TIGHT HAMILTON CYCLES IN RANDOM HYPERGRAPHS

PETER ALLEN*, JULIA BÖTTCHER*, YOSHIHARU KOHAYAKAWA†, AND YURY PERSON‡

Abstract. We give an algorithmic proof for the existence of tight Hamilton cycles in a random r-uniform hypergraph with edge probability $p = n^{-1+\varepsilon}$ for every $\varepsilon > 0$. This partly answers a question of Dudek and Frieze [Random Structures Algorithms], who used a second moment method to show that tight Hamilton cycles exist even for $p = \omega(n)/n$ ($r \geq 3$) where $\omega(n) \to \infty$ arbitrary slowly, and for $p = (e + o(1))/n$ ($r \geq 4$).

The method we develop for proving our result applies to related problems as well.

1. Introduction

The question of when the random graph $G(n,p)$ becomes hamiltonian is well understood. Pósa [19] and Korshunov [15, 16] proved that the hamiltonicity threshold is $\log n/n$, Komlós and Szemerédi [14] determined an exact formula for the probability of the existence of a Hamilton cycle, and Bollobás [4] established an even more powerful hitting time result. The first polynomial time randomised algorithms for finding Hamilton cycles in $G(n,p)$ were developed by Angluin and Valiant [1] and Shamir [21]. Finally, Bollobás, Fenner and Frieze [3] gave a deterministic polynomial time algorithm whose success probability matches the probabilities established by Komlós and Szemerédi.

Date: July 3, 2018.

* Department of Mathematics, London School of Economics, Houghton Street, London WC2A 2AE, U.K. *E-mail*: `p.d.allen|j.boettcher@lse.ac.uk` .

†Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão 1010, 05508–090 São Paulo, Brazil. *E-mail*: `yoshi@ime.usp.br` .

‡Institut für Mathematik, Freie Universität Berlin, Arnimallee 3-5, D-14195 Berlin, Germany. *E-mail*: `person@math.fu-berlin.de` .

PA was partially supported by FAPESP (Proc. 2010/09555-7). JB was partially supported by FAPESP (Proc. 2009/17831-7). YK was partially supported by CNPq (308509/2007-2, 477203/2012-4), CAPES/DAAD (415/ppp-probral/po/D08/11629, 333/09) and NUMEC (Project MaCLinC/USP). YP was partially supported by GIF grant no. I-889-182.6/2005. The cooperation of the authors was supported by a joint CAPES-DAAD project (415/ppp-probral/po/D08/11629, Proj. no. 333/09). The authors are grateful to NUMEC/USP, Núcleo de Modelagem Estocástica e Complexidade of the University of São Paulo, for supporting this research.

For random hypergraphs much less is known. The random r-uniform hypergraph $\mathcal{G}^{(r)}(n,p)$ on vertex set $[n]$ is generated by including each hyperedge from $\binom{[n]}{r}$ independently with probability $p = p(n)$. First, Frieze [9] considered loose Hamilton cycles in random 3-uniform hypergraphs. The *loose r-uniform cycle* on vertex set $[n]$ has edges $\{i+1,\dots,i+r\}$ for exactly all $i = k(r-1)$ with $k \in \mathbb{N}$ and $(r-1) \mid n$, where we calculate modulo n. Frieze showed that the threshold for a loose Hamilton cycle in $\mathcal{G}^{(3)}(n,p)$ is $\Theta(\log n/n^2)$. Dudek and Frieze [6] extended this to r-uniform hypergraphs with $r \geq 4$, where the threshold is $\tilde{\Theta}(\log n/n^{r-1})$. Both results require that n is divisible by $2(r-1)$ (which was recently removed by Dudek, Frieze, Loh and Speiss [7]) and rely on the deep Johansson-Kahn-Vu theorem [13], which makes their proofs non-constructive.

Tight Hamilton cycles, on the other hand, were first considered in connection with packings. The *tight r-uniform cycle* on vertex set $[n]$ has edges $\{i+1,\dots,i+r\}$ for all i calculated modulo n. Frieze, Krivelevich and Loh [11] proved that if $p \gg (\log^{21} n/n)^{1/16}$ and 4 divides n then most edges of $G^{(3)}(n,p)$ can be covered by edge disjoint tight Hamilton cycles. Further packing results were obtained by Frieze and Krivelevich [10] and by Bal and Frieze [2], but the probability range is far from best possible. Subsequently, Dudek and Frieze [5] used a second moment argument to show that the threshold for a tight Hamilton cycle in $\mathcal{G}^{(r)}(n,p)$ is sharp and equals e/n for each $r \geq 4$ and for $r = 3$ they showed that $\mathcal{G}^{(3)}(n,p)$ contains a tight Hamilton cycle when $p = \omega(n)/n$ for any $\omega(n)$ that goes to infinity. Since their method is non-constructive they asked for an algorithm to find a tight Hamilton cycle in a random hypergraph. In this paper we present a randomised algorithm for this problem if p is slightly bigger than in their result.

Theorem 1. *For each integer $r \geq 3$ and $0 < \varepsilon < 1/(4r)$ there is a randomised polynomial time algorithm which for any $n^{-1+\varepsilon} < p \leq 1$ a.a.s. finds a tight Hamilton cycle in the random r-uniform hypergraph $\mathcal{G}^{(r)}(n,p)$.*

The probability referred to in Theorem 1 is with respect to the random bits used by the algorithm as well as by $\mathcal{G}^{(r)}(n,p)$. The running time of the algorithm in the above theorem is polynomial in n, where the degree of the polynomial depends on ε.

Organisation. We first provide some notation and a brief sketch of our proof, formulate the main lemmas and prove Theorem 1 in Section 2. In Sections 3 and 4 we prove the main lemmas, and in Section 5 we end with some remarks and open problems.

2. Lemmas and proof of Theorem 1

2.1. **Notation.** An *s-tuple* $(u_1,\dots,u_s)$ of vertices is an ordered set of vertices. We often denote tuples by bold symbols, and occasionally also omit the brackets and write $\mathbf{u} = u_1, \ldots, u_s$. Additionally, we may also use a tuple as a set and write for example, if S is a set, $S \cup \mathbf{u} := S \cup \{u_i : i \in [s]\}$. The *reverse* of the s-tuple $\mathbf{u}$ is the s-tuple $(u_s, \ldots, u_1)$.

In an r-uniform hypergraph $\mathcal{G}$ the tuple $P = (u_1, \ldots, u_\ell)$ forms a *tight path* if $\{u_{i+1}, \ldots, u_{i+r}\}$ is an edge for every $0 \le i \le \ell - r$. For any $s \in [\ell]$ we say that P *starts* with the s-tuple $(u_1, \ldots, u_s) =: \mathbf{v}$ and *ends* with the s-tuple $(u_{\ell-(s-1)}, \ldots, u_\ell) =: \mathbf{w}$. We also call $\mathbf{v}$ the *start s-tuple* of P, $\mathbf{w}$ the *end s-tuple* of P, and P a $\mathbf{v}-\mathbf{w}$ path. The *interior* of P is formed by all its vertices but its start and end $(r-1)$-tuples. Note that the interior of P is not empty if and only if $\ell > 2(r-1)$.

For a hypergraph $\mathcal{H}$ we define the 1-*density* of $\mathcal{H}$ to be $d^{(1)}(\mathcal{H}) := e(\mathcal{H})/\big(v(\mathcal{H}) - 1\big)$ if $v(\mathcal{H}) > 1$, and $d^{(1)}(\mathcal{H}) := 0$ if $v(\mathcal{H}) = 1$. We set

$$m^{(1)}(\mathcal{H}) := \max\{d^{(1)}(\mathcal{H}') : \mathcal{H}' \subseteq \mathcal{H}\}\,.$$

We denote the r-uniform tight cycle on ℓ vertices by $\mathcal{C}_\ell^{(r)}$. Observe that $m^{(1)}(\mathcal{C}_\ell^{(r)}) = \ell/(\ell-1)$.

2.2. **Outline of the proof.** A simple greedy strategy shows that for $p = n^{\varepsilon-1}$ it is easy to find a tight path (and similarly a tight cycle) in $\mathcal{G}^{(r)}(n,p)$ which covers all but at most $n^{1-\frac{1}{2}\varepsilon}$ of its vertices. Incorporating these few remaining vertices is where the difficulty lies.

To overcome this difficulty we apply the following strategy, which we call the *reservoir method.* We first construct a tight path P of a linear length in n which contains a vertex set W^*, called the *reservoir*, such that for any $W \subseteq W^*$ there is a tight path on $V(P) \setminus W$ whose end $(r-1)$-tuples are the same as that of P. In a second step we use the mentioned greedy strategy to extend P to an almost spanning tight path P', with a leftover set L. The advantage we have gained now is that we are permitted to reuse the vertices in W^*: we will show that, by using a subset W of vertices from W^* to incorporate the vertices from L, we can extend the almost spanning tight path to a spanning tight cycle C. More precisely, we shall delete W from P' (observe that, by construction of P, the hypergraph induced on $V(P) \setminus W$ contains a tight path with the same ends) and use precisely all vertices of W to connect the vertices of L to construct C.

We remark that our method has similarities, in spirit, with the absorbing method for proving extremal results for large structures in dense hypergraphs (see, e.g., Rödl, Ruciński and Szemerédi [20]). The techniques to deal with multi-round exposure in our algorithm is similar to those used by Frieze in [8]. Moreover, a method very similar to ours was used independently by Kühn and Osthus [17] to find bounds on the threshold for the appearance of the square of a Hamilton cycle in a random graph.

2.3. **Lemmas.** We shall rely on the following lemmas. We state these lemmas together with an outline of how they are used, and then give the details of the proof of Theorem 1.

Our first lemma asserts that there are hypergraphs $\mathcal{H}^*$ with density arbitrarily close to 1 which have a spanning tight path and a vertex w^* such that deleting w^* from $\mathcal{H}^*$ leaves a spanning tight path with the same start and end $(r-1)$-tuples.

Lemma 2 (Reservoir lemma)**.** *For all $r \geq 2$ and $0 < \varepsilon < 1/(6r)$, there exist an r-uniform hypergraph $\mathcal{H}^* = \mathcal{H}^*(r,\varepsilon)$ on less than $16/\varepsilon^2$ vertices, a vertex w^*, and two disjoint $(r-1)$-tuples $\mathbf{u} = (u_1, \dots, u_{r-1})$ and $\mathbf{v} = (v_1, \dots, v_{r-1})$ such that*

(i) $m^{(1)}(\mathcal{H}^*) \leq 1 + \varepsilon$,
(ii) $\mathcal{H}^*$ *has a tight Hamilton* $\mathbf{u} - \mathbf{v}$ *path, and*
(iii) $\mathcal{H}^* - w^*$ *has a tight Hamilton* $\mathbf{u} - \mathbf{v}$ *path.*

We provide a proof of Lemma 2 in Section 2. We also call the graph $\mathcal{H}^*$ asserted by this lemma the *reservoir graph* and the vertex w^* the *reservoir vertex*, since they will provide us as follows with the reservoir mentioned in the outline. If we can find many disjoint copies of $\mathcal{H}^*$ in $\mathcal{G}^{(r)}(n,p)$, and if we can connect these copies of $\mathcal{H}^*$ to form a tight path, then the set W^* of reservoir vertices w^* from these $\mathcal{H}^*$-copies forms such a reservoir.

In order to find many disjoint $\mathcal{H}^*$-copies, we use the following standard theorem.

Theorem 3 (see, e.g., [12, Theorem 4.9])**.** *For every r-uniform hypergraph $\mathcal{H}$ there are constants $\nu > 0$ and $C \in \mathbb{N}$ such that if $p \geq Cn^{-1/m^{(1)}(\mathcal{H})}$, then $\mathcal{G}^{(r)}(n,p)$ a.a.s. contains νn vertex disjoint copies of $\mathcal{H}$.*

For connecting the $\mathcal{H}^*$-copies into a long tight path P we use the next lemma.

Lemma 4 (Connection lemma)**.** *Given $r \geq 3$, $0 < \varepsilon < 1/(4r)$ and $\delta > 0$, there exists $\eta > 0$ such that there is a (deterministic) polynomial time algorithm $\mathcal{A}$ which on inputs $\mathcal{G} = \mathcal{G}^{(r)}(n,p)$ with $p = n^{-1+\varepsilon}$ a.a.s. does the following.*

Let $1 \leq k \leq \eta n$, let X be any subset of $[n]$ of size at least δn. Let $\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(k)}$, $\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(k)}$ be any $2k$ pairwise disjoint $(r-1)$-tuples in $[n]$. Then $\mathcal{A}$ finds in $\mathcal{G}$ a collection of vertex disjoint tight paths P_i, $1 \leq i \leq k$, of length at most $\ell := (r-1)/\varepsilon + 2$, such that P_i is a $\mathbf{u}^{(i)} - \mathbf{v}^{(i)}$ path all of whose interior vertices are in X.

We prove this lemma in Section 3. In fact, we will also make use of this lemma after extending P to a maximal tight path P' in order to extend P' (reusing vertices of the reservoir W^*) to cover the leftover vertices L. It is for this reason that we require the lemma to work with a set X which can be quite small.

2.4. **Proof of the main theorem.** Our goal is to describe an algorithm which a.a.s. constructs a tight Hamilton cycle in the r-uniform random hypergraph $\mathcal{G}^{(r)}(n,q)$ in five steps. For convenience we replace $\mathcal{G}^{(r)}(n,p)$ in Theorem 1 by $\mathcal{G}^{(r)}(n,q)$. We make use of a 5-round exposure of the random hypergraph, that is, each of the five algorithm steps will individually a.a.s. succeed on an r-uniform random hypergraph with edge probability somewhat smaller than q. Observe, however, that for the algorithm the input graph is given at once, and not as the union of five graphs. Therefore, in a preprocessing step the algorithm will first split the (random) input hypergraph into five (independent random) hypergraphs. The only probabilistic component of our algorithm is in the preprocessing step.

Our five algorithm steps will then be as follows. Firstly, we apply Theorem 3 in order to find cn vertex disjoint copies of the reservoir graph $\mathcal{H}^*$ from Lemma 2. Secondly, we use the connection lemma, Lemma 4, to connect the $\mathcal{H}^*$ copies to a tight path P of length $c'n$ which contains a set W^* of linearly many reservoir vertices. Thirdly, we greedily extend P until we get a tight path P' on $n - n^{1-(\varepsilon'/2)}$ vertices. In the fourth and fifth step we use W^* and Lemma 4 to connect the remaining vertices to the path constructed so far and to close the path into a cycle.

For technical reasons it will be convenient to assume that the edge probability in each of the last four steps is exactly $q' = n^{-1+\varepsilon'}$ for some ε'. We therefore split our random input hypergraph into five independent random hypergraphs, of which the first has edge probability $q'' \geq q'$ and the remaining four have edge probability q'.

Proof of Theorem 1. Constants: Given $r \geq 3$ and $0 < \varepsilon < 1/(4r)$, set $\varepsilon' := \varepsilon/2$. Suppose in the following that n is sufficiently large and define $q' = n^{-1+\varepsilon'}$. Now let $q > n^{-1+\varepsilon}$ be given and observe that $q \geq 5q' \geq 1-(1-q')^5$. Finally, let $q'' \in (0,1]$ be such that

$$1 - q = (1-q'')(1-q')^4 \tag{1}$$

and note that since $q \geq 1-(1-q')^5$, we have $q'' \geq q'$.

Let $\eta_1 > 0$ be the constant given by Lemma 4 with input r, ε' and $\delta = 1/2$. Let $\mathcal{H}^* = \mathcal{H}^*(r, \varepsilon'/2)$ be the r-uniform reservoir hypergraph given by Lemma 2 and $n^* := v(\mathcal{H}^*)$. Let $\nu > 0$ be the constant given by Theorem 3 with input $\mathcal{H}^*$. We set

$$c := \min\left(\frac{1}{2n^*}, \frac{\nu}{n^*}, \eta_1\right). \tag{2}$$

Finally, let $\eta_2 > 0$ and ℓ_2 be the constants given by Lemma 4 with input r, ε' and $\delta = c/2$.

Preprocessing: We shall use a randomised procedure to split the input graph $\mathcal{G}$ which is distributed according to $\mathcal{G}^{(r)}(n,q)$ into five hypergraphs $\mathcal{G}_1, \dots, \mathcal{G}_5$, such that $\mathcal{G}_1$ is distributed according to $\mathcal{G}^{(r)}(n,q'')$

and $\mathcal{G}_2, \dots, \mathcal{G}_5$ are distributed according to $\mathcal{G}^{(r)}(n, q')$, where the choice of parameters is possible by (1). Moreover these five random hypergraphs are mutually independent.

Our randomised procedure takes a copy $\mathcal{G}$ of $\mathcal{G}^{(r)}(n, q)$ and colours its edges as follows. It colours each edge e of $\mathcal{G}$ independently with a non-empty subset c of $[5]$ such that

$$\Pr(e \text{ receives colour } c) = \begin{cases} q'^{|c|}(1-q')^{4-|c|}(1-q'')/q & \text{if } 1 \notin c \\ q'^{|c|-1}(1-q')^{5-|c|}q''/q & \text{if } 1 \in c\,. \end{cases}$$

Then we let $\mathcal{G}_i$ be the hypergraph with those edges whose colour contains i for each $i \in [5]$.

For justifying that this randomised procedure has the desired effect, let us consider the following second random experiment. We take five independent random hypergraphs, $\mathcal{G}_1 = \mathcal{G}^{(r)}(n, q'')$ and four copies $\mathcal{G}_2, \dots, \mathcal{G}_5$ of $\mathcal{G}^{(r)}(n, q')$, and form an r-uniform hypergraph on n vertices, whose edges are the union of $\mathcal{G}_1, \dots, \mathcal{G}_5$, each receiving a colour which is a subset of $[5]$ identifying the subset of $\mathcal{G}_1, \dots, \mathcal{G}_5$ containing that edge. Observe that we simply obtain $\mathcal{G}^{(r)}(n, q)$, when we ignore the colours in this union.

It is straightforward to check that the two experiments yield identical probability measures on the space of n-vertex coloured hypergraphs. It follows that any algorithm which with some probability finds a tight Hamilton cycle when presented with the five hypergraphs $\mathcal{G}_i$ of the first experiment succeeds with the same probability when presented with five hypergraphs obtained from the second experiment.

Step 1: The first main step of our algorithm finds cn vertex disjoint copies of the reservoir graph $\mathcal{H}^*$ in $\mathcal{G}_1$. To this end we would like to apply Theorem 3, hence we need to check its preconditions. We require that $q'' \geq Cn^{-1/m^{(1)}(\mathcal{H}^*)}$ for some large C. By Lemma 2 we have $m^{(1)}(\mathcal{H}^*) \leq 1 + \frac{1}{2}\varepsilon'$, and $1/(1 + \frac{1}{2}\varepsilon') > 1 - \varepsilon'$. It follows that for all sufficiently large n we have $q' = n^{-1+\varepsilon'} \geq Cn^{-1/m^{(1)}(\mathcal{H}^*)}$, and so the same holds for q'' since $q'' \geq q'$.

By Theorem 3 and (2), a.a.s. $\mathcal{G}_1$ contains at least $\nu n \geq n^* \cdot cn$ vertex disjoint copies of $\mathcal{H}^*$. Hence we can algorithmically *find* a subset of at least cn of them as follows. We search the vertex subsets of size n^* of G_1. Whenever we find a subset that induces $\mathcal{H}^*$ and does not share vertices with a previously chosen $\mathcal{H}^*$-copy, then we choose it. Clearly, we can do this until we chose cn vertex disjoint copies $\mathcal{H}_1, \dots, \mathcal{H}_{cn}$ of $\mathcal{H}^*$. This requires running time $O\big(n^{n^*}\big)$, where $n^* \leq 16\varepsilon^{-2}$ does not depend on n.

Step 2: The second step consists of using $\mathcal{G}_2$ and Lemma 4 with input r, ε' and $\delta = 1/2$ to connect the cn vertex disjoint reservoir graphs into one tight path. Let W^* consist of the cn reservoir vertices, one in each of $\mathcal{H}_1, \dots, \mathcal{H}_{cn}$. By (2), $\mathcal{H}_1, \dots, \mathcal{H}_{cn}$ cover at most $n/2$ vertices. By

Lemma 4 applied with $X = [n] \setminus \big(\bigcup_{i\in[cn]} V(\mathcal{H}_i)\big)$ there is a polynomial time algorithm which a.a.s. for each $1 \le i \le cn-1$ finds a tight path in $\mathcal{G}_2$ connecting the end $(r-1)$-tuple of $\mathcal{H}_i$ with the start $(r-1)$-tuple of $\mathcal{H}_{i+1}$, where these tight paths are disjoint and have their interior in X. This yields a tight path P in $\mathcal{G}_1 \cup \mathcal{G}_2$ containing all of the $\mathcal{H}_i$ with the following property. For any $W \subseteq W^*$, if we remove W from P, then we obtain (using the additional edges of the $\mathcal{H}_i$) a tight path $P(W)$ whose start and end $(r-1)$-tuples are the same as those of P (see Lemma 2(i)).

Step 3: In the third step we use $\mathcal{G}_3$ to greedily extend P to a tight path P' covering all but at most $n^{1-\frac{1}{2}\varepsilon'}$ vertices. Let $P_0 = P$ and do the following for each $i \ge 0$. Let $\mathbf{e}_i$ be the end $(r-1)$-tuple of P_i if there is an edge $\mathbf{e}_i v_i$ in $\mathcal{G}_3$ for some $v_i \in [n] \setminus V(P_i)$ then append v_i to P_i to obtain the tight path P_{i+1}. If no such edge exists, then halt.

Observe that in step i of this procedure, it suffices to reveal the edges $\mathbf{e}_i w$ with $w \in [n] \setminus P_i$. Hence, by the method of deferred decision, the probability that v_i does not exist is at most $(1-q')^{n-|P_i|}$. So, as long as $|P_i| \le n - n^{1-\frac{1}{2}\varepsilon'}$ this probability is at most $\exp(-q' n^{1-\frac{1}{2}\varepsilon'}) \le \exp(-n^{\frac{1}{2}\varepsilon'})$. We take the union bound over all (at most n) i to infer that this procedure a.a.s. indeed terminates with a tight path P' with $|P'| \ge n - n^{1-\frac{1}{2}\varepsilon'}$ which contains P.

Step 4: Now let L' be the set of vertices not covered by P'. Let L be obtained from L' by adding at most $r-2$ vertices of W^*, such that $|L|$ is divisible by $r-1$. Let $Y_1, \dots, Y_t$ be a partition of L into $|L|/(r-1)$ tuples of size $r-1$. Let Y_0 be the reverse of the start $(r-1)$-tuple of P', and Y_{t+1} be the reverse of its end $(r-1)$-tuple.

In the fourth step, we use $\mathcal{G}_4$ and Lemma 4 with input r, ε' and $\delta = \frac{1}{2}c$ to find for each $0 \le i \le \frac{1}{2}t$ a tight path between Y_{2i} and Y_{2i+1} of length at most ℓ_2 using only vertices in $W^* \setminus L$, such that these paths are pairwise disjoint. This is possible since $|W^* \setminus L| \ge \frac{1}{2}cn$ and since $t \le |L| \le n^{1-\frac{1}{2}\varepsilon'} + r - 2$ implies $\frac{t}{2} + 1 \le n^{1-\frac{1}{3}\varepsilon'} \le \eta_2 n$ for n sufficiently large. Let W^{**} be the set of at least $cn - (\frac{t}{2}+1)\ell_2 \ge cn - n^{1-\frac{1}{3}\varepsilon'}\ell_2 \ge \frac{2}{3}cn$ vertices in W^* not used in this step.

Step 5: Similarly, in the fifth step, we use $\mathcal{G}_5$ and Lemma 4, with input r, ε and $\delta = c/2$, to find for each $0 \le i \le \frac{1}{2}(t-1)$ a tight path between Y_{2i+1} and Y_{2i+2} of length at most ℓ_2 using only vertices in $W^{**} \setminus L$, such that these paths are pairwise disjoint. Again, $|W^{**} \setminus L| \ge \frac{1}{2}cn$ and $\frac{t}{2} + 1 \le \eta_2 n$ for n sufficiently large. Thus Lemma 4 guarantees that this step a.a.s. succeeds also and the tight paths can be found in polynomial time.

But now we are done: Let W be the vertices of W^* used in steps 4 and 5. By definition of W^* we can delete the vertices of W from P' and obtain a tight path $P'(W)$ through the remaining vertices of P'

(using additional edges of the reservoir graphs) and with the same start and end $(r-1)$-tuples. Then $P'(W)$ together with the connections constructed in steps 4 and 5 (which incorporated all vertices of L) form a Hamilton cycle in $\mathcal{G}$. $\square$

Remark 5. We note that the only non-deterministic part of the algorithm presented in the above proof concerns the partition of the edges of the input graph into five random subsets at the beginning.

The algorithm in the connection lemma (Lemma 4) is polynomial time, where the power of the polynomial is independent of ε. The same is (obviously) true for the greedy procedure of step 3. Finding many vertex disjoint reservoir graphs in step 1 however, we can only do in time $n^{16\varepsilon^{-2}}$.

3. Proof of the connection lemma

Preliminaries. For a binomially distributed random variable X and a constant γ with $0 < \gamma \le 3/2$ we will use the following Chernoff bound, which can be found, e.g., in [12, Corollary 2.3]:

$$\mathbb{P}\big(|X - \mathbb{E}X| \ge \gamma\mathbb{E}X\big) \le 2\exp(-\gamma^2\mathbb{E}X/3)\,. \tag{3}$$

In addition we apply the following consequence of Janson's inequality (see for example [12], Theorem 2.18): Let $\mathcal{E}$ be a finite set and $\mathcal{P}$ be a family of non-empty subsets of $\mathcal{E}$. Now consider the random experiment where each $e \in \mathcal{E}$ is chosen independently with probability p and define for each $P \in \mathcal{P}$ the indicator variable I_P that each element of P gets chosen. Set $X = \sum_{P\in\mathcal{P}} I_P$ and $\Delta = \frac{1}{2}\sum_{P\neq P', P\cap P'\neq\emptyset} \mathbb{E}(I_P I_{P'})$. Then

$$\mathbb{P}(X = 0) \le \exp(\Delta - \mathbb{E}X)\,. \tag{4}$$

For $e \in \binom{n}{r}$ we say that we *expose the r-set* e in $\mathcal{G}^{(r)}(n,p)$, if we perform (only) the random experiment of including e in $\mathcal{G}^{(r)}$ with probability p (recall that $p := n^{-1+\varepsilon}$). If this random experiment includes e then we say that e *appears*. Clearly, we can iteratively generate (a subgraph of) $\mathcal{G}^{(r)}(n,p)$ by exposing r-sets, as long as we do not expose any r-set twice. For a tuple $\mathbf{u}$ of at most $r-1$ vertices in $[n]$ we say that we *expose the r-sets at* $\mathbf{u}$, if we expose all r-sets $e \in \binom{n}{r}$ with $\mathbf{u} \subseteq e$. Similarly, we expose $\mathcal{H} \subseteq \binom{n}{r}$ if we expose all r-sets $e \in \mathcal{H}$.

In our algorithm we use the following structure. A *fan* $\mathcal{F}(\mathbf{u})$ in an r-uniform hypergraph $\mathcal{H}$ is a set $\{P_1, \dots, P_t\}$ of tight paths in $\mathcal{H}$ which all have length either ℓ or $\ell - 1$, start in the same $(r-1)$-tuple $\mathbf{u}$, and satisfy the following condition. For any set S of at least $r/2$ vertices, let $\{P_j\}_{j\in I}$ be the collection of tight paths in which the set S appears as a consecutive interval. Then the paths $\{P_j\}_{j\in I}$ also coincide between $\mathbf{u}$ and the interval S. The tuple $\mathbf{u}$ is also called the *root* of $\mathcal{F}(\mathbf{u})$. Moreover, ℓ is the *length* of $\mathcal{F}(\mathbf{u})$, and t its *width*. The set of *leaves* $L\big(\mathcal{F}(\mathbf{u})\big)$ of $\mathcal{F}(\mathbf{u})$ is the set of $(r-1)$-tuples $\mathbf{u}'$ such that some path in $\mathcal{F}$ ends in $\mathbf{u}'$. For intuition, observe that in the graph case $r = 2$, a fan is simply a rooted tree all of whose leaves are at distance either ℓ or $\ell - 1$ from the root. For $r \geq 3$, a fan is a more complicated structure.

Idea. We shall consecutively build the $\mathbf{u}^{(i)} - \mathbf{v}^{(i)}$ paths P_i in the set X, starting with P_1. The construction of the path P_i we call *phase* i, and the strategy in this phase is as follows. We shall first expose all the hyperedges at $\mathbf{u}^{(i)}$, excluding a set of 'used' vertices U (like those not in X, or in any $\mathbf{u}^{(i')}$ or $\mathbf{v}^{(i')}$). The edges $\{\mathbf{u}^{(i)}, c\}$ appearing in this process form possible starting edges for a path connecting $\mathbf{u}^{(i)}$ and $\mathbf{v}^{(i)}$. For each such (one edge) path P we next consider the $(r-1)$-endtuple of P and expose all edges at this tuple, excluding edges that were exposed earlier and used vertices (where now we count vertices in P as used). And so on. In this way we obtain a (consecutively growing) fan $\mathcal{F}(\mathbf{u}^{(i)})$ with root $\mathbf{u}^{(i)}$. While growing this fan we shall also insist that no j-tuple of vertices with $j < r$ is used too often. We stop when the fan has width $n^{1-\varepsilon/2}$. We will show that with high probability the fan then has only constant depth. Then we similarly construct a fan $\mathcal{F}(\mathbf{v}^{(i)})$ of width $n^{1-\varepsilon/2}$ with root $\mathbf{v}^{(i)}$ (again avoiding used vertices and exposed edges).

In a last step, for each leaf $\tilde{\mathbf{u}}^{(i)}$ of $\mathcal{F}(\mathbf{u}^{(i)})$ and each leaf $\tilde{\mathbf{v}}^{(i)}$ of $\mathcal{F}(\mathbf{v}^{(i)})$ we expose all $\tilde{\mathbf{u}}^{(i)} - \tilde{\mathbf{v}}^{(i)}$ paths of length $2(r-1)$, avoiding exposed edges. We shall show that with high probability at least one of these paths appears (and the fans $\mathcal{F}(\mathbf{u}^{(i)})$ and $\mathcal{F}(\mathbf{v}^{(i)})$ can be constructed), and hence we have successfully constructed P_i. We shall also show that, in phase i we only exposed much less than a $1/n$ fraction of the r-sets in X. Hence it is plausible that we can avoid these exposed r-sets in future phases. We note that this last statement makes use of the fact $r \geq 3$: our connection algorithm does not work for 2-graphs.

Proof of Lemma 4. Setup: Given $r \geq 3$, $\delta > 0$ and $0 < \varepsilon < 1/(4r)$, we set

$$\xi' := \delta/(48r^2)\,, \quad \xi := (\xi')^r/(r^2(r-1)!) \quad \text{and} \quad \eta = \delta/(16r)\,. \tag{5}$$

Without loss of generality we will assume $|X| = \delta n$: this simplifies our calculations.

In the algorithm described below, we maintain various auxiliary sets. We have a set U of *used vertices*, which contains all vertices in the sets $\mathbf{u}^{(i)}$ and $\mathbf{v}^{(i)}$, and in previously constructed connecting paths. In phase i we maintain additionally a (non-uniform) multihypergraph U_i of *used sets*, which keeps track of the number of times we have so far used a vertex, or pair of vertices, et cetera, consecutively in some path of the fan currently under construction.

Actually, it will greatly simplify the analysis if any such used set can only appear in a unique order on these paths. Hence we choose the following setup. We arbitrarily fix an equipartition

$$X = Y_1 \dot\cup \cdots \dot\cup Y_{2r} \dot\cup Y_1' \dot\cup \cdots \dot\cup Y_{2r}' ,$$

and set $Y := Y_1 \dot\cup \cdots \dot\cup Y_{2r}$ and $Y' := Y_1' \dot\cup \cdots \dot\cup Y_{2r}'$. We shall construct the fan $\mathcal{F}(\mathbf{u}^{(i)})$ with root $\mathbf{u}^{(i)}$ in $Y_1, \ldots, Y_{2r}$, taking successive levels of the fan from successive sets (in cyclic order), and similarly $\mathcal{F}(\mathbf{v}^{(i)})$ in $Y_1', \ldots, Y_{2r}'$.

Further, we maintain an r-uniform *exposé hypergraph* H, which keeps track of the r-sets which we have exposed. We let H_i be the hypergraph with the edges of H at the beginning of phase i.

We define hypergraphs $D_i^{(1)}, \ldots, D_i^{(r-1)}$ of *dangerous sets* for phase i as follows:

$$D_i^{(r-1)} := \Big\{ \mathbf{x} \in \tbinom{X}{r-1} \colon \deg_{H_i}(\mathbf{x}) \geq \xi n \Big\} , \qquad \text{and} \tag{6a}$$

$$D_i^{(j)} := \Big\{ \mathbf{x} \in \tbinom{X}{j} \colon \deg_{D_i^{(j+1)}}(\mathbf{x}) \geq \xi n \Big\} , \qquad r-2 \geq j \geq 1 . \tag{6b}$$

We will not use any set in any $D_i^{(j)}$ consecutively in a path in the fans constructed in phase i.

Given two vertex-disjoint $(r-1)$-sets $\mathbf{u}$ and $\mathbf{v}$, we say that the path $(\mathbf{u}, \mathbf{v})$ of length $2r-2$ is *blocked* by the exposé hypergraph H if any r consecutive vertices of the $(2r-2)$-set $\{\mathbf{u}, \mathbf{v}\}$ is in H. When constructing the fan $\mathcal{F}\big(\mathbf{v}^{(i)}\big)$ with root $\mathbf{v}^{(i)}$, we need to ensure that not too many of its leaves are blocked by H together with too many leaves of the previously constructed fan $\mathcal{F}\big(\mathbf{u}^{(i)}\big)$. For this purpose we define hypergraphs $\tilde{D}_i^{(j)}$ of temporarily dangerous sets in phase i as follows. We call an $(r-1)$-set $\mathbf{y}$ in Y' *temporarily dangerous* if there are at least $\xi' \big| L\big(\mathcal{F}(\mathbf{u}^{(i)})\big) \big|$ leaves $\mathbf{x}$ of $\mathcal{F}(\mathbf{u}^{(i)})$ such that $\{\mathbf{x}, \mathbf{y}\}$ is blocked by H_i. We define

$$\tilde{D}_i^{(r-1)} := \Big\{ \mathbf{y} \in \tbinom{Y'}{r-1} : \mathbf{y} \text{ is temporarily dangerous} \Big\} , \qquad \text{and} \tag{7a}$$

$$\tilde{D}_i^{(j)} := \Big\{ \mathbf{y} \in \tbinom{Y'}{j} \colon \deg_{\tilde{D}_i^{(j+1)}}(\mathbf{y}) \geq \xi' n \Big\} , \qquad \text{for } r-2 \geq j \geq 1 . \tag{7b}$$

Summarising, we do not want to append a vertex $c \in X \setminus U$ to the end $(r-1)$-tuple $\mathbf{a}$ of a path in one of our fans, if for $\mathbf{a}$ or for any end $(j-1)$-tuple $\mathbf{a}_{j-1}$ of $\mathbf{a}$ with $j \in [r-2]$ we have

(*i*) $\{\mathbf{a}, c\}$ is in H,

(*ii*) $\{\mathbf{a}_{j-1}, c\}$ is an edge of $D_i^{(j)}$ or of $\tilde{D}_i^{(j)}$, or

(*iii*) $\{\mathbf{a}_{j-1}, c\}$ has multiplicity greater than $\xi^{r-j} n^{(r-1)/2 - j(1-\varepsilon)}$ in U_i.

Hence we define the set $B(\mathbf{a})$ of *bad vertices* for $\mathbf{a}$ to be the set of vertices in $X \setminus U$ for which at least one of these conditions applies.

Algorithm: The desired paths P_i will be constructed using Algorithm 1. This algorithm constructs for each $i \in [k]$ two fans $\mathcal{F}(\mathbf{u}^{(i)})$ and $\mathcal{F}(\mathbf{v}^{(i)})$, using Algorithm 2 as a subroutine.

Algorithm 1: Connect each pair $\mathbf{u}^{(i)}$, $\mathbf{v}^{(i)}$ with a path P_i

$U := \bigcup_{i\in[k]}\{\mathbf{u}^{(i)},\mathbf{v}^{(i)}\}$; $H := \emptyset$;
foreach $i \in [k]$ **do**
1 construct the fan $\mathcal{F}(\mathbf{u}^{(i)})$ in $Y_1\dot\cup\dots\dot\cup Y_{2r}$;
2 construct the fan $\mathcal{F}(\mathbf{v}^{(i)})$ in $Y_1'\dot\cup\dots\dot\cup Y_{2r}'$;
let $L := L(\mathbf{u}^{(i)})$ be the leaves of $\mathcal{F}(\mathbf{u}^{(i)})$;
let $L' := L(\mathbf{v}^{(i)})$ be the leaves of $\mathcal{F}(\mathbf{v}^{(i)})$ reversed ;
$\mathcal{P} :=$ all $L-L'$-paths of length $2r-2$ not blocked by H ;
3 expose all edges which are in some $P \in \mathcal{P}$;
if one of these paths $\tilde{\mathbf{u}}^{(i)}, \tilde{\mathbf{v}}^{(i)}$ appears **then**
$P(\mathbf{u}^{(\mathbf{i})}) :=$ the path in $\mathcal{F}(\mathbf{u}^{(i)})$ ending with $\tilde{\mathbf{u}}^{(i)}$;
$P(\mathbf{v}^{(\mathbf{i})}) :=$ reversal of the path in $\mathcal{F}(\mathbf{v}^{(i)})$ ending with $\tilde{\mathbf{v}}^{(i)}$;
$P_i := P(\mathbf{u}^{(\mathbf{i})}), \tilde{\mathbf{u}}^{(i)}, \tilde{\mathbf{v}}^{(i)}, P(\mathbf{v}^{(\mathbf{i})})$;
4 **else** halt with **failure**;
5 $U := U \cup V(P_i)$;
6 **foreach** $\mathbf{x} \in L(\mathbf{u}^{(i)}), \mathbf{y} \in L(\mathbf{v}^{(i)})$ **do** $H := H \cup \binom{\mathbf{x}\cup\mathbf{y}}{r}$;
end

Algorithm 2: Construct the fan $\mathcal{F}(\mathbf{u}^{(i)})$

$\mathcal{F}(\mathbf{u}^{(i)}) := \{\mathbf{u}^{(i)}\}$; $U_i := \emptyset$; $t := 1$;
repeat forever
$\mathcal{P} := \mathcal{F}(\mathbf{u}^{(i)})$;
7 **foreach** path $P \in \mathcal{P}$ **do**
let $\mathbf{a}$ be the end $(r-1)$-tuple of P ;
8 expose all edges $\{\mathbf{a}, c\}$ with $c \in C' := Y_t \setminus (V(P) \cup U \cup B(\mathbf{a}))$;
9 $C := \{c\colon \{\mathbf{a}, c\}$ appears in previous step$\}$;
10 **if not** $\delta n^{\varepsilon}/(16r) \le |C| \le \delta n^{\varepsilon}/(2r)$ **then** halt with **failure**;
11 $\mathcal{F}(\mathbf{u}^{(i)}) := \big(\mathcal{F}(\mathbf{u}^{(i)}) \setminus \{P\}\big) \cup \big\{(P, c)\colon c \in C\big\}$;
12 $\mathbf{a}_j :=$ last j vertices of P for $j \in [r-2]$;
13 $U_i := U_i \cup C \cup \bigcup_{c\in C} \big\{\{\mathbf{a}_j, c\}\colon j \in [r-2]\big\}$;
14 $H := H \cup \big\{(\mathbf{a}, c)\colon c \in C'\big\}$;
15 **if** $|\mathcal{F}(\mathbf{u}^{(i)})| \ge n^{(r-1)/2-\varepsilon/2}$ **then return** ;
end
$t := (t \mod 2r) + 1$;
end

It is clear that the running time (whether the algorithm succeeds or fails) is polynomial: Steps 10 and 15 guarantee that in one call, Algorithm 2 runs at most $n^{(r-1)/2}$ times through its repeat loop. Our analysis will show that a.a.s. the algorithm indeed succeeds.

Before we proceed with the analysis, let us remind the reader that H denotes the already exposed hyperedges that appeared so far, H_i consists of the hyperedges of H before the start of phase i, U is the set of already used vertices and U_i is the auxiliary multihypergraph which is maintained through phase i and records those j-tuples ($j \in [r-1]$) that were used for constructing the fan $\mathcal{F}(\mathbf{u}^{(i)})$ ($\mathcal{F}(\mathbf{v}^{(i)})$ resp.).

Analysis: First, we claim that the algorithm is valid in that it does not try to expose any r-set twice. To see this, we need to check that at steps 3 and 8, we do not attempt to re-expose an already exposed r-set. Since we do not expose any r-set in H at either step (by the definition of $B(\mathbf{a})$), it is enough to check that after either step, all exposed r-sets are added to H before the next visit to either step. This takes place in steps 6 and 14.

In order to show that the algorithm succeeds, we need to show that the following hold with sufficiently high probability for each $i \in [k]$.

(A1) Algorithm 2 successfully builds the fans $\mathcal{F}(\mathbf{u}^{(i)})$ and $\mathcal{F}(\mathbf{v}^{(i)})$, that is, the condition in step 15 eventually becomes true, and the condition in step 10 never becomes true.

(A2) If this is the case, then Algorithm 1 successfully constructs P_i, that is, one of the paths exposed in step 3 appears.

(A3) If this is the case, then P_i is of length at most $s = \frac{r-1}{\varepsilon}$, that is, the fans $\mathcal{F}(\mathbf{u}^{(i)})$ and $\mathcal{F}(\mathbf{v}^{(i)})$ have length at most $s/2$.

It is straightforward to see that (A3) holds. Indeed, if Algorithm 2 succeeds in step i, then in the last repetition of the for-loop creating $\mathcal{F}(\mathbf{u}^{(i)})$, the width of $\mathcal{F}(\mathbf{u}^{(i)})$ finally exceeds $n^{(r-1)/2-\varepsilon/2}$. Since by step 10 at most $|C| \leq \delta n^{\varepsilon}/(2r) < n^{(r-1)/2-\varepsilon/2}$ paths are added to $\mathcal{F}(\mathbf{u}^{(i)})$ in this last for-loop (and the same holds for $\mathcal{F}(\mathbf{v}^{(i)})$), we obtain for the width of $\mathcal{F}(\mathbf{u}^{(i)})$ and $\mathcal{F}(\mathbf{v}^{(i)})$ (which equals the number of their leaves) that

$$n^{(r-1)/2-\varepsilon/2} \leq \big|L(\mathbf{u}^{(i)})\big|, \big|L(\mathbf{v}^{(i)})\big| \leq 2n^{(r-1)/2-\varepsilon/2}\,. \tag{8}$$

Now observe that by step 10 the fan $\mathcal{F}(\mathbf{u}^{(i)})$ (and similarly $\mathcal{F}(\mathbf{v}^{(i)})$) has width at least $\big(\delta n^{\varepsilon}/(16r)\big)^{s_i}$, where s_i is the length of $\mathcal{F}(\mathbf{u}^{(i)})$. For $s_i \geq (r-1)/(2\varepsilon)$ this would imply $\big|L(\mathbf{u}^{(i)})\big| \geq \big(\delta n^{\varepsilon}/(16r)\big)^{(r-1)/(2\varepsilon)} > n^{(r-1)/2-\varepsilon/2}$, contradicting (8). Hence we have (A3).

For proving (A1) and (A2), we first show bounds on various quantities during the running of the algorithm. For a set $\mathbf{a}$ in the multiset U_i with $i \in [k]$, we write $\mathrm{mult}_{U_i}(\mathbf{a})$ for the multiplicity of $\mathbf{a}$ in U_i.

Claim 6. If phase i and all phases before succeed, then the following hold throughout phase i.

(a) $|U| \leq k\big(s + 2(r-1)\big) \leq 2kr/\varepsilon$.

(b) For each $j \in [r-1]$ and each j-set $\mathbf{a} \in U_i$ we have
$$\mathrm{mult}_{U_i}(\mathbf{a}) \le \xi^{r-j} n^{((r-1)/2)-j(1-\varepsilon)} + 1\,.$$
(c) For each $j \in [r-1]$ and each $(j-1)$-set $\mathbf{a}$ in $[n]$, for all but ξn vertices $c \in X$ we have
$$\mathrm{mult}_{U_i}(\{\mathbf{a}, c\}) \le \xi^{r-j} n^{((r-1)/2)-j(1-\varepsilon)}\,.$$
(d) $e(H_{i+1}) \le 2^{2r+1}(i+1)n^{r-1-\varepsilon/2}$.

(e) At step 8 in Algorithm 2, we have $|Y_t \setminus (V(P) \cup U \cup B(\mathbf{a}))| \ge \delta n/(8r)$.

Observe that for $j \ge r/2$ Claim 6(b) implies that we always have $\mathrm{mult}_{U_i}(\mathbf{a}) \le 1$ for any j-tuple used in any $\mathcal{F}(\mathbf{u}^{(i)})$ or $\mathcal{F}(\mathbf{v}^{(i)})$. This shows that $\mathcal{F}(\mathbf{u}^{(i)})$ and $\mathcal{F}(\mathbf{v}^{(i)})$ are indeed fans, as we claim.

Proof of Claim 6. We first prove (a). The set U contains the $2k(r-1)$ vertices of the k pairs of $(r-1)$-tuples which we wish to connect, together with all the vertices of the paths thus far constructed. Since by (A3) these paths are of length at most s, it follows that $|U| \le 2k(r-1) + (i-1)s \le k\big(s + 2(r-1)\big)$.

To see that (b) holds, observe that j-sets are added to U_i only at step 13, and at this point the sets added are distinct: two sets either contain different members of Y_t, or they are of different sizes. Moreover, they are added only if their multiplicity in U_i is at most $\xi^{r-j} n^{(r-1)/2-j(1-\varepsilon)}$ by (iii) in the definition of $B(\mathbf{a})$.

For (c) we proceed by induction on j. First consider the case $j=1$. Observe that $c \in X$ is added to U_i in step 13 only if it is added at the end of a path P. Since step 10 guarantees that each fan grows by a factor of at least 2 in each iteration, we have
$$\sum_{c \in X} \mathrm{mult}_{U_i}(c) \le 2\big(|L(\mathbf{u}^{(i)})| + L(\mathbf{v}^{(i)})|\big) \overset{(8)}{\le} 4n^{(r-1)/2-\varepsilon/2} < \xi^r n^{(r-1)/2}\,.$$
We conclude that there are at most
$$\frac{\xi^r n^{(r-1)/2}}{\xi^{r-1} n^{((r-1)/2)-1+\varepsilon}} = \xi n^{1-\varepsilon}$$
vertices $c \in X$ with $\mathrm{mult}_{U_i}(c) > \xi^{r-1} n^{((r-1)/2)-1+\varepsilon}$.

Now assume that (c) holds for $j-1$ and let $\mathbf{a}$ be a $(j-1)$-set in $[n]$. Similarly as before, for $c \in X$ the set $\{\mathbf{a}, c\}$ is in U_i with multiplicity equal to the number of times that $\mathbf{a}$ has appeared as the end of a path P in one of the two fans constructed in this phase and the path (P, c) was subsequently added to the fan in step 11. Since we did not previously halt in step 10, for any P there are at most $\delta n^{\varepsilon}/(2r) \le \frac{1}{2} n^{\varepsilon}$ vertices $c \in X$ such that (P, c) is added in this way. Thus we have
$$\sum_{c \in X} \mathrm{mult}_{U_i}(\mathbf{a}, c) \le \mathrm{mult}_{U_i}(\mathbf{a}) \cdot \tfrac{1}{2} n^{\varepsilon}\,. \tag{9}$$

By (b) we know in addition that

$$\mathrm{mult}_{U_i}(\mathbf{a}) \le \xi^{r-j+1} n^{((r-1)/2)-(j-1)(1-\varepsilon)} + 1\,.$$

Note that if this bound is less than 2 then (9) directly implies that there are at most ξn vertices c with $\mathrm{mult}_{U_i}(\mathbf{a},c) \ge 1$ and we are done. Hence we may assume $\mathrm{mult}_{U_i}(\mathbf{a}) \le 2\xi^{r-j+1} n^{((r-1)/2)-(j-1)(1-\varepsilon)}$. This together with (9) also implies that there are at most

$$\frac{2\xi^{r-j+1} n^{((r-1)/2)-(j-1)(1-\varepsilon)} \cdot \frac{1}{2} n^{\varepsilon}}{\xi^{r-j} n^{((r-1)/2)-j(1-\varepsilon)}} = \xi n$$

vertices $c \in X$ with $\mathrm{mult}_{U_i}(\mathbf{a},c) \ge \xi^{r-j} n^{((r-1)/2)-j(1-\varepsilon)}$, as desired.

For the remaining parts of the claim, we proceed by induction on the phase $i \in [k]$. So assume that the claim holds at the end of the $(i-1)$st phase.

We next prove (d). At the end of phase i, the hypergraph H contains all the r-sets which it had at the end of phase $i-1$, together with all those added in phase i. Now consider the construction of one fan in phase i, say of $\mathcal{F}(\mathbf{u}^{(i)})$. Since we did not halt in step 10, the width of the fan grows exponentially, more than doubling at each step. Thus we can bound the total number of iterations of the for-loop by the number $|L(\mathbf{u}^{(i)})|$ of leaves of this fan (cf. step 11). In each of these iterations, we exposed $|Y_t \setminus (P \cup U \cup B(\mathbf{a}))| < n$ of the r-sets and added them to H. Hence, while constructing $\mathcal{F}(\mathbf{u}^{(i)})$ (and similarly for $\mathcal{F}(\mathbf{v}^{(i)})$), we added at most $|L(\mathbf{u}^{(i)})|n$ new r-sets to H. The only other step where we add r-tuples to H is step 6. In this step, for each pair of leaves of $\mathcal{F}(\mathbf{u}^{(i)})$ and $\mathcal{F}(\mathbf{v}^{(i)})$, we add $\binom{2r-2}{r}$ new r-sets to H. Using the induction hypothesis we thus conclude that at the end of phase i we have

$$\begin{aligned} e(H_{i+1}) &\le e(H_i) + \big(|L(\mathbf{u}^{(i)})| + |L(\mathbf{v}^{(i)})|\big)n + \tbinom{2r-2}{r}|L(\mathbf{u}^{(i)})| \cdot |L(\mathbf{v}^{(i)})| \\ &\overset{(8)}{\le} 2^{2r+1} i \cdot n^{r-1-\varepsilon/2} + 4n^{(r+1)/2-\varepsilon/2} \cdot n + \tbinom{2r-2}{r} 4n^{r-1-\varepsilon} \\ &\le 2^{2r+1}(i+1)n^{r-1-\varepsilon/2}\,, \end{aligned}$$

where for the final inequality we use the fact that $(r+1)/2 \le r-1$, which holds since $r \ge 3$. This is the only step in the analysis where we use $r \ge 3$, but this analysis is reasonably tight: the algorithm does fail for $r = 2$.

Last we prove (e), for which we additionally proceed by induction on the number f of iterations through the for-loop of Algorithm 2 done in the ith phase so far. So we assume that the claim holds at the end of the $(i-1)$st phase and after $f-1$ iterations.

Let P be the path considered in iteration f of this for-loop, and $\mathbf{a}$ the $(r-1)$-tuple ending P. We would like to estimate the size of $B(\mathbf{a}) \cap Y_t$. Keep in mind in the following analysis that for $j \in [r-1]$ the hypergraph D_i^j does not change during phase i, by definition. Similarly, $\tilde{D}_i^j$ does not change once the fan $\mathcal{F}(\mathbf{u}_i)$ is constructed.

Now let us first assess the effect of (i) of the definition of $B(\mathbf{a})$. Since $\mathbf{a}$ is the end of a path constructed by Algorithm 2, step 8 implies that the last vertex b of $\mathbf{a}$ is not contained in $B(\mathbf{b})$ where $\mathbf{b}$ is the end $(r-1)$-tuple of $P-b$. From (ii) in the definition of $B(\mathbf{b})$ we conclude that $\mathbf{a} \notin D_i^{(r-1)}$. Thus, by the definition of $D_i^{(r-1)}$ in (6a), the number of edges in H_i containing $\mathbf{a}$ is smaller than ξn.

But how many edges $\{\mathbf{a}, c\}$ with $c \in Y_t$ did phase i add to H so far? By (b) the set $\mathbf{a}$ has multiplicity at most $\xi n^{(r-1)(2\varepsilon-1)/2} + 1 < 2$ in U_i. It follows that since the start of phase i only one edge containing $\mathbf{a}$ was added to H in step 14: the end r-tuple $\mathbf{a}_r$ of P. However, since $\mathbf{a}_r$ contains no vertices of Y_t because the algorithm takes successive levels of the fan in successive $Y_{t'}$ (or $Y'_{t'}$), we conclude that the current phase did not add any additional edges $\{\mathbf{a}, c\}$ to H with $c \in Y_t$.

Now let us estimate the number of vertices $c \in Y_t$ which (ii) of the definition of $B(\mathbf{a})$ forbids. First, we need to consider the case $j = 1$, and show that the number of vertices in $D_i^{(1)}$ is at most ξn. Suppose not, and observe that by definition (6b), each vertex in $D_i^{(1)}$ extends to at least ξn pairs in $D_i^{(2)}$, and so on, where at the final step each constructed member of $D_i^{(r-1)}$ extends to at least ξn members of H_i. We can construct any given member of H_i in at most $r!$ ways, so we conclude that $e(H_i) \geq (\xi n)^r / r!$, which (for sufficiently large n) contradicts part (d).

Next, again for the case $j = 1$, we need to show that further there are at most $\xi' n$ vertices in $\tilde{D}_i^{(1)}$. Again, suppose not: then as above this implies that the number of pairs of $(r-1)$-tuples $(\mathbf{x}, \mathbf{y})$ with $\mathbf{x} \in L\big(\mathbf{u}^{(i)}\big)$ and $\mathbf{y}$ contained in $Y'_1 \cup \ldots \cup Y'_{2r}$ is at least

$$\xi' \Big| L\big(\mathbf{u}^{(i)}\big) \Big| \cdot (\xi' n)^{r-1} / (r-1)! \overset{(5)}{\geq} r^2 \xi n^{r-1} \Big| L\big(\mathbf{u}^{(i)}\big) \Big| . \tag{10}$$

However, by construction of $\mathcal{F}(\mathbf{u}^{(i)})$, for each $j \in [r-1]$ and each $\mathbf{x} \in L\big(\mathbf{u}^{(i)}\big)$, we have the property that the last j vertices of $\mathbf{x}$ are not in $D_i^{(j)}$. We claim that this implies that the number of $(r-1)$-tuples $\mathbf{y}$ contained in $Y'_1 \cup \ldots \cup Y'_{2r}$ such that $(\mathbf{x}, \mathbf{y})$ is blocked by H_i, is at most $(r-1)^2 \xi n^{r-1}$, which is a contradiction to (10). To see this, consider the following property P of tuples $\mathbf{y}$. For each $r-1 \geq j \geq 1$ and each $1 \leq k \leq r-j$, the tuple consisting of the last j vertices of $\mathbf{x}$ followed by the first k vertices of $\mathbf{y}$ is not in $D_i^{(j+k)}$ (if $j+k < r$) and not in H_i (if $j+k = r$). If $\mathbf{y}$ has property P, then clearly the pair $(\mathbf{x}, \mathbf{y})$ is not blocked by H_i. On the other hand, if $\mathbf{y}$ does not have P, then there is a smallest k for which P fails. By definition of the sets $D_i^{(j+k)}$, for a fixed j given the first $k-1$ vertices of $\mathbf{y}$ there are at most ξn choices for the k-th vertex of $\mathbf{y}$. Hence, in total, given the first $k-1$ vertices of

$\mathbf{y}$ there are at most $(r-1)\xi n$ choices for the k-th vertex of $\mathbf{y}$. Thus the number of $(r-1)$-tuples $\mathbf{y}$ which do not have P is at most $(r-1)^2\xi n^{r-1}$ as desired.

Now given $2 \le j \le r-2$, let $\mathbf{a}_{j-1}$ be the set consisting of the last $j-1$ vertices of $\mathbf{a}$. By construction, $\mathbf{a}_{j-1}$ is in neither $D_i^{(j-1)}$ nor in $\tilde{D}_i^{(j-1)}$. It follows from the definition of these sets in (6b) and (7b) that there are at most ξn vertices c such that $\{\mathbf{a}_{j-1}, c\} \in D_i^{(j)}$, and at most $\xi' n$ such that $\{\mathbf{a}_{j-1}, c\} \in \tilde{D}_i^{(j)}$. Together with the case $j=1$, this gives at most $(r-2)(\xi+\xi')n$ forbidden vertices $c \in Y_t$.

Finally, for (*iii*), observe that by part (*b*), for each $j \in [r-2]$ there are at most ξn vertices $c \in X$ with $\mathrm{mult}_{U_i}\big(\{\mathbf{a}_{j-1}, c\}\big) > \xi^{r-j} n^{\frac{r-1}{2}-j(1-\varepsilon)}$.

Hence, in total, $B(\mathbf{a}) \cap Y_t$ contains at most

$$\xi n + (r-2)(\xi+\xi')n + (r-2)\xi n \overset{(5)}{\le} \frac{\delta n}{4r}$$

vertices. Moreover, it follows from (A3) that $|P| \le r/\varepsilon$, and from (*a*) that $|U| \le 2kr/\varepsilon$. Since we have $|Y_t| = \delta n/(2r)$, we conclude that

$$|Y_t \setminus (P \cup U \cup B(\mathbf{a}))| \ge \frac{\delta n}{2r} - \frac{r}{\varepsilon} - \frac{2kr}{\varepsilon} - \frac{\delta n}{4r} \ge \frac{\delta n}{8r}$$

□

Now we can use a Chernoff bound to show that a.a.s. Algorithm 2 does not fail in step 10.

Claim 7. At any given visit to step 10, Algorithm 2 halts with probability at most $2\exp\big(-\delta n^{\varepsilon}/(96r)\big)$.

Proof. By Claim 6(*e*), we have

$$\delta n/(8r) \le |Y_t \setminus (P \cup U \cup B(\mathbf{a}))| \le |Y_t| = \delta n/(4r)\,.$$

Since C is a p-random subset of $Y_t \setminus (P \cup U \cup B(\mathbf{a}))$ with $p = n^{-1+\varepsilon}$, we obtain $\delta n^{\varepsilon}/(8r) \le \mathbb{E}|C| \le \delta n^{\varepsilon}/(4r)$. Using the Chernoff bound (3) with $\gamma = 1/2$, we conclude that $\delta n^{\varepsilon}/(16r) \le |C| \le \delta n^{\varepsilon}/(2r)$ with probability at least $1 - 2\exp\big(-\delta n^{\varepsilon}/(96r)\big)$. □

We would like to show that also a.a.s. Algorithm 1 does not fail in step 4. Since the events considered in this step are not mutually independent, we use Janson's inequality for this purpose.

Claim 8. At any given visit to step 4, Algorithm 1 halts with probability at most $\exp(-n^{(r-2)\varepsilon}/4)$.

Proof. Let $\mathcal{E} = \bigcup \mathcal{P}$ be the family of r-sets exposed in step 3 in this iteration of the foreach-loop. For each $P \in \mathcal{P}$ let I_P be the indicator variable for the event that the path P appears, which occurs with probability $\tilde{p} = p^{r-1}$. Then $X = \sum_{P\in\mathcal{P}} I_P$ is the random variable counting the number of $L-L'$-paths appearing in this iteration. We would like to use Janson's inequality (4) to show that $X > 0$ with high probability, in which case Algorithm 1 does not halt in step 4.

To this end we first bound $\mathbb{E}X$, for which we need to estimate $|\mathcal{P}|$. Firstly, since $\mathcal{F}(\mathbf{u}^{(i)})$ and $\mathcal{F}(\mathbf{v}^{(i)})$ are disjoint fans, no vertex is in a leaf both of $\mathcal{F}(\mathbf{u}^{(i)})$ and of $\mathcal{F}(\mathbf{v}^{(i)})$, and in particular $L(\mathbf{u}^{(i)})$ and $L(\mathbf{v}^{(i)})$ are disjoint. Now let $\tilde{\mathbf{v}}$ be any $(r-1)$-tuple in $L(\mathbf{v}^{(i)})$. By construction $\tilde{\mathbf{v}}$ is not in $\tilde{D}_i^{(r-1)}$ (see step 8 and the definition of $B(\mathbf{a})$). By (7a) the path $(\tilde{\mathbf{u}}, \tilde{\mathbf{v}})$ therefore is blocked by H for at most $\xi'\big|L(\mathbf{u}^{(i)})\big|$ tuples $\tilde{\mathbf{u}} \in L(\mathbf{u}^{(i)})$. Thus we have

$$|\mathcal{P}| \geq \big|L(\mathbf{v}^{(i)})\big| \cdot (1-\xi')\big|L(\mathbf{u}^{(i)})\big| \overset{(8)}{\geq} (1-\xi')n^{r-1-\varepsilon},$$

which gives

$$\mathbb{E}X = |\mathcal{P}|\tilde{p} \geq (1-\xi')n^{r-1-\varepsilon}n^{(\varepsilon-1)(r-1)} = (1-\xi')n^{(r-2)\varepsilon}. \tag{11}$$

Next we would like to estimate $\mathbb{E}(I_P I_{P'})$ for two distinct paths $P = (\tilde{\mathbf{u}}, \tilde{\mathbf{v}})$ and $P' = (\tilde{\mathbf{u}}', \tilde{\mathbf{v}}')$ which share at least one edge. If P and P' are distinct paths sharing at least one edge, then in particular, either $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{u}}'$ have the same end $r/2$-tuple, or $\tilde{\mathbf{v}}$ and $\tilde{\mathbf{v}}'$ have the same start $r/2$-tuple. Without loss of generality assume the former and suppose that $\tilde{\mathbf{v}}$ and $\tilde{\mathbf{v}}'$ match in the start j-tuple, but not in the $(j+1)$st vertex. Clearly $1 \leq j$, and since $\mathcal{F}(\mathbf{u}^{(i)})$ and $\mathcal{F}(\mathbf{v}^{(i)})$ are fans we have $\tilde{\mathbf{u}} = \tilde{\mathbf{u}}'$ and $j < r/2$. Hence P and P' share precisely an interval of length $r-1+j$, and thus j edges. Therefore $\mathbb{E}(I_P I_{P'}) \leq p^{2r-2-j}$.

In addition, the discussion above shows that for a fixed path $P = (\tilde{\mathbf{u}}, \tilde{\mathbf{v}})$, the number $N_{P,j}$ of paths $P' = (\tilde{\mathbf{u}}', \tilde{\mathbf{v}}')$ such that P and P' share j edges, is at most the number of choices of a leaf $\tilde{\mathbf{v}}' \in L(\mathbf{v}^{(i)})$ such that $\tilde{\mathbf{v}}'$ only has the end $(r-1-j)$-tuple $\mathbf{v}$ different from $\tilde{\mathbf{v}}$, plus the number of choices of a leaf $\tilde{\mathbf{u}}' \in L(\mathbf{u}^{(i)})$ such that $\tilde{\mathbf{u}}'$ only has the start $(r-1-j)$-tuple $\mathbf{u}$ different from $\tilde{\mathbf{u}}$. By Claim 6(b) the start j-tuple of $\tilde{\mathbf{v}}$ and the end j-tuple of $\tilde{\mathbf{u}}$ have multiplicity in U_i at most $n^{(r-1)/2-j(1-\varepsilon)}+1$. By step 13 this implies that there are at most $n^{(r-1)/2-j(1-\varepsilon)}$ choices for $\mathbf{u}$ and for $\mathbf{v}$, and hence $N_{P,j} \leq 2n^{(r-1)/2-j(1-\varepsilon)}$.

With this we are ready to estimate

$$\Delta = \sum_{P \neq P', P \cap P' = \emptyset} \mathbb{E}(I_P I_{P'}) = \sum_{P} \sum_{1 \leq j < r/2} \Big(\sum_{|P' \cap P| = j} \mathbb{E}(I_P I_{P'}) \Big),$$

where $P, P' \in \mathcal{P}$. We have

$$\begin{aligned} \Delta &\leq \sum_{P} \sum_{1 \leq j < r/2} N_{P,j} \cdot p^{2r-2-j} \\ &\leq |L(\mathbf{u}^{(i)})||L(\mathbf{v}^{(i)})| \sum_{1 \leq j < r/2} 2n^{(r-1)/2-j(1-\varepsilon)} p^{2r-2-j}, \end{aligned}$$

which, by (8), is at most

$$n^{(r-1)-\varepsilon}\sum_{1\le j<r/2} 2n^{(r-1)/2-j(1-\varepsilon)}n^{(\varepsilon-1)(2r-2-j)},$$
$$\le n^{(r-1)-\varepsilon}\cdot r\cdot n^{-\frac{3}{2}(r-1)+2\varepsilon(r-1)} < 1.$$

Hence, inequalities (4) and (11) imply that $\mathbb{P}(X=0)\le\exp(\Delta-\mathbb{E}X)\le\exp(-n^{(r-2)\varepsilon}/4)$, and thus Algorithm 1 fails with at most this probability in this visit to step 4 □

Since Algorithm 1 visits step 4 at most $k\le n$ times, we can use Claim 8 and a union bound to infer that (A2) holds with probability at least $1-n\cdot\exp\big(-n^{(r-2)\varepsilon}/4\big)\ge 1-\frac{1}{2}\exp\big(-\delta n^{\varepsilon}/(100r)\big)$. Similarly, step 10 of Algorithm 2 is called at most once per leaf in any of the at most $2k$ constructed fans, which is at most $2k\cdot 2n^{(r-1)/2-\varepsilon/2}\le n^r$ times by (8). It follows from Claim 7 that (A1) holds with probability at least $1-n^r\cdot 2\exp\big(-\delta n^{\varepsilon}/(96r)\big)\ge 1-\frac{1}{2}\exp(-\delta n^{\varepsilon}/(100r)\big)$.

Summarising, we showed that Algorithm 1 constructs the k desired tight paths of length at most ℓ with probability at least $1-\exp(-\delta n^{\varepsilon}/(100r)\big)$. □

4. Proof of the reservoir lemma

In this section we prove Lemma 2.

Proof of Lemma 2. Choose $\ell := \lceil 1/\big(2(r-1)\varepsilon\big)\rceil + 2$. Our strategy will be as follows. We will start by defining an auxiliary r-uniform hypergraph $\mathcal{D}_\ell^{(r)}$ with $2(r-1)(2\ell-1)+1$ vertices and as many edges, which implies

$$d^{(1)}(\mathcal{D}_\ell^{(r)}) = 1+\frac{1}{2(r-1)(2\ell-1)}\le 1+\varepsilon\,. \tag{12}$$

After defining $\mathcal{D}_\ell^{(r)}$ we shall construct a graph $\mathcal{H}^*$ which satisfies (ii) and (iii) and is such that $\mathcal{D}_\ell^{(r)}\subseteq\mathcal{H}^*$ and $\mathcal{D}_\ell^{(r)}$ has maximum 1-density among all subhypergraphs of $\mathcal{H}^*$.

The vertex set of $\mathcal{D}_\ell^{(r)}$ is

$$V(\mathcal{D}_\ell^{(r)}) := U\cup V\cup\bigcup_{i\in[\ell-1]}A_i\cup\bigcup_{i\in[\ell-2]}B_i\,.$$

where $U := (u_1,\dots,u_{r-1},w^*,u_r,\dots,u_{2(r-1)})$, $V := (v'_1,\dots,v'_{2(r-1)})$, $A_i := (a_1^{(i)},\dots a_{2(r-1)}^{(i)})$ for $i\in[\ell-1]$, and $B_i := (b_1^{(i)},\dots b_{2(r-1)}^{(i)})$ for $i\in[\ell-2]$ are ordered sets of vertices. The edge set of $\mathcal{D}_\ell^{(r)}$ contains exactly the edges of the tight paths determined by U, by V, by A_i for

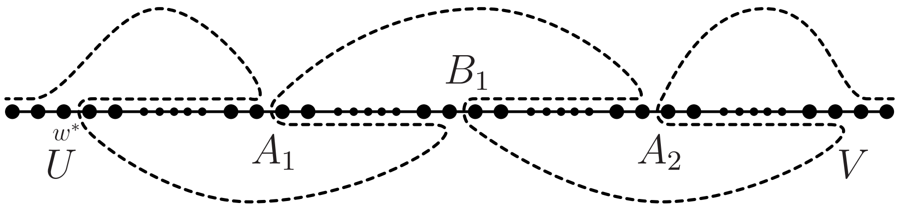

FIGURE 1. $\mathcal{H}^*$ for $r = 3$ and $\ell = 3$. The vertices of $\mathcal{D}_\ell^{(r)}$ are drawn bigger than the vertices newly inserted in $\mathcal{H}$. The continuous line indicates the tight Hamilton path in $\mathcal{H}^*$ from (13), the dashed line the tight Hamilton path in $\mathcal{H}^* - w^*$.

each $i \in [\ell - 1]$, by B_i for each $i \in [\ell - 2]$, as well as by the following vertex sequences:

$$\begin{aligned}
\tilde{U}_A &:= (u_1, \dots, u_{r-1}, a_{r-1}^{(1)}, \dots, a_1^{(1)}) \,, \\
\tilde{V}_A &:= (a_{2(r-1)}^{(\ell-1)}, \dots, a_r^{(\ell-1)}, v'_r, \dots, v'_{2(r-1)}) \,, \\
\tilde{U}_B &:= (u_{2(r-1)}, \dots, u_r, b_{r-1}^{(1)}, \dots, b_1^{(1)}) \,, \\
\tilde{V}_B &:= (b_{2(r-1)}^{(\ell-2)}, \dots, b_r^{(\ell-2)}, v'_{r-1}, \dots, v'_1) \,,
\end{aligned}$$

and

$$\begin{aligned}
\tilde{A}_{i,i+1} &:= (a_{2(r-1)}^{(i)}, \dots, a_r^{(i)}, a_{r-1}^{(i+1)}, \dots, a_1^{(i+1)}) \qquad \text{for all } i \in [\ell - 2] \,, \\
\tilde{B}_{i,i+1} &:= (b_{2(r-1)}^{(i)}, \dots, b_r^{(i)}, b_{r-1}^{(i+1)}, \dots, b_1^{(i+1)}) \qquad \text{for all } i \in [\ell - 3] \,.
\end{aligned}$$

It is not difficult to check that $\mathcal{D}_\ell^{(r)}$ has exactly $2(r-1)(2\ell-1)+1$ vertices and edges as claimed.

In order to obtain $\mathcal{H}^*$ from $\mathcal{D}_\ell^{(r)}$ we first let $v_i := v'_{(r-1)+i}$ for each $i \in [r-1]$. Then we insert

$$k := 3(r-1)^2(2\ell - 1)$$

new vertices 'between' each of the following pairs of vertex sets in $\mathcal{D}_\ell^{(r)}$: U and A_1, $A_{\ell-1}$ and V, A_i and B_i for each $i \in [\ell - 2]$, B_i and A_{i+1} for each $i \in [\ell - 2]$. We let $I(X, Y)$ denote the ordered set of vertices inserted 'between' the sets X and Y in this process (where we choose any ordering). In addition, we add to this graph the tight Hamilton path

$$\begin{aligned}
\text{(13)} \quad & U, I(U, A_1), A_1, I(A_1, B_1), B_1, I(B_1, A_2), A_2, \dots \\
& \qquad \dots, B_{\ell-2}, I(B_{\ell-2}, A_{\ell-1}), A_{\ell-1}, I(A_{\ell-1}, V), V
\end{aligned}$$

running from $\mathbf{u}$ to $\mathbf{v}$ (which uses some edges already present in $\mathcal{D}_\ell^{(r)}$). The resulting hypergraph is $\mathcal{H}^*$ (see also Figure 1).

By construction $v(\mathcal{H}^*) = 2(r-1)(2\ell-1)+1+2k(\ell-1)$ which by definition of k is smaller than $16(r-1)^2\ell^2 \le 16\varepsilon^{-2}$, and $e(\mathcal{H}^*) = 2(r-1)(2\ell-1)+1+2(k+r-1)(\ell-1)$. Since $\ell > 1$ this implies

$$\begin{aligned} d^{(1)}(\mathcal{H}^*) &= 1 + \frac{2(r-1)(\ell-1)+1}{2(r-1)(2\ell-1)+2k(\ell-1)} \\ &< 1 + \frac{2(r-1)(\ell-1)+1}{2k(\ell-1)} \le 1 + \frac{1}{2(r-1)(2\ell-1)} \\ &\overset{(12)}{=} d^{(1)}(\mathcal{D}_\ell^{(r)})\,. \end{aligned} \tag{14}$$

By (13) the hypergraph $\mathcal{H}^*$ satisfies (ii). We define $\tilde{I}(Y,X)$ to be the reversal of $I(X,Y)$. It can be checked that $\mathcal{H}^*$ also contains the tight path

$$\begin{aligned} \tilde{U}_A, \tilde{I}(A_1,U), \tilde{U}_B, \tilde{I}(B_1,A_1), \tilde{A}_{1,2}, \tilde{I}(A_2,B_1), \tilde{B}_{1,2}, \tilde{I}(B_2,A_2), \tilde{A}_{2,3}, \dots \\ \dots, \tilde{A}_{\ell-2,\ell-1}, \tilde{I}(A_{\ell-1},B_{\ell-2}), \tilde{V}_B, \tilde{I}(V,A_{\ell-1}), \tilde{V}_A\,. \end{aligned}$$

This is a tight path from $\mathbf{u}$ to $\mathbf{v}$ running through all vertices of $\mathcal{H}^*$ but w^*, and so $\mathcal{H}^*$ also satisfies (iii). It remains to show that $\mathcal{D}_\ell^{(r)}$ has maximal 1-density among all subgraphs of $\mathcal{H}^*$.

Suppose that $\mathcal{H}$ is a subgraph of $\mathcal{H}^*$ with maximal 1-density. It follows that $\mathcal{H}$ is an induced subgraph of $\mathcal{H}^*$, and that we have $d^{(1)}(\mathcal{H}) \ge d^{(1)}(\mathcal{D}_\ell^{(r)}) > 1$. It follows that $\mathcal{H}$ cannot contain any vertex of degree one (otherwise we could delete it and increase the 1-density). In particular, this means that if $I(X,Y)$ is any of the sets of k vertices which form a tight path in $\mathcal{H}^*$ and which are not present in $\mathcal{D}_\ell^{(r)}$, then either every vertex of $I(X,Y)$ is in $\mathcal{H}$, or none are. Similarly, by the definition of k we have $k \cdot d^{(1)}(\mathcal{D}_\ell^{(r)}) > k + (r-1)$ and so $\mathcal{H}$ cannot contain any k vertices meeting only $k+r-1$ edges. Accordingly $\mathcal{H}$ cannot contain $I(X,Y)$. It follows that $\mathcal{H}$ must be a subgraph of $\mathcal{D}_\ell^{(r)}$.

It is straightforward to check that if any of the vertices

$$\begin{aligned} S := \{u_2, \dots, u_{2(r-1)-1}, w^*, v_2, \dots, v_{2(r-1)-1}, \\ a_2^{(i)}, \dots, a_{2(r-1)-1}^{(i)}, b_2^{(i)}, \dots, b_{2(r-1)-1}^{(i)}\} \end{aligned}$$

of $\mathcal{D}_\ell^{(r)}$ is removed from $\mathcal{D}_\ell^{(r)}$, then we obtain a graph which can be decomposed by successively removing vertices of degree at most one (i.e. it is 1-degenerate) and which therefore has 1-density at most 1. It follows that $S \subseteq V(\mathcal{H})$. Now let x be any vertex of $\mathcal{D}_\ell^{(r)}$ which is not in $\mathcal{H}$. Since $x \notin S$ we have $\deg_{\mathcal{D}_\ell^{(r)}}(x) = 2$, and both edges containing x have all their remaining vertices in S. Thus we have

$$d^{(1)}\big(\mathcal{D}_\ell^{(r)}\big[V(\mathcal{H}) \cup \{x\}\big]\big) \ge \min\big(d^{(1)}(\mathcal{H}), 2\big)$$

and since $d^{(1)}(\mathcal{D}_\ell^{(r)}) < 2$, we conclude that $d^{(1)}(\mathcal{H}) \le d^{(1)}(\mathcal{D}_\ell^{(r)})$ as required. □

5. Concluding remarks

Graphs. We remark that our approach does not work (as such) in the case $r = 2$, even for the sub-optimal edge probability $n^{\varepsilon-1}$. For this case, in the proof of the connection lemma, Lemma 4, when growing a fan we would have to reveal in each iteration of the foreach-loop in Algorithm 2 more than $n^{1-\varepsilon}$ edges at a vertex a. In the construction of one fan we would have to repeat this operation at least $n^{(1/2)-2\varepsilon}$ times: only then we could hope for the fan to have $n^{(1/2)-\varepsilon}$ leaves, which we need in order to get a connection between two such fans at least in expectation. But then we would have revealed at least $n^{1-\varepsilon} \cdot n^{(1/2)-2\varepsilon} = n^{(3/2)-3\varepsilon}$ edges to obtain a single connection. Hence, we cannot obtain a linear number of connections in this way, as required by our strategy.

Vertex disjoint cycles. It is easy to modify our approach to show the following theorem.

Theorem 9. *For every integer $r \geq 3$ and for every $\varepsilon, \delta > 0$ the following holds. Suppose that $n_1, \ldots, n_\ell$ are integers, each at least $2r/\varepsilon$, whose sum is at most n, and $n_1 \geq \delta n$. Then for any $n^{-1+\varepsilon} < p = p(n) \leq 1$, the random r-uniform hypergraph $\mathcal{G}^{(r)}(n,p)$ contains a collection of vertex disjoint tight cycles of lengths $n_1, \ldots, n_\ell$ with probability tending to one as n tends to infinity.*

A proof sketch is as follows. We refer to the steps used in the proof of Theorem 1.

First, we would run step 1 as before, except that we would find reservoir graphs covering only at most $\varepsilon\delta n/(8r)$ vertices. Step 2 remains unchanged. We would then in an extra step (requiring an extra round of probability) to create greedily a collection of vertex disjoint tight paths of lengths slightly shorter than $n_2, \ldots, n_\ell$, and another extra step using Lemma 4 to connect these paths into tight cycles of lengths $n_2, \ldots, n_\ell$. Here we require that the connecting paths always have a precisely specified length. As written, Lemma 4 does not guarantee this (the output paths have lengths differing by at most two, since the paths in each fan can differ in length by one) but it is easy to modify the lemma to obtain this (we would simply extend each of the shorter fan paths by one vertex while avoiding dangerous sets). The remainder of the proof can remain almost unchanged. We extend the reservoir path greedily to cover most of the remaining vertices. Then we apply Lemma 4 twice to cover all the leftover vertices and complete a cycle. Then this cycle has length n_1 as desired. (The only difference is that some of our constants will need to be adapted slightly.)

Again, for fixed r, ε and δ we obtain a randomised polynomial time algorithm from this proof. Note that the condition that the cycles should not be too short cannot be completely removed: in order to have linearly many cycles of length g with high probability, we require that linearly many such cycles exist in expectation. This expectation is of the order $n^g p^g$, which is in $o(n)$ if $p = o\big(n^{-(g-1)/g}\big)$.

Derandomisation. Our approach to Theorem 1 yields a randomised algorithm. However we only actually use the power of randomness in order to preprocess our input hypergraph and 'simulate' multi-round exposure. This motivates the following question.

Question 10. *For a constructive proof which uses multi-round exposure, how can one obtain a* deterministic *algorithm?*

Replacing the randomised preprocessing step with a deterministic splitting of the edges of the complete r-uniform hypergraph into disjoint dense quasirandom subgraphs might be a promising strategy here.

Multi-Round exposure is a very common technique in probabilistic combinatorics. Hence this question might be of interest for other problems as well.

Resilience. A very active recent development in the theory of random graphs is the concept of resilience: under which conditions can one transfer a classical extremal theorem to the random graph setting? Lee and Sudakov [18], improving on previous work of Sudakov and Vu [22], showed that Dirac's theorem can be transferred to random graphs almost as sparse as at the threshold for hamiltonicity. More precisely, they proved that for each $\varepsilon > 0$, if $p \geq C \log n/n$ for some constant $C = C(\varepsilon)$, then almost surely the random graph $G = G(n, p)$ has the following property. Every spanning subgraph of G which has minimum degree $(\frac{1}{2} + \varepsilon)pn$, contains a Hamilton cycle.

It would be interesting to prove a corresponding result for tight Hamilton cycles in subgraphs of random hypergraphs. It is unlikely that the Second Moment Method will provide help for this. Our methods, however, might be robust enough to provide some assistance.

6. Acknowledgements

We would like to thank Klas Markström for suggesting Theorem 9.

References

[1] D. Angluin and L. G. Valiant, *Fast probabilistic algorithms for Hamiltonian circuits and matchings*, J. Comput. System Sci. **18** (1979), no. 2, 155–193.

[2] D. Bal and A. Frieze, *Packing tight Hamilton cycles in uniform hypergraphs*, SIAM Journal on Discrete Mathematics **26** (2012), no. 2, 435–451.

[3] B. Bollobás, T. I. Fenner, and A. Frieze, *An algorithm for finding Hamilton paths and cycles in random graphs*, Combinatorica **7** (1987), no. 4, 327–341.

[4] B. Bollobás, *The evolution of sparse graphs*, Graph theory and combinatorics (Cambridge, 1983), Academic Press, London, 1984, pp. 35–57.

[5] A. Dudek and A. Frieze, *Tight Hamilton cycles in random uniform hypergraphs*, Random Structures Algorithms, to appear.

[6] ———, *Loose Hamilton cycles in random uniform hypergraphs*, Electron. J. Combin. **18** (2011), no. 1, Paper 48, 14.

[7] A. Dudek, A. Frieze, P.-S. Loh, and S. Speiss, *Optimal divisibility conditions for loose Hamilton cycles in random hypergraphs*, Electron. J. Combin. **19** (2012), Note 44, 4.

[8] A. Frieze, *An algorithm for finding Hamilton cycles in random directed graphs*, J. Algorithms **9** (1988), no. 2, 181–204.

[9] ———, *Loose Hamilton cycles in random 3-uniform hypergraphs*, Electron. J. Combin. **17** (2010), no. 1, Note 28, 4.

[10] A. Frieze and M. Krivelevich, *Packing Hamilton cycles in random and pseudo-random hypergraphs*, Random Structures & Algorithms **41** (2012), 1–22.

[11] A. Frieze, M. Krivelevich, and P.-S. Loh, *Packing tight Hamilton cycles in 3-uniform hypergraphs*, Random Structures Algorithms **40** (2012), 269–300.

[12] S. Janson, T. Łuczak, and A. Ruciński, *Random graphs*, Wiley-Interscience, New York, 2000.

[13] A. Johansson, J. Kahn, and V. Vu, *Factors in random graphs*, Random Structures Algorithms **33** (2008), no. 1, 1–28.

[14] J. Komlós and E. Szemerédi, *Limit distribution for the existence of Hamiltonian cycles in a random graph*, Discrete Math. **43** (1983), no. 1, 55–63.

[15] A. Korshunov, *Solution of a problem of Erdős and Renyi on Hamiltonian cycles in nonoriented graphs.*, Sov. Math., Dokl. **17** (1976), 760–764.

[16] ———, *Solution of a problem of P. Erdős and A. Renyi on Hamiltonian cycles in undirected graphs*, Metody Diskretn. Anal. **31** (1977), 17–56.

[17] D. Kühn and D. Osthus, *On Pósa's conjecture for random graphs*, SIAM Journal on Discrete Mathematics **26** (2012), no. 3, 1440–1457.

[18] C. Lee and B. Sudakov, *Dirac's theorem for random graphs*, Random Structures Algorithms **41** (2012), no. 3, 293–305.

[19] L. Pósa, *Hamiltonian circuits in random graphs*, Discrete Math. **14** (1976), no. 4, 359–364.

[20] V. Rödl, A. Ruciński, and E. Szemerédi, *An approximate Dirac-type theorem for k-uniform hypergraphs*, Combinatorica **28** (2008), no. 2, 229–260.

[21] E. Shamir, *How many random edges make a graph Hamiltonian?*, Combinatorica **3** (1983), no. 1, 123–131.

[22] B. Sudakov and V. H. Vu, *Local resilience of graphs*, Random Structures Algorithms **33** (2008), no. 4, 409–433.