Available online at www.sciencedirect.com
SCIENCE DIRECT®
Discrete Applied Mathematics 152 (2005) 139–158
DISCRETE APPLIED MATHEMATICS
www.elsevier.com/locate/dam

Hamiltonian completions of sparse random graphs

David Gamarnik, Maxim Sviridenko
IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA

Received 18 March 2003; received in revised form 5 May 2005; accepted 6 May 2005
Available online 14 July 2005

Abstract

Given a (directed or undirected) graph G, finding the smallest number of additional edges which make the graph Hamiltonian is called the Hamiltonian Completion Problem (HCP). We consider this problem in the context of sparse random graphs $G(n, c/n)$ on n nodes, where each edge is selected independently with probability c/n. We give a complete asymptotic answer to this problem when $c < 1$, by constructing a new linear time algorithm for solving HCP on trees and by using generating function method. We solve the problem both in the cases of undirected and directed graphs.

Keywords: Hamiltonicity; Travelling Salesman Problem; Random graphs

1. Introduction and the main result

Consider a (undirected or directed) graph G on n nodes. How many extra edges, which are not originally present in the graph, do we need to add in order the make the graph Hamiltonian? This is called the Hamiltonian Completion Problem (HCP), and the minimal number of extra edges is defined to be the Hamiltonian Completion Number (HCN). Hamiltonicity itself is then a decision version of this problem—the problem of checking whether the optimal value of HCP is zero. In particular, HCP problem is NP-hard. Several papers studied HCP problem in various graphs with some special structures, for example trees and line graphs of trees [1,6,5,3,10]. Specifically, a linear time algorithm for computing HCN was constructed by Goodman et al. [6] for the case of undirected trees.

To the best of our knowledge HCP was never studied in the context or random graphs. To the contrary, Hamiltonicity was investigated very intensively in a variety of random

E-mail addresses: gamarnik@watson.ibm.com (D. Gamarnik), sviri@us.ibm.com (M. Sviridenko).

0166-218X/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.dam.2005.05.001

graph models, starting from a classical work by Beardwood et al. [2] on optimal Travelling Salesman tours in random planar graphs. We refer the reader to Frieze and Yukich [4] for a very good survey on this subject. One of the most interesting results in this area was obtained by Posa [9]. Solving a problem, which was open for 20 years, he showed that the random graph $G(n, c/n)$, where each edge is selected independently with probability c/n, is Hamiltonian when $c \geqslant 16 \log n$. Komlos and Szemeredi [8] later tightened this bound by proving Hamiltonicity for $c = \log n$. Essentially they showed that, as c is increasing the random graph becomes Hamiltonian with high probability (w.h.p.) precisely when its minimal degree becomes two, w.h.p., which occurs at the threshold $c = \log n$. Interestingly, random regular graphs with degree at least three are Hamiltonian w.h.p. [7]. An interesting open problem remains determining the threshold for Hamiltonicity in random subgraphs of a binary cube $\{0, 1\}^n$, where edges between pairs of nodes with Hamming distance 1 are included with probability p, and between all other pairs with probability 0. It is conjectured that the threshold value is $p = \frac{1}{2}$.

In this paper we study HCP problem in the context of sparse undirected and directed random graphs $G = G(n, c/n)$ on n nodes, where, in the undirected case, every edge $(i, j),\ i, j \leqslant n$ is included with probability c/n, independently for all edges, and $c < 1$ is some fixed constant. For the directed case, we take our undirected random graph model and give every edge a random orientation, with equal probability $\frac{1}{2}$. Again we assume $c < 1$. It is well known [7], that such random graphs are disconnected w.h.p. Moreover, w.h.p., most of the (weak) components of this graph are trees. We obtain a complete asymptotic solution of HCP in these graphs, as $n \to \infty$. It is easy to see that $\mathbb{E}[H(n, c)] = \Theta(n)$, where $H(n, c)$ denotes the optimal value of the HCN and $\mathbb{E}[\cdot]$ is the expectation operator. Indeed, $\mathbb{E}[H(n, c)] = O(n)$, since we can simply plant a Hamiltonian tour. On the other hand, w.h.p. there exists linearly many isolated nodes [7]. As a result, we need at least $\Omega(n)$ extra edges. In this paper we prove the existence and compute the limit of $\lim_n \mathbb{E}[H(n, c)]/n$. Our method of proof is based on constructing a new and simple linear time algorithm for solving HCP on trees. In the case of directed trees, our algorithm is, to the best of our knowledge, the first algorithm for solving HCP on directed trees. As we mentioned above, such algorithms exist for undirected graphs, and some of them have linear complexity. Yet we found that these algorithms are not useful for the analysis of HCP in random graphs. Our algorithms turn our to be far more amenable for the analysis of random instances, thanks to certain recursive properties.

In order to state our main theorem, we need the following technical result.

Proposition 1. *Fix an arbitrary value* $c \in (0, 1)$. *For every pair* $0 \leqslant x < 1, 0 \leqslant y \leqslant 1$ *satisfying*

$$1 \geqslant y \geqslant \max\left\{1 - \frac{x\mathrm{e}^{-c}}{2}, \frac{1+x}{2}\right\} \tag{1}$$

the system of equations and inequalities in variables $g_0, g_1 \geqslant 0$:

$$g_1 = \frac{xy\mathrm{e}^{cg_0 - c}}{1 - cx\mathrm{e}^{cg_0 - c}}, \tag{2}$$

$$g_0 = \frac{x}{y}\,\mathrm{e}^{cg_0-c}(\mathrm{e}^{cg_1} - 1 - cg_1), \tag{3}$$

$$g_0 + g_1 \leqslant 1 \tag{4}$$

has exactly one solution.

Proof. The existence of a solution will follows from the developments in later sections, where we show that a generating function of a certain two-dimensional random variable satisfies (2), (3), (4). We now prove uniqueness. Rewrite the Eq. (2) as

$$g_1 - g_1 cx\mathrm{e}^{cg_0-c} = xy\mathrm{e}^{cg_0-c} \tag{5}$$

and add to (3) multiplied by y to obtain

$$g_1 + yg_0 = x\mathrm{e}^{cg_0+cg_1-c} - x(1-y)\mathrm{e}^{cg_0-c}$$

which we rewrite as

$$g_0 + g_1 = x\mathrm{e}^{cg_0+cg_1-c} - x(1-y)\mathrm{e}^{cg_0-c} + (1-y)g_0.$$

We introduce an independent variable $t = g_0$ and consider a function $h(t)$ implicitly defined by

$$h(t) = x\mathrm{e}^{ch(t)-c} - x(1-y)\mathrm{e}^{ct-c} + (1-y)t. \tag{6}$$

The function $h(t)$ stands for $g_0 + g_1$. Our first claim is that for all $0 \leqslant t \leqslant 1$ there exists exactly one solution $h(t)$ satisfying $h(t) \leqslant 1$, that is satisfying inequality (4). Indeed, the left-hand side of (6) is a linear function of $h = h(t)$ taking values 0 and 1 when $h = 0, 1$ respectively. The right-hand side is a convex function of h. When $h = 0$, its value is $x\mathrm{e}^{-c} - x(1-y)\mathrm{e}^{ct-c} + (1-y)t \geqslant x\mathrm{e}^{-c} - x(1-y) \geqslant x\mathrm{e}^{-c} - (1-y) > 0$, by assumption (1). On the other hand, when $h = 1$, the corresponding value is at most $x + 1 - y$, which is strictly smaller than 1, since $x < 1$ and therefore by assumption (1), $y \geqslant (1+x)/2 > x$. Thus, indeed there exists exactly one solution $h(t) \leqslant 1$ for all $0 \leqslant t \leqslant 1$.

The rest of the argument is structured as follows. We obtained that each value of g_0 uniquely specifies the value of g_1 via $g_1 = h(t) - t = h(g_0) - g_0$. We will show that, moreover, $g_1 = h(t) - t$ is a decreasing function of $g_0 = t$. On the other hand observe that (5) uniquely specifies g_1 as a function of g_0, and, moreover, this function is non-decreasing. Therefore these two functions of g_0 can have at most one intersection, and the proof would be completed.

Our next claim is that the function $h(t)$ satisfies $\dot{h}(t) < 1$. This implies that the function $h(t) - t$ is strictly decreasing, and we would be done. Differentiating both sides of (6) and rearranging we obtain

$$\dot{h}(t)(1 - xc\mathrm{e}^{ch(t)-c}) = -x(1-y)c\mathrm{e}^{ct-c} + (1-y) \leqslant 1 - y.$$

Therefore

$$\dot{h}(t) \leqslant \frac{1-y}{1 - xc\mathrm{e}^{ch(t)-c}} \leqslant \frac{1-y}{1-x} < 1,$$

since $c < 1$, $h(t) \leqslant 1$ and by (1), $y > x$. This completes the proof. □

We will show later that the unique solution $g = g_0 + g_1$ is in fact a generating function of some two-dimensional random vector, corresponding to the HCP in undirected random graphs. The corresponding sequence of equations and inequalities for directed random graphs involves variables g_{00}, g_{01}, g_{11} and is as follows:

$$g_{11} = xy\mathrm{e}^{cg_{00}+cg_{01}-c}, \tag{7}$$

$$g_{01} = x\mathrm{e}^{cg_{00}+\frac{3c}{2}g_{01}+\frac{c}{2}g_{11}-c} - x\mathrm{e}^{cg_{00}+cg_{01}-c}, \tag{8}$$

$$g_{00} = \frac{x}{y}\,\mathrm{e}^{cg_{00}+2cg_{01}+cg_{11}-c} - \frac{2x}{y}\,\mathrm{e}^{cg_{00}+\frac{3c}{2}g_{01}+cg_{11}-c} + \frac{x}{y}\,\mathrm{e}^{cg_{00}+cg_{01}-c}, \tag{9}$$

$$g_{00} + 2g_{01} + g_{11} \leqslant 1. \tag{10}$$

Like in the undirected case, we will show that a certain generating function satisfies these equations and inequalities. Therefore, this system has at least one solution for each $0 < c < 1,\ 0 \leqslant x < 1,\ 0 \leqslant y \leqslant 1$. We were not able, unfortunately, to prove the uniqueness of the solution, but our numerical computations do show the uniqueness. We leave the uniqueness as an open question.

We define functions $g_0(x, y),\ g_1(x, y)$ and $g_{00}(x, y),\ g_{01}(x, y),\ g_{11}(x, y)$ as the unique solutions to the systems of equations and inequalities (2), (3), (4) and (7), (8), (9), (10) respectively, (uniqueness conjectured in the second case) and let

$$g(x, y) \equiv g_0(x, y) + g_1(x, y), \quad \bar{g}(x, y) = g_{00}(x, y) + 2g_{01}(x, y) + g_{11}(x, y). \tag{11}$$

The main result of the paper is stated below.

Theorem 1. *For $c < 1$, the optimal value of the HCP for an undirected random graph $G(n, c/n)$ satisfies*

$$\lim_{n\to\infty} \frac{\mathbb{E}[H(n, c)]}{n} = \int_0^1 \frac{g_y(x, 1)}{x}\,\mathrm{d}x, \tag{12}$$

where $g_y(x, y) = [\partial/\partial y]g(x, y)$. For a directed random graph $G(n, c/n)$ the corresponding value satisfies

$$\lim_{n\to\infty} \frac{\mathbb{E}[H(n, c)]}{n} = \int_0^1 \frac{\bar{g}_y(x, 1)}{x}\,\mathrm{d}x, \tag{13}$$

where $\bar{g}_y(x, y) = [\partial/\partial y]\bar{g}(x, y)$. Both partial derivatives and integrals exist and are finite.

Observe that the value of $[\partial/\partial y]g(x, y)$ for $y = 1$ is completely determined by the values of the function $g(x, y)$ in the region where y is closed to the unity. This region, in particular, is covered by the region specified by the constraint (1). This is why for the purposes of solving the HCP, the uniqueness within the region (1) suffices.

The value of the integrals above can be computed approximately by numerical methods. We will report the results of computations in Section 5. The rest of the paper is organized as follows. In Section 2, we analyze HCN of a fixed deterministic tree. Two subsections correspond to the cases of undirected and directed graphs. We obtain a linear time algorithms for the optimal values of HCP in trees and we show that the optimal value of HCP for a forest is the sum of the optimal values of its individual trees. In Section 3 we use a classical fact from the theory of random graphs that a given fixed node of a random graph $G(n, c/n)$ belongs w.h.p. to a component which, in the limit as $n \to \infty$, is a random Poisson tree. We obtain an exact distribution of the optimal value of HCP of a random Poisson tree, via its generating function. We use this result in Section 4, to complete the proof of Theorem 1. In Section 5 we provide numerical results of the computing the limits (12), (13).

2. Hamiltonian completion of a tree

2.1. Undirected graphs

Let T be a non-random tree with a selected root $r \in T$. We denote by $H(T)$ the HCN of T. Note, that there are possibly several solutions which achieve $H(T)$. We say that the rooted graph (T, r) is type 0, if for *every* optimal solution, both edges incident to r in the resulting Hamiltonian tour belong to T. Otherwise, the pair is defined to be type 1. We also define the root r to be type 0 (type 1), if (T, r) is type 0 (type 1). Any isolated node i is defined to be type 1, and $H(i)$, for convenience, is defined to be 1, by definition.

An example of type 0 tree is a path $T = (r_1, r_2, \ldots, r_t)$, where r is any internal node r_i, $2 \leqslant i \leqslant t-1$. Indeed, the HCN for this graph is 1—add edge (r_t, r_1), and this is a unique optimal solution. The resulting Hamiltonian tour $r_1, r_2, \ldots, r_t, r_1$ (or the reverse tour) uses edges (r_{i-1}, r_i), (r_i, r_{i+1}) incident to $r = r_i$, both of which belong to the tree. On the other hand if $r = r_1$ ($r = r_t$), the pair (T, r) is type 1, since the generated tour T uses a new edge (r_t, r_1) incident to r.

Consider the complete weighted graph G_T on the same vertex set as the tree T, and define weight of an edge to be 0 if this edge belongs to tree T and 1 otherwise. Then the Hamiltonian Completion Problem in tree T is equivalent to the Travelling Salesman Problem in graph G_T and the optimal value of a TSP tour in G_T is equal to a number of edges we need to add to make graph T Hamiltonian. We now prove an auxiliary lemma about a property of the optimal Hamiltonian cycle in G_T. Denote by $T_1, \ldots, T_d$ the subtrees generated by the children $r_1, r_2, \ldots, r_d$ of r in T.

Lemma 2. *For any tour of length H in a G_T which uses $s = 0, 1, 2$ edges of weight 0 incident to r (in other words edges incident to the root r in T) there is a tour in G_T with length at most H, which uses s edges of weight 0 incident to r and visits each subtree $T_1, \ldots, T_d$ exactly ones, i.e. vertices of every subtree $T_1, \ldots, T_d$ form contiguous segments of the Hamiltonian tour in G_T.*

Proof. Indeed, if there are two such contiguous segments $P_1 = (i_1, \ldots, i_p)$ and $P_2 = (i_q, \ldots, i_m)$ belonging to the same subtree T_i and not connected by an edge in a Hamiltonian cycle then at least three out of four edges incident to this segments in a Hamiltonian cycle have weight 1 since there is at most one edge of weight 0 incident to a subtree T_i. Assume, that (i_p, i_{p+1}), (i_{q-1}, i_q) and (i_m, i_{m+1}) are these edges. Therefore, for $P_{12} = i_{p+1}, \ldots, i_{q-1}$, the part of the Hamiltonian cycle between P_1 and P_2, and for

$P_{21} = i_{m+1}, \ldots, i_0$, the part of the Hamiltonian cycle betweenP_2 and P_1, the new tour P_1, P_2, P_{12}, P_{21} has length at most H since we took out three edges (i_p, i_{p+1}), (i_{q-1}, i_q) and (i_m, i_{m+1}) of weight 1 and used instead edges (i_p, i_q), (i_m, i_{p+1}) and (i_{q-1}, i_{m+1}) of weight at most 1. Repeating these procedure we get a Hamiltonian cycle in G_T with the desired properties. □

In the following result we related the optimal values of HCP on trees and forests.

Proposition 2. *The optimal value of HCP on a forest is the sum of the optimal values of HCP of its tree components*.

Proof. The proof is almost exactly the same is of Lemma 2 above. We show that there exists an optimal tour which visits each component of the forest exactly ones. □

The proposition below is the key technical result of this subsection. Here we assume a non-trivial case when the degree of the root r in T is at least 1.

Proposition 3. *The following holds*:

1. *If there are at least two pairs out of* (T_i, r_i), $1 \leqslant i \leqslant d$ *which are type* 1, *then the pair* (T, r) *is type* 0 *and* $H(T) = -1 + \sum_{i=1}^{d} H(T_i)$.
2. *If exactly one of the pairs* (T_i, r_i) *is type* 1, *then* (T, r) *is type* 1 *and* $H(T) = \sum_{i=1}^{d} H(T_i)$.
3. *If all of the pairs* (T_i, r_i) *are type* 0, *then* (T, r) *is type* 1 *and* $H(T) = 1 + \sum_{i=1}^{d} H(T_i)$.

Remark. An immediate corollary of the recursion above is a linear time algorithm (in size n of the tree) for solving HCP on trees.

Proof. We consider the three cases from the claim of the proposition.

Case 1: Assume that there are at least two pairs out of (T_i, r_i) which are type 1. W.l.o.g. let T_1 and T_d be two of these trees. Let $C_1, \ldots, C_d$ be optimal Hamiltonian cycles in G_{T_i}, $i = 1, \ldots, d$ of length $H(T_i)$ such that C_1 and C_d have edges of weight 1 incident to r_1 and r_d. Delete these two edges from C_1 and C_d. Delete one arbitrary edge of length 1 from each $C_2, \ldots, C_{d-1}$. After that connect the path in T_1 obtained from C_1 with the path in T_d obtained from C_d by two edges of weight 0 through the root vertex x and connect remaining paths in any order by $d-1$ edges of weight 1 into a Hamiltonian cycle in G_T. Clearly, these new Hamiltonian cycle has length exactly $-1 + \sum_{i=1}^{d} H(T_i)$ since we deleted one edge of weight 1 from every subtree and added exactly $d-1$ edges of weight one to the Hamiltonian cycle. One the other hand, $-1 + \sum_{i=1}^{d} H(T_i)$ is a lower bound for every Hamiltonian cycle in G_T, since, by Lemma 2 this Hamiltonian cycle must contain a Hamiltonian path for each T_i of length at least $H(T_i) - 1$ and $d-1$ edges of weight 1 between subtrees. Moreover, we can achieve this lower bound only if Hamiltonian path uses two edges of weight 0 incident to the root r to connect two subtrees. Therefore, pair (T, r) has type 0.

Case 2: Assume that exactly one of the pairs (T_i, r_i) is type 1 and assume that this is (T_1, r_1). Then deleting the edge of weight 1 incident to r_1 from C_1, and arbitrary edge of weight 1 from each $C_2, \ldots, C_d$, connecting the Hamiltonian path in C_1 with the root r by the edge of weight 0 and all other Hamiltonian paths into one cycle by d edges of weight 1 we get a cycle of length exactly $\sum_{i=1}^{d} H(T_i)$. Since the constructed tour contains the edge of weight 1 incident to the root r, what we need to prove is that there is no Hamiltonian cycle in G_T of weight $-1+\sum_{i=1}^{d} H(T_i)$. Assume on the contrary that there is such a tour C. Then, as we noticed in the previous paragraph, it must use two edges of weight 0 incident to the root r. Let T_k and T_t be subtrees connected by these edges with the root. Let (T_t, r_t) be the pair of type 0 since both of these pairs (T_k, r_k) and (T_t, r_t) cannot be of type 1. Then subpath through the subtree T_t in a cycle C cannot have length less than $H(T_t)$ since otherwise connecting two endpoints of such path we will either get a tour of weight less than $H(T_t)$ or a tour of weight exactly $H(T_t)$ but with edge of weight 1 incident to r_t and then (T_t, r_t) would be of type 1. Therefore, we have at least one subtree T_t which contributes $H(T_t)$ to the length of C, by adding at least $H(T_i)-1$ for all other trees and $d-1$ to connect all subtours into one Hamiltonian cycle. We get that the tour length is at least $\sum_{i=1}^{d} H(T_i)$. Contradiction.

Case 3: In the last case we assume that all of the pairs (T_i, r_i) are type 0. Then we can easily obtain a Hamiltonian cycle in G_T by deleting one edge of length one in each C_i, $i=1,\ldots,d$ and adding $d+1$ edges of weight 1 connecting resulting paths and x into Hamiltonian cycle in G_T. Clearly, the length of this cycle is $1+\sum_{i=1}^{d} H(T_i)$ and it has an edge of weight 1 incident to r and therefore, what we need to show is that there is no Hamiltonian cycle in G_T of length $\sum_{i=1}^{d} H(T_i)$ or less. The argument is very similar to the one in the previous paragraph. Assume that there is a tour of length smaller than $1+\sum_{i=1}^{d} H(T_i)$. It cannot have two edges of weight 1 incident to r since otherwise this cycle has the weight at least $2+\sum_{i=1}^{d} (H(T_i)-1)+d-1$. Suppose it has one edge of weight 1 incident to r. Let T_t be a subtree connected by this edge with the root. The subpath of Hamiltonian cycle in this subtree must have the length at least $H(T_t)$ and therefore adding $H(T_i)-1$ for all other subtrees and d edges of weight one to connect subtours in different subtrees we get $1+\sum_{i=1}^{d} H(T_i)$, again. Finally, if there are two edges of weight 0 incident to r in a Hamiltonian cycle then let T_k and T_t be subtrees connected by these edges with the root r. Therefore, they will contribute $H(T_k)$ and $H(T_t)$ to the length of the Hamiltonian cycle plus $H(T_i)-1$ for all other trees and $d-1$ to connect all paths in subtrees into one cycle. Again, we obtain $1+\sum_{i=1}^{d} H(T_i)$. □

2.2. Directed graphs

Let T be a non-random directed acyclic rooted graph obtained from some undirected tree by orienting its edges in some way. A graph is defined to be a directed forest if all of its weakly connected components are directed trees. Given a directed tree T, let $r \in T$ denote the root of this tree, and, as above, let $H(T)$ denote the HCN of T. We will say that the rooted graph (T, r) is type $(0, out)$, if for *every* optimal solution, the oriented edge outgoing from r in the resulting Hamiltonian tour, belongs to T. In this case we will also say that the root r is type $(0, out)$. If pair (T, r) is not type $(0, out)$, then it is said to be of type $(1, out)$, i.e. there is an optimal Hamiltonian tour such that the edge outgoing from r does not belong to T. If for *every* optimal solution, the edge incoming to r in the resulting Hamiltonian tour belongs to T then the pair (T, r) has type $(0, in)$ and, otherwise, the pair is said to be of type $(1, in)$.

Any isolated node i is said to be of type $(1, in)$ and $(1, out)$. For a later convenience, $H(i)$ is set to be 1, by definition. An example of type $(0, in)$ $((0, out))$ node is a directed path $T = (r_1, r_2, \ldots, r_t)$, where r is a node r_i, $2 \leqslant i \leqslant t$ $(r_i, 1 \leqslant i \leqslant t-1)$. Indeed, the HCN for this graph is 1—add edge (r_t, r_1), and this is a unique optimal solution. The resulting Hamiltonian tour $r_1, r_2, \ldots, r_t, r_1$ uses edges (r_{i-1}, r_i), (r_i, r_{i+1}) incident to $r = r_i$, both of which belong to the tree for any internal node r_i. On the other hand if $r = r_1$ $(r = r_t)$, the pair (T, r) is of type $(1, in)$ $((1, out))$, since the generated tour T uses a new edge (r_t, r_1) incident to r.

Consider the complete weighted directed graph G_T on the same vertex set as the directed tree T, and define weight of an edge to be 0 if this edge belongs to the directed tree T and 1 otherwise. Then the Hamiltonian Completion Problem for the graph T is equivalent to Travelling Salesman Problem on G_T and the optimal value of TSP tour on G_T is equal to the number of edges we need to add to make the directed graph T Hamiltonian. We now prove an auxiliary lemma analogous to Lemma 2 about certain properties of the optimal Hamiltonian cycles in G_T. Denote by $T_1, \ldots, T_d$ the subtrees generated by children of r in T. (A child of r is any node connected with r by a directed edge oriented either to or from r).

Lemma 3. *For any tour of length H in a G_T which uses $s_{in}, s_{in} = 0, 1$ incoming and $s_{out}, s_{out} = 0, 1$ outgoing edges of weight 0 incident to r (or in other words edges incident to a root r in directed tree T) there is a tour in G_T of length at most H which also uses s_{in} incoming and s_{out} outgoing edges of weight 0 incident to r and visits each subtree $T_1, \ldots, T_d$ exactly ones, i.e. vertices of any subtree $T_1, \ldots, T_d$ form a contiguous segment of Hamiltonian cycle in G_T.*

Proof. Indeed, if there are two such contiguous segments $P_1 = (i_1 \ldots, i_p)$ and $P_2 = (i_q, \ldots, i_m)$ belonging to the same subtree T_i and not connected by an edge in a Hamiltonian cycle then at least three out of four directed edges incident to this segments in a Hamiltonian cycle have weight 1 since there is at most one edge of weight 0 incident to a subtree T_i. Assume, that (i_p, i_{p+1}), (i_{q-1}, i_q) and (i_m, i_{m+1}) are these edges. Therefore, for $P_{12} = i_{p+1}, \ldots, i_{q-1}$, the part of the Hamiltonian cycle between P_1 and P_2, and for $P_{21} = i_{m+1}, \ldots, i_0$, the part of the Hamiltonian cycle between P_2 and P_1, the new tour P_1, P_2, P_{12}, P_{21} has length at most H since we took out three edges (i_p, i_{p+1}), (i_{q-1}, i_q) and (i_m, i_{m+1}) of weight 1 and used instead edges (i_p, i_q), (i_m, i_{p+1}) and (i_{q-1}, i_{m+1}) of weight at most 1. Repeating these procedure we get a Hamiltonian cycle in G_T with the desired properties. □

The following two propositions are analogous to Proposition 2 and 3.

Proposition 4. *The optimal value of HCP problem on a directed forest is the sum of the optimal values of HCP of its individual tree components.*

In the following proposition we assume a non-trivial case when the degree of the root r in the tree T is at least 1. Also, the types of children of the root are assumed to be with respect to the subtrees they generate.

Proposition 5. *Given a rooted directed tree* (T, r), *let* $r_1, \dots, r_d$ *be the children of r and let* $T_1, \dots, T_d$ *be the subtrees emanating from them. Let* X_{in} (X_{out}) *be the set of children connected with r by the edges incoming to* (*outgoing from*) *r. Then*

1. *If there is a child* $r' \in X_{in}$ *of type* $(1, out)$ *and a child* $r'' \in X_{out}$ *of type* $(1, in)$, *then* (T, r) *is type* $(0, in), (0, out)$ *and* $H(T) = -1 + \sum_{i=1}^{d} H(T_i)$.
2. *If there is a child* $r' \in X_{in}$ *of type* $(1, out)$ *and all the children in* X_{out} *are type* $(0, in)$ (*by convention it includes the case when* $X_{out} = \emptyset$), *then* (T, r) *is type* $(0, in), (1, out)$ *and* $H(T) = \sum_{i=1}^{d} H(T_i)$.
3. *If all the children in* X_{in} *are type* $(0, out)$ (*or* $X_{in} = \emptyset$) *but there is a child* $r'' \in X_{out}$ *of type* $(1, in)$, *then* (T, r) *is type* $(1, in), (0, out)$ *and* $H(T) = \sum_{i=1}^{d} H(T_i)$.
4. *Finally, if all the children in* X_{in} *are type* $(0, out)$ (*or* $X_{in} = \emptyset$) *and all the children in* X_{out} *are type* $(0, in)$ (*or* $X_{out} = \emptyset$), *then* (T, r) *is type* $(1, in), (1, out)$ *and* $H(T) = 1 + \sum_{i=1}^{d} H(T_i)$.

Remark. As in the case of undirected graphs, the recursion above leads to a linear time algorithm for solving HCP in directed trees.

Proof. We consider four cases from the claim of the proposition.

Case 1: Assume that there is a child $r' \in X_{in}$ of type $(1, out)$ and a child $r'' \in X_{out}$ of type $(1, in)$. W.l.o.g. let $r' = r_1$ and $r'' = r_d$ be these children. Let $C_1, \dots, C_d$ be optimal Hamiltonian cycles in G_{T_i}, $i = 1, \dots, d$ of length $H(T_i)$ such that C_1 has an edge of weight 1 outgoing from r_1 in T_1 and C_d has an edge of weight 1 incoming to r_d in T_d. Delete these two edges from C_1 and C_d. Delete one arbitrary directed edge of length 1 from each $C_2, \dots, C_{d-1}$. After that connect the path in T_1 obtained from C_1 with the path in T_d obtained from C_d by two edges of weight 0 through the root vertex *r* (such edges exist since $r' \in X_{in}$ and $r'' \in X_{out}$) and connect remaining paths in any order by $d - 1$ edges of weight 1 into a Hamiltonian cycle in G_T. Clearly, these new Hamiltonian cycle has length exactly $-1 + \sum_{i=1}^{d} H(T_i)$ since we deleted one edge of weight 1 from every subtree and added exactly $d - 1$ edges of weight one to the Hamiltonian cycle. One the other hand, $-1 + \sum_{i=1}^{d} H(T_i)$ is a lower bound for every Hamiltonian cycle in G_T, since, by Lemma 2 this Hamiltonian cycle must contain a Hamiltonian path for each T_i of length at least $H(T_i) - 1$ and $d - 1$ edges of weight 1 between subtrees. Moreover, we can achieve this lower bound only if Hamiltonian path uses two edges of weight 0 incident to the root *r* to connect two subtrees. Therefore, pair (T, r) has types $(0, in)$ and $(0, out)$.

Case 2: Assume that there is a child $r' \in X_{in}$ of type $(1, out)$ and all children in X_{out} have type $(0, in)$, assume that $r' = r_1$. Then deleting the edge of weight 1 outgoing from r_1 in C_1, and arbitrary edge of weight 1 from each $C_2, \dots, C_d$, connecting the Hamiltonian path in C_1 with the root *r* by the directed edge (r_1, r) of weight 0 and all other Hamiltonian paths into one cycle by *d* edges of weight 1 we get a cycle of length exactly $\sum_{i=1}^{d} H(T_i)$. To complete the proof we need to prove that

- There is no Hamiltonian cycle in G_T of weight $-1 + \sum_{i=1}^{d} H(T_i)$.
- There is no Hamiltonian cycle in G_T of weight $\sum_{i=1}^{d} H(T_i)$ which uses a directed edge of weight 1 incoming to *r*.

Assume on the contrary that there is a tour C in G_T of weight $-1 + \sum_{i=1}^{d} H(T_i)$. Then as we noticed in the previous paragraph it must use two edges of weight 0 incident to the root r. Let T_t be the subtree connected by edge of weight 0 outgoing from the root r. Then directed subpath through the subtree T_t in a cycle C cannot have length less than $H(T_t)$ since otherwise connecting two endpoints of such path we will either get a tour of weight less than $H(T_t)$ or the tour of weight exactly $H(T_t)$ but with edge of weight 1 incoming to r_t and then (T_t, r_t) would be of type $(1, in)$ and $r_t \in X_{out}$. Therefore, we have at least one subtree T_t which contributes $H(T_t)$ to the length of C. By adding at least $H(T_i) - 1$ for all other trees and $d - 1$ to connect all subtours into one Hamiltonian cycle we get that the tour length is at least $\sum_{i=1}^{d} H(T_i)$. Contradiction.

Using the same argument we can show that there is no Hamiltonian cycle C in G_T of weight $\sum_{i=1}^{d} H(T_i)$ which uses a directed edge of weight 1 incoming to r. Assume that there is such a cycle C. Then it must use the edge of length 0 outgoing from r, since otherwise the weight of C will be at least $1 + \sum_{i=1}^{d} H(T_i)$. Applying previous argument we get that there is at least one subtree T_t whose contribution to weight of C is at least $H(T_t)$. By adding at least $H(T_i) - 1$ for all other trees and d to connect all subtours and the root r into one Hamiltonian cycle we get that the tour length is at least $1 + \sum_{i=1}^{d} H(T_i)$. Contradiction.

Case 3: We omit the proof for this case since it is completely symmetric to the Case 2.

Case 4: Assume that all children in X_{in} are of type $(0, out)$ and all children in X_{out} are of type $(0, in)$. We can easily obtain a Hamiltonian cycle in G_T by deleting one edge of length one in each C_i, $i = 1, \ldots, d$ adding $d + 1$ edges of weight 1 and connecting resulting paths and r into Hamiltonian cycle in G_T. Clearly, the length of this cycle is $1 + \sum_{i=1}^{d} H(T_i)$ and it has edges of weight 1 incoming to and outgoing from r and therefore, what we need to show is that there is no Hamiltonian cycle in G_T of length $\sum_{i=1}^{d} H(T_i)$ or less. The argument is very similar to the one in the previous cases. Assume that there is a tour of length smaller than $1 + \sum_{i=1}^{d} H(T_i)$. It cannot have two edges of weight 1 incident to r since otherwise this cycle has the weight at least $2 + \sum_{i=1}^{d}(H(T_i) - 1) + d - 1$. Suppose it has exactly one edge of weight 0 incident to r. Let T_t be a subtree connected by this edge with the root and assume that $r_t \in X_{in}$. The subpath of Hamiltonian cycle in this subtree must have the length at least $H(T_t)$ since r_t has type $(0, out)$ and therefore adding $H(T_i) - 1$ for all other subtrees and d edges of weight one to connect subtours in different subtrees we get $1 + \sum_{i=1}^{d} H(T_i)$, again. Finally, if there are two edges of weight 0 incident to r in a Hamiltonian cycle then let T_k and T_t be subtrees connected by these edges with the root r. Therefore, they will contribute $H(T_k)$ and $H(T_t)$ to the length of the Hamiltonian cycle plus $H(T_i) - 1$ for all other trees and $d - 1$ to connect all paths in subtrees into one cycle. Again, we obtain $1 + \sum_{i=1}^{d} H(T_i)$. $\square$

The following symmetry property will be useful in analyzing the random instances of HCP.

Proposition 6. *Given a directed rooted tree (T, r) consider the tree $(\hat{T}, r)$ obtained by reversing the direction of every edge in T. Then the optimal value of the HCP for T and $\hat{T}$ are the same and, moreover, an optimal solution for $\bar{T}$ can be obtained from an optimal solution for T by reversing the directions of all the newly added edges. Finally, for every $s = 0, 1$, if (T, r) is type (s, in) $((s, out))$, then $(\hat{T}, r)$ is type (s, out) $((s, in))$.*

Proof. The proof follows immediately from the definition of HCP and types.

3. Hamiltonian completion of a Poisson tree

3.1. Undirected graphs

One of the classical results of the theory of random graphs states that, w.h.p., a random graph $G(n, c/n)$ for $c<1$ consists mostly of disconnected trees and some small cycles, with only constantly many nodes belonging to cycles [7]. In other words, w.h.p., a node i which is selected randomly and uniformly from the set of all nodes, belongs to a component which is a tree. Moreover, if we take i as a root of this tree, each node of this tree has outdegree distributed according to a Poisson distribution with parameter c (denoted $\mathrm{Pois}(c)$), in the limit as $n \to \infty$. Namely, if j is any node of this tree, then j has $k \geqslant 0$ children with the probability $(c^k/k!)\mathrm{e}^{-c}$, in the limit as $n \to \infty$. Then the expected outdegree for each node is c and the expected size of this tree is $1 + c + c^2 + \ldots = 1/(1-c)$.

Motivated by this, in the present section we analyze the Hamiltonian completion of a random Poisson tree T—a randomly generated tree with outdegree distribution $\mathrm{Pois}(c)$. When $c<1$ such a Poisson tree is finite with probability one and therefore its optimal value of the HCP is also finite, with probability one. Let $H = H(T)$ denote the optimal (random) value of the HCP of a Poisson tree T with parameter c. Let also $N = N(T)$ denote the number of nodes in the Poisson tree T, and let $t \in \{0, 1\}$ be the type of this tree. We denote by $g_0(x, y)$ and $g_1(x, y)$ the generating function of the joint distribution of (N, H), when the root of the tree is type 0 or type 1 respectively. That is

$$g_0(x, y) = \sum_{m \geqslant 1, h \geqslant 1} x^m y^h \mathrm{Prob}\{N = m, H = h, t = 0\}, \tag{14}$$

$$g_1(x, y) = \sum_{m \geqslant 1, h \geqslant 1} x^m y^h \mathrm{Prob}\{N = m, H = h, t = 1\}. \tag{15}$$

The summation starts with $h \geqslant 1$ since, by assumption, Hamiltonian completion of an isolated node is 1. Given an arbitrary two-dimensional random variable Z in $\mathscr{Z}^2$ with a generating function $g_Z(x, y) = \sum_{-\infty<m,h<\infty} x^m y^h \mathrm{Prob}\{Z = (m, h)\}$, observe then, that the deterministic variables $Z = (1, 1)$, $Z = (1, 0)$ and $Z = (1, -1)$ have generating functions xy, x and x/y, respectively. The following fact is a classical result from the probability theory.

Proposition 7. *Let $Z_1, \ldots, Z_i \in \mathscr{Z}^2$ be independent random variables with generating functions $g_{Z_1}(x, y), \ldots, g_{Z_i}(x, y)$, respectively. Then the generating function $g_Z(x, y)$ of $Z = \sum_{1 \leqslant j \leqslant i} Z_i$ is $\prod_{1 \leqslant j \leqslant i} g_{Z_i}(x, y)$.*

We now state and prove the main result of this subsection.

Theorem 4. *The generating functions $g_0 = g_0(x, y)$, $g_1 = g_1(x, y)$ defined in* (14) *and* (15) *satisfy the functional equations* (2), (3) *and the functional inequality* (4), *for all* $0 < c < 1$, $x, y \in [0, 1]$.

Proof. The inequality (4) follows from

$$g_0(x,y)+g_1(x,y)=\sum_{m\geqslant 1,h\geqslant 1} x^m y^h \mathrm{Prob}\{N=m, H=h\} \leqslant \sum_{m,h\geqslant 1} \mathrm{Prob}\{N=m, H=h\}=1.$$

We now prove (2), (3). Let r and $r_1,\ldots,r_K$ denote the root and the children of the root of our Poisson tree with parameter c, respectively. Let N, H, t denote the number of nodes, the HCN and the type of the root node r, respectively. Also let N_i, H_i, t_i denote the number of nodes, HCN and the type of the rooted subtree (T_i, r_i), generated by nodes r_i, respectively, for $i=1,2,\ldots,K$, assuming $K>0$. When $K=0$ these quantities are not defined. Then $N=1+\sum_{i=1}^{K} N_i$. Note, that conditioned on $K=k>0$, each triplet (N_i, H_i, t_i) has the same distribution as (N, H, t), and, moreover, these triplets (N_i, H_i, t_i) have independent probability distributions. When $K=0$, we have by convention $N=1$, $H=1$ and $t=1$. We now fix $k>0$ and condition on the event $K=k$. We will consider the case $K=0$ later. We have, $K=k$, with probability $(c^k/k!)\mathrm{e}^{-c}$. Let p_0 (p_1) be the probability that the root r has type 0 (1). We consider the following cases:

1. $t_i=0$ for all $1\leqslant i\leqslant k$. This event occurs with probability p_0^k. In this case, applying Theorem 3, $H=1+\sum_{i=1}^{k} H_i$ and $t=1$. We also have $N=1+\sum_{i=1}^{K} N_i$. Thus, conditioning on this event we have $(N,H)=(1,1)+\sum_{i=1}^{k}(N_i,H_i)$. Applying Proposition 7, and recalling that the generating function of the deterministic vector $(1,1)$ is xy, we obtain

$$\begin{aligned} &g_1(x,y|K=k, t_1=t_2=\cdots=t_k=0)\\ &\quad = xy\prod_{1\leqslant i\leqslant k} g_0(x,y|t_i=0)=xyg_0^k(x,y|t=0). \end{aligned} \tag{16}$$

   The last equality follows from the fact that the generating function g_0 conditioned on the event that the tree T_i is type 0 is the same for all children $r_1,\ldots,r_k$ and is the same as the generating function g_0 of the entire rooted tree T conditioned on the type $t=0$. Moreover, in this case $g_0(x,y|K=k, t_1=\cdots=t_k=0)=0$, since, from Proposition 3, the root r is type 1.
2. $t_{i_0}=1, t_i=0, i\neq i_0$ for some i_0. This event occurs with probability $kp_1p_0^{k-1}$. Then, from Theorem 3 we have $H=\sum_{i=1}^{k} H_i$ and $t=1$. Thus $(N,H)=(1,0)+\sum_{i=1}^{k}(N_i,H_i)$. Using Proposition 7 we obtain

$$g_1(x,y|K=k, t_{i_0}=1, t_i=0, i\neq i_0)=xg_1(x,y|t=1)g_0^{k-1}(x,y|t=0). \tag{17}$$

   Again, in this case $g_0(x,y|K=k, t_{i_0}=1, t_i=0, i\neq i_0)=0$.
3. There exists exactly $j\geqslant 2$ children for which $t_j=1$. This event, which we denote by E_j, occurs with probability $\binom{k}{j}p_1^j p_0^{k-j}$. Note that this event can only occur when $k\geqslant 2$. From Theorem 3 we have $H=-1+\sum_{i=1}^{k} H_i$ and $t=0$. Using Proposition 7 we

obtain

$$g_0(x, y|K = k, E_j) = \frac{x}{y} g_1^j(x, y|t = 1) g_0^{k-j}(x, y|t = 0) \tag{18}$$

and $g_1(x, y|K = k, E_j) = 0$.

If $K = 0$, which occurs with probability e^{-c}, we have by definition $N = 1$, $H = 1$, $t = 1$. Then $g_1(x, y) = xy$ and $g_0(x, y) = 0$. We now combine this with (16)–(18) and uncondition the event $K = k$ to obtain

$$\begin{aligned} g_1(x, y) = xy\mathrm{e}^{-c} + \sum_{k \geqslant 1} \frac{c^k}{k!} \mathrm{e}^{-c} (xy g_0^k(x, y|t = 0) p_0^k \\ + x g_1(x, y|t = 1) g_0^{k-1}(x, y|t = 0) k p_1 p_0^{k-1}), \end{aligned} \tag{19}$$

$$g_0(x, y) = \sum_{k \geqslant 2} \frac{c^k}{k!} \mathrm{e}^{-c} \sum_{2 \leqslant j \leqslant k} \frac{x}{y} g_1^j(x, y|t = 1) g_0^{k-j}(x, y|t = 0) \binom{k}{j} p_1^j p_0^{k-j}. \tag{20}$$

Note, that $g_0(x, y|t = 0) p_0 = g_0(x, y)$ and $g_1(x, y|t = 1) p_1 = g_1(x, y)$. Using binomial expansion for the (20), we obtain that

$$\begin{aligned} &\sum_{2 \leqslant j \leqslant k} g_1^j(x, y|t = 1) g_0^{k-j}(x, y|t = 0) \binom{k}{j} p_1^j p_0^{k-j} \\ &\quad = (g_0(x, y) + g_1(x, y))^k - g_0^k(x, y) - k g_1(x, y) g_0^{k-1}(x, y). \end{aligned}$$

Then we obtain from (19), (20)

$$g_1(x, y) = xy\mathrm{e}^{c g_0(x,y) - c} + x c g_1(x, y) \mathrm{e}^{c g_0(x,y) - c} \tag{21}$$

and

$$\begin{aligned} g_0(x, y) &= \frac{x}{y} \sum_{k \geqslant 2} \frac{c^k}{k!} \mathrm{e}^{-c} ((g_0(x, y) + g_1(x, y))^k - g_0^k(x, y) - k g_1(x, y) g_0^{k-1}(x, y)) \\ &= \frac{x}{y} (\mathrm{e}^{c g_0(x,y) + c g_1(x,y) - c} - \mathrm{e}^{-c} - c(g_0(x, y) + g_1(x, y)) \mathrm{e}^{-c}) \\ &\quad - \frac{x}{y} (\mathrm{e}^{c g_0(x,y) - c} - \mathrm{e}^{-c} - c g_0(x, y) \mathrm{e}^{-c}) \\ &\quad - \frac{x}{y} c g_1(x, y) (\mathrm{e}^{c g_0(x,y) - c} - \mathrm{e}^{-c}) \\ &= \frac{x}{y} \mathrm{e}^{c g_0(x,y) + c g_1(x,y) - c} - \frac{x}{y} \mathrm{e}^{c g_0(x,y) - c} - \frac{x}{y} c g_1(x, y) \mathrm{e}^{c g_0(x,y) - c}. \end{aligned} \tag{22}$$

We rewrite the results as

$$g_1(x, y) = \frac{xy\mathrm{e}^{c g_0(x,y) - c}}{1 - xc\mathrm{e}^{c g_0(x,y) - c}}, \tag{23}$$

$$g_0(x, y) = \frac{x}{y} \mathrm{e}^{c g_0(x,y) - c} (\mathrm{e}^{c g_1(x,y)} - 1 - c g_1(x, y)). \tag{24}$$

This completes the proof of the theorem. □

3.2. Directed graphs

We now analyze the case when our randomly generated tree is a directed graph. The setup is the same as in Subsection 3.1, except for every edge is directed. The direction is chosen at random equiprobably from each of the two possibilities and independently for all the edges and independently from other randomness in the tree. As in the undirected case, N, H, t denote the number of nodes, the value of the Hamiltonian completion and the type of the root of the tree T, respectively, and N_i, H_i, t_i stand for the same for children of the root. Again we have $N = 1 + \sum_i N_i$. The type t takes one of the four values $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$ which are short-hand notations for $((0, in), (0, out))$, $((0, in), (1, out))$, $((1, in), (0, out))$, $((1, in), (1, out))$, respectively.

For every pair $(v, w) \in \{0,1\}^2$, let p_{vw} denote the probability that the root r has type (v, w). Also for every $(v, w) \in \{0,1\}^2$ we introduce the generating function

$$g_{vw}(x, y) = \sum_{m \geqslant 1, h \geqslant 1} x^m y^h \text{Prob}\{N = m, H = h, t = (v, w)\}. \tag{25}$$

From Proposition 6 and since the two directions of each edge are equiprobable, it follows that $g_{01}(x, y) = g_{10}(x, y)$, As in Subsection 3.1, our next goal is deriving equations which bind the three generating functions.

Theorem 5. *The generating functions* $g_{00}(x, y)$, $g_{01}(x, y)$, $g_{11}(x, y)$ *satisfy the functional Eqs.* (7), (8), (9), *for all* $x \in [0, 1]$, $y \in (0, 1]$ *and the functional inequality* (10).

Proof. Let r and $r_1, \ldots, r_K$ denote the root and the children of our random tree T, with the possibility $K = 0$. Conditioned on $K = 0$, we have, by convention from Subsection 2.2, $g_{11}(x, y|K = 0) = xy$ and $g_{vw}(x, y|K = 0) = 0$ for all other $vw \in \{0, 1\}$.

We now fix $k > 0$ and consider the event $K = k$. Further, we fix $k_1, k_2 \geqslant 0$ with $k_1 + k_2 = k$ and consider the event $|X_{in}| = k_1$, $|X_{out}| = k_2$. These two events occur with probability $(c^k/k!)\mathrm{e}^{-c}\binom{k}{k_1}2^{-k}$. Furthermore, for every pair $(v, w) \in \{0, 1\}^2$, consider the event that the number of nodes in X_{in} (X_{out}) of type (v, w) (the type is with respect to the generated subtrees) is j^{in}_{vw} (j^{out}_{vw}). This event occurs with probability

$$\binom{k_1}{j^{in}_{00} \quad j^{in}_{01} \quad j^{in}_{10} \quad j^{in}_{11}} \binom{k_2}{j^{out}_{00} \quad j^{out}_{01} \quad j^{out}_{10} \quad j^{out}_{11}} \prod_{v,w=0,1} p_{vw}^{j^{in}_{vw} + j^{out}_{vw}}, \tag{26}$$

were for any non-negative integers a, a_1, a_2, a_3, a_4, with $a_1 + a_2 + a_3 + a_4 = a$, $\binom{a}{a_1 \quad a_2 \quad a_3 \quad a_4}$ denotes the standard combinatorial term $a!/(a_1!a_2!a_3!a_4!)$.

Next we consider four cases corresponding to the cases in Proposition 5. The argument is very similar to the one in the proof of Theorem 4.

1. $j^{in}_{01} + j^{in}_{11} > 0$, $j^{out}_{10} + j^{out}_{11} > 0$. Applying Proposition 5, r is type $(0, 0)$, $H = -1 + \sum_{1 \leqslant i \leqslant k} H_i$, and using Proposition 7 the corresponding conditioned generating

function satisfies

$$g_{00}(x, y|\cdot) = \frac{x}{y} \prod_{v,w=0,1} g_{vw}^{j_{vw}^{in}+j_{vw}^{out}}(x, y|t = (v, w)) \tag{27}$$

and $g_{vw}(x, y|\cdot) = 0$ for all $(vw) \neq (0, 0)$.

2. $j_{01}^{in} + j_{11}^{in} > 0$, $j_{10}^{out} + j_{11}^{out} = 0$. Then r is type $(0, 1)$ and

$$g_{01}(x, y|\cdot) = x \prod_{v,w=0,1} g_{vw}^{j_{vw}^{in}+j_{vw}^{out}}(x, y|t = (v, w)) \tag{28}$$

and $g_{vw}(x, y|\cdot) = 0$ for all $(vw) \neq (0, 1)$.

3. $j_{01}^{in} + j_{11}^{in} = 0$, $j_{10}^{out} + j_{11}^{out} > 0$. The analysis of this case is skipped since it corresponds to computing $g_{10}(x, y)$, which is equal to $g_{01}(x, y)$, as we observed above.
4. $j_{01}^{in} + j_{11}^{in} = 0$, $j_{10}^{out} + j_{11}^{out} = 0$. Then r is type $(1, 1)$ and

$$g_{11}(x, y|\cdot) = xy \prod_{v,w=0,1} g_{vw}^{j_{vw}^{in}+j_{vw}^{out}}(x, y|t = (v, w)) \tag{29}$$

and $g_{vw}(x, y|\cdot) = 0$ for all $(vw) \neq (1, 1)$.

Next, we combine these equations to obtain defining on $g_{vw}(x, y)$. For convenience, it is easier to start with $g_{11}(x, y)$. From (26) and (29) and recalling $g_{11}(x, y|K = 0) = xy$, we obtain

$$\begin{aligned} g_{11}(x, y) = xy\mathrm{e}^{-c} + \sum_{k=k_1+k_2 \geqslant 1} \frac{c^k}{k!} \mathrm{e}^{-c} \binom{k}{k_1} 2^{-k} xy \sum_{j_{vw}^{in}, j_{vw}^{out} : j_{01}^{in} = j_{11}^{in} = j_{10}^{out} = j_{11}^{out} = 0}, \\ \binom{k_1}{j_{00}^{in} \quad j_{01}^{in} \quad j_{10}^{in} \quad j_{11}^{in}} \binom{k_2}{j_{00}^{out} \quad j_{01}^{out} \quad j_{10}^{out} \quad j_{11}^{out}} \\ \times \prod_{v,w=0,1} p_{vw}^{j_{vw}^{in}+j_{vw}^{out}} g_{vw}^{j_{vw}^{in}+j_{vw}^{out}}(x, y|t = (v, w)). \end{aligned}$$

We have $g_{vw}(x, y|t = (v, w))\, p_{vw} = g_{vw}(x, y)$. Cancelling $k!$, $k_1!$, $k_2!$ and using elementary computations, we obtain

$$\begin{aligned} g_{11}(x, y) &= \sum_{k \geqslant 0} \frac{c^k}{k!} \mathrm{e}^{-c} \frac{1}{2^k} xy (2g_{00}(x, y) + g_{10}(x, y) + g_{01}(x, y))^k \\ &= xy\mathrm{e}^{c(2g_{00}(x,y)+g_{10}(x,y)+g_{01}(x,y))/2-c}. \end{aligned}$$

We do a similar computation for g_{01} using (28). The sum corresponding to constraints $j_{01}^{in} + j_{11}^{in} > 0$, $j_{10}^{out} + j_{11}^{out} = 0$ we represent as a difference between the sum corresponding to just $j_{10}^{out} + j_{11}^{out} = 0$ and $j_{01}^{in} + j_{11}^{in} = 0$, $j_{10}^{out} + j_{11}^{out} = 0$. Simplifying as we did for g_{11}, we obtain

$$
\begin{aligned}
g_{01}(x,y) = \sum_{k\geqslant 1} \frac{c^k}{k!} e^{-c} \frac{1}{2^k} x[&(2g_{00}(x,y)+2g_{01}(x,y)+g_{10}(x,y)+g_{11}(x,y))^k \\
&-(2g_{00}(x,y)+g_{01}(x,y)+g_{10}(x,y))^k] \\
= x e^{c(2g_{00}(x,y)+2g_{01}(x,y)+g_{10}(x,y)+g_{11}(x,y))/2-c}& \\
- x e^{c(2g_{00}(x,y)+g_{01}(x,y)+g_{10}(x,y))/2-c}&.
\end{aligned}
$$

To compute g_{00} note that the condition $j_{01}^{in}+j_{11}^{in}, j_{10}^{out}+j_{11}^{out}>0$ implies $k\geqslant 2$. The sum $\sum_{j_{vw}^{in}, j_{vw}^{out}: j_{01}^{in}+j_{11}^{in}, j_{10}^{out}+j_{11}^{out}>0}$ we represent as $\sum_{j_{vw}^{in}, j_{vw}^{out}} - \sum_{j_{vw}^{in}, j_{vw}^{out}: j_{01}^{in}=j_{11}^{in}=0} - \sum_{j_{vw}^{in}, j_{vw}^{out}: j_{10}^{out}=j_{11}^{out}=0} + \sum_{j_{vw}^{in}, j_{vw}^{out}: j_{01}^{in}=j_{11}^{in}=j_{10}^{out}=j_{11}^{out}=0}$. Applying (27), we obtain

$$
\begin{aligned}
g_{00}(x,y) = \sum_{k\geqslant 2} \frac{c^k}{k!} e^{-c} \frac{1}{2^k} \frac{x}{y}[&(2g_{00}(x,y)+2g_{01}(x,y)+2g_{10}(x,y)+2g_{11}(x,y))^k \\
&-(2g_{00}(x,y)+2g_{01}(x,y)+g_{10}(x,y)+g_{11}(x,y))^k \\
&-(2g_{00}(x,y)+g_{01}(x,y)+2g_{10}(x,y)+g_{11}(x,y))^k \\
&+(2g_{00}(x,y)+g_{01}(x,y)+g_{10}(x,y))^k].
\end{aligned}
$$

The terms in the right-hand side corresponding to $k=0,1$ are equal to zero. Therefore,

$$
\begin{aligned}
g_{00}(x,y) = &\frac{x}{y} e^{c(2g_{00}(x,y)+2g_{01}(x,y)+2g_{10}(x,y)+2g_{11}(x,y))/2-c} \\
&-\frac{x}{y} e^{c(2g_{00}(x,y)+g_{01}(x,y)+2g_{10}(x,y)+g_{11}(x,y))/2-c} \\
&-\frac{x}{y} e^{c(2g_{00}(x,y)+2g_{01}(x,y)+g_{10}(x,y)+g_{11}(x,y))/2-c} \\
&+\frac{x}{y} e^{c(2g_{00}(x,y)+g_{01}(x,y)+g_{10}(x,y))/2-c}. \qquad \square
\end{aligned}
$$

4. Hamiltonian completion of a random graph $G(n, c/n)$

In this section we complete the proof of Theorem 1. We do this by relating the HCP on $G(n, c/n)$ to the HCP on Poisson trees and applying the results of the previous section. Let T denote a random Poisson tree, introduced in the previous section. As before, N, H, t denote the number of nodes, Hamiltonian completion and the type of T. The proposition below relates the HCP on a sparse random graph $G(n, c/n), c<1$ to the HCP on a tree T. The statement and the derivation below applies to both the undirected and the directed cases. We will indicate the distinctions when appropriate. We recall that $H(n, c)$ denotes the HCN of $G(n, c/n)$.

Proposition 8. *The following convergence holds as* $n \to \infty$:

$$\frac{\mathbb{E}[H(n,c)]}{n} \to \sum_{m=1}^{\infty} \frac{\mathbb{E}[H|N=m]\mathrm{Prob}\{N=m\}}{m} < \infty, \tag{30}$$

where the expectation and the probability on the right-hand side are with respect to the (*undirected or directed*) *random Poisson tree T.*

Proof. We decompose G into its connected (weakly connected in case of directed graphs) components. Denote the tree components by $T_1, \ldots, T_R$ and let P be the union of all non-tree components. We know that the expected number of nodes in P is $O(1)$. As a result $H(P) = O(1)$. For every node $i = 1, 2, \ldots, n$, if i belongs to a tree component, denote the component by $T(i)$. Otherwise set $T(i) = \emptyset$. By convention we put $H(\emptyset) = 0$. From Propositions 2, 4,

$$\begin{aligned} H(G) &= \sum_{1 \leqslant t \leqslant R} H(T_t) + H(P) = \sum_{1 \leqslant t \leqslant R} \left[\sum_{m=1}^{n} H(T_t) 1\{|T_t| = m\} \right] + H(P) \\ &= \sum_{i=1}^{n} \sum_{m=1}^{n} \frac{H(T(i)) 1\{|T(i) = m|\}}{m} + H(P), \end{aligned} \tag{31}$$

where we simply decompose the sum into the parts corresponding to the same size of the tree, and in the last equality the division by m comes from the fact that each node of the tree was counted m times. After taking expectations and using symmetry we obtain

$$\mathbb{E}[H(G)] = n \sum_{m=1}^{n} \frac{\mathbb{E}[H(T(1)) 1\{|T(1)| = m\}]}{m} + O(1),$$

since the value $\mathbb{E}[H(T(1))1\{|T(1)| = m\}] = \mathbb{E}[H(T(1))||T(1)| = m\}]\mathrm{Prob}(|T(1)| = m)$ is the same for all nodes i. But, w.h.p., the component containing node 1 is a tree, and, in particular, it is a Poisson tree $T(1)$, in the limit as $n \to \infty$. Therefore, its number of nodes, Hamiltonian completion and type are distributed as N, H, t of a random Poisson tree with parameter c, introduced in the previous section. It now remains to show that the infinite sum in the right-hand side of (30) is finite. Note that, trivially, for any tree T, $H(T) \leqslant |T|$. Then

$$\frac{\mathbb{E}[H 1\{N = m\}]}{m} \leqslant \mathrm{Prob}\{N = m\}.$$

As a result, the partial sum in (30) starting from $m = m_0$ is at most $\mathrm{Prob}\{N \geqslant m_0\}$. We know that the expected size of a Poisson tree is $1/(1-c) < \infty$. Using Markov inequality, $\mathrm{Prob}\{N \geqslant m_0\} \leqslant 1/((1-c)m_0)$, which converges to 0 as m_0 increases. This completes the proof.

We now complete the proof of Theorem 1 by expressing the sum in (30) through the generating functions $g(x, y)$, $\bar{g}(x, y)$ defined by (11) corresponding to the pair (N, H).

Note that

$$g(x, y) = \sum x^m y^h \mathrm{Prob}\{H = h, N = m\},$$
$$\bar{g}(x, y) = \sum x^m y^h \mathrm{Prob}\{H = h, N = m\},$$

although they are not equal since the distribution of H is different in the undirected in directed cases. Nevertheless, the remaining part of the proof is identical for both cases and, therefore, we only consider the undirected case corresponding to generating function $g(x, y)$. The sum in (30) is equal to

$$\sum_{m=1}^{\infty} \sum_{h=1}^{\infty} \frac{h\mathrm{Prob}\{H = h, N = m\}}{m}. \tag{32}$$

Functions $g(x, y)$, $\bar{g}(x, y)$ when defined over $0 \leqslant x, y \leqslant 1$ are uniform limits of polynomials, and, as such, are infinitely often differentiable in this domain. Differentiating function $g(x, y)$ with respect to y and interchanging the order of summation and integration we obtain $\partial g(x, y)/\partial y = \sum_{m,h \geqslant 1} hx^m y^{h-1} \mathrm{Prob}\{N = m, H = h\}$. We now divide this function by x to obtain $(1/x)(\partial g(x, y)/\partial y) = \sum_{m,h \geqslant 1} hx^{m-1} y^{h-1} \mathrm{Prob}\{N = m, H = h\}$. For a large constant $C > 0$ we have $\sum_{m,h \geqslant C} hx^{m-1} y^{h-1} \mathrm{Prob}\{N = m, H = h\} \leqslant \sum_{h \geqslant C} h\mathrm{Prob}\{H = h\} \to 0$ as $C \to \infty$, since $E[H] \leqslant E[N] < \infty$. Therefore, the function $\sum_{m,h \geqslant 1} hx^{m-1} y^{h-1} \mathrm{Prob}\{N = m, H = h\}$ is also a uniform limit of its partial sums. Interchanging the integration and summation we obtain $\int (1/x)(\partial g(x, y)/\partial y)\, \mathrm{d}x = \sum_{m,h \geqslant 1} (h/m) x^m y^{h-1} \mathrm{Prob}\{N = m, H = h\}$. For $x = y = 1$, the value of this function is exactly the right-hand side of(32). Since the value of this function is 0 when $x = 0$, we set $y = 1$ and obtain

$$\int_0^1 \frac{1}{x} \frac{\partial g(x, y)}{\partial y}\bigg|_{y=1} \mathrm{d}x = \sum_{m,h \geqslant 1} (h/m) \mathrm{Prob}\{N = m, H = h\}.$$

The left-hand side of this equation is the integral in (12). This completes the proof of Theorem 1. □

5. Numerical computations

This section is devoted to numerical computations of the integral in the right-hand sides of (12), (13). We only do the computations for the case of undirected graphs. The computations for the case of directed graphs is similar, except for we check, in addition, the uniqueness of the solutions to (7), (8), (9), (10).

Fixing $0 < c < 1$, we perform the following computations. We let $X = \{i/K, i = 1, \ldots, K\}$ and $Y = \{i/K, i = 1, \ldots, K\}$ be the set of points of discretized the interval $[0, 1]$ where K is some large integer (we took $K = 1000$). To compute (12) we compute functions $g_0(x, y)$, $g_1(x, y)$ using functional equations (2), (3), on a discretized unit square $(x, y) \in X \times Y$. The numerical computations of g_0, g_1 is straightforward from (2), (3). We also check that inequality (4) is satisfied (which is guaranteed by Proposition 1). The Fig. 1 displays the solution of the system of equations (2), (3) and inequality (4) for $c = 0.9$.

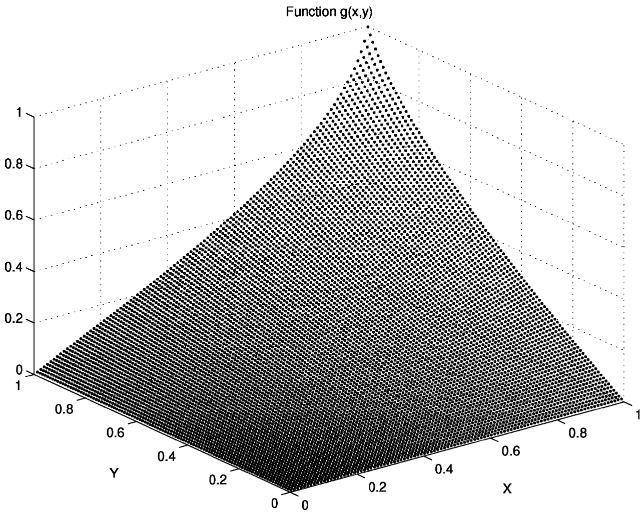

Fig. 1. Generating function for the value $c = 0.9$.

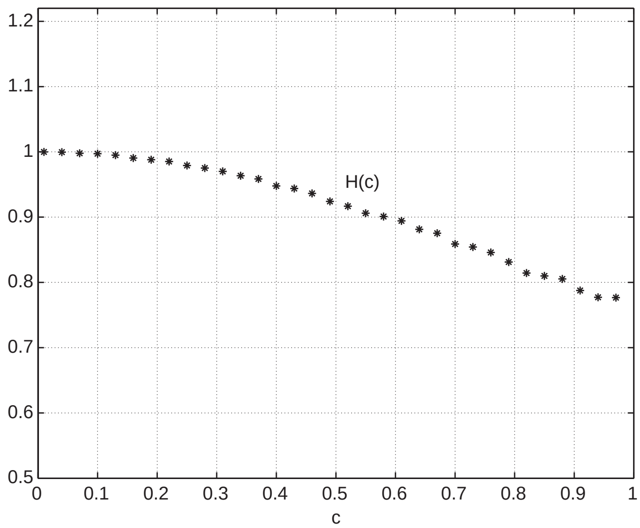

Fig. 2. The value of the HCN for different values of c.

For each $x \in X$ we approximately compute the value of derivative $\partial g(x, y)/\partial y$ for $y = 1$ using formula

$$\frac{g(x, 1) - g\left(x, \frac{K-1}{K}\right)}{1/K}.$$

Next we define a step function

$$f(x) = \frac{K\left(g(x,1) - g\left(x, \frac{K-1}{K}\right)\right)}{i/K} = K^2 \frac{g(x,1) - g\left(x, \frac{K-1}{K}\right)}{i}$$

for $x \in [(i-1)/K, i/K]$ and $i = 1, \ldots, K$. This function approximates $1/x\,(\partial g(x, y)/\partial y)$. Then the integral of interest is approximately equal to the $\int_0^1 f(x)\,\mathrm{d}x$ and can be computed by the formula

$$\sum_{i=1}^{K} K^2 \cdot \frac{g(x,1) - g\left(x, \frac{K-1}{K}\right)}{i}.$$

Fig. 2 shows values of the integral for various values of c between 0 and 1. The value approaches 0.77 as c approaches 1.

References

[1] A. Agnetis, P. Detti, C. Meloni, D. Pacciarelli, A linear algorithm for the Hamiltonian completion number of the line graph of a tree, Inform. Process. Lett. 79 (1) (2001) 17–24.

[2] J. Beardwood, J.H. Halton, J.M. Hammersley, The shortest path through many points, Proc. Camb. Philos. Soc. 55 (1959) 299–327.

[3] F.T. Boesch, S. Chen, J.A.M. McHugh, On covering the points of a graph with point disjoint paths, in: R.A. Bari, F. Harary (Eds.), Graphs and Combinatorics, Springer, Berlin, 1974.

[4] A. Frieze, J. Yukich, Probabilistic analysis of the traveling salesman problem, in: G. Gutin, A.P. Punnen (Eds.), The Traveling Salesman Problem and its Variations, Kluwer Academic Publisher, Dordrecht, 2002.

[5] S.E. Goodman, S.T. Hedetniemi, On the Hamiltonian completion problem, in: R.A. Bari, F. Harary (Eds.), Graphs and Combinatorics, Springer, Berlin, 1974.

[6] S.E. Goodman, S.T. Hedetniemi, P.J. Slater, Advances on the Hamiltonian completion problem, J. ACM 22 (3) (1975) 352–360.

[7] S. Janson, T. Luczak, A. Rucinski, Random Graphs, Wiley, New York, 2000.

[8] J. Komlos, E. Szemeredi, Limit distributions for the existence of Hamilton circuits in a random graph, Discrete Math. 43 (1983) 55–63.

[9] L. Posa, Hamiltonian circuits in random graphs, Discrete Math. 14 (1976) 359–364.

[10] A. Raychaudhuri, The total interval number of a tree and the Hamiltonian completion number of its line graph, Inform. Process. Lett. 56 (1995) 299–306.