Hamiltonicity of expanders: optimal bounds and applications

Nemanja Draganić* Richard Montgomery † David Munhá Correia‡
Alexey Pokrovskiy§ Benny Sudakov‡

Abstract

An n-vertex graph G is a C-expander if $|N(X)| \geq C|X|$ for every $X \subseteq V(G)$ with $|X| < n/2C$ and there is an edge between every two disjoint sets of at least $n/2C$ vertices. We show that there is some constant $C > 0$ for which every C-expander is Hamiltonian. In particular, this implies the well known conjecture of Krivelevich and Sudakov from 2003 on Hamilton cycles in (n, d, λ)-graphs. This completes a long line of research on the Hamiltonicity of sparse graphs, and has many applications.

1 Introduction

A Hamilton cycle in a graph G is a cycle that contains all the vertices of G. The presence of such a cycle categorizes G as Hamiltonian. This fundamental concept in Graph Theory has been extensively studied, for example see [2, 12, 16, 19, 20, 28, 35, 43, 45, 49, 50, 57] and the surveys [33, 51]. Deciding whether a graph is Hamiltonian or not is an NP-complete problem, and thus it is an important area of research to find simple conditions which imply Hamiltonicity. One classic example is Dirac's theorem [22] that any graph with $n \geq 3$ vertices and minimum degree at least $n/2$ is Hamiltonian. Another beautiful condition, by Chvátal and Erdős [16], is that if the connectivity of a graph is at least its independence number, then the graph is Hamiltonian. Other famous Hamiltonicity conditions include those of Chvátal [15], Jackson [39] and Nash-Williams [55]. Most known conditions for Hamiltonicity, however, require the graph to be very dense. All the results mentioned above imply that the graph has linear minimum degree, except the Chvátal-Erdős result, which still implies minimum degree $\Omega(\sqrt{n})$. Hence, it is of particular interest to find Hamiltonicity conditions which also apply to sparse graphs.

A key area of research towards this in the last 50 years has been on Hamiltonicity in sparse random graphs, and, in particular, which binomial random graphs $G(n, p)$ and which random regular graphs $G_{n,d}$ are likely to be Hamiltonian. Pósa [57] was the first to determine the threshold for Hamiltonicity in $G(n, p)$, introducing the famous rotation-extension technique which quickly become a widely used tool with innumerable applications (including in this paper). After a refinement of Pósa's result by Korshunov [42], in 1983 Bollobás [11] and Komlós and Szemerédi [41] independently showed that, if $p = (\log n + \log\log n + \omega(1))/n$, then $G(n, p)$ is almost certainly Hamiltonian. As is well-known, when $p = (\log n + \log\log n - \omega(1))/n$, $G(n, p)$ almost certainly has a vertex with degree 1, and hence no Hamilton cycle. Random regular graphs, then, may be far sparser yet plausibly Hamiltonian with high probability, and, after significant focus on the problem, it is now known that $G_{n,d}$ will almost surely have a Hamilton cycle for all $3 \leq d \leq n - 1$. For further details on this, see the work of Cooper, Frieze, and Reed [18] and Krivelevich, Sudakov, Vu, and Wormald [48].

*Mathematical Institute, University of Oxford, UK. Research supported by SNSF project 217926.
Email: **nemanja.draganic@maths.ox.ac.uk**

†Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK. Research supported by the European Research Council (ERC) under the European Union Horizon 2020 research and innovation programme (grant agreement No. 947978).
Email: **richard.montgomery@warwick.ac.uk.**

‡Department of Mathematics, ETH, Zürich, Switzerland. Research supported in part by SNSF grant 200021_196965.
Emails: **{david.munhacanascorreia, benjamin.sudakov}@math.ethz.ch**.

§Department of Mathematics, University College London, Gower Street, London WC1E 6BT, UK.
Email: **dralexeypokrovskiy@gmail.com.**

The well-established understanding of Hamiltonicity in random graphs presents an important step towards the search for simple properties of sparse graphs which imply Hamiltonicity. It points to considering natural 'pseudorandom' conditions which are required by a deterministic graph to resemble a random graph. However, forgoing the randomness of $G(n,p)$ and relying only on these pseudorandom properties to find a Hamilton cycle presents a significantly firmer challenge, similar to the generalisation of other problems from random to pseudorandom graphs. Pseudorandom graphs have been systematically studied since work by Thomason [60, 61] in the 1980's, a history that can be found in the survey by Krivelevich and Sudakov [47]. The most studied class of pseudorandom *regular* graphs was introduced by Alon and is defined using spectral properties. Recalling that if a graph G is d-regular then its largest eigenvalue is d, we denote the second largest eigenvalue of G in absolute value by $\lambda(G)$. Then, a graph G is an (n,d,λ)-*graph* if it is d-regular with n vertices and satisfies $|\lambda(G)| \leq \lambda$.

The first major step towards understanding the Hamiltonicity of such pseudorandom graphs was made by Krivelevich and Sudakov in 2003 in their influential paper [46]. They showed that if d is sufficiently larger than λ, then the graph is Hamiltonian. More precisely,

$$\frac{d}{\lambda} \geq 1000 \cdot \frac{\log n \cdot (\log\log\log n)}{(\log\log n)^2} \tag{1}$$

implies that every (n,d,λ)-graph is Hamiltonian. In the same paper, Krivelevich and Sudakov made the beautiful conjecture that (1) can be replaced by $\frac{d}{\lambda} \geq C$ for some large constant C, as follows.

Conjecture 1.1. *There exists $C > 0$ such that if $\frac{d}{\lambda} \geq C$, then every (n,d,λ)-graph is Hamiltonian.*

That is, if the absolute value of every other eigenvalue of a regular graph is at most a small constant fraction of the largest eigenvalue, then the graph should be Hamiltonian.

Considering the result in [46] in the context of random graphs allows us to benchmark, broadly, this progress. That is to say, the result of [46] is strong enough to prove the likely Hamiltonicity of the random regular graph $G_{n,d}$ when $d \geq \log^{2-o(1)} n$, while the ultimate goal, Conjecture 1.1, would be strong enough to prove the likely Hamiltonicity of the random regular graph $G_{n,d}$ when d is at least some large constant. Despite a great deal of attention, for example seen by the various relaxations and generalisations of the problem studied in [4, 13, 35, 47, 44], and many incentivising applications (see Section 1.1), the bound at (1) established in [46] remained unchallenged for 20 years. Only recently, Glock, Munhá Correia and Sudakov [32] finally improved this, significantly strengthening the result of [46] by showing that, for some large constant $C > 0$, $d/\lambda \geq C \log^{1/3} n$ suffices to imply Hamiltonicity. Moreover, they showed that Conjecture 1.1 holds in the special case where $d \geq n^{\alpha}$, for any fixed α, that is, in this case (1) can be weakened to $d/\lambda \geq C$.

Krivelevich and Sudakov applied their bound at (1) to other problems on Hamiltonicity in sparse graphs; these and other applications are discussed in Section 1.1. To allow applications to non-regular graphs, other pseudorandom conditions which imply Hamiltonicity were also studied. Motivated by this, shortly after Conjecture 1.1 was stated, several papers considered an even stronger conjecture, singling out the key properties of (n,d,λ)-graphs thought to give some potential for proving Hamiltonicity. To state this even stronger conjecture, whose variant appeared for example in [13], we need the following definition.

Definition 1.2. An n-vertex graph G with $n \geq 3$ is a C-*expander* if,

(a) $|N(X)| \geq C|X|$ for all vertex sets $X \subseteq V(G)$ with $|X| < n/2C$, and,

(b) there is an edge between any disjoint vertex sets $X, Y \subseteq V(G)$ with $|X|, |Y| \geq n/2C$.

Conjecture 1.3. *For every sufficiently large $C > 0$, every C-expander is Hamiltonian.*

In 2012, Hefetz, Krivelevich and Szabó [35] made progress on this problem; the precise expansion conditions used in their result can be found in Theorem 1.1 of [35] (in particular weakening (a) in our Definition 1.2), but imply that every $(\log^{1-o(1)} n)$-expander is Hamiltonian.

In this paper, we prove Conjecture 1.3, thus completing an extensive line of research on Hamiltonicity problems.

Theorem 1.4. *For every sufficiently large $C > 0$, every C-expander is Hamiltonian.*

This result has a large number of applications, as discussed below. In particular, it is a standard exercise to show that for every $C > 0$ there exists a constant C_0 such that, if $\frac{d}{\lambda} \geq C_0$, then every (n, d, λ)-graph is a C-expander. Thus, clearly Conjecture 1.1 is implied by Theorem 1.4, giving the following.

Theorem 1.5. *There is a constant $C > 0$ such that if $\frac{d}{\lambda} \geq C$ then every (n, d, λ)-graph is Hamiltonian.*

To prove Theorem 1.4, we take a very different approach to the previous work on this problem, and so give a detailed outline of the proof in Section 2.2. We note here that in fact a stronger result than Theorem 1.4 holds – such graphs are not only Hamiltonian, but Hamilton-connected. Furthermore, the Hamilton cycle in Theorem 1.5 can be found in polynomial time. We discuss these strengthenings in the concluding remarks. We finish this section by giving some examples of the applications of Theorems 1.4 and 1.5.

1.1 Applications

Pseudorandom conditions for Hamiltonicity have found a large variety of applications, and Theorems 1.4 and 1.5 immediately improve the bounds required for many such applications. We will discuss here three applications: Hamiltonicity in random Cayley graphs, Hamiltonicity in well-connected graphs and Hamilton cycles with few colours in edge-coloured graphs. Further applications, many of them discussed in [32, 46], can be found in various other fields, ranging from problems in positional games (see, e.g., [25, 17, 35, 34]), to questions about finding coverings and packings of Hamilton cycles in random and pseudorandom graphs (see, e.g., [27, 23, 37]), Hamiltonicity thresholds in different random graph models (see, e.g., [30, 6, 31]), and for various other problems (see, e.g., [36, 40]), including as far afield as Alon and Bourgain's work on additive patterns in multiplicative subgroups [5].

Hamiltonicity in random Cayley graphs. In 1969, Lovász [52] made the following famous conjecture about the Hamiltonicity of vertex-transitive graphs, which are graphs in which any vertex can be mapped to any other vertex by an automorphism.

Conjecture 1.6. *Every connected vertex-transitive graph contains a Hamilton path, and, except for five known examples, a Hamilton cycle.*

As Cayley graphs are vertex-transitive and none of the five known exceptions in Conjecture 1.6 are Cayley graphs, Lovász's conjecture implies the following earlier conjecture, posed in 1959, by Strasser [58].

Conjecture 1.7. *Every connected Cayley graph is Hamiltonian.*

Conjecture 1.7 is known to be true when the underlying group is abelian, but the only progress towards the conjectures in general is a result of Babai [8] that every vertex-transitive n-vertex graph contains a cycle of length $\Omega(\sqrt{n})$ (see [21] for a recent improvement by DeVos) and a result of Christofides, Hladký and Máthé [14] that every vertex-transitive graph of linear minimum degree contains a Hamilton cycle.

The "random version" of Conjecture 1.7 is a natural relaxation of the original problem. Alon and Roichman [7] showed that there is a constant $C > 0$ for which, for any group G, the Cayley graph generated by a random set S of $C \log |G|$ elements, $\Gamma(G, S)$ say, is almost surely connected. Hence, an important instance of Conjecture 1.7 is to show that $\Gamma(G, S)$ is almost surely Hamiltonian. This problem was also stated as a conjecture by Pak and Radoičić [56]. Theorem 1.5 resolves this conjecture. Indeed, Alon and Roichman [7] showed that if $|S| \geq C \log |G|$ for some large constant C, then $\Gamma(G, S)$ is almost surely an (n, d, λ)-graph with $d/\lambda \geq K$ for some large constant K. Thus, we obtain the following.

Theorem 1.8. *Let C be a sufficiently large constant. Let G be a group of order n and $d \geq C \log n$. If $S \subseteq G$ is a set of size d chosen uniformly at random, then, with high probability, $\Gamma(G, S)$ is Hamiltonian.*

Hamiltonicity in well-connected graphs. For any function $f : \mathbb{Z}^+ \to \mathbb{R}$, we say that a graph G is *f-connected* if $|A \cap B| \geq f(\min(|A \setminus B|, |B \setminus A|))$ for every two subsets $A, B \setminus V(G)$ such that $V(G) = A \cup B$ and there is no edge between $A \setminus B$ and $B \setminus A$. In [13], Brandt, Broersma, Diestel, and Kriesell proved that if $f(k) \geq 2(k+1)^2$ for every $k \in \mathbb{N}$ then G is Hamiltonian. This bound was improved to $k \log k + O(1)$ in [35]. Brandt et al. also conjectured that there exists a function f which is linear in k yet ensures Hamiltonicity. It is not difficult to check that if $f(k) = Ck$ for all k, then an f-connected graph is a $(C/2)$-expander. Indeed, given a set $X \subseteq V(G)$ of size at most n/C and its neighbourhood, $N(X)$, applying the f-connected condition with $A = X \cup N(X), B = V(G) \setminus X$ shows that $|N(X)| \geq C|X|/2$, while, given two disjoint sets $X, Y \subseteq V(G)$ of size at least n/C, setting $A = V(G) \setminus X, B = V(G) \setminus Y$ and applying the f-connected condition shows that there is an edge between X, Y. Therefore, Theorem 1.4 implies the conjecture, as follows.

Theorem 1.9. *Let C be a sufficiently large constant and let $f(k) = Ck$ for every $k \in \mathbb{Z}^+$. Then, any f-connected graph with at least 3 vertices is Hamiltonian.*

Hamilton cycles using few colours. An *optimally* edge-coloured graph G is one which is properly coloured using the fewest possible number of colours. Akbari, Etesami, Mahini, and Mahmoody [3] proved that any optimally coloured n-vertex complete graph K_n has a Hamilton cycle containing edges of at most $8\sqrt{n}$ colours, and conjectured there should always be such a cycle using only $O(\log n)$ colours, which would be best possible up to a multiplicative constant. The bound from [3] on the number of colours was later improved to $O(\log^3 n)$ by Balla, Pokrovskiy and Sudakov [10]. Their strategy is to randomly pick $d = O(\log^3 n)$ colours and show that the subgraph of these colours is an (n, d, λ)-graph with high probability, for some appropriate λ, and apply the result from [46] on the Hamiltonicity of (n, d, λ)-graphs. Using their improved condition, Glock, Munhá Correia and Sudakov [32] showed that this is possible with only $O(\log^{5/3} n)$ colours using the same method. Applying instead the bound in Theorem 1.5 immediately proves the conjecture (see [32] for more details).

Theorem 1.10. *Every optimally coloured K_n has a Hamilton cycle with $O(\log n)$ colours.*

2 Preliminaries

2.1 Notation.

Our notation is standard, and we recall here only the most important. A graph G has vertex $V(G)$ and edge set $E(G)$, and we set $|G| = |V(G)|$. A *linear forest* is a graph consisting of a collection of vertex disjoint paths. We denote by $\mathrm{End}(\mathcal{F})$ the set of vertices which are endpoints of paths in $\mathcal{F}$. Two vertices are *at distance ℓ in G* if the length of the shortest path between them in G is ℓ. For a subset of vertices X, we denote by $G[X]$ the induced subgraph of G with vertex set X, by $\Gamma_G(X)$ the set of vertices adjacent to at least one vertex in X, and by $N_G(X)$ the (outer) neighbourhood $N_G(X) = \Gamma_G(X) \setminus X$. We omit G in the subscript when it is clear from the context which graph we are working with.

2.2 Proof outline

Where G is an n-vertex C-expander, for some large constant C, we wish to find a Hamilton cycle in G. Before describing our methods, it is instructive to briefly recall the well-known Pósa rotation-extension approach, and how the approach works on random graphs but not on pseudorandom graphs (for a more detailed approach than this sketch, see, for example, [11]).

Take a maximal length path P in G, with endvertices x and x_0, say. If x has a neighbour y in G on P whose neighbour on P which is nearest to x, z say, is not equal to x, then the path P can be rotated

Figure 1: On the left, a rotation of an xx_0-path P by removing zy and adding xy to get an xx_0-path highlighted in blue. On the right, an example rotation of a linear forest $\mathcal{F}$ to get the 4 paths highlighted in blue, thus removing x from the set of endvertices and adding z.

by adding xy and removing yz (see Figure 1). This gives us a path with the same length as P, where one endvertex is x_0 and the other is a new endvertex, z. As Pósa showed, if G is a C-expander, then this can be done iteratively to show that there are at least (say) $n/3$ different new endvertices v, say those in E_{x_0}, so that you can rotate P repeatedly to get an x_0v-path P_v with the same length as P. Then, each of these paths P_v can be rotated without changing the endvertex v, to get at least $n/3$ new endvertices, say those in F_v. If G contains any edge in $\{vu : v \in E_{x_0}, u \in F_v\}$, G has a cycle with length $|P|$, whence, as G is connected (a trivial consequence of the expansion condition), if $|P| < n$ then we can find a longer path than P, a contradiction. Thus, G has a Hamilton cycle. Of course, G may not contain any edge in $\{vu : v \in E_{x_0}, u \in F_v\}$. When working with random graphs, however, we can reserve some random edges and sprinkle them in to find an edge in $\{vu : v \in E_{x_0}, u \in F_v\}$ and hence a longer path than P – doing this iteratively can get a Hamilton path and thus a Hamilton cycle.

In pseudorandom graphs we cannot do this sprinkling. Thus, the dream strategy would be to find two large sets A, B with $A \times B \subset \{vu : v \in E_{x_0}, u \in F_v\}$, so that we can apply Definition 1.2b) to find the edge. Indeed, for their result discussed in the introduction, Hefetz, Krivelevich and Szabó [35] found such sets A, B with $|A|, |B| \geq n/\log^{1-o(1)} n$. However, large linear sets A and B (as would be required for Theorem 1.4) seem too hard to find, essentially because the rotations performed around each endvertex of the path interfere with each other by rotating sections of the paths.

For our approach, instead of rotating only paths, we perform rotations of disjoint union of paths, i.e., of linear forests. The details of these rotations can be found in Section 3, but, analogously, we alter a linear forest and preserve all but one endvertex of these paths (one sample rotation is depicted in Figure 1), iteratively developing many different possibilities for the endvertex we are changing. The implementation is much more intricate, but, essentially, if most of the vertices in a linear forest $\mathcal{F}$ are not in long paths then we can take two endvertices x and x' of different paths, divide the paths in $\mathcal{F}$ into two groups $\mathcal{F}_x$ and $\mathcal{F}_{x'}$, and perform rotations just within these groups to find new endvertices in place of x and x'. As these two sets of rotations are done using disjoint linear forests, they can be done independently, to get sets of new endvertices A and B respectively, such that any edge from A to B will allow us to connect two paths together, reducing the number of paths in the linear forest by 1 and losing the endvertices x and x'. This is done in Section 3.

We then wish to apply this iteratively to reduce the number of paths in our linear forest. However these rotations may create long paths (so that it is hard to find the partition into forests $\mathcal{F}_x$ and $\mathcal{F}_{x'}$). Thus, we combine this with an argument showing if $\mathcal{F}$ has too many long paths then we can replace it with a linear forest $\mathcal{F}'$ whose lengths are more evenly spread yet whose endvertices are still a subset of the endvertices of $\mathcal{F}$. This is also done using rotations, and is carried out in Section 4 (which includes a sketch of the proof once the necessary definitions have been introduced). Starting with a spanning linear forest $\mathcal{F}$, we can alternately perform rotations and add an edge (reducing the number of paths by one) while losing any two endvertices (where, moreover, with some more effort, we can specify these two endvertices) and then replace the current linear forest with one with more evenly spread path lengths (so that the first step keeps working). Ultimately, this allows us to reach $\mathcal{F}'$, a spanning linear forest of at most $n^{0.8}$ paths, all of whose endvertices were also endvertices of paths in $\mathcal{F}$.

This argument can be pushed further to end up with fewer paths, but cannot produce a Hamilton

path let alone a Hamilton cycle. Thus, we need some way to connect the paths we produce into a Hamilton cycle. We do this by setting aside at the outset an (A, B)-*linking structure* H in G, where $|A| = |B|$ and this has the property that H contains a spanning linear forest connecting any desired partition of pairs a, b with $a \in A$ and $b \in B$. More precisely, this is defined as follows.

Definition 2.1. A graph H with disjoint sets $A, B \subseteq V(H)$ of equal size is said to be an (A, B)-linking structure if for every bijection $\varphi : A \to B$ there exist vertex disjoint paths $P_1, \ldots, P_{|A|}$ of equal length such that the following hold.

1. The paths cover all the vertices of H, so that $V(H) = V(P_1) \cup \ldots \cup V(P_{|A|})$.
2. For each $i \in [|A|]$, the path P_i has endpoints a and $\varphi(a)$ for some $a \in A$.

By using properties of sorting networks, adapting an approach of Hyde, Morrison, Müyesser, and Pavez-Signé [38], we can find an (A, B)-linking structure H in G with $|A| = |B| = n^{0.9}$ and $|H| = o(n^{0.95})$. Taking then an initial linear forest $\mathcal{F}$ which covers $G - V(H) \setminus (A \cup B)$ and has every vertex in $A \cup B$ among its endvertices, we then apply our methods to reach a linear forest $\mathcal{F}'$ with the same vertex set as $\mathcal{F}$, but whose endvertices are exactly those in $A \cup B$ (with no isolated vertices in $\mathcal{F}'$). Then, applying the linking property of H allows us to connect the paths in $\mathcal{F}''$ together into our desired Hamilton cycle (see Figure 2).

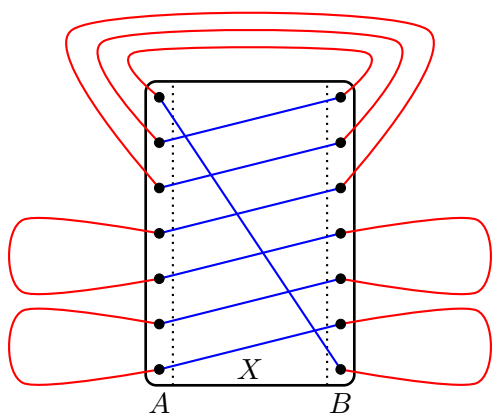

Figure 2: Given a linear forest $\mathcal{F}'$ (depicted in red) which spans $G - (X \setminus (A \cup B))$ with no isolated vertices and endvertices $\mathrm{End}(\mathcal{F}') = A \cup B$, if $G[X]$ is an (A, B)-linking structure, then the paths in blue can be found disjointly while using all the vertices in X, thus linking $\mathcal{F}'$ into a Hamilton cycle.

2.3 Expansion and linear forests

In order to carry out Pósa rotations in linear forests, we need to define the interior of sets with respect to the linear forest, and a variation of expansion where we only consider the neighbourhood of sets within the interior of the linear forest. For this, we will use the following definitions.

Definition 2.2. Let G be a graph, and $U, V \subseteq V(G)$ with $|U| \geq 10^5 C$. We say that U *C-expands into V (in G)* if all $X \subseteq U$ with $|X| \leq |U|/5000C$ have $|N_G(X) \cap V| \geq C|X|$.

Definition 2.3. Let G be a graph and $\mathcal{F}$ a linear forest in G. For a set $U \subseteq V(\mathcal{F})$, we define the *interior of U in $\mathcal{F}$* to be the set $\mathrm{int}_{\mathcal{F}}(U) := \{u \in U : N_{\mathcal{F}}(u) \subseteq U\}$. Furthermore, we say that U is an $(\mathcal{F}, C)$-*expander* if U C-expands into $\mathrm{int}_{\mathcal{F}}(U)$.

The following observation is immediate from the previous definition.

Observation 2.4. *For all linear forests $\mathcal{F}$ and subsets $X, Y \subseteq V(\mathcal{F})$ it must be that* $\mathrm{int}_{\mathcal{F}}(X \setminus Y) \geq \mathrm{int}_{\mathcal{F}}(X) - 3|Y|$.

To carry out rotations within only some of the paths in a linear forest, we will need to find expanding subgraphs sitting within the interior of these specific paths. Often expander subgraphs can be found using only Definition 1.2b) by removing a maximal set of vertices which does not expand and is not too large (see, for example, [53, Section 3.7]). Here, we work similarly, but remove multiple sets $B_{i,j}$ to find a collection of 4 sets U_i' with expansion properties (with respect to some linear forest $\mathcal{F}$) as follows.

Lemma 2.5. *Let $C \geq 10^5$ and $C' = C/5000$. Let G be a C-expander containing a linear forest $\mathcal{F}$, and let $U_1, U_2, U_3, U_4 \subset V(G)$ be disjoint sets with $|\mathrm{int}_{\mathcal{F}}(U_i)| \geq n/500$ for each $i \in [4]$. Then, there are subsets $U_i' \subseteq U_i$, $i \in [4]$, such that*

- *for each $i, j \in [4]$, U_i' C'-expands into $\mathrm{int}_{\mathcal{F}}(U_j')$, and*
- *$|U_i'| \geq |U_i| - 2n/C$ for each $i \in [4]$.*

In particular, for each $i \in [4]$, U_i' is an $(\mathcal{F}, C')$-expander.

Proof. Pick sets $B_{i,j} \subseteq U_i$, $i, j \in [4]$, satisfying the following:

(1) $|B_{i,j}| \leq 2n/C$ for all $i, j \in [4]$.

(2) $\left|N(B_{i,j}) \cap \mathrm{int}_{\mathcal{F}}\left(U_j \setminus \bigcup_{t=1}^{4} B_{j,t}\right)\right| \leq C'|B_{i,j}|$ for all $i, j \in [4]$.

(3) $\sum_{i,j \in [4]} |B_{i,j}|$ is as large as possible, subject to (1) and (2).

First note that this is indeed possible, as the sets $B_{i,j} = \emptyset$, $i, j \in [4]$, satisfy (1) and (2). Let $B := \bigcup_{i,j \in [4]} B_{i,j}$.

Claim 1. *For each $i, j \in [4]$, $|B_{i,j}| < n/2C$.*

Proof. Suppose otherwise. Then, the definition of a C-expander implies that $B_{i,j}$ is adjacent to all but at most $n/2C$ vertices of $V(G) \setminus B_{i,j}$ and, thus, to all but at most $n/2C$ vertices in $\mathrm{int}_{\mathcal{F}}(U_j \setminus B)$. Hence,

$$\begin{aligned}|N(B_{i,j}) \cap \mathrm{int}_{\mathcal{F}}(U_j \setminus B)| &\geq |\mathrm{int}_{\mathcal{F}}(U_j \setminus B)| - n/2C \geq |\mathrm{int}_{\mathcal{F}}(U_j)| - 3|B \cap U_j| - n/2C \\ &\geq |\mathrm{int}_{\mathcal{F}}(U_j)| - 3 \cdot 16 \cdot 2n/C - n/2C \geq (10C' - 96.5)n/C > C'|B_{i,j}|,\end{aligned}$$

contradicting (2). Note that in the previous inequalities we used Observation 2.4, and that $|\mathrm{int}_{\mathcal{F}}(U_i)| \geq n/500$ for each $i \in [4]$ and $C' = C/5000$. ⊡

We now show that setting $U_i' := U_i \setminus B$ for each $i \in [4]$ gives the first item of the desired outcome of the lemma.

Claim 2. *For each $i, j \in [4]$, $U_i \setminus B$ C'-expands into $\mathrm{int}_{\mathcal{F}}(U_j \setminus B)$.*

Proof. Suppose for contradiction that there is some $i, j \in [4]$ for which there is some $X \subseteq U_i \setminus B$ with $|N(X) \cap \mathrm{int}_{\mathcal{F}}(U_j \setminus B)| < C'|X|$ and $|X| \leq n/5000C' \leq n/C$. Then, $|B_{i,j} \cup X| \leq |B_{i,j}| + |X| \leq 2n/C$ by Claim 1 and, furthermore,

$$\begin{aligned}|N(B_{i,j} \cup X) \cap \mathrm{int}_{\mathcal{F}}(U_j \setminus B)| &\leq |N(B_{i,j}) \cap \mathrm{int}_{\mathcal{F}}(U_j \setminus B)| + |N(X) \cap \mathrm{int}_{\mathcal{F}}(U_j \setminus B)| \\ &\leq C'|B_{i,j}| + C'|X| = C'|B_{i,j} \cup X|.\end{aligned}$$

For each $s, t \in [4]$, we have $|N(B_{s,t}) \cap \mathrm{int}_{\mathcal{F}}(U_t \setminus (B \cup X))| \leq |N(B_{s,t}) \cap \mathrm{int}_{\mathcal{F}}(U_t \setminus B)| \leq C'|B_{s,t}|$. Thus, replacing $B_{i,j}$ by $B_{i,j} \cup X$ in $\{B_{s,t} : s, t \in [4]\}$ gives a family of sets satisfying (1) and (2), but with a larger total number of vertices, contradicting (3). ⊡

Finally, setting $U_i' = U_i \setminus B$ for each $i \in [4]$ satisfies the lemma. Indeed, note that since $|B \cap U_i| \leq 4 \cdot n/2C$ for each $i \in [4]$ by Claim 1, we have that $|U_i'| \geq |U_i| - 2n/C$. □

3 Rotations of linear forests

In this section we give our methods for reducing the number of paths in a spanning linear forest in an expander, by rotating the linear forest. After defining our rotations and making some simple observations in Section 3.1, we will prove some intermediary results for these rotations in Section 3.2 before proving the key result of this section, Lemma 3.7, in Section 3.3.

3.1 Rotation definition and simple observations

We first define rotations of linear forests, which can be compared to rotations of paths by lining up the paths of the linear forest as in Figure 1. Rotations of linear forests can, however, look quite different depending on whether the edge used to rotate is between two different paths, or between vertices in the same path, and on whether these paths are isolated vertices, or not; the different possiblities are illustrated in Figure 3.

Definition 3.1. Let P_1 and P_2 be two (possibly equal) paths of a linear forest $\mathcal{F}$ in a graph G. Let x be an endpoint of P_1 and let z be a neighbour (in G) of x in P_2. Let y be a vertex adjacent to z in P_2 if $|P_2| \geq 1$ (if $P_1 = P_2$, then let y be the vertex closer to x in P_1, where we may have $y = x$), and let $y = z$ otherwise. Then, the linear forest $\mathcal{F}'$ obtained from $\mathcal{F}$ by removing the edge yz (if $y \neq z$) and adding the edge $xz \in E(G)$ is called a 1-*rotation of* $\mathcal{F}$ *in* G.

We call the vertex x the *old endpoint*, y the *new endpoint* (indeed, note that $y \in \mathrm{End}(\mathcal{F}')$), and z the *pivot*. The edge zy is called the *broken edge* of the rotation. We will refer to the process of removing yz (if it exists) and adding xz to $\mathcal{F}$ as a 1-rotation, as well as the resulting linear forest $\mathcal{F}'$, for example, saying that, by performing a 1-rotation on $\mathcal{F}$ with old endpoint x, pivot z and new endpoint y, we get the 1-rotation $\mathcal{F}'$ of $\mathcal{F}$.

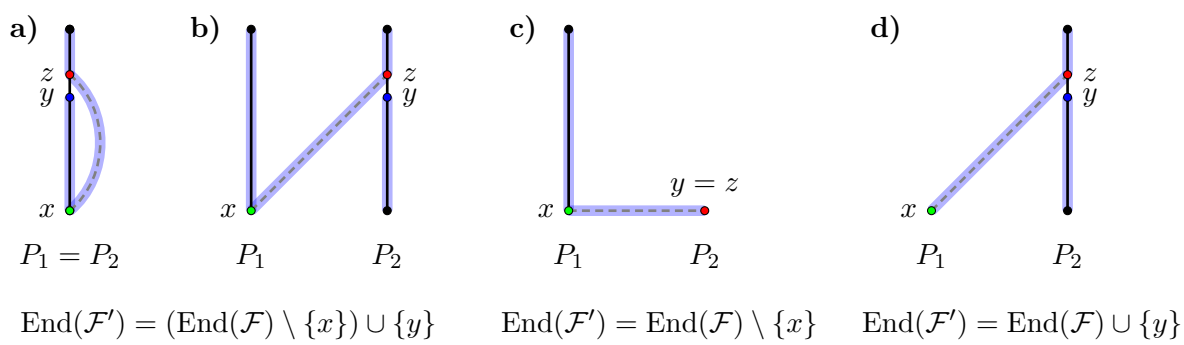

Figure 3: Different 1-rotations of $\mathcal{F} = \{P_1, P_2\}$ to get $\mathcal{F}'$ by removing the edge yz (if it exists) and adding the dashed edge xz, with the new paths highlighted in blue and the change in endvertices given underneath. In **a)** $P_1 = P_2$, and in **b)**–**d)** $P_1 \neq P_2$. In **b)** neither x or z are isolated, in **c)** $y = z$ is isolated but not x and in **d)** x is isolated but not z.

As in Pósa's original rotation-extension method on paths, we will repeatedly rotate linear forests to find many different new endvertices. For this, we will now define a *k-rotation*.

Definition 3.2. A *k-rotation* of a forest $\mathcal{F}$ in a graph G is any forest which is obtained by a sequence of k consecutive 1-rotations starting with $\mathcal{F}$ in G, where the old endpoint of each 1-rotation is the new endpoint of the previous one and the pivot is at least at distance 3 in $\mathcal{F}$ from the old endpoint of the first 1-rotation and the pivot in the i-th rotation for all $i < k$ (and thus the broken edge belongs to $\mathcal{F}$

as well). We say that the k-rotation has *old endpoint* x, or *starts at* x, and *new endpoint* y if x is the old endpoint of the first 1-rotation and y is the new endpoint of the last 1-rotation.

For a set $U \subseteq V(G)$ we denote by $E_U^k(v, \mathcal{F})$ the set of new endpoints obtained by i-rotations of $\mathcal{F}$ starting from v (i.e., with old endpoint v) with $i \leq k$, such that the pivots of each 1-rotation are contained in $\mathrm{int}_{\mathcal{F}}(U)$, so that each broken edge belongs to $\mathcal{F}[U]$. We call such rotations (U, i)*-rotations*, and where i is not specified call them U*-rotations*.

Next we prove some simple results on k-rotations.

Lemma 3.3. *Let $\mathcal{F}'$ be a (U, k)-rotation of $\mathcal{F}$, with $k \geq 1$, starting with vertex v and let u be the new endpoint. Then, the following holds.*

(A) *If $E_U^k(v, \mathcal{F}) \subseteq S \subseteq E_U^{k+1}(v, \mathcal{F})$, then the set of pivots used in the 1-rotations to get new endpoints in S is of size at most $2|S|$.*

Furthermore, if $\mathcal{F}$ has no isolated vertices, then the following also hold.

(B) *The only possible isolated vertex in $\mathcal{F}'$ is u, which furthermore can only be isolated in $\mathcal{F}'$ if $u \in End(\mathcal{F})$.*

(C) *$End(\mathcal{F}') = (End(\mathcal{F}) \cup \{u\}) \setminus \{v\}$.*

Proof. For (A) note that every pivot used for a i-rotation in S is adjacent to a vertex in S in the forest $\mathcal{F}$, or is equal to a vertex in S in the case this pivot is isolated in $\mathcal{F}$. This is because all broken edges in a k-rotation $\mathcal{F}'$ are always edges in the initial forest $\mathcal{F}$. As the maximum degree of the linear forest $\mathcal{F}$ is 2, there are at most $2|S|$ pivots used.

We prove part (B) by induction on k and under the weaker assumption that $\mathcal{F}$ has no isolated vertices *except for, possibly,* v. For $k = 1$, let z be the pivot of the rotation, and note that v is adjacent to z in $\mathcal{F}'$, hence it is not isolated in $\mathcal{F}'$. Since $\mathcal{F}$ contains no other isolated vertices, note that the only possible isolated vertices in $\mathcal{F}'$ are z and u, as zu is the only deleted edge in $\mathcal{F}$. But z is adjacent to v in $\mathcal{F}'$, so only u can be isolated in $\mathcal{F}$, and this happens if and only if it did not have any other adjacent vertex in $\mathcal{F}$ besides z, i.e. if and only if it is an endpoint of $\mathcal{F}$. Suppose now the statement holds for $k - 1 \geq 1$. Consider a k-rotation $\mathcal{F}_k$ of $\mathcal{F}$ with old endpoint v and new endpoint u. By definition of k-rotation, we get a $(k-1)$-rotation $\mathcal{F}_{k-1}$ of $\mathcal{F}$ with old endpoint v and new endpoint w, such that $\mathcal{F}_k$ is a 1-rotation of $\mathcal{F}_{k-1}$ with old endpoint w and new endpoint u and a pivot z. By the induction hypothesis, the only possible isolated vertex in $\mathcal{F}_{k-1}$ is w. Applying the induction hypothesis to $\mathcal{F}_{k-1}$ and then to its 1-rotation $\mathcal{F}_k$ implies that u is the only possible isolated vertex in $\mathcal{F}_k$, and that this is the case if and only if u is an endpoint of $\mathcal{F}_{k-1}$. As z is at distance at most 3 from all the previous pivots (used for $\mathcal{F}_{k-1}$), we know that u is an endpoint of $\mathcal{F}_{k-1}$ if and only if it is an endpoint of $\mathcal{F}$, which completes this part of the proof.

For part (C), we also proceed by induction. Let $k = 1$ and let z be the pivot of the 1-rotation $\mathcal{F}'$. If z is the neighbouring vertex of v in $\mathcal{F}$, the statement trivially holds as $\mathcal{F} = \mathcal{F}'$ and $u = v$. Otherwise, note that v has precisely one neighbour in $\mathcal{F}$, while it has another neighbour z in $\mathcal{F}'$, so v is not an endpoint in $\mathcal{F}'$. On the other hand, the edge zu is removed from $\mathcal{F}$, so u has at most one neighbour in $\mathcal{F}'$, so it is an endpoint. Assume the statement holds for $k - 1 \geq 1$. Consider a k-rotation $\mathcal{F}_k$ of $\mathcal{F}$ with old endpoint v and new endpoint u. As before, we get a $(k-1)$-rotation $\mathcal{F}_{k-1}$ of $\mathcal{F}$ with old endpoint v, new endpoint w, such that $\mathcal{F}_k$ is a 1-rotation of $\mathcal{F}_{k-1}$ with old endpoint w and new endpoint u and a pivot z. By the induction hypothesis, we know that $\mathrm{End}(\mathcal{F}_{k-1}) = (\mathrm{End}(\mathcal{F}) \cup \{w\}) \setminus \{v\}$. We distinguish two cases. If w is not an endpoint of $\mathcal{F}$, then by (B) it is not isolated in $\mathcal{F}_{k-1}$, and hence by applying the induction hypothesis to the 1-rotation $\mathcal{F}_k$ of $\mathcal{F}_{k-1}$, we see that $\mathrm{End}(\mathcal{F}_k) = (\mathrm{End}(\mathcal{F}_{k-1}) \cup \{u\}) \setminus \{w\} = (\mathrm{End}(\mathcal{F}) \cup \{u\}) \setminus \{v\}$. In the second case, when w is an endpoint of $\mathcal{F}$, again by (B) we have that w is the only isolated vertex in $\mathcal{F}_{k-1}$. Thus, the 1-rotation $\mathcal{F}_k$ of $\mathcal{F}_{k-1}$ still contains w as an endpoint, while similarly to the case $k = 1$ the only new created endpoint is u. Hence $\mathrm{End}(\mathcal{F}_k) = ((\mathrm{End}(\mathcal{F}) \cup \{w\}) \setminus \{v\}) \cup \{u\} = (\mathrm{End}(\mathcal{F}) \setminus \{v\}) \cup \{u\}$ as required. $\square$

3.2 Intermediate results on rotations

As mentioned in the proof sketch, a key part of Pósa's rotation-extension technique is to show that performing rotations iteratively in an expander creates a linear-sized set of potential endvertices. We now show the corresponding result for our linear forest rotations.

Lemma 3.4. *Let $C > 100$ and let $\mathcal{F}$ be a linear forest in a graph G, v an endpoint of $\mathcal{F}$, and U an $(\mathcal{F}, C)$-expander containing v. Then $|E_U^k(v, \mathcal{F})| \geq |U|/10^5$ for $k = 2\log_C n$.*

Proof. Suppose, to the contrary, that $|E_U^k(v, \mathcal{F})| < |U|/10^5$. For each $i \in [k-1]$, we have $|E_U^i(v, \mathcal{F})| \leq |E_U^k(v, \mathcal{F})| < |U|/10^5$. Now, for each $i \in [k-1]$, and each pair (x, y) with $x \in E_U^i(v, \mathcal{F}), y \in \mathrm{int}_{\mathcal{F}}(U)$ and $xy \in E(G)$, y is a good candidate to use as a pivot for a 1-rotation of a linear forest which has x as an endvertex, created by rotating $\mathcal{F}$ up to $k-1$ times. However, by our definition, we need to have that this pivot is at least at distance 3 in $\mathcal{F}$ from all the previous pivots in the same rotations. In total, though, as we have at most $2|E_U^i(v, \mathcal{F})|$ pivots used by part (A) of Lemma 3.3, this will rule out at most $10|E_U^i(v, \mathcal{F})|$ potential pivots y. Finally, we note that two pivots in $N(E_U^i(v, \mathcal{F})) \cap \mathrm{int}_{\mathcal{F}}(U)$ may possibly correspond to the same new endpoint. Therefore, we have

$$|E_U^{i+1}(v, \mathcal{F})| \geq \frac{|N(E_U^i(v, \mathcal{F})) \cap \mathrm{int}_{\mathcal{F}}(U)| - 10|E_U^i(v, \mathcal{F})|}{2}. \tag{2}$$

If $|E_U^i(v, \mathcal{F})| \geq |U|/5000C$, then, by considering a subset $S \subseteq E_U^i(v, \mathcal{F})$ of size $|U|/5000C$, we have that since U is an $(\mathcal{F}, C)$-expander, $|N(S) \cap \mathrm{int}_{\mathcal{F}}(U)| \geq C|S|$ and so, $|N(E_U^i(v, \mathcal{F})) \cap \mathrm{int}_{\mathcal{F}}(U)| \geq C|S| - |E_U^i(v, \mathcal{F})| \geq |U|/5000 - |U|/10^5 \geq 9|U|/10^5 + 10|E_U^i(v, \mathcal{F})|$, where we are using that $|E_U^i(v, \mathcal{F})| \leq |U|/10^5$. Then, by (2), $|E_U^{i+1}(v, \mathcal{F})| \geq |U|/10^5$, a contradiction. Therefore, $|E_U^i(v, \mathcal{F})| \leq |U|/5000C$, whence we have $|N(E_U^i(v, \mathcal{F})) \cap \mathrm{int}_{\mathcal{F}}(U)| \geq C|E_U^i(v, \mathcal{F})|$ as U is an $(\mathcal{F}, C)$-expander, so that (2) implies

$$|E_U^{i+1}(v, \mathcal{F})| \geq \frac{C|E_U^i(v, \mathcal{F})|}{3}.$$

As this holds for every $0 \leq i \leq k-1$, we have $|E_U^k(v, \mathcal{F})| \geq (C/3)^k > n$, a contradiction. □

The next lemma essentially says that if a linear forest $\mathcal{F}$ in an n-vertex expander can be split into two so that rotations can be done on each subcollection of paths independently starting from two vertices x and y respectively, then we can alter $O(\log n)$ edges in $\mathcal{F}$ to decrease the number of paths by 1 while losing exactly x and y as endvertices, as follows.

Lemma 3.5. *Let $C > C' > 10^{10}$ and let G be an n-vertex C-expander and $\mathcal{F}$ a spanning linear forest in G. Suppose $\mathcal{F}$ has no isolated vertices and let x, y be two endpoints of $\mathcal{F}$. Let X, Y be $(\mathcal{F}, C')$-expanders of size at least $0.0001n$ such that $x \in X, y \in Y$, no path in $\mathcal{F}$ intersects both X and Y, and there is no vertex in $(X \setminus \{x\}) \cup (Y \setminus \{y\})$ which is an endpoint in $\mathcal{F}$. Then, there is a spanning linear forest $\mathcal{F}'$ in G for which $|E(\mathcal{F}) \Delta E(\mathcal{F}')| = O(\log n)$, $\mathrm{End}(\mathcal{F}') = \mathrm{End}(\mathcal{F}) \setminus \{x, y\}$, and $\mathcal{F}'$ contains no isolated vertices.*

Proof. Let $\mathcal{F}_X$ be the collection of paths in $\mathcal{F}$ which intersect X and $\mathcal{F}_Y$ be the collection of paths in $\mathcal{F}$ which intersect Y, so that $\mathcal{F}_X \cap \mathcal{F}_Y = \emptyset$. By Lemma 3.4, there exists a subset $X' \subseteq X$ of size at least $|X|/10^5 > n/C$ such that for all $x' \in X'$ there is an X-rotation $\mathcal{F}_{x'}$ of $\mathcal{F}$ with new endpoint x'. Note crucially that since $\mathcal{F}_{x'}$ is an X-rotation, we have that $\mathcal{F}_Y \subseteq \mathcal{F}_{x'}$. Similarly, there is a subset $Y' \subseteq Y$ of size at least $|Y|/10^5 > n/C$ such that for all $y' \in Y'$ there is a Y-rotation $\mathcal{F}_{y'}$ of $\mathcal{F}$ with new endpoint y'. Let $x'y'$ be an edge between X' and Y', which exists as G is a C-expander.

Now, the forest $\mathcal{F}_{x'}$ is such that $\mathcal{F}_Y \subseteq \mathcal{F}_{x'}$. Therefore, any Y-rotation of $\mathcal{F}$ is identical when restricted to Y to any Y-rotation of $\mathcal{F}_{x'}$ with the same change in endpoints. Therefore, since y' is a new endpoint created by a Y-rotation $\mathcal{F}_{y'}$ of $\mathcal{F}$, there is then a linear forest $\mathcal{F}_{x',y'}$ which is obtained by performing the X-rotation and Y-rotation consecutively. Note that $\mathcal{F}_{x',y'}$ has no isolated vertices by Lemma 3.3 (B).

In particular, then, x', y' are not isolated in $\mathcal{F}_{x',y'}$ and so adding the edge $x'y'$ to $\mathcal{F}_{x',y'}$ we obtain the forest $\mathcal{F}'$ which has neither x' or y' as endvertices. More precisely, $\mathrm{End}(\mathcal{F}') = \mathrm{End}(\mathcal{F}) \setminus \{x, y\}$. Since $\mathcal{F}_x$ and $\mathcal{F}_y$ are $O(\log n)$-rotations, we conclude that $\mathcal{F}$ and $\mathcal{F}'$ differ in at most $O(\log n)$ edges, i.e., $|E(\mathcal{F}) \Delta E(\mathcal{F}')| = O(\log n)$. □

Unfortunately, even when we can split our linear forest into two subcollections of paths with many vertices, we cannot guarantee that we can rotate within them independently (as for Lemma 3.5) from two arbitrary endvertices x and y. Therefore, in our proof, we will first need to rotate in the whole linear forest to replace x and y by vertices for which we can do this. This is carried out in Section 3.3, using the following lemma.

Lemma 3.6. *Let $C \geq C' > 10^6$ and let G be an n-vertex C-expander which contains a linear forest $\mathcal{F}$ with no isolated vertices. Let $U, V \subseteq V(G)$ be two subsets of vertices. Let $u \in U$ be an endpoint of $\mathcal{F}$ and U an $(\mathcal{F}, C')$-expander with $|U| \geq 10^5 n/C$. Let V satisfy $|\mathrm{int}_{\mathcal{F}}(V)| \geq 11n/C$ and contain no endpoints of $\mathcal{F}$. Then, there is an $(U \cup V, O(\log n))$-rotation $\mathcal{F}'$ of $\mathcal{F}$ and $v \in V$ such that $\mathrm{End}(\mathcal{F}') = (\mathrm{End}(\mathcal{F}) \setminus \{u\}) \cup \{v\}$. Furthermore, we have that $\mathcal{F}'[V \setminus v] = \mathcal{F}[V \setminus v]$, all the edges broken in the successive 1-rotations except the last one are not in $\mathcal{F}[V]$, and $\mathcal{F}'$ has no isolated vertices.*

Proof. First apply Lemma 3.4 to see that $|E_U^{2\log_{C'} n}(u, \mathcal{F})| \geq |U|/10^5 \geq n/C$. We consider two cases.

Case 1: $E_U^{2\log_{C'} n}(u, \mathcal{F}) \cap V \neq \emptyset$.

Let $k \leq 2\log_{C'} n$ be the smallest integer such that $E_U^k(u, \mathcal{F}) \cap V \neq \emptyset$ and $v \in E_U^k(u, \mathcal{F}) \cap V$. Let $\mathcal{F}'$ be the rotation corresponding to the endpoint v. Then $\mathcal{F}'$ is a $O(\log n)$-rotation of $\mathcal{F}$. Furthermore, by the minimality of k, none of the new endpoints, except for v, in the successive 1-rotations creating $\mathcal{F}'$ belong to V. Hence, none of the broken edges in these successive rotations are contained in V. Therefore, the only difference between $\mathcal{F}'[V]$ and $\mathcal{F}[V]$ occurs in the last 1-rotation, with a broken edge incident to v. Hence, $\mathcal{F}'[V \setminus v] = \mathcal{F}[V \setminus v]$, as desired. It is easy to check that $\mathcal{F}'$ has no isolated vertices. Indeed, note that by Lemma 3.3 (B), the only candidate is possibly v, but since V has no endpoints in $\mathcal{F}$, by Lemma 3.3 (B) we are done.

Case 2: $E_U^{2\log_{C'} n}(u, \mathcal{F}) \cap V = \emptyset$.

Let $k \leq 2\log_{C'} n$ be the smallest integer such that $|E_U^k(u, \mathcal{F})| \geq n/C$, and let $E_U^{k-1}(u, \mathcal{F}) \subseteq X \subseteq E_U^k(u, \mathcal{F})$ satisfy $|X| = n/C$. Let P be the set consisting of u and all the pivots used in 1-rotations creating endpoints in X and note that by part (A) of Lemma 3.3 we have $|P| \leq 2|X| + 1$. Let P^2 be set of vertices which are at distance at most 2 from P in the forest $\mathcal{F}$. Now we have that $|\mathrm{int}_{\mathcal{F}}(V) \setminus P^2| \geq 11n/C - 5|P| \geq 11n/C - 10n/C - 5 \geq n/2C$. Since G is a C-expander, there is an edge xy with $x \in X$ and $y \in \mathrm{int}_{\mathcal{F}}(V) \setminus P^2$. Consider the (U, i)-rotation of $\mathcal{F}$ for which x is the last new endpoint. Combined with the 1-rotation with old endpoint x, pivot y and a new endpoint $v \in V$, we obtain a new rotation which satisfies two properties: its last new endpoint is in V, and all the previous pivots except maybe the last one are in U. Denote this rotation by $\mathcal{F}'$.

The above argument shows that the obtained i-rotation $\mathcal{F}'$ satisfies $i \leq 2\log_{C'} n + 1 < \log n$. Note that, by construction, all the pivots used to obtain $\mathcal{F}'$ are at least at distance 3 in $\mathcal{F}$. Also note that since $\mathcal{F}'$ is a combination of a (U, i)-rotation and another 1-rotation, such that the (U, i)-rotation uses no new endpoints in V, the only edge which is possibly broken in $\mathcal{F}[V]$ is in the last 1-rotation when the edge yv is broken. This immediately implies $\mathcal{F}'[V \setminus v] = \mathcal{F}[V \setminus v]$. Finally, again by Lemma 3.3 (B) there are no isolated vertices in $\mathcal{F}'$. □

3.3 The main rotation lemma

The next lemma is one of our key results. As it is the most technical part of our proof, we will now outline its proof briefly. Given an n-vertex C-expander G containing a spanning linear forest, with most

of its vertices on paths that are not too long, in which x and y are endvertices, we wish to perform rotations on $\mathcal{F}$ until we have enough vertices potentially replacing x and y as endpoints that can use Definition 1.2b) to connect two paths so that we end up with one fewer path while losing exactly x and y as endpoints (see Lemma 3.7). Using Lemma 3.6, we can perform rotations on $\mathcal{F}$ to replace x and y each with linearly many different new endpoints. However, we want to do this independently, so that Definition 1.2b) can be applied. We manage this in two stages. First, we use rotations to replace x and y by two new endpoints x_1 and x_2 so that we can use rotations on two different sets of vertices (U_1 and U_2 respectively) to change x_1 or to change x_2. We then use this property to replace x_1 and x_2 by two new endpoints x_3 and x_4, so that these endpoints can be rotated independently by using vertices on two different sets of paths (by rotating with vertex sets U_3 and U_4, whose vertices never appear together on the same path). The two sets of endpoints created by doing this will have some edge between them by Definition 1.2b), allowing us to join two paths together. However, this might create a cycle if x_3 and x_4 were endpoints of the same path. Therefore, instead, we first rotate to replace x_1 by some other vertex h (lying in a fifth set U_{hop}) before performing one extra rotation to change h to a new endpoint, x_3, where this brief 'hop' allows us to ensure that x_3 is then not an endpoint of the same path as x_4.

Lemma 3.7. *Let $C > 10^{10}$, let G be an n-vertex C-expander and let $\mathcal{F}$ be a spanning linear forest in G with no isolated vertices and such that at least $0.1n$ vertices of G belong to paths in $\mathcal{F}$ of lengths between 100 and $\sqrt{n}$. Let also $x, y \in \mathrm{End}(\mathcal{F})$. Then, there is a spanning linear forest $\mathcal{F}'$ in G with no isolated vertices such that $|E(\mathcal{F})\Delta E(\mathcal{F}')| = O(\log n)$ and $\mathrm{End}(\mathcal{F}') = \mathrm{End}(\mathcal{F}) \setminus \{x, y\}$.*

Proof. Let $P_1, \dots, P_t$ be the paths in $\mathcal{F}$ of length between 100 and $\sqrt{n}$, so that $\sum_{i\in[t]} |P_i| \geq 0.1n$. Let $\mathcal{H} := \{P'_1, \dots, P'_t\}$, such that each path P'_i is obtained from P_i by removing the two endpoints of the path P_i, so that now $\sum_{i\in[t]} |P'_i| \geq 0.1n(1 - \frac{1}{50}) \geq 0.098n$. Take a partition of this linear forest $\mathcal{H}$ into $\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3, \mathcal{H}_4, \mathcal{H}_{\text{hop}}$, each of which span at least $0.015n$ vertices. Hence, the interior of each of these forests (since they consist of paths of length at least 98) is at least of size $\frac{96}{98} \cdot 0.015n \geq 0.01n$.

Let $C' = C/5000$. By Lemma 2.5, we can find subsets $U_1 \subseteq V(\mathcal{H}_1)$ and $U_2 \subseteq V(\mathcal{H}_2)$, where U_i is such that it C'-expands into $\mathrm{int}_{\mathcal{H}}(U_j) \subseteq \mathrm{int}_{\mathcal{F}}(U_j)$ for all $i, j \in \{1, 2\}$. In particular, U_1 and U_2 are $(\mathcal{F}, C')$-expanders. Furthermore, every vertex in U_1 has at least $C' > 100$ neighbours in $\mathrm{int}_{\mathcal{F}}(U_2)$, and every vertex in U_2 has at least $C' > 100$ neighbours in $\mathrm{int}_{\mathcal{F}}(U_1)$. We also can get from Lemma 2.5 that, for each $i \in \{1, 2\}$, it holds that $|\mathrm{int}_{\mathcal{F}}(U_i)| \geq |V(\mathcal{H}_i)| - 2n/C \geq 0.01n - 2n/C \geq 0.005n$.

We now use Lemma 3.6 to prove that the following holds.

Claim 3. *There is a linear forest $\mathcal{F}'$ and $x_1 \in U_1, x_2 \in U_2$ such that $\mathrm{End}(\mathcal{F}') = (\mathrm{End}(\mathcal{F}) \setminus \{x, y\}) \cup \{x_1, x_2\}$, $|E(\mathcal{F}'[U_1 \cup U_2]) \Delta E(\mathcal{F}[U_1 \cup U_2])| \leq 20$ and $|E(\mathcal{F})\Delta E(\mathcal{F}')| = O(\log n)$. Furthermore, $\mathcal{F}'$ has no isolated vertices.*

Proof. First, apply Lemma 3.6 with $u := x, U := V(G)$, and $V := U_1 \cup U_2$ to give a $(V(G), O(\log n))$-rotation $\mathcal{F}_1$ of $\mathcal{F}$, where x is replaced by a new endpoint $x_1 \in U_1 \cup U_2$, so that $\mathrm{End}(\mathcal{F}_1) = (\mathrm{End}(\mathcal{F}) \setminus \{x\}) \cup \{x_1\}$, and $\mathcal{F}[(U_1 \cup U_2) \setminus x_1] = \mathcal{F}_1[(U_1 \cup U_2) \setminus x_1]$. Furthermore, $\mathcal{F}_1$ has no isolated vertices. Without loss of generality, let $x_1 \in U_1$.

Now, we apply again Lemma 3.6 to $\mathcal{F}_1$ with $u := y, U := V(G) \setminus T$, and $V := (U_1 \cup U_2) \setminus T$, where T is the set of vertices which are at most at distance 2 from x_1 in $\mathcal{F}$, so that $|T| \leq 5$. Thus, we get a $O(\log n)$-rotation $\mathcal{F}_2$ of $\mathcal{F}_1$ where y is replaced by a new endpoint $w \in (U_1 \cup U_2) \setminus T$, so that $\mathrm{End}(\mathcal{F}_2) = (\mathrm{End}(\mathcal{F}) \setminus \{x, y\}) \cup \{x_1, w\}$ and $\mathcal{F}_1[V \setminus \{w\}] = \mathcal{F}_2[V \setminus \{w\}]$. Furthermore, since the endpoint is not in T, it is different to x_1. Finally, $\mathcal{F}_2$ has no isolated vertices.

Now, in the case that $w \in U_2$, the linear forest $\mathcal{F}' := \mathcal{F}_2$ is as desired with $x_2 := w$ and x_1 since $\mathcal{F}[(U_1 \cup U_2) \setminus (\{x_1, w\} \cup T)] = \mathcal{F}_2[(U_1 \cup U_2) \setminus (\{x_1, w\} \cup T)]$. Otherwise, if $w \in U_1$, then, by the previous condition that every vertex in U_1 has at least 100 neighbours in $\mathrm{int}_{\mathcal{F}}(U_2)$, there is an edge wz for some $z \in \mathrm{int}_{\mathcal{F}}(U_2) \cap \mathrm{int}_{\mathcal{F}_2}(U_2)$. We can then use this edge to create a 1-rotation $\mathcal{F}'$ of $\mathcal{F}_2$, replacing the old endpoint w with a new endpoint $x_2 \in U_2 \setminus \{x_1\}$, that is, $\mathrm{End}(\mathcal{F}_2) = (\mathrm{End}(\mathcal{F}) \setminus \{x, y\}) \cup \{x_1, x_2\}$. The linear forest

$\mathcal{F}'$ is then as desired since, in particular, $\mathcal{F}[(U_1 \cup U_2) \setminus (\{w, z, x_2\} \cup T)] = \mathcal{F}'[(U_1 \cup U_2) \setminus (\{w, z, x_2\} \cup T)]$ and again, by Lemma 3.3 (B), since $\mathcal{F}_2$ has no isolated vertices and x_2 is not an endpoint of $\mathcal{F}$, and so, also not one of $\mathcal{F}_2$ given that $\mathrm{End}(\mathcal{F}_2) = (\mathrm{End}(\mathcal{F}) \setminus \{x, y\}) \cup \{x_1, w\}$, $\mathcal{F}'$ has no isolated vertices.

Finally, we have $|E(\mathcal{F}'[U_1 \cup U_2]) \Delta E(\mathcal{F}[U_1 \cup U_2])| \leq 20$, since these linear forests differ in at most 8 vertices, either $T \cup \{x_1, w\}$ in the first case, or $T \cup \{w, z, x_2\}$ in the second. ⊡

Let us make two remarks. First, the future rotations of $\mathcal{F}'$ that we will perform in the rest of the proof will all be $V(\mathcal{H})$-rotations, and, therefore, none of those linear forests will contain any isolated vertices; this is important since there will be applications of Lemma 3.3, Lemma 3.5 and Lemma 3.6 throughout the rest of the proof. Secondly, note that to obtain $\mathcal{F}'$ we rearranged some paths, hence, it is entirely possible that, in $\mathcal{F}'$, vertices in U_1 live in the same paths as vertices in U_2 (although this was not the case in $\mathcal{F}$). Thus, we cannot yet apply Lemma 3.5 to the new endpoints x_1, x_2.

Now, let $\mathcal{H}'_3 \subseteq \mathcal{H}_3, \mathcal{H}'_4 \subseteq \mathcal{H}_4$, and $\mathcal{H}'_{\mathrm{hop}} \subseteq \mathcal{H}_{\mathrm{hop}}$, be the linear sub-forests of $\mathcal{H}_3, \mathcal{H}_4, \mathcal{H}_{\mathrm{hop}}$ consisting of paths P'_i such that $P_i \in \mathcal{F} \cap \mathcal{F}'$. By the previous claim, we have that $|E(\mathcal{F}) \Delta E(\mathcal{F}')| = O(\log n)$. Since each path in $\mathcal{H}$ has size at most $\sqrt{n} = o(n/\log n)$, we must then have that $|V(\mathcal{H}'_j)| \geq 0.009n$ for each $j \in \{3, 4, hop\}$. By Lemma 2.5, there exist an $(\mathcal{H}'_3, C')$-expander U_3, an $(\mathcal{H}'_4, C')$-expander U_4, and an $(\mathcal{H}'_{\mathrm{hop}}, C')$-expander U_5, each of size at least $0.009n - n/2C \geq 0.005n$. As before, since we also can get that each U_i is such that it C'-expands into $\mathrm{int}_{\mathcal{F}'}(U_j)$ (for all $i, j \in \{3, 4, hop\}$), we can have that every vertex in U_{hop} has at least $C' \geq 1$ neighbours in $\mathrm{int}_{\mathcal{F}'}(U_3)$. Also, note the following crucial observation.

Claim 4. *U_1 and U_2 are $(\mathcal{F}', C' - 40)$-expanders.*

Indeed, this is easy to see since the claim before implies that $|E(\mathcal{F}'[U_1 \cup U_2]) \Delta E(\mathcal{F}[U_1 \cup U_2])| \leq 20$. We now apply Lemma 3.6 again. First, we apply the lemma to $\mathcal{F}'$ with $x := x_1, U := U_1$, and $V := U_{\mathrm{hop}}$ (recall that U_{hop} does not contain endpoints of $\mathcal{F}$, and so also of $\mathcal{F}'$), which gives a $(U_1 \cup U_{\mathrm{hop}}, O(\log n))$-rotation $\mathcal{F}'_1$ of $\mathcal{F}'$ in which the endpoint x_1 is replaced by an endpoint $h \in U_{\mathrm{hop}}$. Moreover, we have that $\mathcal{F}'_1[U_{\mathrm{hop}} \setminus h] = \mathcal{F}'[U_{\mathrm{hop}} \setminus h]$ and $\mathcal{F}'_1$ is a $(U_1 \cup U_{\mathrm{hop}})$-rotation, so that all broken edges in the successive rotations belong to $U_1 \cup U_{\mathrm{hop}}$ and, thus, U_2 is still an $(\mathcal{F}'_1, C' - 40)$-expander, $\mathrm{int}_{\mathcal{F}'_1}(U_4) = \mathrm{int}_{\mathcal{F}}(U_4)$ is of size at least $0.001n$, and $\mathcal{F}'_1[U_4] = \mathcal{F}'[U_4]$. Furthermore, the lemma implies that no vertex of U_4 is an endpoint of $\mathcal{F}'_1$. Therefore, we can apply again Lemma 3.6 to $\mathcal{F}'_1$, now with $x := x_2, U := U_2$, and $V := U_4$. This gives a $(U_2 \cup U_4, O(\log n))$-rotation $\mathcal{F}'_2$ with a new endpoint $x_4 \in U_4$ replacing x_2.

Now, note that for all paths $P'_i \in \mathcal{H}'_3$, we have that $P_i \in \mathcal{F}'_2$. Note also that since the previous rotations forming $\mathcal{F}'_1$ and $\mathcal{F}'_2$ are such that all the broken edges in the successive 1-rotations apart from the last ones are not in $\mathcal{F}[U_{\mathrm{hop}}]$ or $\mathcal{F}[U_4]$, we have that the path in $\mathcal{F}'_2$ containing h is a sub-path of a path $P_i \in \mathcal{F}$ such that P'_i is contained in $\mathcal{H}'_{\mathrm{hop}}$ and the path containing x_4 is a sub-path of a path $P_j \in \mathcal{F}$ such that P'_j is contained in $\mathcal{H}'_4$. Thus, h, x_4 are endpoints of different paths in $\mathcal{F}'_2$. Now, by assumption, $h \in U_{\mathrm{hop}}$ has a neighbour in $\mathrm{int}_{\mathcal{F}'}(U_3) = \mathrm{int}_{\mathcal{F}'_2}(U_3)$. This implies that we can perform a 1-rotation on $\mathcal{F}'_2$ to get $\mathcal{F}'_3$ and a vertex $x_3 \in U_3$ with $\mathrm{End}(\mathcal{F}'_3) = (\mathrm{End}(\mathcal{F}'_2) \setminus \{h\}) \cup \{x_3\} = (\mathrm{End}(\mathcal{F}) \setminus \{x, y\}) \cup \{x_3, x_4\}$.

Now, since h, x_4 were endpoints of different paths in $\mathcal{F}'_2$, it is easy to check that there can be no path of $\mathcal{F}'_3$ which intersects both U_3 and U_4. Since U_3 and U_4 are still $(\mathcal{F}'_3, C' - 2)$-expanders and have size at least $0.005n$, Lemma 3.5 implies the existence of a linear forest $\mathcal{F}'_4$ so that $\mathrm{End}(\mathcal{F}'_4) = \mathrm{End}(\mathcal{F}) \setminus \{x, y\}$. Moreover, since every rotation performed was an $O(\log n)$-rotation, we have that $|E(\mathcal{F}) \Delta E(\mathcal{F}'_4)| = O(\log n)$, as desired. □

4 Linear forests with few paths

The goal of this section is to find a spanning linear forest in an expander with a relatively small number of paths. That is, we will prove the following result.

Lemma 4.1. *Let G be an n-vertex C-expander for $C > 10^{10}$. Then, it contains a spanning linear forest with at most $n^{0.8}$ paths and no isolated vertices.*

We will find the required linear forest for Lemma 4.1 by taking a minimal such linear forest under a certain ordering, which is based on the lengths of the paths in the forest. In this section we will define the length of a path to be the number of vertices it has (which makes calculations a bit easier).

Definition 4.2. Given a linear forest $\mathcal{F}$ and integers a, b, we define $S_{\mathcal{F}}(a, b)$ to be the set of vertices in the collection of paths in $\mathcal{F}$ of length at least a and at most b. For each $v \in V(\mathcal{F})$, we denote by $\mathrm{seg}_{\mathcal{F}}(v)$ the path in $\mathcal{F}$ which contains v. We omit $\mathcal{F}$ in the subscript when it is clear from the context.

Definition 4.3. Let $<_{\text{lex}}$ be the ordering on the family of all linear forests, where for any two linear forests $\mathcal{F}_1$ and $\mathcal{F}_2$ we have $\mathcal{F}_1 <_{\text{lex}} \mathcal{F}_2$ if

- $\mathcal{F}_1$ has fewer paths than $\mathcal{F}_2$, or
- $\mathcal{F}_1$ and $\mathcal{F}_2$ have the same number of paths and the vector of path lengths of $\mathcal{F}_1$ in decreasing order is lexicographically smaller than that of $\mathcal{F}_2$.

To illustrate this definition, for example, if $\mathcal{F}_1$ consists of paths of length $8, 5, 3, 2, 2$ and $\mathcal{F}_2$ consists of paths of length $8, 5, 4, 2, 1$, then we have $F_1 <_{\text{lex}} F_2$ as they both have 5 paths but $(8, 5, 3, 2, 2)$ is smaller than $(8, 5, 4, 2, 1)$ in the lexicographic ordering.

As discussed in Section 3, it will be convenient to use spanning linear forests with no isolated vertices. We will be able to apply the following lemma to show that a $<_{\text{lex}}$-minimal spanning linear forest in an expander has no isolated vertices.

Lemma 4.4. *Let $\mathcal{F}$ be a $<_{\text{lex}}$-minimal spanning linear forest in an n-vertex C-expander G for some $C > 10^6$. Suppose that $\mathcal{F}$ contains an isolated vertex v, and $\mathcal{F}'$ is a k-rotation of $\mathcal{F}$ with old endpoint v. Then $\mathcal{F}'$ is $<_{\text{lex}}$-minimal, the new endpoint u of $\mathcal{F}'$ is isolated in $\mathcal{F}'$, and* $\mathrm{End}(\mathcal{F}') = \mathrm{End}(\mathcal{F})$.

Proof. We prove this by induction on k. For the initial case, "$k = 1$", let w be the pivot of the 1-rotation, and let P be the path of $\mathcal{F}$ containing w. Let P_1 and P_2 be the two (possibly empty) subpaths of $P - w$, labelled so that $|P_1| \leq |P_2|$. If $|P_2| \geq 2$, then replacing $\{v\}$ and P in $\mathcal{F}$ with the paths P_1wv and P_2 gives a linear forest $\mathcal{F}'$ with $\mathcal{F}' <_{\text{lex}} \mathcal{F}$, which is a contradiction. Similarly, if $|P_1| = 0$, then replacing $\{v\}$ and P in $\mathcal{F}$ with the path P_2wv gives a linear forest $\mathcal{F}'$ with $\mathcal{F}' <_{\text{lex}} \mathcal{F}$, which is a contradiction. Thus we can suppose that P_1, P_2 are both single vertices, i.e. that $P = xwy$ for some x, y. Then, a 1-rotation with pivot w replaces the paths P and $\{v\}$ with either the paths vwx, y or the paths vwy, x. In both cases the new linear forest has the same path lengths as $\mathcal{F}$ did (and so stays $<_{\text{lex}}$-minimal), the same set of endpoints as $\mathcal{F}$, and has new endpoint either x or y, which is now isolated.

For the induction step, consider a k-rotation $\mathcal{F}_k$ of $\mathcal{F}$ with old endpoint v. By the definition of a k-rotation, we get a $(k-1)$-rotation $\mathcal{F}_{k-1}$ of $\mathcal{F}$ with old endpoint v, new endpoint u, such that $\mathcal{F}_k$ is a 1-rotation of $\mathcal{F}_{k-1}$ with old endpoint u. By induction we get that $\mathcal{F}_{k-1}$ is $<_{\text{lex}}$-minimal, has u isolated, and has $\mathrm{End}(\mathcal{F}_{k-1}) = \mathrm{End}(\mathcal{F})$. By the "$k = 1$" case, we have that F_k is $<_{\text{lex}}$-minimal, its new endpoint is isolated, and $\mathrm{End}(\mathcal{F}_k) = \mathrm{End}(\mathcal{F}_{k-1}) = \mathrm{End}(\mathcal{F})$. □

We now sketch out the proof of Lemma 4.1. Given a $<_{\text{lex}}$-minimal spanning linear forest $\mathcal{F}$ in a C-expander G, with C large, we will first observe that there cannot be an edge in G from the end of a path in $\mathcal{F}$ to a path in $\mathcal{F}$ which is more than five times as long (see Lemma 4.5). We will then argue, for any $i \geq 0$, that this implies that if we perform a sequence of i 1-rotations starting with an endpoint of a smallest path in a $<_{\text{lex}}$-minimal spanning linear forest $\mathcal{F}$, then the new endvertex must be in a path at most 6^i times as long as a smallest path (see Lemma 4.6). However, Lemma 3.4 implies that within at most $k := 2\log_C n$ rotations we reach at least $n/10^5$ new endpoints, which then collectively have edges to all but at most $n/2C$ vertices in the graph by the definition of a C-expander. After using Lemma 4.4 and assuming, for sake of contradiction that $\mathcal{F}$ has more than $n^{0.8}$ paths, we can deduce that $\mathcal{F}$ has no isolated vertices and also note that $\mathcal{F}$ must have at most $100n/C$ vertices contained in paths of length

less than 100, so that the $<_{\text{lex}}$-minimality of $\mathcal{F}$ then implies by Lemma 3.7 that at least $0.8n$ of the vertices of G are in paths in $\mathcal{F}$ with more than $\sqrt{n}$ vertices. In combination, this all implies that the shortest path in $\mathcal{F}$ must have at least $6^{-(k+1)}\sqrt{n}$ vertices, so that $\mathcal{F}$ contains at most $\sqrt{n} \cdot 6^{(k+1)} \leq n^{0.8}$ paths, a contradiction, as required.

First, then, we show endvertices of paths cannot have an edge to much longer paths in a $<_{\text{lex}}$-minimal spanning linear forest, in the following stronger form.

Lemma 4.5. *Let $t \geq t' > 0$ be integers. Let $\mathcal{F}$ be a minimal spanning linear forest in G with respect to $<_{\text{lex}}$, with paths of decreasing lengths $\ell_1, \ldots, \ell_t$. Let $\mathcal{F}'$ be another forest with paths of lengths $s_1, \ldots, s_t$ in decreasing order, where the first $t'-1$ lengths are the same as in $\mathcal{F}$, i.e. $\ell_i = s_i$ for each $i < t'$. Let x be an endpoint in $\mathcal{F}'$ in a path of length at most $s_{t'}$. Then, $N_G(x) \subseteq S_{\mathcal{F}'}(0, 5s_{t'})$.*

Proof. Suppose for contradiction there is a path P_i (of length s_i) in $\mathcal{F}'$ with $s_i \leq s_{t'}$, whose endpoint x has a neighbour y in another path P_j of length at least $5s_{t'} \geq 5|P_i| \geq 5$. Then, we can obtain a new forest $\mathcal{F}''$ from $\mathcal{F}'$ such that $\mathcal{F}'' <_{\text{lex}} \mathcal{F}$, as follows. We replace the paths P_i and P_j in $\mathcal{F}'$ by two new paths. The first new path is the concatenation of P_i, the edge xy and the shorter subpath of P_j which starts at y and ends in an endpoint of P_j. The other path is the remaining part of P_j. Note that both of those paths are shorter than P_j as the first path is of length at most $|P_i| + 1 + |P_j|/2 \leq |P_j|/5 + 1 + |P_j|/2 < |P_j|$, and the second one is a strict subpath of P_j. This contradicts the $<_{\text{lex}}$-minimality of $\mathcal{F}$. □

Next, we iterate Lemma 4.5, showing that repeatedly applying 1-rotations cannot create a new endpoint with a neighbour in a considerably longer path than the path containing the old endpoint, as follows.

Lemma 4.6. *Let $\mathcal{F}$ be a $<_{\text{lex}}$-minimal spanning linear forest of G. Let v be an endpoint of a path in $\mathcal{F}$. Then $N(E_G^k(v, \mathcal{F})) \subseteq S_{\mathcal{F}}(0, 6^{k+1}|\text{seg}_{\mathcal{F}}(v)|)$ for every positive integer k.*

Proof. Let $\mathcal{F}_k$ be a k-rotation of $\mathcal{F}$ and let u_k be its new endpoint and $v = u_0$ its old endpoint. Consider the linear forests $\mathcal{F} = \mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_k$, where $\mathcal{F}_i$ is a 1-rotation of $\mathcal{F}_{i-1}$, and where the new endpoint of $\mathcal{F}_i$ is u_i, the old endpoint is u_{i-1}, and the pivot is z_i. We show that the following two conditions hold for every $i \geq 0$:

(1) $N(u_i) \subseteq S_{\mathcal{F}}(0, 6^{i+1}|\text{seg}_{\mathcal{F}}(v)|)$

(2) The paths of length $> 6^{i+1}|\text{seg}_{\mathcal{F}}(v)|$ are the same in $\mathcal{F}$ and $\mathcal{F}_{i+1}$.

Hence, the particular case $i = k$ proves the statement of the lemma. For the base case $i = 0$, let $\ell_1, \ldots, \ell_t$ be the lengths of the paths in $\mathcal{F}$ in decreasing order. Let t' be such that v is in a path of length $\ell_{t'}$. Note that by Lemma 4.5 applied to $\mathcal{F} = \mathcal{F}'$ with t', we know that v only has neighbours in $S_{\mathcal{F}}(0, 5|\text{seg}_{\mathcal{F}}(v)|) \subseteq S_{\mathcal{F}}(0, 6|\text{seg}_{\mathcal{F}}(v)|)$, as required for part (1). In the resulting linear forest $\mathcal{F}_1$ only two paths have changed compared to $\mathcal{F}$ – one of length $|\text{seg}_{\mathcal{F}}(v)|$ and one of length at most $5|\text{seg}_{\mathcal{F}}(v)|$, and so we obtained two new paths of length at most $6|\text{seg}_{\mathcal{F}}(v)|$ each. Hence, all the paths in $\mathcal{F}_1$ of length larger than $6|\text{seg}_{\mathcal{F}}(v)|$ are the same as in $\mathcal{F}$, proving part (2).

Suppose now the claim holds for all $j < i$, and let us prove it for i. Note that the first condition implies that all the pivots z_j for $j \leq i$ are in $S_{\mathcal{F}}(0, 6^j|\text{seg}_{\mathcal{F}}(v)|)$, as each z_j lives in $N(u_{j-1})$. In particular, $z_i \in S_{\mathcal{F}}(0, 6^i|\text{seg}_{\mathcal{F}}(v)|)$, meaning that $u_i \in S_{\mathcal{F}}(0, 6^i|\text{seg}_{\mathcal{F}}(v)|)$, because either $z_i = u_i$ or the broken edge z_iu_i is in $\mathcal{F}$, so u_i and z_i are on the same path in $\mathcal{F}$. Additionally, note that by the induction hypothesis (part (2)) all the paths of length larger than $6^i|\text{seg}_{\mathcal{F}}(v)|$ are the same in $\mathcal{F}$ and $\mathcal{F}_i$. Hence, since $u_i \in S_{\mathcal{F}}(0, 6^i|\text{seg}_{\mathcal{F}}(v)|)$, then $u_i \in S_{\mathcal{F}_i}(0, 6^i|\text{seg}_{\mathcal{F}}(v)|)$. Therefore we can apply Lemma 4.5 to $\mathcal{F}, \mathcal{F}_i$ and u_i to get that $N(u_i) \subseteq S_{\mathcal{F}_i}(0, 5|\text{seg}_{\mathcal{F}_i}(u_i)|)$, which by the previous discussion gives that $N(u_i) \subseteq S_{\mathcal{F}_i}(0, 5 \cdot 6^i|\text{seg}_{\mathcal{F}}(v)|)$. Now, we show that (2) holds. Note that, in order to get $\mathcal{F}_{i+1}$, in $\mathcal{F}_i$ the path which contains u_i and a path which contains a vertex in $N(u_i)$ are replaced with two paths on the same set of vertices. Thus the two newly obtained paths are of length at most

$6^i|\text{seg}_{\mathcal{F}}(v)| + 5 \cdot 6^i|\text{seg}_{\mathcal{F}}(v)| \leq 6^{i+1}|\text{seg}_{\mathcal{F}}(v)|$. As the paths of length $> 6^i|\text{seg}_{\mathcal{F}}(v)|$ are the same in $\mathcal{F}$ and $\mathcal{F}_i$, this means that the paths of length $> 6^{i+1}|\text{seg}_{\mathcal{F}}(v)|$ are the same in $\mathcal{F}$ and $\mathcal{F}_{i+1}$, completing part (2).

For part (1), recall that $N(u_i) \subseteq S_{\mathcal{F}_i}(0, 5 \cdot 6^i|\text{seg}_{\mathcal{F}}(v)|)$. Thus, $N(u_i) \subseteq S_{\mathcal{F}}(0, 6^{i+1}|\text{seg}_{\mathcal{F}}(v)|)$. Indeed, otherwise a vertex $x \in N(u_i)$ is in a path of length $> 6^{i+1}|\text{seg}_{\mathcal{F}}(v)| \geq 6^i|\text{seg}_{\mathcal{F}}(v)|$ in $\mathcal{F}$, but in $\mathcal{F}_i$ it is in a path of length at most $6^{i+1}|\text{seg}_{\mathcal{F}}(v)|$. By the induction hypothesis for part (2) for $i-1$, the paths of length $> 6^i|\text{seg}_{\mathcal{F}}(v)|$ are the same in $\mathcal{F}$ and $\mathcal{F}_i$ and thus, this is a contradiction. This completes the proof. □

Next, we show that in a $<_{\text{lex}}$-minimal spanning linear forest with quite a lot of paths, only a small proportion of the vertices can lie in very long paths.

Lemma 4.7. *Let $C > 10^{10}$. Let $\mathcal{F}$ be a $<_{\text{lex}}$-minimal spanning linear forest in an n-vertex C-expander G. Suppose that the number of paths in $\mathcal{F}$ is at least n^{ε}, for some $\varepsilon > 0$. Then, $|S(6n^{1-\varepsilon+2\log_C 6}, n)| \leq n/2C$.*

Proof. Since we have n^{ε} paths, there is one of length at most $n^{1-\varepsilon}$. Set $k := 2\log_C n$, let v be an endpoint of a shortest path, which is of length $m \leq n^{1-\varepsilon}$, and suppose for contradiction that $|S(6mn^{2\log_C 6}, n)| > n/2C$. By Lemma 3.4, we have $|E_G^k(v, \mathcal{F})| \geq n/2C$. We now have two cases. If $E_G^k(v, \mathcal{F}) \cap S(6mn^{2\log_C 6}, n) \neq \emptyset$, then let i be the smallest such that there exists $x \in E_G^i(v, \mathcal{F}) \cap S(6mn^{2\log_C 6}, n)$. Let y be the pivot for x in the corresponding i-rotation and recall that x, y are in the same path in $\mathcal{F}$, hence $y \in S(6mn^{2\log_C 6}, n)$ as well. Note also that $y \in N(E_G^{i-1}(v, \mathcal{F}))$, so that $S(6mn^{2\log_C 6}, n) \cap N(E_G^{i-1}(v, \mathcal{F})) \neq \emptyset$. On the other hand, if $E_G^k(v, \mathcal{F}) \cap S(6mn^{2\log_C 6}, n) = \emptyset$, then by the definition of a C-expander, we have that $N(E_G^k(v, \mathcal{F})) \cap S(6mn^{2\log_C 6}, n) \neq \emptyset$. Both of these cases contradict Lemma 4.6, which says that for $i \leq k$ we have

$$N(E_G^i(v, \mathcal{F})) \subseteq S(0, 6^{k+1}|\text{seg}_{\mathcal{F}}(v)|) = S(0, 6^{1+2\log_C n}m) = S(0, 6^{1+2\log_6 n/\log_6 C}m) = S(0, 6mn^{2\log_C 6}).$$

Hence, $|S(6n^{1-\varepsilon+2\log_C 6}, n)| \leq |S(6mn^{2\log_C 6}, n)| \leq n/2C$. □

Finally, then, in this section, we can put this all together to prove Lemma 4.1.

Proof of Lemma 4.1. Let $\mathcal{F}$ be a $<_{\text{lex}}$-minimal spanning forest in G. First we will show that it must contain at most n/C paths. Indeed, suppose otherwise and let S be a set which contains exactly one endpoint from each path in $\mathcal{F}$. Split S into disjoint sets $S = S_1 \cup S_2$, both of size $|S_1|, |S_2| \geq n/2C$. Since G is a C-expander, there is an edge between S_1 and S_2. Adding this edge to $\mathcal{F}$ creates a linear forest with one fewer path than $\mathcal{F}$, contradicting the $<_{\text{lex}}$-minimality of $\mathcal{F}$.

We now show that $\mathcal{F}$ has no isolated vertices. Again, suppose otherwise and let v be an isolated vertex in $\mathcal{F}$. By Lemma 4.4, for all k, we have that $E^k(v, \mathcal{F}) \subseteq \text{End}(\mathcal{F})$. Lemma 3.4 then tells us that $|\text{End}(\mathcal{F})| \geq |E^{2\log_C n}(v, \mathcal{F})| \geq n/10^5 > 2n/C$. This contradicts the previous assertion that $\mathcal{F}$ has at most n/C paths.

Now that we know that $\mathcal{F}$ has at most n/C paths and no isolated vertices, it must be that at most $100n/C$ vertices in $\mathcal{F}$ are contained in paths of length less than 100. To conclude, assume for contradiction that this forest has more than $n^{0.8}$ paths. Now, apply Lemma 4.7 with $\varepsilon = 0.8$ to conclude that $\mathcal{F}$ satisfies $|S_{\mathcal{F}}(6n^{1-\varepsilon+2\log_C 6}, n)| \leq n/2C$. Since $1 - \varepsilon + 2\log_C 6 < 1/2$, the number of vertices in paths longer than $\sqrt{n}$ is less than $n/2C < n/2$. Hence, at least $n/2 - 100n/C \geq 0.1n$ of the vertices are in $S_{\mathcal{F}}(100, \sqrt{n})$. Now, apply Lemma 3.7 to $\mathcal{F}$ to obtain a new linear forest with fewer paths, contradicting the $<_{\text{lex}}$-minimality of $\mathcal{F}$. □

5 Linking structures and decomposing the expander graph

In this section, we will prove the main result we need which will find an appropriate linking structure in an expander. For convenience for our application, we will additionally find relatively long paths connecting the two special sets in the linking structure, as in the following result.

Lemma 5.1. *Let $C > 10^{15}$, let n be sufficiently large and let G be a C-expander. Then, there is a partition of $V(G)$ into three sets X, Y, Z, and disjoint sets $A, B \subset X$, with the following properties.*

- *$G[X]$ is an (A, B)-linking structure and $|A| = |B| = n^{0.9}$.*
- *$G[Y \cup A \cup B]$ has a spanning linear forest with $|A| = |B|$ paths of size $n^{0.1}/5$ whose endpoints are in $A \cup B$.*
- *$G[Z]$ and $G[Z \cup Y \cup A \cup B]$ are both $C/100$-expanders.*

For [38, Proposition 4.4], Hyde, Morrison, Müyesser and Pavez-Signé showed that there exist linking structures with convenient properties, which allow for them to be constructed in pseudorandom graphs. Our methods would be content with a linking structure with much smaller sets A, B, than in the efficiently constructed linking structure in [38], but we will use the same construction, briefly outlining the main arguments for the sake of completeness. We will start in Section 5.1 by recalling some 'extendability' methods, before proving in Section 5.2 that there is a linking structure that can be constructed using these methods, and using this to prove Lemma 5.1.

5.1 Embedding in expanders

Here we briefly discuss a very useful 'extendability' method for embedding sparse graphs in expander graphs. The method combines a technique Friedman and Pippenger [29] used to inductively embed trees leaf-by-leaf into expanding graphs with a 'roll-back' idea of Johannsen (used to prove Lemma 5.3 below). For more on this technique, see [24]. The key definition defining a type of 'good' embedding is the following.

Definition 5.2. Let D, m be positive integers with $D \geq 3$. Let G be a graph and let $H \subset G$ be a subgraph with $\Delta(H) \leq D$. Then H is (D, m)-*extendable (in G)* if for every $S \subset V(G)$ with $1 \leq |S| \leq 2m$ we have

$$|\Gamma_G(S) \setminus V(H)| \geq (D-1)|S| - \sum_{u \in S \cap V(H)} (d_H(u) - 1). \tag{3}$$

The main result to state here is the following, which allows us to add paths to existing (D, m)-extendable subgraphs of an expander such that the resulting subgraph stays (D, m)-extendable. Hence, we will be able to embed in expanders any type of sparse graphs which can be recursively built by adding path by path. This lemma is, for example, a weaker version of [53, Corollary 3.12].

Lemma 5.3. *Let $C > 10D$ and let G be an n-vertex C-expander with a $(D, n/2C)$-extendable subgraph H with $|H| \leq n - 5nD/C - \ell$, where $\ell \geq \log n$. Then, the following hold for all vertices $y \in V(H)$ with $deg_H(y) \leq D/2$.*

- *There exists a path P in G with endpoint y of length ℓ such that all its vertices except for y lie outside of H and $H \cup P$ is $(D, n/2C)$-extendable.*
- *For every $x \in V(H) \setminus \{y\}$ with $deg_H(x) \leq D/2$, there exists an xy-path P in G of length ℓ such that all its internal vertices lie outside of H and $H \cup P$ is $(D, n/2C)$-extendable.*

5.2 Linking structures in expanders

We first need the following definition, as introduced in [38].

Definition 5.4. Let G be a graph and $A \subseteq V(G)$. We say that G is *A-path-constructible* if there exists a sequence of edge-disjoint paths $P_1, \ldots, P_t$ in G with the following properties.

1. $E(G) = \bigcup_{j \in [t]} E(P_j)$.

2. For each $i \in [t]$, the internal vertices of P_i are disjoint from $A \cup \bigcup_{j\in[i-1]} V(P_j)$.

3. For each $i \in [t]$, at least one of the endpoints of P_i belongs to $A \cup \bigcup_{j\in[i-1]} V(P_j)$.

We will say that G is A-path-constructible *with paths of length between* ℓ_1 *and* ℓ_2 if all the paths P_t have lengths between ℓ_1 and ℓ_2.

The crucial property of path-constructible graphs is simple to observe given Section 5.1: if they have low maximum degree and the paths P_i are sufficiently long (that is, of size $\Omega(\log n)$), then Lemma 5.3 implies that they can be embedded in expanders (and, in particular, in pseudorandom graphs, as done in [38]). As mentioned before, we now state the main lemma for finding our linking structure, which is a weaker version of [38, Proposition 4.4].

Lemma 5.5. *For all sufficiently large N, there exists a graph H with at most $N^{1.1}$ vertices and disjoint sets $A, B \subseteq V(H)$ such that H is an (A, B)-linking structure and the following hold.*

1. *$|A| = |B| = N$ and $A \cup B$ is an independent set in H.*

2. *$\Delta(H) \leq 4$.*

3. *H is $(A \cup B)$-path-constructible with paths of length between $10 \log N$ and $40 \log N$.*

In order to obtain their stronger version of Lemma 5.5, the authors in [38] rely on a result concerning sorting networks with optimal depth. Earlier uses of sorting networks in extremal combinatorics include in the work of Kühn, Lapinskas, Osthus and Patel [49] on Hamilton cycles in highly connected tournaments and in the work of Müyesser and Pokrovskiy [54] giving, among other results, a combinatorial proof of the Hall-Paige conjecture. We will define directly the graph theoretic counterpart of a sorting network, for convenience calling this itself a *sorting network*. For a more detailed discussion concerning sorting networks and their graph theoretic counterpart see, for example, [38, Section 4].

Definition 5.6. A graph G is an (N, ℓ)-*sorting network* if there exist $A, B \subseteq V(G)$ for which G is an (A, B)-linking structure and disjoint sets $V_0 = A, V_1, \ldots, V_{\ell-1}, V_\ell = B$ with the following properties.

1. $V(G) = V_0 \cup \ldots \cup V_\ell$.

2. For every $0 \leq i \leq \ell$, V_i is an independent set in G with $|V_i| = |A| = |B| = N$.

3. For every $0 \leq i \leq \ell$, the bipartite graph $G[V_i, V_{i+1}]$ is the disjoint union of $K_{2,2}$'s and edges.

Sorting networks which are efficient enough for our purposes are simple to construct, and results as early as 1959 ([59]) imply that there exist $(N, O(\log^2 N))$-sorting networks. Moreover, in 1983, Ajtai, Komlós and Szemerédi [1] showed that, for each N, there exist (N, ℓ)-sorting networks with $\ell = O(\log N)$, where ℓ can easily seen to be optimal up to a constant multiple. The sorting network in the above definition is not necessarily the linking structure we require for Lemma 5.5 – although it has a low maximum degree, it might not be $(A \cup B)$-path-constructible. As observed in [38], it is possible to 'transform' a sorting network into an $(A \cup B)$-path-constructible linking structure. Briefly, to do this, start by taking the graph G with the properties above with $|A| = |B| = N$ and $\ell = O(\log N)$. Then, for each $0 \leq i < \ell$, take the bipartite graph $G[V_i, V_{i+1}]$ and, using part 3 of the definition above, substitute each single edge by a path of an appropriately chosen length m and each $K_{2,2}$ with vertices $a_1, a_2 \in V_i$ and $b_1, b_2 \in V_{i+1}$ by the $(2m)$-vertex graphs in Figure 4.
Notice that the graph in Figure 4 is an $(\{a_1, a_2\}, \{b_1, b_2\})$-linking structure. Indeed, the two horizontal paths are vertex disjoint paths linking a_1 to b_1 and a_2 to b_2, and they span all the vertices, and the blue and red paths are vertex disjoint paths linking a_1 to b_2 and a_2 to b_1, and they span all the vertices. Since each of these graphs is a linking structure, substituting each $K_{2,2}$ in each $G[V_i, V_{i+1}]$ with one of these graphs and each edge in each $G[V_i, V_{i+1}]$ with a path, maintains that G is an (A, B)-linking structure.

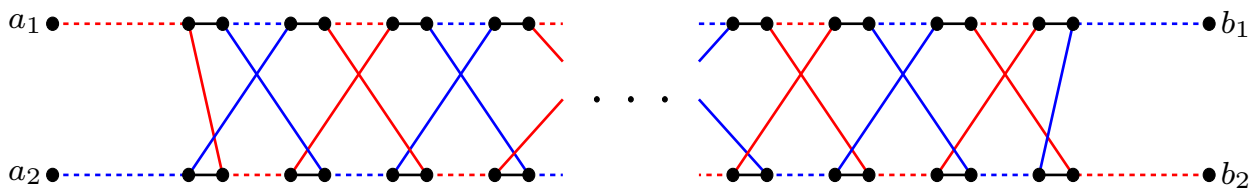

Figure 4: Each $K_{2,2}$ in the sorting network with parts $\{a_1, a_2\}$ and $\{b_1, b_2\}$ is substituted with a graph of the type above. Each full line represents an edge and each dotted line represents a path of an appropriately chosen length. The pattern in the figure is such that we can always construct such a graph with any number of marked vertices (i.e., those drawn with a circle in the figure) which is sufficiently large and divisible by 4. Note that the union of the full lines is simply an even cycle.

The crucial property of the graph in Figure 4 is that it is also $\{a_1, a_2\}$-path-constructible (as well as $\{a_1, a_2, b_1, b_2\}$-path-constructible). Indeed, it can be constructed by first taking a_1 and adding its incident path represented by a dotted line, then, constructing the even cycle in the middle formed by the full lines (which represent edges), and, finally, adding all of the remaining paths represented by dotted lines. Furthermore, substituting each $K_{2,2}$ in each $G[V_i, V_{i+1}]$ with one of the graphs in Figure 4 and each edge in each $G[V_i, V_{i+1}]$ with a path, transforms G into an $A \cup B$-path-constructible graph.

We can now complete our sketch proof of Lemma 5.5 by taking the illustrated graph to be such that it has between $10 \log N$ and $40 \log N$ marked vertices (which as is explained in the caption of the figure is possible) and the dotted lines to represent equally-sized paths with length also between $10 \log N$ and $40 \log N$. We then have that the illustrated graph has $2m = \Theta(\log^2 N)$ vertices and it is $\{a_1, a_2\}$-path-constructible (and $\{a_1, a_2, b_1, b_2\}$-path-constructible) with paths of length between $10 \log N$ and $40 \log N$. Then, substituting each $K_{2,2}$ in each $G[V_i, V_{i+1}]$ with one of these illustrated graphs and each edge in each $G[V_i, V_{i+1}]$ with a path of length m makes it so that the resulting graph is an (A, B)-linking structure with $O(N \log^3 N) \leq N^{1.1}$ vertices and is $(A \cup B)$-path-constructible with paths of length between $10 \log N$ and $40 \log N$. Thus, it satisfies Lemma 5.5.

Using Lemma 5.5, Lemma 5.3 and Definition 5.2, we can now prove Lemma 5.1.

Proof of Lemma 5.1. First, we will find $A, B \subseteq V(G)$ and a subgraph $H \subseteq G$ such that H is an (A, B)-linking structure with the following properties.

1. $|A| = |B| = n^{0.9}$.
2. $|H| \leq n/100$.
3. $\Delta(H) \leq 4$.
4. H is $(C/50, n/2C)$-extendable.

We will then set $X := V(H)$. Let then $N := n^{0.9}$ and apply Lemma 5.5 to find a linking structure with the given properties. Now we embed H in G so that it is extendable. First, we embed $H[A \cup B]$ in G such that $H[A \cup B]$ is $(C/50, n/2C)$-extendable. For this, note that Lemma 5.3 implies that there exists a path P in G of length $|A| + |B|$ which is $(C/50, n/2C)$-extendable. Indeed, we can start with a $(C/50, n/2C)$-extendable edge xy in G (since G is a C-expander, every edge is $(C/50, n/2C)$-extendable) and then apply the lemma with $\ell = |A| + |B|$. Since P is $(C/50, n/2C)$-extendable, the subgraph on $V(P)$ formed by removing all edges (i.e., the empty subgraph with vertex set $V(P)$) is also $(C/50, n/2C)$-extendable and hence, we can partition P into two sets A, B of equal size and we have an embedding of $H[A \cup B]$ in G which is $(C/50, n/2C)$-extendable.

Since H is $A \cup B$-path-constructible, there exists a sequence of paths $P_1, \ldots, P_t$ and graphs $H_0 := H[A \cup B], H_1, \ldots, H_t := H$ such that for each i, the following hold: $H_{i+1} = H_i \cup P_{i+1}$; and P_{i+1} is a path whose internal vertices are not in H_i and with at least one endpoint in H_i. Furthermore, each path P_i is of length between $10 \log N$ and $40 \log N$. Since $|V(H)| \leq N^{1.1} \leq n/100 \leq n - n/10 - 40 \log n$, we can then iteratively apply Lemma 5.3 embedding the paths P_i one by one while maintaining the $(C/50, n/2C)$-extendability property. At the end, we have $H \subseteq G$ satisfying all desired properties.

We now set $X := V(H)$ and note that the definition of a $(C/50, n/2C)$-extendable subgraph implies that $G[V(G) \setminus X]$ and $G[(V(G) \setminus X) \cup A \cup B]$ are C'-expanders for $C' = C/100$. Indeed, by definition, for every $S \subseteq V(G)$ of size at most n/C we have

$$|\Gamma(S) \setminus V(H)| \geq (C/50 - 1)|S| - \sum_{u \in S \cap V(H)} (d_H(u) - 1) \geq C|S|/50 - 5|S|,$$

which implies that S satisfies $|N_{G-X}(S)| \geq C|S|/50 - 5|S| - |S| \geq C|S|/100 = C'|S|$.

By repeated application of Lemma 5.3, we can then construct a linear forest $\mathcal{F}$ with $|A| = |B|$ equally-sized paths whose endpoints are in $A \cup B$, interior vertices are in $V(G) \setminus X$, and which have length $n^{0.1}/5$, such that $H \cup \mathcal{F}$ is a $(C/50, n/2C)$-extendable subgraph of G. Let $Y \cup A \cup B$ be the set of vertices spanned by $\mathcal{F}$, where Y is disjoint to X. Finally, let Z denote the rest of the vertices. Since $H \cup \mathcal{F}$ is $(C/50, n/2C)$-extendable, by the same argument as above we have that $G[Z]$ is a C'-expander for $C' = C/100$. Finally, since $(V(G) \setminus X) \cup A \cup B = Z \cup Y \cup A \cup B$, as mentioned above we also have that $G[Z \cup Y \cup A \cup B]$ is a $C/100$-expander. □

6 Proof of Theorem 1.4

Finally, we can put our work together to prove Theorem 1.4.

Proof of Theorem 1.4. Let G be a C-expander and apply Lemma 5.1 to find X, Y, Z with the stated properties. The following holds due to the properties of the linking structure in $G[X]$ and is crucial to observe. If there exists a spanning linear forest $\mathcal{F}$ of $G' := G - (X \setminus (A \cup B))$ with no isolated vertices and such that $\mathrm{End}(\mathcal{F}) = A \cup B$, then G contains a Hamilton cycle (see Figure 2). More formally, we can relabel the vertices so that the pairs of endpoints in $\mathcal{F}$ are $(a_1, b_1), (a_2, b_2), \ldots, (a_t, b_t)$ for some $t \leq |A|$, as well as the pairs $(a_{t+1}, a_{t+2}), (a_{t+3}, a_{t+4}) \ldots, (a_{|A|-1}, a_{|A|})$ and $(b_{t+1}, b_{t+2}), (b_{t+3}, b_{t+4}) \ldots, (b_{|A|-1}, b_{|A|})$, then use the linking structure to find paths which span $G[X]$, such that the pairs of endpoints are (a_{i+1}, b_i) for all i, with indices taken modulo $|A|$.

All we need now is to find such an $\mathcal{F}$. First apply Lemma 4.1 to $G[Z]$, which is an expander by the property from Lemma 5.1, giving a spanning linear forest in $G[Z]$ with at most $n^{0.8}$ paths and no isolated vertices. Now, we define the linear forest $\mathcal{F}_0$ in G' to be the union of two linear forests: the first is the linear forest in $G[Y \cup A \cup B]$ with endpoints in $A \cup B$ given by Lemma 5.1, and the second is the linear forest found in $G[Z]$ with at most $n^{0.8}$ paths. Let us denote the paths in the first forest by $P_1, \ldots, P_t$ (with $t = n^{0.9}$) and the second linear forest as $\mathcal{H}$. Recall that the paths $P_1, \ldots, P_t$ have size $n^{0.1}/5$.

To find the linear forest $\mathcal{F}$, we now apply Lemma 3.7 repeatedly as follows, starting with $\mathcal{F}_0$. At each step, we take the current linear forest $\mathcal{F}_i$, and if there are at least $t/2 = n^{0.9}/2$ paths P_j (so that they span at least $0.1n$ vertices) which are still in $\mathcal{F}_i$, we do the following (maintaining the invariants that there are no isolated vertices in $\mathcal{F}_i$ and that $A \cup B \subseteq \mathrm{End}(\mathcal{F}_i)$):

- If $|\mathrm{End}(\mathcal{F}_i)| \geq |A \cup B| + 1$, then note that since every path in $\mathcal{F}_i$ has two endpoints, the total number of endpoints is even. Since $|A \cup B|$ is even, then there are at least two endpoints outside of this set, call them x and y. By applying Lemma 3.7 with x, y and $\mathcal{F} := \mathcal{F}_i$ we obtain a forest $\mathcal{F}_{i+1}$ such that $\mathrm{End}(\mathcal{F}_{i+1}) = \mathrm{End}(\mathcal{F}_i) \setminus \{x, y\}$ and $|E(\mathcal{F}_{i+1}) \Delta E(\mathcal{F}_i)| = O(\log n)$.

- Otherwise, $\mathcal{F} := \mathcal{F}_i$ is the desired linear forest and we finish the process.

Now, clearly, we can have only at most $2|\mathrm{End}(\mathcal{F}_0) \setminus A \cup B| = 2|\mathcal{H}| \leq 2n^{0.8}$ steps in this process. Moreover, it is possible to perform each step because at each step we have $|E(\mathcal{F}_{i+1})\Delta E(\mathcal{F}_i)| = O(\log n)$, and therefore $O(\log n)$ paths P_i are changed for the next linear forest. This implies that, at each step, $\mathcal{F}_i$ still contains at least $t - O(i \log n) \geq t - O(n^{0.8} \log n) \geq t/2$ paths P_j. Thus, we can find such an $\mathcal{F}$ at the end of the process and, hence, a Hamilton cycle. □

7 Concluding remarks

We have shown that every C-expander is Hamiltonian for large enough C. In fact a stronger statement holds that the graph is *Hamilton-connected*, i.e. that between every two vertices there exists a Hamilton path. This result is especially interesting for applications, as oftentimes one needs to complete the embedding of a structure by constructing a spanning path between two specified vertices in an expanding subgraph of the host graph (see, for example, [23, 35, 37, 53]). We will only comment on the parts of the proof which have to be changed to prove the following result.

Theorem 7.1. *For every sufficiently large $C > 0$, every C-expander is Hamilton-connected.*

Proof sketch. Given a pair of vertices x, y in a C-expander G, we wish to find a Hamilton cycle whose endpoints are x, y. To this end, we may assume that G contains the edge xy and then show that G has a Hamilton cycle which contains xy. Indeed, if G does not contain xy, then we can simply add it to G and the resulting graph will still be a C-expander.

The only change needed in our proofs to yield this is the following. In Lemma 5.1, we can also require that one of the paths which in $G[Y \cup A \cup B]$ contains the edge xy. This is a straightforward application of the extendability method.

Note that the only place where we might change the paths in $G[Y \cup A \cup B]$ is in the proof of Theorem 1.4, by applying Lemma 3.7. But if we modify Lemma 3.7 slightly so that the rotations never break the edge xy, then the resulting structure obtained in Theorem 1.4 is a Hamilton path between x and y. And it is indeed not a problem to always avoid a small absolute constant number of vertices when performing rotations, as we have seen many times in the proof of Theorem 1.4. □

As mentioned in the outline, Hamiltonicity is an NP-complete problem. Nevertheless, our proof yields an efficient (polynomial time) algorithm for finding a Hamilton cycle in spectral expanders with a modest bound on the spectral ratio. This complements the classic result by Bollobás, Fenner, and Frieze [12] for random graphs at the Hamiltonicity threshold.

Theorem 7.2. *There exists $C > 0$ such that $\frac{d}{\lambda} > C$ implies that every (n, d, λ)-graph is Hamiltonian, and the Hamilton cycle can be found in polynomial time.*

Proof sketch. We will discuss how each relevant part of the proof can be made algorithmic. First, in Section 2, Lemma 2.5 is currently not algorithmic, but in the case that the graph is an (n, d, λ)-graph, having a fraction of the degrees of each vertex inside of a set already implies expansion, so the cleaning procedure can be done efficiently. Indeed, we can simplify the current proof by constructing the sets $B_{i,j}$ by always adding one vertex v at a time from U_i to $B_{i,j}$ if v has a small number of neighbours in $\mathrm{int}_{\mathcal{F}}\left(U_j \setminus \bigcup_{t=1}^{4} B_{j,t}\right)$. Now, no $B_{i,j}$ cannot become too large, as this would violate the Expander mixing lemma (see, e.g., Theorem 2.11 in [47]).

Every result in Section 3 can also be implemented in polynomial time, and this is not hard to see that performing the relevant rotations can be done efficiently with slightly careful bookkeeping.

The main result of Section 4 shows that in a C-expander there exists a linear forest with $n^{0.8}$ paths and no isolated vertices. The proof shows that a $<_{lex}$-minimal forest has those properties. Moreover, the

proofs are structured in such a way that they show that either a $<_{lex}$-minimal forest has a certain property, or it can be slightly changed to get a smaller forest in the $<_{lex}$-ordering to obtain a contradiction. To make the proofs algorithmic, we can simply start with an arbitrary linear forest, and if the current forest does not satisfy the required property, then we replace it with another forest in polynomial time. Since we always replace two paths of lengths a, b with another two paths with integer lengths c, d in the open interval (a, b) so that $a + b = c + d$, the sum of the squares of the lengths of the paths drops by at least 1. Since the sum of the squares is at most n^2, we need to perform the mentioned procedures only at most n^2 many times, resulting in a polynomial time algorithm.

For Section 5, note that we use the extendability method to find the required structures. Although the extendability method as quoted is based on a non-constructive result, and no constructive version is known for C-expanders, there is a version developed in [24] which works for robust expanders, and in particular for (n, d, λ)-graphs, which can be used to construct the linking structure with desired properties in polynomial time.

Section 6 applies the results from the previous sections only at most a linear number of times, so we only need polynomially many steps in total. □

Note added to proof: Simultaneously with the present paper, Ferber, Han, Mao, and Vershynin posted the arXiv preprint [26], where they used tools from random matrix theory to show that (n, d, λ)-graphs with $d/\lambda > C$ are Hamiltonian if $d > \log^{10} n$.

References

[1] M. Ajtai, J. Komlós, and E. Szemerédi. An $O(n \log n)$ sorting network. In *Proceedings of the fifteenth annual ACM symposium on Theory of computing*, pages 1–9, 1983.

[2] M. Ajtai, J. Komlós, and E. Szemerédi. First occurrence of Hamilton cycles in random graphs. *North-Holland Math. Stud.*, 115(C):173–178, 1985.

[3] S. Akbari, O. Etesami, H. Mahini, and M. Mahmoody. On rainbow cycles in edge colored complete graphs. *Australas. J. Combin.*, 37:33, 2007.

[4] P. Allen, J. Böttcher, Y. K. Hàn, Hiep, and Y. Person. Powers of Hamilton cycles in pseudorandom graphs. *Combinatorica*, 37:573–616, 2017.

[5] N. Alon and J. Bourgain. Additive patterns in multiplicative subgroups. *Geom. Funct. Anal.*, 24:721–739, 2014.

[6] N. Alon and A. Nussboim. K-wise independent random graphs. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 813–822. IEEE, 2008.

[7] N. Alon and Y. Roichman. Random Cayley graphs and expanders. *Random Structures & Algorithms*, 5(2):271–284, 1994.

[8] L. Babai. Long cycles in vertex-transitive graphs. *J. Graph Theory*, 3(3):301–304, 1979.

[9] B. Bollobas, T. I. Fenner, and A. M. Frieze, "An algorithm for finding Hamilton paths and cycles in random graphs," *Combinatorica*, vol. 7, pp. 327–341, 1987, Springer.

[10] I. Balla, A. Pokrovskiy, and B. Sudakov. A remark on Hamilton cycles with few colors. *Moscow Journal of Combinatorics and Number Theory*, 7(3):73–77, 2017.

[11] B. Bollobás. The evolution of sparse graphs. In *Graph theory and combinatorics (Cambridge, 1983)*, pages 35–57, 1984.

[12] B. Bollobás, T. I. Fenner, and A. M. Frieze. An algorithm for finding Hamilton paths and cycles in random graphs. *Combinatorica*, 7:327–341, 1987.

[13] S. Brandt, H. Broersma, R. Diestel, and M. Kriesell. Global connectivity and expansion: long cycles and factors in f-connected graphs. *Combinatorica*, 26:17–36, 2006.

[14] D. Christofides, J. Hladký, and A. Máthé. Hamilton cycles in dense vertex-transitive graphs. *J. Combin. Theory Ser. B*, 109:34–72, 2014.

[15] V. Chvátal. On Hamilton's ideals. *Journal of Combinatorial Theory, Series B*, 12(2):163–168, 1972.

[16] V. Chvátal and P. Erdős. A note on Hamiltonian circuits. *Discrete Math.*, 2(2):111–113, 1972.

[17] D. Clemens, A. Ferber, M. Krivelevich, and A. Liebenau. Fast strategies in Maker-Breaker games played on random boards. *Combinatorics, Probability and Computing*, 21(6):897–915, 2012.

[18] C. Cooper, A. Frieze, and B. Reed. Random regular graphs of non-constant degree: connectivity and Hamiltonicity. *Combin. Probab. Comput.*, 11(3):249–261, 2002.

[19] B. Csaba, D. Kühn, A. Lo, D. Osthus, and A. Treglown. Proof of the 1-factorization and Hamilton decomposition conjectures. *Mem. Amer. Math. Soc.*, 244(monograph 1154):164, 2016.

[20] B. Cuckler and J. Kahn. Hamiltonian cycles in Dirac graphs. *Combinatorica*, 29:299–326, 2009.

[21] M. DeVos. Longer cycles in vertex transitive graphs. arXiv:2302.04255, 2023.

[22] G. A. Dirac. Some theorems on abstract graphs. *Proc. Lond. Math. Soc.*, 3(1):69–81, 1952.

[23] N. Draganić, S. Glock, D. Munhá Correia, and B. Sudakov. Optimal Hamilton covers and linear arboricity for random graphs. *arXiv preprint arXiv:2310.11580*, 2023.

[24] N. Draganić, M. Krivelevich, and R. Nenadov. Rolling backwards can move you forward: on embedding problems in sparse expanders. *Trans. Amer. Math. Soc.*, 375(7):5195–5216, 2022.

[25] A. Ferber, R. Glebov, M. Krivelevich, and A. Naor. Biased games on random boards. *Random Structures & Algorithms*, 46(4):651–676, 2015.

[26] A. Ferber, J. Han. D. Mao, and R. Vershynin. Hamiltonicity of sparse pseudorandom graphs. *arXiv:2402.06177*, 2024.

[27] A. Ferber and E. Long. Packing and counting arbitrary Hamilton cycles in random digraphs. *Random Structures & Algorithms*, 54(3):499–514, 2019.

[28] A. Ferber, E. Long, and B. Sudakov. Counting Hamilton decompositions of oriented graphs. *Int. Math. Res. Not. IMRN*, (22):6908–6933, 2018.

[29] J. Friedman and N. Pippenger. Expanding graphs contain all small trees. *Combinatorica*, 7:71–76, 1987.

[30] A. Frieze, M. Krivelevich, P. Michaeli, and R. Peled. On the trace of random walks on random graphs. *Proceedings of the London Mathematical Society*, 116(4):847–877, 2018.

[31] A. Frieze, S. Vempala, and J. Vera. Logconcave random graphs. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 779–788, 2008.

[32] S. Glock, D. Munhá Correia, and B. Sudakov. Hamilton cycles in pseudorandom graphs. *arXiv preprint arXiv:2303.05356*, 2023.

[33] R. J. Gould. Recent advances on the Hamiltonian problem: Survey iii. *Graphs Combin.*, 30:1–46, 2014.

[34] D. Hefetz, M. Krivelevich, M. Stojaković, and T. Szabó. *Positional games*, volume 44. Springer, 2014.

[35] D. Hefetz, M. Krivelevich, and T. Szabó. Hamilton cycles in highly connected and expanding graphs. *Combinatorica*, 29(5):547–568, 2009.

[36] D. Hefetz, M. Krivelevich, and T. Szabó. Sharp threshold for the appearance of certain spanning trees in random graphs. *Random Structures & Algorithms*, 41(4):391–412, 2012.

[37] D. Hefetz, D. Kühn, J. Lapinskas, and D. Osthus. Optimal covers with Hamilton cycles in random graphs. *Combinatorica*, 34(5):573–596, 2014.

[38] J. Hyde, N. Morrison, A. Müyesser, and M. Pavez-Signé. Spanning trees in pseudorandom graphs via sorting networks. *arXiv preprint arXiv:2311.03185*, 2023.

[39] B. Jackson. Hamilton cycles in regular 2-connected graphs. *Journal of Combinatorial Theory, Series B*, 29(1):27–46, 1980.

[40] D. Johannsen, M. Krivelevich, and W. Samotij. Expanders are universal for the class of all spanning trees. *Combinatorics, Probability and Computing*, 22(2):253–281, 2013.

[41] J. Komlós and E. Szemerédi. Limit distribution for the existence of Hamiltonian cycles in a random graph. *Discrete Math.*, 43(1):55–63, 1983.

[42] A. Korshunov. Solution of a problem of Erdős and Rényi on Hamilton cycles in non-oriented graphs. *Soviet Math Dokl*, 17:760–764, 1976.

[43] M. Krivelevich. The critical bias for the Hamiltonicity game is $(1+o(1))$. *J. Amer. Math. Soc.*, 24(1):125–131, 2011.

[44] M. Krivelevich. On the number of Hamilton cycles in pseudo-random graphs. *arXiv preprint arXiv:1111.6261*, 2011.

[45] M. Krivelevich, C. Lee, and B. Sudakov. Robust Hamiltonicity of Dirac graphs. *Trans. Amer. Math. Soc.*, 366(6):3095–3130, 2014.

[46] M. Krivelevich and B. Sudakov. Sparse pseudo-random graphs are Hamiltonian. *J. Graph Theory*, 42(1):17–33, 2003.

[47] M. Krivelevich and B. Sudakov. Pseudo-random graphs. In *More sets, graphs and numbers: A Salute to Vera Sos and András Hajnal*, pages 199–262. Springer, 2006.

[48] M. Krivelevich, B. Sudakov, V. H. Vu, and N. C. Wormald. Random regular graphs of high degree. *Random Structures Algorithms*, 18(4):346–363, 2001.

[49] D. Kühn, J. Lapinskas, D. Osthus, and V. Patel. Proof of a conjecture of Thomassen on Hamilton cycles in highly connected tournaments. *Proceedings of the London Mathematical Society*, 109(3):733–762, 2014.

[50] D. Kühn and D. Osthus. Hamilton decompositions of regular expanders: a proof of Kelly's conjecture for large tournaments. *Adv. Math.*, 237:62–146, 2013.

[51] D. Kühn and D. Osthus. Hamilton cycles in graphs and hypergraphs: an extremal perspective. In *Proceedings of the International Congress of Mathematicians, Seoul, Korea.*, volume 4, pages 381–406, 2014.

[52] L. Lovász. Combinatorial structures and their applications. In *Proc. Calgary Internat. Conf., Calgary, Alberta*, pages 243–246, 1970.

[53] R. Montgomery. Spanning trees in random graphs. *Advances in Mathematics*, 356:106793, 2019.

[54] A. Müyesser and A. Pokrovskiy. A random Hall-Paige conjecture. *arXiv preprint arXiv:2204.09666*, 2022.

[55] C. S. J. Nash-Williams. Edge-disjoint Hamiltonian circuits in graphs with vertices of large valency. In *Studies in Pure Mathematics (Presented to Richard Rado)*, pages 157–183. Academic Press London, 1971.

[56] I. Pak and R. Radoičić. Hamiltonian paths in Cayley graphs. *Discrete Math.*, 309(17):5501–5508, 2009.

[57] L. Pósa. Hamiltonian circuits in random graphs. *Discrete Math.*, 14(4):359–364, 1976.

[58] E. Rapaport Strasser. Cayley color groups and Hamilton lines. *Scripta Math.*, 24:51–58, 1959.

[59] D. L. Shell. A high-speed sorting procedure. *Communications of the ACM*, 2(7):30–32, 1959.

[60] A. Thomason. Pseudo-random graphs. In *Proceedings of Random Graphs, Poznań 1985, M. Karoński, ed., Annals of Discrete Math. 33 (North Holland 1987)*, volume 144, pages 307–331.

[61] A. Thomason. Random graphs, strongly regular graphs and pseudorandom graphs. *Surveys in Combinatorics*, 123(173-195):1, 1987.