Approximately Counting Hamilton Paths and Cycles in Dense Graphs

Martin Dyer* Alan Frieze† Mark Jerrum‡

August 27, 1995

Abstract

We describe fully polynomial randomized approximation schemes for the problems of determining the number of Hamilton paths and cycles in an n-vertex graph with minimum degree $(\frac{1}{2}+\alpha)n$, for any fixed $\alpha > 0$. We show that the exact counting problems are #P-complete. We also describe fully polynomial randomized approximation schemes for counting paths and cycles of all sizes in such graphs.

1 Introduction

Combinatorial counting problems have a long history, even from the computational viewpoint. For example, the classical matrix-tree theorem provides

*School of Computer Studies, University of Leeds, Leeds LS2 9JT, UK. Supported in part by Esprit Working Group "RAND."

†Mathematics Department, Carnegie Mellon University, Pittsburgh PA 15213, USA. Supported in part by NSF Grant 9002435.

‡Department of Computer Science, University of Edinburgh, The King's Buildings, Edinburgh EH9 3JZ, UK. Supported in part by grant GR/F 90363 of the UK Science and Engineering Research Council, and Esprit Working Group "RAND;" work done while visiting the NEC Research Institute, Princeton NJ, USA.

a good algorithm for determining the number of trees in a graph. However, it seems that few interesting combinatorial structures possess good counting algorithms. This intuition was made precise by Valiant [21] using the class #P. He showed that many problems for which the decision counterpart is easy were nevertheless complete for this class. Since it is unlikely that #P = P, exact counting is apparently intractable for many natural problems. For example Valiant [20] showed that 0-1 permanent evaluation, and counting the number of bases of a (suitably presented) matroid [21] were #P-complete. Many other problems have since been added to this list, for example volume computation for polyhedra [6], counting linear extensions of a partial order [3] and counting Eulerian orientations of a graph [17].

The hardness of most counting problems has led to an interest in *approximate* counting. The most fruitful approach in this respect has been *randomized approximation*. This is based on the idea of a fully polynomial randomized approximation scheme (*fpras*) due to Karp and Luby [15]. Thus if N is the true value, we must determine an estimate $\widehat{N}$ such that for given $\varepsilon, \delta > 0$

$$\Pr\left(1/(1+\varepsilon) < \widehat{N}/N < (1+\varepsilon)\right) > 1-\delta,$$

in time polynomial in the size of the input, ε^{-1} and $\log(\delta^{-1})$. Examples of problems amenable to this type of approximation are dense 0-1 permanent [4, 12], matchings [12], volume computation [7], counting Eulerian orientations [17], counting linear extensions of a partial order [16] and computing the partition function for the ferromagnetic Ising model [13]. The algorithms in the papers cited use a *random walk* to generate an almost uniform random solution to the problem (e.g., a random matching), and then apply multi-stage statistical sampling methods to obtain the desired estimate.

One obvious requirement for such approximate counting to be possible is that the associated decision problem be easy. In fact, it appears from experience that it must be "very easy" in order to have a realistic hope that a randomized approximation scheme can be found.

In this paper, we add further entries to the small but growing list of randomly approximable hard counting problems: that of counting the number of Hamilton paths and cycles in "dense" graphs. Let $G = (V, E)$ be a graph, where $V = \{v_1, v_2, \ldots, v_n\}$. Denote the degree of vertex v_i by d_i, for $i = 1, 2, \ldots, n$. We will say that G is *dense* if $\min_i d_i \geq (\frac{1}{2}+\alpha)n$, where $0 < \alpha \leq \frac{1}{2}$ is a fixed constant. Under these circumstances it is known [5] that G must contain a Hamilton cycle. Moreover, the proof of this fact is easily modified to give a simple polynomial-time algorithm for constructing such a Hamilton cycle. This algorithm, which uses edges whose existence is guaranteed by the pigeonhole principle to "patch together" disjoint cycles, provides the required easy decision procedure.

We consider here the natural but more difficult problems of counting the *number* of Hamilton paths and cycles in such graphs. We show in Section 4 that these problems are in fact #P-complete, so exact counting is presumably intractable. More positively, our main results in Sections 2 and 3 establish the existence of *fpras's* for these counting problems when $\alpha > 0$. We may observe that if the degree condition is relaxed to $\min_i d_i \geq (\frac{1}{2} - \alpha_n)n$ with $\alpha_n = \Omega(n^{\kappa-1})$ for any fixed $\kappa > 0$, then the question of the existence of any Hamilton path or cycle becomes NP-Complete,[1] and approximate counting is NP-hard. Thus our results establish quite precisely the difficulty of the counting problem except in the region where α is close to zero. Section 5 extends the positive results of the earlier sections to cover self-avoiding paths and cycles of all lengths.

The natural approach given previous successes in this area is to try to find a rapidly mixing Markov chain with state space the set of Hamilton cycles of a given dense graph, and possibly its Hamilton paths as well. Earlier attempts with this approach have proved fruitless. Somewhat surprisingly, the key lies in the fact that in dense graphs, Hamilton cycles form a substantial fraction of the set of 2-factors, a *2-factor* being defined as a set of vertex-disjoint cycles which together contain all vertices of G. This is not obvious a priori and the main technical difficulty in the approach lies in obtaining a good upper bound on the ratio of 2-factors to Hamilton cycles in a dense graph. A direct attack — relating the number of 2-factors with k cycles to the number with $k+1$ cycles — appears unworkable. Instead, we introduce a weight function on 2-factors that allows us to argue about the distribution of total

[1] This is true even if we insist on G being k-connected for any $k = o(n)$. The construction is from Bollobás [2]. Start with an arbitrary graph G and add a clique C of size $m = n^{1/\kappa}$ and an independent set I of size $m-1$ and then join every vertex in C to every other vertex, to produce a graph Γ. Then G has a Hamilton path if and only if Γ has a Hamilton cycle. Also Γ contains a Hamilton path if and only if G contains two vertex disjoint paths that cover all its vertices.

weight as a function of the number of cycles. By a rather delicate analysis, we are able to show that the Hamilton cycles carry sufficient weight for our purpose. In summary we prove

Theorem 1.1 *If G is dense then there are fpras's for*

(a) *approximating its number of Hamilton cycles,*

(b) *approximating its number of Hamilton paths,*

(c) *approximating its number of cycles of all sizes,*

(d) *approximating its number of paths of all sizes.*

2 Outline approach

Our approach to constructing an fpras for Hamilton cycles in a dense graph G is via a randomized reduction to sampling and estimating *2-factors* in G. An *almost uniform sampler* for 2-factors in a graph is a randomized algorithm that takes as input a graph G and $\delta > 0$ and outputs a 2-factor Z (a random variable) such that

$$1/(1+\delta)N \leq \Pr(Z = F) \leq (1+\delta)/N,$$

where F is any 2-factor in G and N is the total number of 2-factors. The sampler is said to be *fully polynomial* if it runs in time polynomial in the size of G and $\log \delta^{-1}$. Using known techniques, 2-factors in a dense graph G may be efficiently sampled, and their number estimated.

Theorem 2.1 *There exist both a fully polynomial randomized approximation scheme and a fully polynomial almost uniform sampler for the set of 2-factors in a dense graph.*

This result follows immediately from Corollary 4.2 of Jerrum and Sinclair [14], as will become clear once the notation used there has been explained. For convenience, the corollary in question is repeated below as Proposition 2.2.

Proposition 2.2 *There exists a fully polynomial almost uniform sampler for $\mathcal{G}(\mathbf{d}, X)$ and a fully polynomial randomized approximation scheme for $|\mathcal{G}(\mathbf{d}, X)|$, provided the pair $(\mathbf{d}, X)$ satisfies $e(\mathbf{d}) > d_{\max}(d_{\max} + x_{\max} - 1)$.*

In Proposition 2.2, $\mathbf{d} = (d_1, \ldots, d_n)$ stands for a degree sequence on $V = \{v_1, v_2, \ldots, v_n\}$, and $X \subseteq V^{(2)}$ for the edge set of an "excluded" graph on vertex set V. The notation $\mathcal{G}(\mathbf{d}, X)$ stands for the set of graphs on vertex set V that have degree sequence $\mathbf{d}$ and avoid all edges in X. Finally, $e(\mathbf{d})$ is the number of edges in any graph with $\mathbf{d}$ as degree sequence, $d_{\max}$ is the largest component of $\mathbf{d}$, and $x_{\max}$ is the largest degree of any vertex in the excluded graph (V, X).

Proof of Proposition 2.2 (Sketch) Our aim here is merely to indicate the algorithmic techniques used to sample from, and estimate the size of $\mathcal{G}(\mathbf{d}, X)$. For a full proof, the reader is directed to [14].

Using a reduction due to Tutte [19], a graph Γ is constructed whose perfect matchings are in (constant) many-one correspondence with elements of $\mathcal{G}(\mathbf{d}, X)$. An algorithm of Jerrum and Sinclair [12], based on the the simulation of a rapidly mixing Markov chain, is then used to sample or estimate the number of perfect matchings in Γ, as required. For this algorithm to be applicable, we require that Γ satisfy a certain condition; it is this condition, translated back through the reduction to the pair $(\mathbf{d}, X)$, that gives rise to the condition $e(\mathbf{d}) > d_{\max}(d_{\max} + x_{\max} - 1)$. $\square$

Proof of Theorem 2.1 The set of 2-factors in a graph $G = (V, E)$ is equal to $\mathcal{G}(\mathbf{d}, X)$, where $\mathbf{d} = (2, 2, \ldots, 2)$, and $X = V^{(2)} - E$ is the complementary edge set to E. The result now follows from Proposition 2.2, since, for a dense G and n sufficiently large, $d_{\max} = 2$, $x_{\max} < \frac{1}{2}n - 1$, and $d_{\max}(d_{\max} + x_{\max} - 1) < n = e(\mathbf{d})$. $\square$

Given Theorem 2.1, the reduction from Hamilton cycles to perfect matchings is easy to describe. We estimate first the number of 2-factors in G, and then the number of Hamilton cycles by standard sampling methods as a proportion of the number of 2-factors. Both counting and sampling phases run in polynomial time, by Theorem 2.1, provided only that G is dense. For the sampling phase to produce an accurate estimate, it is necessary that the ratio of 2-factors to Hamilton cycles in G not be too large. This will be established in Section 3.

We remark that it would be sufficient to be able to generate a random Hamilton cycle. We could then proceed alternatively by adding one edge at a time, giving a sequence of $M = O(n^2)$ graphs $G = G_0, G_1, \dots, G_M = K_n$. We could then estimate the ratio of the number of Hamilton cycles in G_{i-1} to those in G_i for $i = 1, 2, \dots, M$. The degree conditions can be used to show that each of these ratios is not too small and hence can be estimated efficiently. (This is similar to an idea in [4].)

The method of using random 2-factors to generate random Hamilton cycles was previously used by Frieze and Suen [9] in the context of *random digraphs* and more recently by Frieze, Jerrum and Molloy [8] with regard to random regular graphs. It is interesting that the same method should be successful here also. It raises the intriguing possibility of using existing approaches to other random graph problems to guide the design of new randomized algorithms for restricted versions of the corresponding deterministic problem.

3 Many 2-factors are Hamiltonian

Let n be a natural number and $\beta = 10/\alpha^2$. Let $k_0 = \lfloor \beta \ln n \rfloor$, and for $1 \le k \le n$, define $g(k) = n^\beta k! (\beta \ln n)^{-k}$, and

$$f(k) = \begin{cases} g(k), & \text{if } k \le k_0; \\ g(k_0), & \text{otherwise.} \end{cases}$$

Lemma 3.1 *Let f be the function defined above. Then*

1. *f is non-increasing and satisfies*

$$\min\{f(k-1), f(k-2)\} = f(k-1) \ge (\beta \ln n) k^{-1} f(k);$$

2. *$f(k) \ge 1$, for all k.*

Proof Observe that g is unimodal, and that k_0 is the value of k minimizing $g(k)$; it follows that f is non-increasing. When $k \le k_0$, we have $f(k-1) = g(k-1) = (\beta \ln n) k^{-1} g(k) = (\beta \ln n) k^{-1} f(k)$; otherwise, $f(k-1) = g(k_0) =$

$f(k) \geq (\beta \ln n)k^{-1} f(k)$. In either case, the inequality in part 1 of the lemma holds.

Part 2 of the lemma follows from the chain of inequalities

$$\frac{1}{f(k)} \leq \frac{1}{g(k_0)} \leq \frac{(\beta \ln n)^{k_0}}{n^\beta k_0!} \leq n^{-\beta} \sum_{k=0}^{\infty} \frac{(\beta \ln n)^k}{k!} = n^{-\beta} \exp(\beta \ln n) = 1.$$

□

Lemma 3.2 *Suppose α is constant greater than 0. Let $G = (V, E)$ be an undirected graph of order n and minimum degree $(\frac{1}{2} + \alpha)n$. Then the number of 2-factors in G exceeds the number of Hamilton cycles by at most a polynomial (in n) factor, the degree of the polynomial depending only on α.*

Proof For $1 \leq k \leq \lfloor n/3 \rfloor$, let Φ_k be the set of all 2-factors in G containing exactly k cycles, and let $\Phi = \cup_k \Phi_k$ be the set of all 2-factors. Define

$$\begin{aligned} \Psi \;\; = \;\; & \{(F, F') : F \in \Phi_k, F' \in \Phi_{k'}, k' < k, \\ & \text{and } F \oplus F' \cong C_6\}, \end{aligned}$$

where $\oplus$ denotes symmetric difference, and C_6 is the cycle on 6 vertices. Observe that (Φ, Ψ) is an acyclic directed graph; let us agree to call its component parts *nodes* and *arcs* to avoid confusion with the vertices and edges of G. Observe also that if $(F, F') \in \Psi$ is an arc, then F' can be obtained from F by deleting three edges and adding three others, and that this operation can decrease the number of cycles by at most two. Thus every arc $(F, F') \in \Psi$ is directed from a node F in some Φ_k to a node F' in Φ_{k-1} or Φ_{k-2}.

Our proof strategy is to define a positive weight function on the arc set Ψ such that the total weight of arcs leaving each node (2-factor) $F \in \Phi \setminus \Phi_1$ is at least one greater than the total weight of arcs entering F. This will imply that the total weight of arcs entering Φ_1 is an upper bound on the number of non-Hamilton 2-factors in G, and that the maximum total weight of arcs entering a single node in Φ_1 is an upper bound on the ratio $|\Phi \setminus \Phi_1|/|\Phi_1|$.

The weight function $w : \Psi \to \mathbb{R}^+$ we employ is defined as follows. For any arc (F, F') with $F' \in \Phi_k$: if the 2-factor F' is obtained from F by coalescing two cycles of lengths l_1 and l_2 into a single cycle of length $l_1 + l_2$, then $w(F, F') = (l_1^{-1} + l_2^{-1}) f(k)$; if F' results from coalescing three cycles of length l_1, l_2 and l_3 into a single one of length $l_1 + l_2 + l_3$, then $w(F, F') = (l_1^{-1} + l_2^{-1} + l_3^{-1}) f(k)$.

Let $F \in \Phi_k$ be a 2-factor with $k > 1$ cycles $\gamma_1, \gamma_2, \ldots, \gamma_k$, of lengths n_1, $n_2, \ldots, n_k$. We proceed to bound from below the total weight of arcs *leaving* F. For this purpose imagine that the cycles $\gamma_1, \gamma_2, \ldots, \gamma_k$ are oriented in some way, so that we can speak of each oriented edge (u, u') in some cycle γ_i as being "forward" or "backward". Since we are interested in obtaining a *lower* bound, it is enough to consider only arcs (F, F^+) from F of a certain kind: namely, those for which the 6-cycle $\gamma = F \oplus F^+$ is of the form $\gamma = (x, x', y, y', z, z')$, where $(x, x') \in F$ is a forward cycle edge, $(y, y') \in F$ is a forward edge in a cycle distinct from the first, and $(z, z') \in F$ is a backward cycle edge. The edge (z, z') may be in the same cycle as either (x, x') or (y, y'), or in a third cycle. Observe that (x', y), (y', z) and (z', x) must necessarily be edges of F^+. It is routine to check that any cycle $\gamma = (x, x', y, y', z, z')$ satisfying the above constraints does correspond to a valid arc from F. The fact that (z, z') is oriented in the opposite sense to (x, x') and (y, y') plays a crucial role in ensuring that the number of cycles decreases in the passage to F^+ when only two cycles involved.

First, we estimate the number of cycles γ for which (x, x') is contained in a particular cycle γ_i of F. We might say that γ is *rooted* at γ_i. Assume, for a moment, that the vertices x, x', y, y' have already been chosen. There are at least $(\frac{1}{2} + \alpha)n - 5$ ways to extend the path (x, x', y, y'), first to z and then to z', which are consistent with the rules given above; let Z' be the set of all vertices z' so reachable. Denote by $G(x)$ the set of vertices adjacent to x. The number of ways of completing the path (x, x', y, y') to a valid 6-cycle is at least

$$\begin{aligned} |G(x) \cap Z'| &\geq |G(x)| + |Z'| - n \\ &\geq (\tfrac{1}{2} + \alpha)n + [(\tfrac{1}{2} + \alpha)n - 5] - n \\ &= 2\alpha n - 5 \\ &\geq \alpha n, \end{aligned}$$

for n sufficiently large. A lower bound on the number of 6-cycles γ rooted at γ_i now follows easily: there are n_i choices for (x, x'); then at least $(\frac{1}{2} + \alpha)n - n_i$ choices for (y, y'); and finally — as we have just argued — at least αn ways to complete the cycle. Thus the total number of 6-cycles rooted at γ_i is at least $\alpha n n_i[(\frac{1}{2}+\alpha)n - n_i]$.

We are now poised to bound the total weight of arcs leaving F. Each arc (F, F^+) defined by a cycle γ rooted at γ_i has weight at least $n_i^{-1}\min\{f(k-1), f(k-2)\}$, which, by Lemma 3.1, is bounded below by $(\beta \ln n)(kn_i)^{-1}f(k)$. Thus the total weight of arcs leaving F is bounded as follows:

$$
\begin{aligned}
\sum_{F^+:(F,F^+)\in\Psi} w(F,F^+) &\geq \sum_{i=1}^{k} \alpha n n_i[(\tfrac{1}{2}+\alpha)n - n_i]\frac{(\beta\ln n)f(k)}{kn_i} && (1)\\
&= \alpha n^2[(\tfrac{1}{2}+\alpha)k - 1]\frac{(\beta\ln n)f(k)}{k}\\
&\geq \alpha^2\beta\, f(k)n^2\ln n\\
&\geq 10f(k)n^2\ln n, && (2)
\end{aligned}
$$

where we have used the fact that $k \geq 2$. Note that the presence of a unique backward edge, namely (z, z'), ensures that each cycle γ has a distinguishable root, and hence that the arcs (F, F^+) were not overcounted in summation (1).

We now turn to the corresponding *upper* bound on the total weight of arcs $(F^-, F) \in \Psi$ *entering* F. It is straightforward to verify that the cycle $\gamma = (x, x', y, y', z, z') = F^- \oplus F$ must contain three edges — (x, x'), (y, y') and (z, z') — from a single cycle γ_i of F, the remaining edges coming from F^-. The labeling of vertices in γ can be made canonical in the following way: assume an ordering on vertices in V, and assign label x to the smallest vertex. The condition $(x, x') \in F$ uniquely identifies vertex x', and the labeling of the other vertices in the cycle γ follows.

Removing the three edges (x, x'), (y, y') and (z, z') from γ_i leaves a triple of simple paths of lengths (say) $a-1$, $b-1$ and $c-1$: these lengths correspond (respectively) to the segment containing x, the segment containing x', and the remaining segment. Going round the cycle γ_i, starting at x' and ending at x, the vertices x, x', y, y', z, z' may appear in one of eight possible sequences:

$$
\begin{aligned}
&x', y', y, z', z, x;\\
&x', z, z', y, y', x;\\
&x', z, z', y', y, x;\\
&x', z', z, y, y', x;
\end{aligned}
$$

$$
\begin{array}{c}
x', y', y, z, z', x; \\
x', y, y', z', z, x; \\
x', z', z, y', y, x; \\
x', y, y', z, z', x.
\end{array}
$$

For a given triple of lengths (a, b, c), each of the above sequences corresponds to at most n_i possible choices for the edges (x, x'), (y, y') and (z, z'), yielding a maximum of $8n_i$ in total. To see this, observe that the edge (x, x') may be chosen in n_i ways (minimality of x fixes the orientation of the edge), and that the choice of (x, x') combined with the information provided by the sequence completely determines the triple of edges.

The eight sequences divide into five possible cases, as the first four sequences lead to equivalent outcomes (covered by case 1 below). Taken in order, the five cases are:

1. For at most $4n_i$ of the choices for the edges (x, x'), (y, y') and (z, z'), $\gamma_i \oplus \gamma$ is a single cycle;
2. for at most n_i choices, $\gamma_i \oplus \gamma$ is a pair of cycles of lengths a and $b + c$;
3. for at most n_i choices, $\gamma_i \oplus \gamma$ is a pair of cycles of lengths b and $a + c$;
4. for at most n_i choices, $\gamma_i \oplus \gamma$ is a pair of cycles of lengths c and $a + b$;
5. for at most n_i choices, $\gamma_i \oplus \gamma$ is a triple of cycles of lengths a, b and c.

The first case does not yield an arc (F^-, F), since the number of cycles does not decrease when passing from $F^- = F \oplus \gamma$ to F, but the other four cases do have to be reckoned with.

The total weight of arcs entering F can be bounded above as follows:

$$
\begin{aligned}
\sum_{F^- : (F^-, F) \in \Psi} w(F^-, F) \quad \leq \quad & \sum_{i=1}^{k} n_i f(k) \sum_{\substack{a,b,c \geq 1 \\ a+b+c=n_i}} \left[\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} \right) + \right. \\
& \left. \left(\frac{1}{a} + \frac{1}{b+c} \right) + \left(\frac{1}{b} + \frac{1}{a+c} \right) + \left(\frac{1}{c} + \frac{1}{a+b} \right) \right]
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{k} n_i f(k) \sum_{\substack{a,b,c\geq 1\\ a+b+c=n_i}} \left[\frac{6}{a} + \frac{3}{b+c}\right] \\
&\leq \sum_{i=1}^{k} n_i f(k) n \sum_{a=1}^{n_i-1} \left[\frac{6}{a} + \frac{3}{n_i - a}\right] \\
&\leq 9 f(k) n^2 H_n \qquad (3)
\end{aligned}
$$

where $H_n = \sum_{i=1}^{n} i^{-1} \leq \ln n + 1$ is the nth harmonic number [10, eq. (6.60)]. Combining inequalities (2) and (3), we have

$$
\begin{aligned}
\sum_{F^+:(F,F^+)\in\Psi} w(F,F^+) - \sum_{F^-:(F^-,F)\in\Psi} w(F^-,F) &\geq 10 f(k) n^2 \ln n - 9 f(k) n^2 H_n \\
&\geq f(k) n^2 (\ln n - 9) \\
&\geq n^2 (\ln n - 9),
\end{aligned}
$$

where the final inequality is by Lemma 3.1. Thus the total weight of arcs leaving F exceeds the total weight of arcs entering by at least 1, provided n is sufficiently large. The number of non-Hamilton 2-factors $|\Phi \setminus \Phi_1|$ is bounded above by the total weight of arcs entering Φ_1, which in turn is bounded — see inequality (3) — by $|\Phi_1| \times 9 f(1) n^2 H_n = |\Phi_1| \times O(n^{2+\beta})$. This establishes the lemma. □

4 Exact counting is #P-complete

Let #HC (resp. #HP) be the problem of counting the number of Hamilton cycles (resp. paths) in an undirected graph. It is known [21, 18] that #HC is #P-complete, and it follows, by an easy reduction, that #HP is also #P-complete.

Theorem 4.1 *Both #HC and #HP are #P-complete when restricted to graphs G of minimum degree at least $(1-\alpha)n$, where n is the number of vertices in G, and $\alpha > 0$.*

Proof We first present a Turing reduction from #HP to #HP such that all the target instances of #HP satisfy the required minimum degree condition.

Let $G = (V, E)$ be a undirected graph of order n with vertex set V and edge set E, considered as an instance of #HP. A typical target instance of #HP is a graph G_t constructed from G by forming the disjoint union of G with the complete graph K_t on $t \geq n$ vertices, and connecting every vertex in G with every vertex in K_t.

Assume $t \geq 3$. For $1 \leq k \leq n$, denote by P_k the set of all covers of G by k vertex-disjoint oriented paths, where paths of length 0 are allowed. Each oriented Hamilton path P in G_t induces an element of $\cup_k P_k$ by restriction to G. Conversely, each element of P_k may be extended in precisely $t!\binom{t+1}{k}k! = (t+1)!\binom{t}{k-1}(k-1)!$ ways to an oriented Hamilton path in G_t: the vertices in K_t may be visited in $t!$ orders; there are $\binom{t+1}{k}$ ways to choose k positions in that order during which excursions to G can be made, including the two positions prior to and following the order, and $k!$ ways to match those positions to the k oriented paths covering G. Thus the number p_t of oriented Hamilton paths in G_t can be expressed as the sum

$$p_t = \sum_{k=1}^{n} (t+1)! \binom{t}{k-1} (k-1)! |P_k|.$$

Note that p_t may be evaluated by one call to an oracle for #HP, since the number of oriented Hamilton paths is twice the number of unoriented paths. Using n such calls we may evaluate p_t for $t = t_0 + j$ and $j = 1, 2, \ldots, n$, where $t_0 = \lceil \alpha^{-1} n \rceil$ is chosen sufficiently large that every graph G_t with $t > t_0$ satisfies the minimum degree constraint. Recovering the values $\{(k-1)!|P_k| : 1 \leq k \leq n\}$ from $\{p_{t_0+j}/(t_0+j+1)! : 1 \leq j \leq n\}$ amounts to inverting the matrix

$$A = (A_{jk}) = \left(\binom{t_0 + j}{k-1} : 1 \leq j, k \leq n \right),$$

which may be expressed as the product $A = LU$ of a lower triangular matrix $L = (L_{jh})$ and upper triangular matrix $U = (U_{hk})$ defined as follows:

$$L_{jh} = \binom{j-1}{h-1} \quad \text{and} \quad U_{hk} = \binom{t_0+1}{k-h}.$$

The equality $A = LU$ is a direct consequence of the "Vandermonde convolution" formula [10, eq. (5.22)]

$$\sum_{h=1}^{n} \binom{j-1}{h-1}\binom{t_0+1}{k-h} = \binom{t_0+j}{k-1}.$$

Both L and U have unit diagonals and are hence non-singular: indeed their inverses have the following simple explicit forms, as can be verified by direct multiplication using standard identities involving sums of products of binomial coefficients [10, eqs (5.24), (5.25)]:

$$(L^{-1})_{hk} \quad = \quad (-1)^{h+k}\binom{h-1}{k-1}$$

and

$$(U^{-1})_{jh} \quad = \quad (-1)^{j+h}\binom{t_0+h-j}{t_0+1}.$$

Since $A^{-1} = U^{-1}L^{-1}$, the values $\{|P_k| : 1 \leq k \leq n\}$ may be computed in polynomial time using two matrix multiplications involving integers of $O(n \log n)$ bits. Observe that $\frac{1}{2}|P_1|$ gives the number of (unoriented) Hamilton paths in G.

The hardness of #HC is now simple to verify. Given a graph $G = (V, E)$ with the minimum degree condition we add a new vertex x and edges (x, v) for all $v \in V$ to create G'. Note the G' satisfies the minimum degree condition as well. Removing x from a Hamilton cycle in G' creates a Hamilton path in G. This defines a bijection from the set of Hamilton cycles in G' to the set of Hamilton paths in G. □

5 Counting the number of cycles of all sizes

We will first consider approximating the total number of cycles in graphs with minimum degree $(\frac{1}{2} + \alpha)n$.

We first note that if we add a loop to each vertex and extend the definition of 2-factor to include loops as cycles of length one, then the argument of [14] may be extended to this case (note that we still forbid cycles of length two i.e. double edges). Thus there exists both a fully polynomial randomized approximation scheme and a fully polynomial almost uniform sampler for the set of *partial* 2-factors in a dense graph. Let a partial 2-factor be *cyclic* if it consists of a single cycle of length at least three and a collection of loops. Clearly the number of cyclic partial 2-factors is the same as the number of cycles.

The procedure for approximating the number of cycles of all sizes is as follows: we estimate first the number of partial 2-factors in G, and then the number of cyclic partial 2-factors by standard sampling methods as a proportion of the number of partial 2-factors. To produce an accurate estimate in polynomial time it is only necessary to show that the ratio of partial 2-factors to cyclic partial 2-factors is not too large. Let

$$\mathcal{F}_\ell = \{\text{partial 2-factors with } \ell \text{ loops}\}, \quad \text{and} \quad f_\ell = |\mathcal{F}_\ell|.$$

For a given $F \in \mathcal{F}_\ell$ let $L = \{\text{loops of } F\}$, which we will now identify with the corresponding set of vertices. For $v \in L$ let d_v denote the number of neighbours of v in L and $D = \sum_{v \in L} d_v$.

If $v \in L$ then there are at least $2\alpha n - 2d_v$ ways of adding v to a cycle C of F by deleting an edge (a, b) of C and adding edges $(a, v), (v, b)$. Indeed we go round each cycle C of F; if the successor b of a vertex a neighbouring v is also a neighbour of v, then it forms an (a, b, v) triangle. The number of such triangles is at least $2\alpha n - 2d_v$.

So in total there are at least

$$\begin{aligned} \sum_{v \in L} (2\alpha n - 2d_v) &= 2\ell\alpha n - 2D & (4) \\ &\geq 2\ell(\alpha n - (\ell - 1)) & (5) \end{aligned}$$

such augmentations.

Suppose first that $\ell \leq \ell_1 = \lfloor \alpha n/2 \rfloor$. Then (5) gives at least $\ell \alpha n$ augmentations of $F \in \mathcal{F}_\ell$ to an $F' \in \mathcal{F}_{\ell-1}$. Each $F' \in \mathcal{F}_{\ell-1}$ arises in at most n ways and so

$$\frac{f_{\ell-1}}{f_\ell} \geq \alpha \ell.$$

Putting $\ell_0 = \lceil 2/\alpha \rceil$ we see that

$$f_{\ell_1} + f_{\ell_1 - 1} + \cdots + f_{\ell_0+1} \leq f_{\ell_0} \leq f_{\ell_0} + f_{\ell_0 - 1} + \cdots + f_0. \tag{6}$$

Suppose next that $\ell > \ell_1$. Note first that since a graph with r vertices and s edges contains at least $r - s + 1$ distinct cycles, we see that L contains at least

$$\frac{D}{2} - \ell + 1 \tag{7}$$

distinct cycles.

Adding a cycle C contained in L to F and removing $|C|$ loops gives us a 2-factor in $\mathcal{F}_{\ell'}$ where $\ell' < \ell$. From (4) and (7) we see that there are at least[2]

$$\begin{aligned} \left(\frac{2\ell\alpha n - 2D}{4}\right)^+ + \left(\frac{D}{2} - \ell\right)^+ &\geq \ell\left(\frac{\alpha n}{2} - 1\right) && (8) \\ &\geq \frac{\ell\alpha n}{3} && (9) \end{aligned}$$

augmentations of either sort from F. Each $F' \in \mathcal{F}_{<\ell}$ arises in at most $n + n$ ways (accounting for both ways of reducing L) and so

$$\begin{aligned} f_\ell &\leq \frac{6}{\alpha\ell}(f_{\ell-1} + f_{\ell-2} + \cdots + f_0) \\ &\leq \theta(f_{\ell-1} + f_{\ell-2} + \cdots + f_0), \end{aligned}$$

where $\theta = 12/(\alpha^2 n)$, assuming $\ell > \ell_1$.

Thus

$$\frac{f_\ell + f_{\ell-1} + \cdots + f_0}{f_{\ell-1} + f_{\ell-2} + \cdots + f_0} \leq 1 + \theta$$

and so

$$f_\ell + f_{\ell-1} + \cdots + f_0 \leq (1+\theta)^{\ell - \ell_1}\Sigma_1, \tag{10}$$

where $\Sigma_1 = f_{\ell_1} + f_{\ell_1 - 1} + \cdots + f_0$. We weaken (10) to

$$\begin{aligned} f_{\ell_1 + k} &\leq (1+\theta)^k \Sigma_1 \\ &\leq e^{12\alpha^{-2}}\Sigma_1. && (11) \end{aligned}$$

[2] $x^+ = \max\{0, x\}$

It follows from (6) and (11) that

$$\frac{f_0 + f_1 + \cdots + f_n}{f_0 + f_1 + \cdots + f_{\ell_0}} \leq 2 + 2ne^{12\alpha^{-2}}. \tag{12}$$

Now take an $F \in \mathcal{F}_\ell$ where $\ell \leq \ell_0$ and fix its set of loops L. The number of partial 2-factors with this same L is at most a polynomial factor, $p(n)$ say, of the number of cycles of size $n - \ell$ through $V \setminus L$, by the results of Section 3. (It is clear that because ℓ is small here, the required degree conditions are satisfied.) Thus, by (12), the ratio of partial 2-factors to cyclic partial 2-factors is $O(np(n))$ and we have proved the existence of an fpras for the number of cycles.

6 Paths and Hamilton Paths

We obtain an fpras for counting the number of Hamilton paths in the following way. We add a vertex v_0 and join it by an edge to every vertex of G. Call this new graph G^*. The number of Hamilton cycles in G^* is equal to the number of Hamilton paths in G. Since G^* is dense we can approximate the latter quantity by approximating the former.

Similarly, to estimate the number of paths of all lengths, we compute an estimate c^* for the number of cycles in G^* and an estimate ρ^* for the proportion ρ of cycles which contain v_0. Since the number of cycles containing v_0 is the number of paths in G, this provides an estimate $\rho^* c^*$ for the number of paths. Also, this will give us an fpras provided ρ is not too small. But clearly $\rho \geq 3/4$ and we are done.

7 Concluding remarks

We remark that it is not difficult to adapt the above methods to the corresponding *directed* case. Here we will have both minimum indegree and outdegree at each vertex guaranteed to be at least $(\frac{1}{2} + \alpha)n$. Also we may similarly count the number of connected k-factors in G for any $k = o(n)$. (Hamilton cycles are, of course, connected 2-factors.)

We leave open the following questions. First, is it possible to count approximately as $\alpha \to 0$ in any fashion? Secondly, is there a random walk on Hamilton cycles and (in some sense) "near-Hamilton-cycles" which is rapidly mixing? In other words, can we avoid the Tutte construction and the need for 2-factors with many cycles?

Finally, are there other interesting counting problems which are tractable on such dense graphs? Note that Annan [1] has recently found an fpras for the number of spanning forests in a dense graph. This can easily be modified to approximate the total number of (not necessarily spanning) trees in a dense graph. On the other hand Jerrum [11] has recently shown that the problem of computing this number for a general graph is #P-Complete.

Acknowledgement We thank Alistair Sinclair for his comments on an earlier draft.

References

[1] J.D.Annan, *A randomised approximation algorithm for counting the number of forests in dense graphs*, to appear.

[2] B.Bollobás, *Extremal graph theory*, Academic Press, 1978.

[3] G. R. Brightwell and P. M. Winkler, *Counting linear extensions*, Order **8** (1991) 225–242.

[4] A. Z. Broder, *How hard is it to marry at random? (On the approximation of the permanent)*, Proceedings of the 18th ACM Symposium on Theory of Computing, 1986, 50–58. Erratum in Proceedings of the 20th ACM Symposium on Theory of Computing, 1988, p. 551.

[5] G. A. Dirac, *Some theorems on abstract graphs*, Proceedings of the London Mathematical Society **2** (1952) 69–81.

[6] M. E. Dyer and A. M. Frieze, *On the complexity of computing the volume of a polyhedron*, SIAM Journal on Computing **17** (1988) 967–974.

[7] M. E. Dyer, A. M. Frieze, and R. Kannan, *A random polynomial time algorithm for approximating the volume of convex bodies*, Journal of the ACM **38** (1991) 1–17.

[8] A. M. Frieze, M. R. Jerrum, M. Molloy, R. Robinson and N. Wormald *Generating and counting Hamilton cycles in random regular graphs*, Report ECS-LFCS-94-313, Department of Computer Science, University of Edinburgh, December 1994.

[9] A. M. Frieze and S. Suen, *Counting Hamilton cycles in random directed graphs*, Random Structures and Algorithms **3** (1992) 235–242.

[10] R. L. Graham, D. E. Knuth and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, Addison-Wesley, Reading MA, 1989.

[11] M. R. Jerrum, *Counting trees in a graph is #P-Complete*, Information Processing Letters **51** (1994) 111–116.

[12] M. R. Jerrum and A. J. Sinclair, *Approximating the permanent*, SIAM Journal on Computing **18** (1989) 1149–1178.

[13] M. R. Jerrum and A. J. Sinclair, *Polynomial-time approximation algorithms for the Ising model*, SIAM Journal on Computing **22** (1993) 1087–1116.

[14] M. Jerrum and A. Sinclair, *Fast uniform generation of regular graphs*, Theoretical Computer Science **73** (1990) 91–100.

[15] R. M. Karp and M. Luby, *Monte-Carlo algorithms for enumeration and reliability problems*, Proceedings of the 24th IEEE Symposium on Foundations of Computer Science, 1983, 56–64.

[16] A. Karzanov and L. G. Khachiyan, *On the conductance of order Markov chains*, Technical Report DCS TR 268, Rutgers University, 1990.

[17] M. Mihail and P. Winkler, *On the number of Euler orientations of a graph*, Proceedings of the 3rd Annual ACM-SIAM Symposium on Discrete Algorithms, 1992 138–145.

[18] J. Simon, *On the difference between one and many*, Proceedings of the Fourth International Colloquium on Automata, Languages and Programming, Lecture Notes in Computer Science 52, Springer-Verlag, 1977, 480–491.

[19] W. T. Tutte, *A short proof of the factor theorem for finite graphs*, Canadian Journal of Mathematics **6** (1954) 347–352.

[20] L. G. Valiant, *The complexity of computing the permanent*, Theoretical Computer Science **8** (1979) 189–201.

[21] L. G. Valiant, *The complexity of enumeration and reliability problems*, SIAM Journal on Computing **8** (1979) 410–421.