s. finds a hamilton cycle in \( D(n, m+1) \).
HAMilton Cycles in Random GrApHes

Proof. We use a variation of the colouring argument of [9]. Let $\alpha = \beta \log \pi$ and $\varepsilon > 0$ be small. Let $Y$ be a random $\alpha$-subset of $V$. For $y \in Y$ let $t(y)$ be the vertex chosen by $y$ in the construction of $E'$. Set $\mathcal{X} = \{(y, t(y)): y \in Y\}$. We define two events.

$E_1 = \{(3.1.a) \text{ holds for } H_y, (3.1.b) \text{ holds and } \text{HAM fails on } D(n, m + 1)\}$

$E_2 = E_1 \cap \{\mathcal{X} \text{ is deletable and } |Y \cap END(H)| \geq \varepsilon n\}$

where $\varepsilon = (1 - \varepsilon)(\gamma - \beta)(1 - \beta)$.

The lemma follows from two inequalities plus the fact that $\varepsilon$ is arbitrary.

\[ \Pr(E_1 | E_2) \geq (1 - o(1))(1 - \beta)^{\varepsilon}. \quad (3.2.a) \]

\[ \Pr(E_2) \leq (1 - x)^{\varepsilon}. \quad (3.2.b) \]

For then $\Pr(E_1) = o(1)$ and $\Pr(\text{HAM fails}) \leq 1 - \Pr((3.1.a) \text{ and (3.1.b)}) + \Pr(E_2)$.

Proof of (3.2.a). Given $D(n, m + 1)$ and an ordering of the adjacency lists such that $E_1$ occurs, we have $|\text{END}(H)| \geq \varepsilon n$ from Lemma 2.3. The probability of $E_2$ is then easily seen to be at least $(1 - o(1))(1 - \beta)^{\varepsilon}$.

Proof of (3.2.b). We prove this by showing that

\[ \Pr(E_2 | H_y) \leq (1 - x)^{\varepsilon}. \quad (3.3) \]

(by $(1 - x)^{\varepsilon}$ we mean that we are given $Y$, the $m + 1$ vertex choices for $x \notin Y$ and the $m$ vertex choices for $x \in Y$). If $E_2$ occurs then Lemmas 2.1 and 2.2 and (3.1.a) imply that

\[ \text{HAM fails on } H_y, \quad |Y \cap \text{END}(H_y)| \geq 70 \text{ and } |\text{END}(H_y, w)| \geq 3n \text{ for } w \in \text{END}(H_y). \quad (3.4.a) \]

\[ y \in Y \cap \text{END}(H_y) \text{ implies } x(y) \notin \text{END}(H_y, x). \quad (3.4.b) \]

It is important to note that, given $H_y$, although $Y$ is determined the choices $x(y), y \in Y$ are arbitrary and hence equally likely.

Now $\Pr(E_1 | H_y) = 0$ if $H_y$ does not satisfy (3.4.a) and so assume that (3.4.a) is satisfied. In this case

\[ \Pr(E_2 | H_y) \leq \Pr((3.4.b) | H_y) \leq (1 - x)^{\varepsilon_{\text{END}(H_y)}} \]

and (3.3) follows.
Lemma 3.2. Let 
\[ \phi_0(x) = \frac{3^{n/2} n^{n-3} (1-x)^{n(n-3)/2}}{2^{3(n-3)} n^{n(n-3)/2}}. \]

If \( m > 4, n < (2m-4)/(6m-3), \) and \( \phi_0(x) < 1 \) then (3.1a) holds a.s.

Proof. Let 
\[ u_n = \frac{n!}{k!(2k)!((n-3)!)^2} \left( \frac{3}{n} \right)^m \left( \frac{1}{n} \right)^{n(n-3)/2}. \]

Then \( \Pr(3.1a) \) fails \( \leq \sum_{k=1}^n u_k \leq \sum_{k=1}^n (\phi_0(n))^{-k} \) on using Stirling's inequalities.

Now one can see that for some constant \( A_c, \phi_0(x) \leq (A_c x)^x \). Thus 
\[ \sum_{k=1}^n (\phi_0(n))^{-k} \leq \sum_{k=1}^n (A_c n)^{n(n-3)/2} = O(1/n). \] On the other hand, by differentiating \( \phi_0(x) \) twice we find 
\[ \frac{\phi_0''(x)}{\phi_0'(x)} \left( \frac{\phi_0'(x)}{\phi_0(x)} \right)^2 = \frac{2m}{(1-x)^3} + \frac{3m}{1-x} + \frac{m-\frac{9}{x}}{1-3x} \]
\[ \geq \frac{(6m-3)}{9(1-3x)}. \]

Thus the lemma's assumptions imply \( \phi_0 \) is convex in \([0, x] \). Since \( \phi_0(1/2,1,0) < 1 \) the result follows. \( \square \)

Lemma 3.3. Let \( \varepsilon > 0 \) be arbitrary. Then \( \text{HAM} \) a.s. makes no more than \( (1 + \varepsilon)(n/(2m-1)) \) cycle extensions.

Proof. Consider the start of stage \( k \), in particular the first execution of L2. Here \( \varepsilon = \omega_l \), and 
\[ \Pr(X = \mathcal{G} \mid \text{previous history}) \leq \frac{k}{n} \left( \frac{1}{n} - \frac{1}{n} \right)^{m-k} \]
\[ \leq u_n = \frac{k}{n} \frac{1}{n} e^{-m/k} \tag{3.5} \]

To see this we observe: at any stage of the algorithm \( P_l \) contains two types of vertex, live and dead. A vertex \( w \) is dead if at some time previously, at statement L2 we found \( X = \mathcal{G} \) for \( \varepsilon = w \). For such a vertex we know that it has no neighbours in \( Q_k = P_l - P_l \). If \( w \) is a live vertex than all we know is that each time \( w \) has appeared as \( \varepsilon = l \). \( X \neq \mathcal{G} \). Let \( L_0 = [\text{live vertices}] \). Consider now the choices made by vertices in \( Q_k \). The only way they have been conditioned is that they do not choose in \( P_l - L_0 \), i.e., their chances
of choosing in $L_k$ have increased. Furthermore, we have no information about which vertices in $L_k$ are chosen by $x \in Q_3$, for when the algorithm establishes such a choice $x$ immediately becomes an endpoint of the current path. Now $w_i$ has just been added to $P_k$. By the above we know that given the previous history the probability that the $(x)$-th $m-1$ choices of $w_i$ that are not yet known to us are all in $P_k$ is $\Pr\{\exists w_i \text{ chosen in } P_k\}$ and independently the probability that no vertex in $Q_3$ chooses $w_i$ is $\Pr\{\exists w_i \text{ chosen in } Q_3\}$. Thus, using Theorem 1 of Hoeffding [14],

$$\Pr\{\sum_{i=1}^{n} Z_i \geq (1 + \epsilon) \sum_{i=1}^{n} u_i\} \leq e^{-\epsilon^2 \sum_{i=1}^{n} u_i^2} = o(1).$$

Hence for large $n$ HAM a.s. makes fewer than

$$(1 + \epsilon) \sum_{i=1}^{n} u_i \leq (1 + 2\epsilon) \int_{0}^{1} x^{m-1} e^{-x} \, dx$$

(3.6)

$$= (1 + 2\epsilon) \int_{0}^{1} (1 - x)^{m-1} e^{-x} \, dx$$

(3.7)

$$\leq (1 + 2\epsilon) \int_{0}^{1} e^{-m(1+x)} \, dx$$

and the result follows as $c$ is arbitrary.

To prove Theorem 1.1 we take $m = 9$, $\epsilon = 0.27$ satisfies the conditions of Lemma 3.2. Taking $\beta = 1/17$ from Lemma 3.3 and applying Lemma 3.1 yields the theorem. (Note that using the exact value for the integral in (3.7) does not reduce the value of $m$. Now if $X = \emptyset$ on the first execution of 1.2 in stage $k$ then $w_i$'s choices are random in $L_k$. Using this we can reduce $m$ to 7 and replace 10 by 8 in Theorem 1.1. This requires us to obtain a.s. upper bounds for the number of dead vertices at any stage and use a computer to estimate integrals numerically. We judge that it is not worth reproducing the entire argument here.)

**Regular Graphs**

This section is devoted to the proof of Theorem 1.2. We must first describe how the edges of $R(n, \epsilon)$ are partitioned into $E_u$ and $E_v$. We let
each \( v \in V_r \) independently choose one edge randomly from its \( r \) incident edges and place it in \( E^* \). Thus the same edge can be chosen twice.

Next let \( W = \{ \{ v, v' \} : v \text{ is incident with } > r/2 \text{ edges of } E^* \} \) and then \( E_* = \{ e \in E^* : e \cap W = \emptyset \} \). Having made this partition we put the adjacency lists in random order and apply HAM. In order to prove Theorem 1.2 we need a model for studying \( R(n, r) \). Let \( \text{REG}(n, r) \) be the set of \( r \)-regular graphs with vertex set \( V_r \) and consider the model defined in Bollobás [3]. Let \( D_1, D_2, \ldots, D_m \) be disjoint sets with \( |D_i| = r \) and set \( D = \{ D_1, D_2, \ldots, D_m \} \). A configuration \( C \) is a partition of \( D \) into \( m \) pairs, the edges of \( C \). Let \( \Phi \) be the set of all \( \{\{m\} = (2m)/(2^m m!) \} \) configurations. Turn \( \Phi \) into a probability space by giving all members of \( \Phi \) the same probability. For \( C \in \Phi \) let \( g(C) \) be the multi-graph with vertex set \( V_r \) in which \( i \) is joined to \( j \) whenever \( C \) has an edge with one end-vertex in \( D_i \) and the other in \( D_j \). Clearly \( \text{REG}(n, r) \subseteq g(\Phi) \) and \( |g^{-1}(G)| = n^r \) for every \( R(n, r) \subseteq \text{REG}(n, r) \).

Let \( Q \) be a property of the graphs in \( \text{REG}(n, r) \) and let \( Q^* \) be a property of the configurations in \( \Phi \). Suppose these properties are such that for \( G \in \text{REG}(n, r) \) and \( C \in g^{-1}(G) \) the configuration \( C \) has \( Q^* \) if and only if \( G \) has \( Q \). All we shall need from [3] is that if almost every \( C \) has \( Q^* \) then almost every \( G \) has \( Q \).

We shall thus be able to prove the theorem if we can show that HAM applied to a multigraph \( g(C) \), \( C \) chosen randomly from \( \Phi \), almost surely finds a hamilton cycle.

In terms of configurations our partition of the edges of \( C \) is done as follows: suppose that \( A_i = \{(i-1)r + 1, 2, \ldots, r \} \) for \( i = 1, 2, \ldots, n \). Let \( A_* = \{(i-1)r + 1, 2, \ldots, n \} \) and \( C^* = \{ e \in C : e \notin A_* \} \).

Let \( D^* = D \setminus \bigcup_{e \in C} e \) and \( C_* = \{ e \in C^* : e \cap D^* = \emptyset \} \) for \( e \in V_r \). Let \( W = \{ \{ v, v' \} : |D^*| > r/2 \} \) and then let \( C_* = \bigcup_{e \in C} e \) and \( C = C_* \setminus C_*. \)

For \( C \) chosen randomly from \( g^{-1}(\text{REG}(n, r)) \), taking \( E_* = g(C_c) \) and \( E_* \) yields an \( R(n, r) \) with the same random edge partition as that given at the start of this section. We now prove the equivalent of Lemmas 3.1–3.3.

**Lemma 4.1.** Suppose that the following are true a.s. for \( x = \mu(r), \beta = \beta(r), \gamma = \gamma(r), 0 < \beta < x < 1 \).

\[
S \subseteq V_r, |S| \leq n \text{ implies that } |N(S, g(C_c))| \geq 2|S|; \quad (4.1a)
\]

**HAM applied to** \( g(C) \) **makes fewer than** \( m \) **cycle extensions:**

\[
|R_2| \geq (1 - \gamma/2); \quad (4.1b)
\]

\[
1 - \frac{\beta}{1 - \gamma} > \left(1 - \frac{\beta}{2}\right)^{2^2}. \quad (4.1d)
\]
where \( \delta = (x - \beta - \gamma)(1 - \beta - \gamma) \). Then HAM a.s. finds a hamilton cycle in \( R(n, r) \).

Proof. Let \( \omega = \lfloor \log n \rfloor \) and \( \varepsilon > 0 \) be small. For \( x \in D \) let \( p(x, C) \) be the element of \( D \) paired with \( x \) by \( C \). Also let \( h(x) \) be defined by \( x \in D_{h(x)} \).

Let \( Y \) be a random \( \omega \)-subset of \( B_C \), \( Z = \{ h(y) : y \in Y \} \) and \( X = \{ x \in C : x \cap Y \neq \emptyset \} \). We define two events:

- \( E_1 = \{ (4.1a) \text{ holds for } C, \ (4.1b), (4.1c) \text{ hold and HAM fails on } C \} \)
- \( E_2 = E_1 \cap \{ X \text{ is deleteable and } (i) \ |Z \cap \text{ END}(g(C))| \geqslant \xi_0, \\
\quad (ii) |Z \cap \text{ END}(g(C), w)| \geqslant \xi_0 \text{ for } w \in \text{ END}(g(C)) \} \),

where \( \xi = \xi(n) \) is such that \( \xi_0 \) is the smallest even integer \( \geqslant (1 - \varepsilon) \delta_n \).

The lemma follows from two inequalities as before.

\[
\Pr(E_2|E_1) \leq (1 - o(1))(1 - \beta - \eta)^v \tag{4.2a}
\]

\[
\Pr(E_2) \leq (1 + o(1))(1 - \xi)(\xi/2)^v \tag{4.2b}
\]

The proof of (4.2a) is essentially the same as that for (3.2a).

To prove (4.2b) we show

\[
\Pr(E_2|C_x) \leq (1 + o(1))(1 - \xi/2)^v \tag{4.2b'}, \quad \text{where } C_x = C - X.
\]

If \( E_2 \) occurs then Lemmas 2.1–2.3 and (4.1a) imply that

- HAM fails on \( g(C_x) \), \( |Z \cap \text{ END}(g(C_x))| \geqslant \xi_0 \) and \( |Z \cap \text{ END}(g(C), w)| \geqslant \xi_0 \) for \( w \in \text{ END}(g(C)) \),

\[
\xi = \xi(n) \implies h \neq h, \quad (\text{4.3a})
\]

It is important now to note that, given \( C_x, Y \) is determined but the elements of \( X \) are paired up arbitrarily.

Now \( \Pr(E_2|C_x) = 0 \) if \( C_x \) does not satisfy (4.3a) and so assume that (4.3a) is satisfied. In this case

\[
\Pr(E_2|C_x) \leq \Pr(4.3b|C_x)
\]

\[
\leq (2\xi - \xi)\xi(1 - \xi)(2\xi - \xi/2)^v \tag{4.3b'}, \quad \text{where}
\]

\[
(1 + o(1))(1 - \xi/2)^v \tag{4.3b''.}
\]

using \( |X| \leq \xi_0 \) and that the “first” \( \xi_0/2 \) points of \( Z \cap \text{ END}(g(C_x)) \) have at most \( 2\xi - \xi_0 \) choices of points to be paired with. The lemma follows.
Lemma 4.2. Let
\[ \phi(s) = \frac{(3x)^{3x-s-1} e^x}{s^s (2\pi s)^{3/2}} (1 - 3x)^{3x-s-1}. \]

If \( r \geq 60, 0 \leq s \leq (r-7)(3r-3) \) and \( \phi(s) < 1 \) then (4.1a) holds a.s.

Proof. Now
\[ \Pr(4.1a \text{ fails}) \leq \sum_{s=0}^{r} \frac{n^s}{s! (2\pi s)^{3/2} (1 - 3x)^{3x-s-1}}. \]

where \( \pi_s = \Pr(N(F \cap g(C \cup \{ k+1, k+2, \ldots, 3k \}) \neq \emptyset) \).

Small \( k \). Suppose first that \( 1 \leq k \leq \sqrt{3r} \), where \( s_c = (2^{3r} (3r-3)^3)/3r \).

Since the minimum degree in \( g(C \cup \{ k+1, k+2, \ldots, 3k \}) \) is at least \( r/2 \) we have
\[ \pi_s \leq \binom{rk}{r/2} \Pr(\{ 1, 2, \ldots, \lfloor r/2 \rfloor \} \text{ are paired by } C \text{ in } \{ 1, 2, \ldots, 3k \}) \]
\[ = \binom{rk}{r/2} \frac{M^{r/2}}{n^r} \]
\[ = \left( \frac{2k^3}{n} \right)^{r/2}. \]

A routine calculation using
\[ \frac{n^r}{kn! (2\pi s)^{1/2}} \preceq n^{1/2} \]

now yields
\[ \sum_{s=0}^{r} \frac{n^s}{s! (2\pi s)^{3/2} (1 - 3x)^{3x-s-1}} \pi_s \sim (1). \]

Large \( k \). We consider the pairings made by all the points in \( V_{\cup_{k \in A} B_k \setminus A_k} \). This yields
\[ \pi_s \leq \binom{3r-1+k+x}{r} \frac{M^{r/2}}{n^r} \]
\[ \leq \left( \frac{M}{n} \right)^{r/2} e^{(3r-1+k+x)/6} \]

and hence, using Stirling's inequalities,
\[ \sum_{s=0}^{r} \frac{n^s}{s! (2\pi s)^{3/2} (1 - 3x)^{3x-s-1}} \pi_s \leq \sum_{s=0}^{r} \phi(s/n)^r. \quad (4.4) \]
We next use
\[ \frac{dG_n(x)}{d\phi_n(x)} = \frac{G_n(x)}{\phi_n(x)} = \frac{r - 7}{2x} - \frac{9}{1 - 3x} \]
to show that \( r \) is convex in \([0, (r - 7)/3(r - 3)]\). One then checks that \( \phi_n(x) \leq \phi_n(x) = (3x)^{r-1} (3x)^{r-7} \) and \( \phi_n(x) < 1 \) if \( x \geq 3 \). The lemma now follows from (4.4) provided \( 1/r < c \), and this holds if \( r \geq 60 \).

**Lemma 4.3.**
\[ |B_{r1}| \geq \left( 1 - 2r \left( \frac{2e}{r-2} \right)^{r^2-1} \right) n \quad \text{a.s.} \]

**Proof.** If \( |D_{r1}| > r/2 \) then at least \( r/2 - 1 \) out of \( \{2, 3, \ldots, r\} \) are paired by \( C \) with elements of \( A_r \). We deduce therefore that
\[ \Pr(|D_{r1}| > r/2) \leq \left( \frac{r}{r/2 - 1} \right) e^{-r^2/2} \leq \left( \frac{2e}{r-2} \right)^{r^2-1}. \]
Hence
\[ E(|W|) \leq \left( \frac{2e}{r-2} \right)^{r^2-1} n. \]
One may similarly show that \( \text{Var}(W) = O(n) \). Thus the Chebyshev inequality shows that \( |W| \leq 2E(|W|) \) a.s. Now use the fact that \( |B_{r1}| \geq n - r|W| \).

**Lemma 4.4.** Let \( \epsilon > 0 \) be arbitrary. Then \( \text{HAM} \) a.s. makes not more than \( (1 + 1/(n-3)) \) cycle extensions.

**Proof.** Consider the start of stage \( k \), in particular the first execution of L2. Here \( r = u_1 \), and
\[ \Pr(X = 0/\{\text{previous history}\}) \leq \left( \frac{r^2 - 2k + (n-k)}{(r-2)k + n(n-k)} \right)^{-2} . \] (4.5)
To see this we can assume that the previous history gives us all pairings in \( C \) that only involve elements of \( D_{r1} \), \( D_{r2} \) is accounted for. One point of \( D_{r1} \) is known to be paired with a point of \( D_{r2} \). The remaining points of \( D \) are paired arbitrarily. We consider \( r-2 \) such points of \( D_{r2} \). The probability that each of these is paired with one of the \( d \) points previously mentioned or the \( n-d \) points of \( A \), associated with vertices not in \( P_3 \), is bounded above by the RHS of (4.5).
Arguing as in Lemma 3.3 we then see that HAM a.s. makes fewer
\[
(1 + \epsilon) \int_0^1 \frac{(r-2)x + (1-x) y}{(r-2)x + y(1-x)} \, dx
\]
cycle extensions. The result follows on substituting $x = 1 - y$ in the integral
and using
\[
\frac{(r-2)-(r-3) y}{(r-2)+2y} \leq \left( 1 - \frac{(r-3)y}{r-2} \right) \leq e^{-(r-3)y(r-2)}.
\]
To obtain the theorem we use $\epsilon = 0.309$ in Lemma 4.1. Everything goes
through for $r \geq 85$.

**Sparse Random Graphs**

This section is devoted to the proof of Theorem 1.3. Let $G = G_{n,p}$. The
idea is to define a "large" set $V^* \subseteq V_r$ such that the graph $H = G[V^*]$ is
a.s.
hamiltonian. This is what is done in [12].

**Construction of $V^*$**

**Step 1.** The 2-core of $G$ is the largest set $S \subseteq V$ such that $\delta(G[S]) \geq 2$.
It exists because if $\delta(G[S_i]) \geq 2$ for $i = 1, 2$ then $\delta(G[S_1 \cup S_2]) \geq 2$.

The following algorithm constructs the 2-core \textsc{TWOC}:

\begin{algorithm}
begin
\textsc{TWOC} := $V_r$;
while $\delta(G[\textsc{TWOC}]) < 2$ do
$\textsc{TWOC} := \textsc{TWOC} - \{v \in \textsc{TWOV} : d_G[\textsc{TWOC}](v) < 2\}$
end.
\end{algorithm}

On termination \textsc{TWOC} is the 2-core. This is because one can easily show
inductively that each iteration removes vertices not in the 2-core. Note also
that no cycle of $G$ contains a vertex of $V$ or \textsc{TWOC}.

The remaining steps remove vertices so that $S \subseteq V^*$, $|N[S,H]| < 2|S|$ implies
that $S$ is large.

Let $v \in V_r$ be small if $d_G(v) \leq \epsilon/10$ and large otherwise.

**Step 2.**

\begin{algorithm}
begin
\textsc{small} := \{small vertices\}, $X := \emptyset$;
repeat
$S := \{v \in V_r : |N[v,G] \cap (X \cup \textsc{small})| \geq 2\}$;
$X := X \cup S$
until $S = \emptyset$
end
end.
Then let $Y = \{ y \in V, d_G(y) = 2 \text{ and } N(y, G) \cap X \neq \emptyset \} \text{ and}$

$$W = \bigcup_{i=1}^{t} W_i,$$

where $W_i = \{ v \in \text{SMALL} : i \in v \} \text{ and a path of length } t \text{ from } v \text{ to } w \in G \}$ (We allow $v = w$ when $t = 3$ or 4).

We finally define $V^* = \text{TWOC} - (W \cup X \cup Y)$ and $H = G[V^*]$. The reasons for the exact definition of $W, X, Y$ are made clear by the proofs in [12]. One (co) time is ample for the construction of $H$. The following results are proved in [12].

**Lemma 5.1.** For large enough $c$ we a.s. have

(a) $|V^*| \geq n\left(1 - (1 - 1/c(e))^{1/2}\right)$, where $\lim_{n \to \infty} t(c) = 0$.

(b) $|\text{SMALL}| \leq n^{1/3}$.

(c) $S \subseteq \text{LARGE} \Rightarrow \exists \text{ a path of length } t \text{ from } s \in S \text{ to } t \in S$.

(d) $S \subseteq V^*$, $n/12 \leq |S| \leq n/2 \Rightarrow |\{e, \ u \in E(G) : e \in S, u \notin S\}| \geq c|S|/15$.

(e) $S \subseteq V^*$, $|S| \leq n/12 \Rightarrow \exists S \subseteq H \Rightarrow |S| \geq 2|S|$.

(f) $S \subseteq V^*$, $|S| \geq n^{1/3} \Rightarrow |\{e \in E(G) : e \cap S \neq \emptyset\}| < 4c|S|$.

Note now that by construction

$$d_G(e) \geq d_G(v) - 1 \text{ for } v \in V^* \quad (5.1)$$

To complete the description of CYCLEFIND we must describe the partition of $E(G)$ into $E_1$ and $E_2$. To do this we first construct $\hat{E} \subseteq E(G)$ as follows: independently for each $e = (v, w) \in E(G)$ we do a $v$-experiment and a $w$-experiment both of which have probability $1/\sqrt{2}$ of success. If both succeed we include $e$ in $\hat{E}$. Thus we can view $\hat{E}$ as $\hat{\Delta}(G)$, where $\hat{\Delta} = \hat{\Delta}_e \leq 0$. Next let $\hat{H} = \hat{G}[V^*]$, where $V^*$ is defined in terms of $G$. Let

** Large $\quad \hat{H} \subseteq \hat{V}^* \Rightarrow \hat{d}_G(e) \geq c/20$.

We let $E_2 = \{ e \in E, e \in \text{LARGE} \}$.

**Lemma 5.2.** For large enough $c$ we a.s. have

(a) $S \subseteq V^*$, $|S| \leq n/12 \Rightarrow |N(S, H)| \geq 2|S|$.

(b) $H_1$ is connected.

(c) $|E_1| \geq c n^3$. 

**HAMILTON CYCLES IN RANDOM GRAPHS**

245
Proof. (a) Let now $S \in F^*$ with $|S| \leq n/12$. Let $S = S_1 \cup S_2$, where $S_2 = S \setminus \text{LARGE}$. Now

\[
|N(S, H_4)| = |N(S_1, H_4)| + |N(S_2, H_4)| - |N(S_1, H_4) \cap S_2| - |N(S_2, H_4) \cap N(S_1, H_4)|. \tag{5.2}
\]

But, assuming the conditions of Lemma 4.1 hold,

\[
|N(S_1, H_4)| = |N(S_1, H_4)| \geq 2|S_1| \quad \text{by Lemma 5.1(a).} \tag{5.3a}
\]

\[
|N(S_2, H_4)| \geq |N(S_2, H_4)| \geq |N(S_2, G)| \geq |S_2| \quad \text{as } F^* \setminus X = \emptyset. \tag{5.3b}
\]

\[
|N(S_1, H_4) \cap S_2| \leq |S_2| \quad \text{by Lemma 5.1(c) with } C \text{ replaced by } c/2. \tag{5.3c}
\]

\[
|N(S_2, H_4) \cap S_2| \leq |S_2| \quad \text{as } F^* \cap W = \emptyset. \tag{5.3d}
\]

\[
|N(S_1, H_4) \cap N(S_2, H_4)| \leq |S_2| \quad \text{as } F^* \cap W = \emptyset. \tag{5.3e}
\]

Equations (5.2) and (5.3) together imply (a).

(b) Suppose that $H_4$ contains a component $A$, where $|A| \leq |F^*|/2$ and let $B = F^* - A$. By (a) of this lemma we know that $|A| \geq n/4$ a.s. Lemma 5.1(a) and (f) imply that for large $c$

\[
|E(G) - E(H)| \leq 5c\text{e}^{-cn} \text{ a.s.} \tag{5.4}
\]

But then Lemma 5.1(d) implies that $H$ a.s. contains at least $c|A|/15 \geq 5/\text{e}^{cn}$ (large $c$) edges joining $A$ and $B$. Conditional on this event the probability that none of these edges is included in $E$ is no more than $2^{-cn}$. Thus

\[
\Pr(H_4 \text{ is not connected}) \leq 2^{2^c} \text{e}^{-cn} + o(1) = o(1) \quad \text{large } c.
\]

(c) An application of the Chebyshev inequality shows that

\[
\Pr\left(|\{v \in F^* : d_G(v) \leq c/2 + 1\}| \leq 5c\text{e}^{-cn/2}\right) \leq \text{e}^{-cn/2} \text{ a.s.} \tag{5.5}
\]

Equations (5.1), (5.5) and Lemma 5.1(a) imply that for large $c$

\[
\Pr\left(|\{v \in F^* : d_G(v) > c/2\}| \geq n(1 - 2\text{e}^{-cn/2})\right) \geq n(1 - 2\text{e}^{-cn/2}) \text{ a.s.} \tag{5.6}
\]

Since $|E(G)|$ is binomially distributed with mean $cn/2$ it is easy to show

\[
|E(G)| \geq 7n/15 \text{ a.s.}
\]
and using (5.4) we have, for large $c$,
\[ |E(H)| \geq 6cn/13 \quad \text{a.s.} \quad (5.7) \]

Let $\hat{A} = \{ v \in V(H) : d_{\hat{H}}(v) > c/2 \}$. It follows from (5.6) and (5.7) that for large $c$
\[ |\hat{A}| \geq 5cn/11 \quad \text{a.s.} \quad (5.8) \]

Now let $\tilde{A} = \hat{A} \cap \hat{E}$. We have
\[ |E(H)| \geq |\hat{A}| - \sum_{v \in \tilde{A}} d(v) \quad x(v) \quad (5.9) \]

where
\[
\begin{align*}
    x(v) &= 1 & \text{if } d_{\hat{H}}(v) > 3c \text{ or } d_{\hat{H}}(v) \leq c/2 \text{ or there are at least } d_{\hat{H}}(v) - c/20 \text{ successful } v-\text{experiments} \\
    &= 0 & \text{otherwise.}
\end{align*}
\]

Now (5.8) plus the fact that $v \in \hat{A}$ is independently placed in $\hat{A}$ with probability $1/2$ yields
\[ |\hat{A}| \geq 5cn/23 \quad \text{a.s.} \quad (5.10) \]

Now, by construction, the random variables $x(v)$, $v \in \hat{V}^*$ are independent and
\[
\begin{align*}
    \Pr(x(v) = 1) &\leq \left( \frac{d_{\hat{H}}(v)}{c/20} \right)^{c/20 - c/2} \quad \text{for } c/2 < d_{\hat{H}}(v) < 3c. \\
    &\leq \left( \frac{20d_{\hat{H}}(v)}{c} \right)^{c/20 - c/2} \quad \text{for } c/20 < d_{\hat{H}}(v) < 3c/20. \\
    &\leq 2^{-c/20} \quad \text{for large } c.
\end{align*}
\]

The independence of the $x(v)\'s$ then implies
\[ \sum_{v : d_{\hat{H}}(v) < c/20 \text{ or } d_{\hat{H}}(v) > 3c} d_{\hat{H}}(v) x(v) \leq 2^{c/20} n \quad \text{a.s.} \quad (5.11) \]

Now
\[ \sum_{v : d_{\hat{H}}(v) = 3c} d_{\hat{H}}(v) x(v) = \sum_{v : d_{\hat{H}}(v) = 3c} d_{\hat{H}}(v). \quad (5.12) \]
But
\[ E(d) = \sum_{k=0}^{n-1} k \binom{n-1}{k} p^k (1-p)^{n-1-k} \]
\[ \leq 6\alpha \left( \frac{n}{\ln(1-p)} \right)^2 n = \frac{\alpha n^2}{(1-p)^2} \]

Also it is not difficult to show that \( \text{Var}(d) = O(n) \) and thus, using the Chebyshev inequality,
\[ d \leq 2(c/3)^n n \quad \text{a.s.} \quad (5.13) \]

Equation (5.5) plus (5.9)–(5.13) yield (c) for large \( n \).

Having constructed \( H \) we will of course apply HAM. In this case the adjacency lists need not be randomised. Indeed we can assume that they are in increasing order. We finish our proof as before. Having generated \( \tilde{E} \) we generate \( X \subset \tilde{E} \) by independently including \( e \) in \( X \) with probability \( p_1 = \log n/n \). Our two events are
\[ E_1 = \{ \text{the conditions of Lemmas 5.1 and HAM fails on } H \} \]
\[ E_2 = E_1 \cap \{ X \text{ is deletable and no edge of } X \text{ is incident with any } v \in V^* - V^* \} \]

or \( \tau \) such that \( d_\tau (e) \leq c/2 + 1 \)

The theorem follows from
\[ \Pr(E_1 | E_2) \leq (1-p)^{n_1} \quad \text{for } \epsilon \text{ large} \quad (5.14) \]
\[ \Pr(E_2) \leq (1-p_1/2)^{n_1^{1/2}} \quad (5.15) \]

**Proof of (5.14).** HAM makes fewer than \( n \) cycle extensions and given \( E_1 \), Lemma 5.3(a) and (f) and (5.5) imply there are fewer than \( 5e^{-c/2} n \) edges incident with a vertex of \( V^* - V^* \).

**Proof of (5.15).** As usual we show that
\[ \Pr(E_2 | G_x) \leq (1-p_1/2)^{n_1^{1/2}} \text{, where } G_x = (V_x, E(G) - X). \quad (5.16) \]

Now if \( E_3 \) occurs then applying the method used to construct \( H \) from \( G \) will produce \( H_x \) from \( G_x \). Furthermore HAM will fail on \( H_x \) and
\[ (i) |\text{END}(H_x)| \geq n/12 \]
\[ (ii) |\text{END}(X_x, v)| \geq n/12 \quad \text{for } v \in \text{END}(H_x) \]

and of course
\[ e \in Y = \{ (e, w) : e \in \text{END}(H_x), w \in \text{END}(H_x, v) \} \quad (5.18) \]
implies $i \neq X$. Note that even (5.17) is determined by $G_x$, not $G$. Now $\Pr(E_1 \, | \, G_x) = 0$ if (5.17) does not hold and so assume it does. Note next that if $G_x$ is given then $X$ is a random subset of $E(G_x)$, where $e \in E(G_x)$ is independently included with probability $pp_e(1 - p) \geq pp_e/2$. But then

$$\Pr(E_1 \, | \, G_x) \leq \Pr((5.18) \, | \, G_x) \leq (1 - pp_e/2)^{3/2} \leq (1 - pp_e/2)^{1/1000}$$

and the theorem follows.

**CONCLUSION**

We have extended the results of [7, 13] to random graphs with constant average degree. The most important open problems are to (1) reduce the values of 10 and 85 in Theorems 1.1 and 1.2 to 3; (2) modify CYCLEFIND so that it a.s. finds the longest cycle in $G_{\infty}$ and get an asymptotic expression for this length; (3) remove the necessity for partitioning $E(H)$ into $E_+$ and $E_-$; and (4) extend all these results to digraphs.

*Note added in proof.* T. Luczak and the author have now reduced the 10 of Theorem 1 to 5.

**REFERENCES**


