COMBINATORICA 7 (4) (1987) 327—341

AN ALGORITHM FOR FINDING HAMILTON PATHS AND CYCLES IN RANDOM GRAPHS

B. BOLLOBÁS, T. I. FENNER and A. M. FRIEZE

Received 1 August 1985
Revised 19 September 1986

This paper describes a polynomial time algorithm HAM that searches for Hamilton cycles in undirected graphs. On a random graph its asymptotic probability of success is that of the existence of such a cycle. If all graphs with n vertices are considered equally likely, then using dynamic programming on failure leads to an algorithm with polynomial expected time. The algorithm HAM is also used to solve the symmetric bottleneck travelling salesman problem with probability tending to 1, as n tends to ∞.

Various modifications of HAM are shown to solve several Hamilton path problems.

1. Introduction

We shall give a polynomial time algorithm HAM that searches for Hamilton cycles in undirected graphs. As one would expect this algirthm is not perfectly reliable, i.e. a graph G may have a Hamilton cycle but our algorithm may fail to find one. However, if G is chosen at random then our algorithm has an asymptotically small probability of faliure.

To be precise: let Γ_0 denote the set of graphs wih vertex set $V_n=\{1, 2, ..., n\}$ and m edges.

We turn Γ_0 into a probability space by given each $G\in\Gamma_0$ the probability $1/|\Gamma_0|=1/\binom{N}{m}$ where $N=\binom{n}{2}$. Let $G_{n,m}$ denote a graph chosen randomly from Γ_0.

Now let

$$m = n\log n/2+n\log\log n/2+c_n n \tag{1.1}$$

for some sequence c_n. Komlós and Szemerédi [10] have shown that

$$\lim_{n\to\infty} \Pr(G_{n,m} \text{ is hamiltonian}) = \begin{cases} 0 & \text{if } c_n\to-\infty \\ e^{-e^{-2c}} & \text{if } c_n\to c \\ 1 & \text{if } c_n\to+\infty \end{cases}$$

$$= \lim_{n\to\infty} \Pr(G_{n,m} \text{ has minimum degree at least 2}).$$

Supported by NSF Grant MCS 810 4854.
AMS subject classification: 05 C 45, 68 E 10.

Their proof is essentially non-constructive (see also Bollobás [2] or Frieze [7] for alternative non-constructive proofs). Polynomial time algorithms which have a high probability of finding Hamilton cycles have been described by Angluin and Valiant [1] and Shamir [12]. The algorithm due to Shamir [12], HAM1, say, satisfies

$$\lim_{n\to\infty} \Pr(\text{HAM1 finds a Hamilton cyle in } G_{n,m}) = 1$$

if $c_n > (1+\varepsilon)\log\log n$ for some fixed $\varepsilon > 0$. We first improve this to obtain the essentially best possible result.

Theorem 1.1. (a) *Let m be defined as in* (1.1). *Then*

$$\lim_{n\to\infty} \Pr(\text{HAM finds a Hamilton cycle in } G_{n,m}) = \begin{cases} 0 & \text{if } c_n \to -\infty \\ e^{-e^{-2c}} & \text{if } c_n \to c \\ 1 & \text{if } c_n \to \infty \end{cases}.$$

(b) HAM *runs in* $0(n^{4+\varepsilon})$ *time.*

(Note that result (a) cannot be improved, although (b) possibly could.)

We note that the algorithms described in [1] and [12] require the input graph to have its adjacency lists given in a random order. In our algorithm this is not necessary. Thus the previous algorithms can be viewed as randomised algorithms that work well on random inputs while our algorithm is a deterministic algorithm that works well on random inputs.

• We next consider the case where each of the 2^N graphs with vertices V_n is equally likely to be chosen. Under this model the probability of failure is so small that, if we apply dynamic programming [9] when HAM fails, we obtain the following result as a corollary of the proof of Theorem 1.1.

Theorem 1.2. *There is an algorithm for solving the Hamilton cycle problem with polynomial expected running time.*

Our algorithm also has an application in solving the symmetric bottleneck travelling salesman problem (BTSP). An instance of BTSP is specified by the assignment of a numerical weight to the edges of a complete graph K_n on n vertices. The objective is to find a hamiltonian circuit for which the maximum edge-weight is minimised.

Let us assume that edge-weights are drawn independently from the uniform distribution over [0, 1]. Karp and Steele [11] remark that Shamir's algorithm can be used to find a *near optimal* solution with probability tending to 1. (Throughout this paper all limits are taken as $n\to\infty$ and this is implied if it is not explicitly stated). A modification of our proof of Theorem 1.1 gives

Theorem 1.3. *There is a polynomial time algorithm* BOT *satisfying*

$$\lim_{n\to\infty} \Pr(\text{BOT solves BTSP exactly}) = 1.$$

We shall also consider the use of HAM is finding hamilton paths. There are 3 cases to consider:

(1) find a Hamilton path from vertex 1 to vertex n,

(2) find a Hamilton path from vertex 1 to any other vertex,

(3) find a Hamilton path without specifying either endpoint.

Let a Hamilton path satisfying the condition (i), $i=1,2,3$, be of type i. Then we can construct simple modifications HAMPi to HAM, $i=1,2,3$, such that the following result holds.

Theorem 1.4. *Let* m *be defined as in* (1.1). *Then*

$$\lim_{n\to\infty} \Pr(\text{HAMPi finds a Hamilton path of type } i \text{ in } G_{n,m}) =$$

$$= \lim_{n\to\infty} \Pr(G_{n,m} \text{ contains a Hamilton path of type } i) =$$

$$= \lim_{n\to\infty} \Pr(G_{n,m} \text{ contains} \leqq i-1 \text{ vertices of degree 1}) =$$

$$= \begin{cases} 0 & \text{if } c_n \to -\infty \\ e^{-\lambda}(1+\lambda+\ldots+\lambda^{i-1}) & \text{if } c_n \to c \quad \text{where} \quad \lambda = e^{-2c} \\ 1 & \text{if } c_n \to +\infty. \end{cases}$$

We can also use the algorithm to find Hamilton paths between all pairs of vertices. A graph is said to be *Hamilton connected* if it contains a Hamilton path joining each distinct pair of vertices.

Theorem 1.5. *Let* $m = n\log n/2 + n\log\log n + c_n n$. *Then*

$$\lim_{n\to\infty} \Pr(G_{n,m} \text{ is Hamilton connected}) =$$

$$= \lim_{n\to\infty} \Pr(G_{n,m} \text{ has minimum degree at least 3}) =$$

$$= \begin{cases} 0 & \text{if } c_n \to -\infty \\ e^{-\lambda/2} & \text{if } c_n \to c, \quad \text{where} \quad \lambda = e^{-2c} \\ 1 & \text{if } c_n \to +\infty. \end{cases}$$

2. Algorithm HAM

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

$$\text{ROTATE}(P, e) = (v_1, v_2, \ldots, v_i, v_k, v_{k-1}, \ldots, v_{i+1}).$$

The algorithm we describe is based on ideas in the proof used in [6]. It proceeds by a sequence of stages. At the beginning of the k^{th} stage we have a path P_k of length k, with endpoints w_0 and w_1. We try to extend P_k from either w_0 or w_1. If we fail but $\{w_0, w_1\}\in E$ then connectivity tells us that we can find a longer path. Failing this, we do a sequence of rotations 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 either we have succeeded in obtaining a path of length $k+1$ or we have exceeded a certain length of rotation sequence. We now give a formal description:

Algorithm HAM

Input: a connected graph $G=(V_n, E)$ of minimum degree at least 2.

begin
    let P be the path $(1, w)$ where $w=\min\{v: \{1, v\}\in E\}$;
    $k:=1$
L1 **begin** {stage k begins here}
    $Q_1:=P_k$; $s:=1$; $t:=1$; $\delta(Q_1):=0$;
{Rermark: $\delta(Q_s)$ is the number of rotations in the sequence constructing Q_s from Q_1}
    repeat
      let path Q_s have endpoints w_0, w_1 where $w_0<w_1$;
    for $i=0, 1$ **do**
    begin
      Suppose that the edges incident with w_i and not contained in Q_s are $\{w_i, x_1\}, \ldots, \{w_i, x_p\}$ where $x_1<x_2<\ldots<x_p$;
      for $j=1$ **to** p **do**
      if x_j is not on Q_s **then**
      begin
        $P_{k+1}:=Q_s+\{w_i, x_j\}$; {*extension*}
        $k:=k+1$; **goto** L1
      end
      else if $x_j=w_{1-i}$ **then**
      begin
        let C be the cycle $Q_s+\{w_0, w_1\}$;
        if C is a Hamiltonian cycle **then** terminate successfully
        else
        begin
          starting from w_0, let u be the first vertex along Q_s which is adjacent to some vertex not in C; let v be the lowest numbered neighbour of u not in C and let u_1 and u_2 be the neighbours of u on C where $u_1<u_2$, then
          $P_{k+1}:=C+\{u, v\}-\{u, u_1\}$; $k:=k+1$;
          goto L1 {**cycle extension**}
        end
      end else
      begin
        $t:=t+1$; $Q_t:=\text{ROTATE}(Q_s, \{w_i, x_j\})$; $\delta(Q_t):=\delta(Q_s)+1$
      end
      {next j}
    end {next i}
    $s:=s+1$
    until $\delta(Q_s)\geqq 2T+1$;
        {where $T=\lceil \log n/(\log d-\log\log d)\rceil+1$ and $d=2m/n$}
    terminate unsuccessfully
  end
end

We now introduce some notation used in the analysis of HAM. Suppose that HAM terminates unsuccessfully in stage k on input G. Let $\mathrm{END}(G)=\{v$: there exists, in stage k, a path Q_s with v as an endpoint and $\delta(Q_s)=t$, $1\leqq t\leqq T\}$. For $x\in\mathrm{END}(G)$ let $\mathrm{END}(G,x)=\{v$: there exists, in stage k, a path Q_s with x, v as endpoints and $\delta(Q_s)=t$, $1\leqq t\leqq 2T\}$. We note that

(2.1) G cannot contain an edge $\{x, y\}$ where $x\in\mathrm{END}(G)$ and $y\in\mathrm{END}(G,x)$.

Consider P_k, the initial path in stage k. It is the final path in a sequence $P^{(0)}=P_1$, $P^{(1)}, P^{(2)}, \ldots, P^{(M)}=P_k$ where $P^{(i+1)}$ is obtained from $P^{(i)}$ by a single *extension, cycle extension,* or *rotation*. Let $W(G)=\{\text{edges in } P^{(1)}, P^{(2)}, \ldots, P^{(M)}\}\cup\cup\{\{w_0, w_1\}$: HAM executes a cycle extension on a path with endpoints w_0 and $w_1\}$. For $X\subseteqq E$ let $G_X=(V_n, E-X)$, we can then deduce.

Lemma 2.1. *Suppose that* HAM *terminates unsuccessfully in stage k on input G. If* $X\subseteqq E-W(G)$ *then* HAM *will also terminate unsuccessfully in stage k on* G_X.

[On input G_X HAM will actually generate P_k at the start of stage k via the same sequence $P^{(0)}, P^{(1)}, \ldots, P^{(M)}$].

The following inequality is straightforward:

(2.2) $$|W(G)| \leqq n(2T+2).$$

3. Proof of Theorems 1.1 and 1.2

We say that an event A_n, depending on n, occurs almost surely (a.s.) if $\lim_{n\to\infty} \Pr(A_n)=1$.

We now prove a structural lemma concerning $G=G_{n,m}$. Let $d=2m/n$ as in HAM. A vertex is *small* if $\deg(v)\leqq d/20$ and *large* otherwise. For $S\subseteqq V_n$ let $N(S,G)=\{w\in V_n-S$: there exists $v\in S$ such that $\{v,w\}\in E\}$.

Lemma 3.1. *The following statements hold a.s., provided* $c_n\not\to-\infty$:

(a) $G_{n,m}$ *contains no more than* $n^{1/3}$ *small vertices.*

(b) $G_{n,m}$ *does not contain* 2 *small vertices at a distance of* 4 *or less apart.*

(c) $G_{n,m}$ *contains no vertex of degree exceeding* 5*d*.

(d) *There does not exist a set of large verices S with* $|S|\leqq n/d$ *and* $|N(S,G)|\leqq\leqq d|S|/300$.

Proof. It is much easier to work with the independent model $G_{n,p}$ which is a random graph with vertices V_n, in which each possible edge is independently included with probability p and excluded with probability $1-p$. It is well known that if $p=m/N$ then $G_{n,m}$ and $G_{n,p}$ have similar properties. We shall calculate with $G_{n,p}$, $p=m/N$ and translate our results to $G_{n,m}$.

Let $E_{n,p}$ denote the (random) set of edges in $G_{n,p}$. We note first that $|E_{n,p}|$ is distributed as a binomial random variable with parameters N, p and that conditional on $|E_{n,p}|=m$, $G_{n,p}$ is distributed as $G_{n,m}$. It follows from Stirling's inequalities for factorials that

(3.1) $$\Pr(|E_{n,p}|=m) = (1-o(1))\big(N/(2\pi m(N-m))\big)^{1/2} \geqq \geqq (1-o(1))(2/\pi N)^{1/2}.$$

Also for a graph property A

$$\Pr(G_{n,p} \text{ has } A) = \sum_{m'} \Pr(G_{n,p} \text{ has } A \,\big|\, |E_{n,p}| = m') \Pr(|E_{n,p}| = m') = \tag{3.2}$$

$$= \sum_{m'} \Pr(G_{n,m} \text{ has } A) \Pr(|E_{n,p}| = m').$$

Thus from (3.1) and (3.2) we have

$$\Pr(G_{n,m} \text{ has } A) \leqq \big(1+o(1)\big)(\pi N/2)^{1/2} \Pr(G_{n,p} \text{ has } A). \tag{3.3}$$

In our proof we will often use non-integral quantities where we should really round up or down. It will be clear that such aberrations do not affect the validity of our arguments.

(a) If $G_{n,p}$ has $\geqq n^{1/3}$ small vertices then there exists a set S, $|S|=n^{1/3}$ such that each vertex of S is adjacent to no more than $d/20$ vertices of V_n-S. Thus

$$\Pr(G_{n,p} \text{ has } \geqq s = n^{1/3} \text{ small vertices}) \leqq \binom{n}{s}\left(\sum_{k=0}^{d/20}\binom{n-s}{k}p^k(1-p)^{(n-s-k)}\right)^s \leqq \tag{3.4}$$

$$\leqq (ne/s)^s\big(c((n-s)20ep/d)^{d/20}\exp(-19d/20+sd/n)\big)^s$$

$$\leqq \exp(-n^{1/3}d/12) \tag{3.5}$$

using $d \geqq \log n$. (Note that the summation in (3.4) is dominated by its last term). Thus, using (3.3) $\Pr\big(G_{n,m} \text{ has} \geqq n^{1/3}, \text{ small vertices}\big) = O\big(n \exp(-n^{1/3}d/12)\big)$.

(b) Let A_1 denote 'there exist 2 small vertices at a distance $\leqq 4$ apart'. Then

(3.6)

$$\Pr(G_{n,p} \text{ has } A_1) \leqq \binom{n}{2}\left(\sum_{k=0}^{d/20}\binom{n-2}{k}p^k(1-p)^{n-k-2}\right)^2 (n^3p^4+n^2p^3+np^2+p) \leqq n^5p^4e^{-1.5d}.$$

For $p \geqq 2\log n/n$ we can use (3.3) and (3.6). For smaller p in the range $\log n/n \leqq p \leqq 2\log n/n$ we need a bit more work. It follows from (3.2) that there exists m', $m-(n\log^3 n)^{1/2} \leqq m' \leqq m$ such that $\Pr(G_{n,m'} \text{ has } A_1) \leqq 3n^5p^4e^{-1.5d}$.

Now $G_{n,m}$ is obtained from $G_{n,m'}$ by adding $m-m'$ random edges. Thus $\Pr(G_{n,m} \text{ has } A_1) \leqq \Pr(G_{n,m'} \text{ has } A_1) + \pi_1 \Pr(G_{n,m'} \text{ does not have } A_1)$ where $\pi_1 = \Pr$ (one of the $m-m'$ added edges meets a vertex that is within distance 1 of a small vertex). That π_1 is $o(1)$ follows from (a) and $m-m' \leqq (n\log^3 n)^{1/2}$.

(c) $\Pr(G_{n,p} \text{ has a vertex of degree} \geqq 5d) \leqq n\binom{n}{5d}p^{5d} \leqq n(e/5)^{5d}$.

Now use (3.3).

(d) We prove the result for $G=G_{n,p}$, the result for $G_{n,m}$ follows from (3.3). For a set $K \subseteq V_n$ let A_K be the event that $|N(K,G)| \leqq \alpha|K|d$, where $\alpha = 1/300$.

We first consider $|K|$ large. Suppose first that $n^{\cdot 1} \leqq k = |K| \leqq n/d$. We prove a stronger result than needed. Now

$$\Pr(\text{there exists } K,\ |K| = k, \text{ and } A_k) \leqq$$

$$\leqq \Delta = \binom{n}{k}\sum_{t=0}^{\alpha kd}\binom{n-k}{t}p_1^t(1-p_1)^{n-k-t}$$

where $p_1=(1-(1-p)^k)\leqq kp\leqq 1$ is the probability that a vertex not in K is adjacent to at least one vertex of K, if $|K|=k$. For large n, $\alpha kd\leqq(n-k)p_1/2$ and so for some constant $c>0$

$$\begin{aligned}\Delta &\leqq c\binom{n}{k}\binom{n-k}{\alpha kd}p_1^{\alpha kd}(1-p)^{k(n-k-\alpha kd)}\leqq\\ &\leqq c(ne/k)^k(ne/\alpha kd)^{\alpha kd}(kp)^{\alpha kd}\exp(-kd+k+\alpha kd)\leqq\\ &\leqq c\big((ne^2/k)(e^2/\alpha)^{\alpha d}e^{-d}\big)^k=\\ &=O(n^{-\gamma})\quad\text{for any constant } \gamma>0,\end{aligned}$$

provided $n^{0.1}\leqq k\leqq n/d$.

For $1\leqq k\leqq\min(n^{0.1}, n/d)$ we use two methods of proof which cover the range of possibilties. We first assume that $p=O(n^{-0.75})$. If there exists a set K ot large vertices such that A_K occurs then, by considering $T=K\cup N(K, G)$, there exists a set T, $|T|=t$, $d/20\leqq t\leqq n^{0.1}(1+\alpha d)$ which contains at least $2t$ edges. Then for some constant $c>0$

$$\begin{aligned}\Pr(\text{there exist such a } T) &\leqq c\sum_t\binom{n}{t}\binom{t^2/2}{2t}p^{2t}(1-p)^{t^2/2-2t}\leqq\\ &\leqq\sum_t(ne^4tp^2/16)^t=\\ &=O(n^{-\gamma})\quad\text{for any constant } \gamma>0.\end{aligned}$$

We finally consider $p\geqq n^{-0.8}$. We independently orient the edges of G randomly to obtain a digraph G', i.e. if $\{u, v\}\in E_{n,p}$ then we direct from u to v with probability 1/2 and from v to u with probability 1/2. Let B be the event 'there exists $v\in V_n$ such that v is large in G but has outdegree $\leqq d/50$ in G'. Since $d\geqq n^{0.2}$ we find $\Pr(B)=O(n^{-\gamma})$ for any constant $\gamma>0$. Suppose then that B does not occur and A_K occurs for some small set K of large vertices. Then there exists a set K all of whose vertices have outdegree $\geqq d/50$ in G' for which the outdegree of the set K is no more than $\alpha|K|\,d$. Then if $k=|K|$

$$\begin{aligned}\Pr(\text{there exists such a } K) &\leqq\binom{n}{k}\binom{n}{\alpha kd}\left\{\binom{\alpha kd}{d/50}(p/2)^{d/50}\right\}^k=\\ &=O(n^{-\gamma})\quad\text{for any constant } \gamma>0.\end{aligned}$$

Let $\Gamma_1=\{G\in\Gamma_0\colon G$ is connected, has minimum degree at least 2 and satisfies all the conditions listed in Lemma 3.1$\}$. Suppose that HAM terminates unsuccessfully in stage k on $G_{n,m}$. Now let $X\subseteq E$ be *deletable* if:

(i) no edge of X meets a small vertex;
(ii) no large vertex meets more than $d/1000$ edges of X;
(iii) $X\cap W(G)=\emptyset$.

Lemma 3.2. *Suppose* HAM *terminates unsuccessfully in stage k on $G=G_{n,m}\in\Gamma_1$. Suppose $X\subseteq E$ is deletable. Then for, n large,*

$$|\mathrm{END}(G_X)|\geqq n/1000; \tag{3.6}$$

$$|\mathrm{END}(G_X, x)|\geqq n/1000\quad \textit{for}\quad x\in\mathrm{END}(G_X). \tag{3.7}$$

Proof. Consider the execution of HAM on G_X. From Lemma 2.1 we know that HAM will start stage k with the same P_k as for G and terminate unsuccessfully in this stage. Suppose P_k has endpoints w_0 and w_1. Let $S_t=\{v$: v is large (in G) and there exists a path Q_s with endpoints w_0, v such that $\delta(Q_s)=t\}$. We prove (3.6) by showing that

$$\left|\bigcup_{t=1}^{T} S_t\right| \geqq n/1000. \tag{3.8}$$

We show first that, for n large, $S_1\neq\emptyset$. Let $x_0, x_1, \ldots, x_h$ be the neighbours of w_1 in G_X where $\{x_0, w_1\}$ is an edge of $P_k=Q_1$. Let y_i be the endpoint, other than w_0, of ROTATE $(P_k, \{w_1, w_i\})$ for $i=1, 2, \ldots, h$.

Case 1: w_1 is small.

Then $h\geqq 1$ as $G\in\Gamma_1$ and X is deletable. Also y_1 is large (Lemma 3.1(b)) and so $y_1\in S_1$.

Case 2: w_1 is large.

$h\geqq d/20-d/1000-1$ for n large. Also at most one of $y_1, y_2, \ldots, y_h$ can be small (Lemma 3.1(b)) and so $|S_1|\geqq h-1>0$ for n large. We show next that, for n large,

$$|S_t| \leqq n/d \quad \text{implies} \quad |S_{t+1}| \geqq d|S_t|/1000. \tag{3.9}$$

For each vertex $v\in S_t$ choose one path $Q_{s(v)}$ with endpoints w_0 and v such that $\delta(Q_{s(v)})=t$. Consider now pairs (v, w) where $v\in S_t$ and $w\in W(v)=N(\{v\}, G_X)$. If $\{v, w\}$ is not an edge of $Q_{s(v)}$ let $x(v, w)$ be the endpoint of ROTATE $(Q_{s(v)}, \{v, w\})$ other than w_0. If $x(v, w)$ is large then $x\in S_{t+1}$. Let

$$\alpha(v, w) = \begin{cases} 1 & \text{if} \quad \begin{array}{l} \text{(a)} \quad \{v, w\} \text{ is an edge of } Q_{s(v)} \\ \qquad\qquad \text{OR} \\ \text{(b)} \quad x = x(v, w) \text{ is small} \\ \qquad\qquad \text{OR} \\ \text{(c)} \quad \{x, w\} \text{ is not an edge of } P_k. \end{array} \\ 0 & \text{otherwise} \end{cases}$$

Now for each $v\in S_t$ there are at most $t+2$ w's such that $\alpha(v, w)=1$ (1 for ech of (a) and (b) and t for (c) as $Q_{s(v)}$ is obtained from P_k by t rotations and hence contains at most t edges not in P_k). On the other hand, for each $w\in N(S_t, G_X)$ there can be at most 2 $x\in S_{t+1}$ such that for some $v\in S_t$, $x=x(v, w)$ and $\alpha(v, w)=0$, since x will be a neighbour of w on P_k. Thus

$$\begin{aligned} |S_{t+1}| &= |\{x(v, w)\colon v\in S_t, w\in W(v) \text{ and } x(v, w) \text{ is large}\}| \geqq \\ &\geqq |\{x(v, w)\colon v\in S_t, w\in W(v) \text{ and } \alpha(v, w) = 0\}| \leqq \\ &\geqq |\{w\in N(S_t, G_X)\colon \text{there exists } v\in S_t \text{ with } \alpha(v, w) = 0\}|/2 \geqq \\ &\geqq (|N(S_t, G_X)|-(t+2)|S_t|)/2 \geqq \\ &\geqq (|N(S_t, G)|-(d/1000+t+2)|S_t|)/2 \geqq \\ &\geqq (d/300-(d/1000+T+2))|S_t|/2 \geqq \\ &\geqq d|S_t|/999 \quad \text{for } n \text{ large.} \end{aligned}$$

Since $S_1 \neq \emptyset$ and (3.9) holds, we know that for some $\tau \leqq T-1$ that $|S_\tau| \geqq n/d$. Let $S' \subseteq S_\tau$ be of size $\lceil n/d \rceil$. Applying the same argument as used to prove (3.9), using S' in place of S_τ we have $|S_{\tau+1}| \geqq d|S'|/999 \geqq n/1000$ for n large. This verifies (3.6). To prove (3.7) consider $x \in \mathrm{END}(G_X)$, choose a path $Q = Q_\sigma$ having x as one of its endpoints and $\delta(Q_\sigma) \leqq T$ and then redefine $S_t = \{v$: v is large (in G) and there exists a path Q_s with endpoints x, v such that Q_s is obtained from Q_σ using t rotations with x as a fixed endpoint$\}$. Now apply the argument used to prove (3.9), using Q_σ in place of P_k, to prove (3.7).

We can now prove Theorem 1.1. Now it is known (see for example Erdős and Rényi [5]) that if $c_n \not\to -\infty$ then $G_{n,m}$ is a.s. connected and in general

$$\Pr(G_{n,m} \text{ has a vertex of degree 1}) \approx 1 - e^{-e^{-2c_n}}.$$

Thus if $c_n \to -\infty$, $G_{n,m}$ a.s. has a vertex of degree 1 and so there is nothing to prove. If $c_n \not\to -\infty$ then, using Lemma 3.1, we have

$$|\Gamma_1| = (1-o(1))e^{-e^{-2c_n}}|\Gamma_0|. \tag{3.10}$$

Now let $\Gamma_2 = \{G\colon G \in \Gamma_1$ and HAM terminates unsuccessfully on $G\}$. It follows from (3.10) that to prove Theorem 1.1. we need only show that

$$\lim_{n\to\infty} |\Gamma_2|/|\Gamma_0| = 0. \tag{3.11}$$

To prove (3.11) we use a colouring argument developed in Fenner and Frieze [6]. Let now $w = \lceil \lambda d \rceil$ for some constant $\lambda > 0$. For each $G \in \Gamma_0$ let (G, j), $j = 1, 2, \ldots,$ $J = \binom{m}{w}$ enumerate all the possible ways of colouring w edges of G green and the remaining $m-w$ edges blue. Let $X = X(G, j)$ denote the set of green edges. Let

$$a(G,j) = \begin{cases} 1 & \text{if} \quad \text{(3.12a) HAM terminates unsuccessfully on } G \text{ and } G_X; \\ & \qquad \text{(3.12b) there does not exist } e = \{x,y\} \in X \text{ such that} \\ & \qquad\qquad x \in \mathrm{END}(G_X) \text{ and } y \in \mathrm{END}(x); \\ & \qquad \text{(3.12c) } |\mathrm{END}(G_X)| \geqq n/1000 \text{ and } |\mathrm{END}(G_X, x)| \geqq n/1000 \\ & \qquad\qquad \text{for all } x \in \mathrm{END}(G_X) \\ 0 & \text{otherwise.} \end{cases}$$

We show first that for $G \in \Gamma_2$

$$\sum_{j=1}^{J} a(G,j) \geqq (1-o(1))\binom{m_1}{w} \quad \text{where} \quad m_1 = m - (2T+2)n = (1-o(1))m. \tag{3.13}$$

To see this let $G \in \Gamma_2$ and let HAM terminate unsuccessfully in stage k on G. As $G \in \Gamma_1$, it follows from (2.1), Lemma 2.1 and Lemma 3.2 that if $X = X(G, j)$ is deletable then $a(G, j) = 1$. Let $G' = (V', E')$ be the subgraph of G induced by the large vertices and those edges not in $W(G)$. Then $(V') \geqq n - n^{1/3}$ and $|E'| = m' \geqq m - n(2T+2)$. The number of deletable sets is the number of ways of choosing w edges from E' subject to the condition that no vertex in V' has more than $d/1000$

of its incident vertices chosen. Using Lemma 3.1(c) it is not difficult to show that this is $(1-o(1))\binom{m'}{w}$ which implies (3.13). (Choose edges of E' independently with probability $4\lambda/n$. One almost surely chooses more than w edges. Furthermore the number of edges chosen incident with a given vertex is dominated stochastically by a binomial random variable with parameters $\lceil 5d\rceil$ and $4\lambda/n$).

On the other hand, let H be a fixed graph with vertices V_n and $m-w$ edges. Let $b(H)=|\{(G,j): H=G_X,\ G\in\Gamma_0 \text{ and } a(G,j)=1\}|$. We see that

$$b(H) \leqq \binom{N'-m+w}{w} \quad \text{where} \quad N' = \binom{n}{2}-\binom{\lceil n/1000\rceil}{2}. \tag{3.14}$$

If (3.12a) or (3.12b) do not hold for H (replace G_X by H in these statements) then $b_H=0$. Given (3.12a), (3.12b) there are at most $N'-m+w$ edges to choose from in order to unsure (3.12c).

Now

$$\sum_{G\in\Gamma_0}\sum_{j=1}^{J} a(G,j) = \sum_{H} b(H).$$

Thus

$$\begin{aligned}(1-o(1))\binom{n_1}{w}|\Gamma_2| &\leqq \sum_{G\in\Gamma_2}\sum_{j=1}^{J} a(G,j) \quad \text{by (3.13)}\\ &\leqq \sum_{G\in\Gamma_0}\sum_{j=1}^{J} a(G,j) =\\ &= \sum_{H} b(H) \leqq\\ &\leqq \binom{N'-m+w}{w}\binom{N}{m-w} \quad \text{by (3.14)}\end{aligned}$$

Thus

$$\begin{aligned}|\Gamma_2|/|\Gamma_0| &\leqq (1+o(1))\binom{N'-m+w}{w}\binom{N}{m-w}\Big/\left\{\binom{m_1}{w}\binom{N}{m}\right\} \leqq\\ &\leqq (1+o(1))\big((N'-m+w)(m-w)\big)/\big((N-m+w)(m_1-w)\big)\big)^w \leqq\\ &\leqq (1+o(1))e^{-w(N-N')/N}(1+o(1))^w \leqq\\ &\leqq e^{-\lambda d/1000001} \quad \text{for } n \text{ large.}\end{aligned} \tag{3.15}$$

We can take any constant value $\lambda>0$ here and this will complete the proof of Theorem 1.1(a). To prove part (b) we note that on $G\in\Gamma_1$ HAM executes $O(n(5d)^{2T})=O(n^{3+\varepsilon})$ rotations. Thus, as given, HAM runs in time $O(n^{4+\varepsilon})$ with probability $1-o(1)$. We can easily make it run in time $O(n^{4+\varepsilon})$ by imposing a suitable time limit.

We now turn to the proof of Theorem 1.2. If all graphs are equally likely to be chosen then this is the same model as $G_{n,p}$, $p=1/2$. We use (3.2) where property A will mean that the given graph is connected, has minimum degree at least 2 and yet HAM terminates *unsuccessfully*. We will show that

$$\Pr(G_{n,1/2} \text{ has } A) = o(1/2^n). \tag{3.16}$$

Since Dynamic Programming requires time $O(n^2 2^n)$, this will prove the theorem. Using Theorem I. 7(i) of [3], for the tail of the Binomial Distribution we see that

$$\Pr\left(\big||E_{n,1/2}| - n^2/4\big| \geqq (n^3 \log n)^{1/2}\right) = o(1/3^n). \tag{3.17}$$

Thus, using (3.2), we need only prove

$$\Pr(G_{n,m'} \text{ has } A) = o(1/2^n) \quad \text{for} \quad |m' - n^2/4| < (n^3 \log n)^{1/2}. \tag{3.18}$$

Letting Γ_i, $i=0, 1, 2$ refer to $G_{n,m'}$ we define $\Gamma_1' = \{G \in \Gamma_0\colon G$ does not satisfy all the conditions of Lemma 3.1$\}$ and $\Gamma_A = \{G \in \Gamma_0\colon G$ is connected, has minimum degree at least 2 and yet HAM terminates unsuccessfully on $G\}$. Then $\Pr(G_{n,m'} \text{ has } A) = |\Gamma_A'|/|\Gamma_0| \leqq |\Gamma_1|/|\Gamma_0| + |\Gamma_A - \Gamma_1'|/|\Gamma_0| = |\Gamma_1'|/|\Gamma_0| + |\Gamma_2|/|\Gamma_0|$. Now let $d' = 2m'/n = (1-o(1))n/2$ in our range of interest. We see that, for large n, conditions (b) and (d) of Lemma 3.1 are always true for $G_{n,m'}$. It follows from (3.3), (3.5) and (3.6) that $|\Gamma_1'|/|\Gamma_0| = \Pr(G_{n,m'} \in \Gamma_1') = O(n \exp(-n^{4/3}/25) + n^6 \exp(-0{,}74n)) = o(1/2^n)$. Putting $\lambda = 2000002$ in (3.15) shows that $|\Gamma_2|/|\Gamma_0| = o(e^{-n})$ and this completes the proof of Theorem 1.2. ∎

4. Algorithm BOT

We turn now to BTSP. Given an instance of this problem, let the edges of K_n be ordered $e_1, e_2, \ldots, e_N$ where $c(e_1) \leqq c(e_2) \leqq \ldots \leqq c(e_N)$, $c(e)$ being the 'cost' of edge e. Let $E_t = \{e_1, e_2, \ldots, e_t\}$ and let $G_t = (V_n, E_t)$. Note that G_t has the same distribution as $G_{n,t}$.

Algorithm BOT
begin
let $\mu = \min\{t\colon G_t$ has minimum degree at least 2$\}$;
apply HAM to G_μ
end

It is clear that if HAM terminates successfully on G_μ then BOT solves BTSP exactly. We cannot apply Theorem 1.1. directly as G_μ as defined in BOT above has a slightly different distribution to $G_{n,\mu}$ conditional on minimum degree 2. Let $m' = \lfloor n \log n/2 + n \log\log n/2 - n \log\log n/2 \rfloor$ and $m'' = m' + \lceil n \log\log\log n \rceil$. It is known that $G_{m'}$ a.s. is connected and has minimum degree 1 and that $G_{m''}$ a.s. has minimum degree 2. Thus $m' < \mu < m''$ a.s. Now for $m' < m \leqq m''$ define the events:

$A_m = \{G_m$ is connected and satisfies the conditions of Lemma 3.1, where $d = \log n\}$,

$B_m = \{A_m$ and G_m has minimum degree at least 2$\}$,

$C_m = \{G_m$ is connected, has minimum degree at least 2 and HAM terminates unsuccessfully on $G_{m'}\}$

Then, where, $M=\{m\colon m'<m\leqq m''\}$,

$$\Pr(\text{BOT fails}) = \Pr\Big(\bigcup_{m\in M} C_m\Big)+o(1) \leqq \tag{4.1}$$

$$\leqq \Pr\Big(\Big(\bigcup_{m\in M} C_m\Big)\cap\Big(\bigcap_{m\in M} A_m\Big)\Big)+\Pr\Big(\bigcup_{m\in M}\bar{A}_m\Big)+o(1) \leqq$$

$$\leqq \sum_{m\in M}\Pr(C_m\cap A_m)+\Pr\Big(\bigcup_{m\in M}\bar{A}_m\Big)+o(1) =$$

$$= \sum_{m\in M}\Pr(C_m\cap B_m)+\Pr\Big(\bigcup_{m\in M}\bar{A}_m\Big)+o(1) \leqq$$

$$\leqq \sum_{m\in M}\Pr(C_m| B_m)+\Pr\Big(\bigcup_{m\in M}\bar{A}_m\Big)+o(1).$$

For $x\in\{a, b, c, d\}$ let $A_m(x)=\{G_m$ satisfies condition x of Lemma 3.1$\}$ and let $D_m=\{G_m$ is connected$\}$. Now

$$\Pr\Big(\bigcup_{m\in M}\bar{D}_m\Big) = \Pr(\bar{D}_{m+1}) = o(1). \tag{4.2}$$

The calculations in Lemma 3.1 show that

$$\Pr\Big(\bigcup_{m\in M}(\bar{A}_m(a)\cup\bar{A}_m(d))\Big) = O(n^{-\alpha}) \quad \text{for any constant } \alpha>0. \tag{4.3}$$

(Although Lemma 3.1 specifically excludes $c_n\to-\infty$, the calculations are still valid for $c_4\geqq-\log\log\log n$.)

$$\Pr\Big(\bigcup_{m\in M}\bar{A}_m(c)\Big) = \Pr(\bar{A}_{m''}(c)) = o(1). \tag{4.4}$$

By considering the addition of the $m+1$'st edge we obtain $\Pr(\bar{A}_{m+1}(b)\cap A_m(\mathrm{b})\cap \cap A_m(a)\cap A_m(c))=O((\log n)^4 n^{-4/3})$. Thus $\Pr\Big(\bigcup_{m\in M}(\bar{A}_{m+1}(b)\cap A_m(b)\cap A_m(a)\cap \cap A_m(c)\Big)=o(1)$. It then follows from (4.3) that $\Pr\Big(\bigcup_{m\in M}(\bar{A}_{m+1}(b)\cap A_m(\mathrm{b}))\Big)=o(1)$ and hence

$$\Pr\Big(\bigcup_{m\in M}\bar{A}_m(b)\Big) \leqq \Pr(\bar{A}_{m'+1}(b))\Pr\Big(\bigcup_{m\in M}(\bar{A}_{m+1}(b)\cap A_m(b))\Big) \tag{4.5}$$

(4.2)—(4.5) yield

$$\Pr\Big(\bigcup_{m\in M}\bar{A}_m\Big) = o(1). \tag{4.6}$$

Now in the notation used to prove Theorem 1.1, we have $\Pr(C_m|B_m)= =|\Gamma_2|/|\Gamma_1|=(|\Gamma_2|/|\Gamma_0|)(|\Gamma_0|/|\Gamma_1|)\leqq n^{-\lambda/1000001}\log n$ for any constant $\lambda>0$, using (3.15) and (3.10). By choosing $\lambda=2000002$ we obtain

$$\sum_{m\in M}\Pr(C_m| B_m) = o(1). \tag{4.7}$$

Theorem 1.3 now follows from (4.1), (4.6) and (4.7). ∎

5. Hamilton Paths

We shall view the problem of finding a Hamilton path from vertex a to vertex b in a graph G as that of finding a Hamilton cycle in the graph $G(a, b)=(V(G), E(G)\cup\{\{a, b\}\})$ that contains the edge $\{a, b\}$. To ensure that HAM searches for a Hamilton cycle containing a particular edge $\{a, b\}$ we make some minor modifications:

(1) *Initialisation*

Let $G=G_{n,m}(a, b)$. Let H_2 be the graph induced by the edge $\{a, b\}$ and all edges incident with vertices of degree 2 in G. If H_2 contains a cycle or a vertex of degree 3 or more then HAM terminates successfully (success means that HAM has been able to decide as to whether or not G contains a Hamilton cycle using the edge $\{a, b\}$). Otherwise, if H_2 consists of vertex disjoint paths, then we say that the degree 2 vertices of $G_{n,m}$ are compatible with $\{a, b\}$. In this case we initialise P_1 to be the component of H_2 containing $\{a, b\}$.

(2) *Rotations*

We omit any rotation that involves deleting an edge of P_1.

(3) *Cycle Extensions*

If HAM wishes to do a cycle extension but can only do so by deleting an edge of P_1 then HAM will terminate unsuccessfully, otherwise HAM will choose the 'first' possible 'legal' cycle extension. Note that HAM will not terminate unsuccessfully in (3) if

(5.1) $G_{n,m}$ does not contain a path P of length 3 or more which has at most 2 vertices which have neighbours not in P.

We call HAM with the above modifications HAM (a, b).

Next let $\Gamma_1(a, b)=\{G_{n,m}$: (1) $G_{n,m}$ satisfies all the conditions of Lemma 3.1 as well as (5.1), (2) $G_{n,m}(a, b)$ has minimum degree at least 2$\}$. Note that it is straightforward to show that $G_{n,m}$ a.s. satisfies (5.1).

We next indicate the proof of

Lemma 5.1. $\Pr(\text{HAM}(1, n)$ *terminates unsuccessfully* $|G_{n,m}\in\Gamma_1(1, n))=O(n^{-\gamma})$ *for any constant* $\gamma>0$.

Proof. (Outline) The proof of Lemma 3.2 requires only small modifications. Consider first the proof that $S_1\neq\emptyset$. Only Case 1 requires a mention. If $P_1\neq(1, n)$ then by condition (b) of Lemma 3.1 all neighbours of w_1 yield legal rotations. If $P_1=(1, n)$ there is only a problem if w_1 is of degree 2 and 1 or n is a neighbour of w_1. But then we have the contraction that w_1 is a vertex of P_1 or $1, n$ are both neighbours of w_1.

For the rest of the proof of Lemma 3.2 the only change is that $t+2$ is replaced by $t+3$. The colouring argument that follows Lemma 3.2 goes through with only trivial changes, particularly if we separate the cases $\{1, n\}\in E(G_{n,m})$ and $\{1, n\}\notin E(G_{n,m})$.

Since λ can be chosen arbitrarily large in this argument, we have the stated $O(n^{-\gamma})$ probability of failure.

Theorem 1.4 and 1.5 are consequences of

Lemma 5.2. *Let* $m=(n\log n+n\log\log n)/2+c_n n$ *where* $c_n\geqq -O(1)$. *For* $1\leqq i\leqq j\leqq n$ *let* $A(i,j)$ *denote the event that* $G=G_{n,m}(i,j)$ *has minimum degree at least* 2, *the degree* 2 *vertices of* $G_{n,m}$ *are compatible with* $\{i,j\}$ *and yet* HAM (i,j) *fails to find a Hamilton cycle in* G *using* $\{i,j\}$.
Let

$$A=\bigcup_{i=1}^{n-1}\bigcup_{j=i+1}^{n}A(i,j).$$

Then

$$\lim_{n\to\infty}\Pr(A)=0.$$

Proof. $\Pr(A)\leqq\Pr\big(A$ and $G_{n,m}$ satisfies the conditions of Lemma 3.1 and (5.1)$\big)+$ $+o(1)=\Pr\big(\bigcup_{i=1}^{n-1}\bigcup_{j=i+1}^{n}(G_{n,m}\in\Gamma_1(i,j)$ and HAM terminates unsuccessfully$)\big)+$ $+o(1)\leqq\binom{r}{2}\Pr\big(\text{HAM}(1,n)$ terminates unsuccessfully $|G_{n,m}\in\Gamma_1(1,n)\big)+o(1)=o(1)$. ∎

Theorem 1.5 follows immediately.

Theorem 1.4(1) follows from the fact that vertices 1 and n are a.s. large, and have no degree 2 neighbours.

Theorem 1.4(2) and (3) follows from the above and the fact that a.s. no vertex of degree 1 has a degree 2 neighbour.

6. Conclusions

The results of this paper show that the hamiltonian cycle problem can be considered to be well-solved in a prohabilistic sense. They can be extended to cover the problem of finding disjoint hamiltonian cycles by following the approach described in Bollobás and Frieze [4].

Indeed all one has to do is to repeatedly apply HAM and remove Hamilton cycles. The proof that this works with the same limiting probability as that of having minimum degree $2k$ can be obtained from the proof of Theorem 1.1 with only slight modifications.

With a few minor changes to the proofs one can show that HAM combined with Dynamic Programming has polynomial expected running time for $G_{n,p}$ if $p>1-1/\sqrt{2}$. For smaller p we find that the probability that there are 2 vertices of degree 2 which are close exceeds $(1/2+\varepsilon)^n$. However, by initially covering vertices of small degree with vertex disjoint paths, if possible, we can obtain a polynominal expected running time algorithm whenever p is a positive constant. Independently, Gurevich and Shelah [8] have constructed an $O(n)$ expected running time Hamilton path algorithm for p constant. Subsequently, a similar result has been obtained by Thomason [13]. In a future note we hope to show that HAM combined with Dynamic Programming has $O(n)$ expected running time for $G_{n,p}$ if $p>1-1/\sqrt{2}$.

References

[1] D. Angluin and L. G. Valiant, Fast probabilistic algorithms for Hamilton circuits and matechings, *J. Computer Syst.* **18**, (1979), 155—193.
[2] B. Bollobás, Almost all regular graphs are Hamiltonian, *European Journal of Combinatorics* **4**, (1983), 311—316.
[3] B. Bollobás, *Random Graphs*, Academic Press, London, (1985).
[4] B. Bollobás and A. M. Frieze, On matchings and Hamilton cycles in random graphs, *Annals of Discrete Mathematics*, **28** (1985), 23—46.
[5] P. Erdős and A. Rényi, On the strength of connectedness of a random graph. *Acta. Math. Acad. Sci. Hungar.* **12**, (1961), 261—267.
[6] 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.
[7] A. M. Frieze, Limit distribution for the existence of Hamiltonian cycles in a random bipartite graph, *European Journal of Combinatorics*, **6** (1985), 327—334.
[8] Y. Gurevich and S. Shelah, Technical Report of the University of Michigan, 1985.
[9] M. Held and R. M. Karp, A dynamic programming approach to sequencing problems, *SIAM Journal on Applied Mathematics* **10**, (1962), 196—210.
[10] M. Komlós and E. Szemerédi, Limit distribution for the existence of a Hamilton cycle in a random graph, *Discrete Mathematics* **43**, (1983), 55—63.
[11] R. M. Karp and J. M. Steele, Probabilistic analysis of the travelling salesman problem, in *The Travelling Salesman Problem: A Guided Tour* (E. L. Lawler, J. K. Lenstra, A. H. G. Rinnoy-Kan and D. B. Shmoys, eds), John Wiley and Sons (1985).
[12] E. Shamir, How many edges are needed to make a random graph Hamiltonian? *Combinatorica* **3**, (1983), 123—132.
[13] A. G. Thomason, A simple linear expected time algorithm for Hamilton cycles, to appear.

B. Bollobás

Department of Pure Mathematics
University of Cambridge, Cabridge, England
Department of Mathematics
LSU, Baton Rouge, LA, USA

T. I. Fenner

Department of Computer Science
Birkbeck College, University of London, England

A. M. Frieze

Department of Computer Science
Queen Mary College, University of London, England

Department of Mathematics
Carnegie Mellon University,
Pittsburgh, PA, USA.