Towards the Erdős-Gallai Cycle Decomposition Conjecture

Matija Bucić[*]   Richard Montgomery[†]

Abstract

In the 1960's, Erdős and Gallai conjectured that the edges of any n-vertex graph can be decomposed into $O(n)$ cycles and edges. We improve upon the previous best bound of $O(n \log \log n)$ cycles and edges due to Conlon, Fox and Sudakov, by showing an n-vertex graph can always be decomposed into $O(n \log^\star n)$ cycles and edges, where $\log^\star n$ is the iterated logarithm function.

1   Introduction

When is it possible to decompose a graph into edge disjoint subgraphs with certain properties? Many classical problems in extremal combinatorics fall within this framework and its natural hypergraph generalisation, while decomposition problems have strong links to many other fields, including the design of experiments, coding theory, complexity theory and distributed computing (see, for example, [3, 10, 32]). The particular case where we seek to decompose a graph into cycles has a long history, dating back to the 18th century and Euler's result on the existence of Euler tours. As Veblen [53, 54] observed for his algebraic approach to the Four-Colour Theorem, Euler's result immediately implies that any graph with even vertex degrees (i.e., any *Eulerian* graph) has a decomposition into cycles. As it is immediate that any graph with a vertex of odd degree cannot be decomposed into cycles, this exactly characterises which graphs have cycle decompositions.

Another very classical cycle decomposition result is due to Walecki [44] from 1892, who showed it is possible to decompose any complete graph with an odd number of vertices into Hamilton cycles. This gives a cycle decomposition into few cycles, indeed, into optimally few cycles. This raises a very natural question of whether every Eulerian graph has a cycle decomposition into few cycles? That only $O(n)$ cycles might be needed to decompose any n-vertex Eulerian graph is easily seen to be equivalent to the following classical conjecture of Erdős and Gallai [18] dating back to the 1960's, which is one of the major open problems on graph decompositions.

Conjecture 1. *Any n-vertex graph can be decomposed into $O(n)$ cycles and edges.*

While Conjecture 1 is equivalent to conjecturing that every n-vertex Eulerian graph can be decomposed into $O(n)$ cycles, as noted above, if they both hold then the optimal implicit constants in these conjectures seem likely to be different. For the Eulerian problem, Hajós conjectured that $\frac{n}{2}$ cycles should be sufficient [43] (see also [6, 9, 12, 20, 23]), while the best known lower bound for the number of cycles and edges required in Conjecture 1 is $(\frac{3}{2} - o(1))n$, as observed by Erdős in 1983 [17], improving on a previous construction of Gallai [18] (see Section 6).

Since its formulation, the Erdős-Gallai Conjecture has often been highlighted (see, for example, [6, 11, 23, 25, 36, 48, 49]), with Erdős himself mentioning it in many of his open problem collections [14, 15, 16, 17]. Despite this attention, and a lot of work on related problems over the years, direct progress towards the Erdős-Gallai Conjecture has only been made within the last decade. The previous related results, which we discuss first, are mostly on the analogous path decomposition problem and the covering version of the Erdős-Gallai conjecture.

Path decompositions. In the 1960's, Gallai [43] posed the analogous path decomposition version of Conjecture 1. In particular, he conjectured that any connected n-vertex graph can be decomposed into at most $\frac{n+1}{2}$ paths. Lovász [43] in 1968 proved that any graph can be decomposed into at most $n-1$ paths. This follows easily from his complete solution to the problem of how many paths *or* cycles one needs to decompose an n-vertex graph, to which the answer is $\lfloor \frac{n}{2} \rfloor$. Currently the best general bound in the path decomposition problem is due independently to Dean and Kouider [13] and Yan [56], who showed that any graph can be decomposed into at most $\lfloor \frac{2}{3}n \rfloor$ paths. Gallai's path decomposition conjecture is known to hold for quite a few special classes of graphs, with connected planar graphs being the most recent addition to the list. This latest result is due to Blanché, Bonamy and Bonichon in [5], where a more exhaustive list of partial results can also be found.

Covering problems. Another interesting direction which has attracted a lot of attention is the covering version of Conjecture 1, in which we do not insist that the cycles we find should be disjoint, only that together they contain all the edges of the host graph. In 1985, Pyber [47] proved the covering version of the Erdős-Gallai conjecture, showing that the edges of any n-vertex graph can be covered with $n-1$ cycles and edges. The analogous covering version of Gallai's conjecture, raised by Chung [9] in 1980, has been settled first approximately by Pyber [49] in 1996 and then completely by Fan [19] in 2002, who showed that the edges of any connected graph can be covered by $\lceil \frac{n}{2} \rceil$ paths. The covering version of Hajós's conjecture was also solved by Fan [20], who showed that any n-vertex Eulerian graph can be covered by at most $\lfloor \frac{n-1}{2} \rfloor$ cycles, settling another conjecture of Chung. As with the other two covering results above, this bound is best possible.

Results on the Erdős-Gallai conjecture. In more recent years, the Erdős-Gallai conjecture (along with more accurate results on the implicit bounds) has been shown to hold for two large specific classes of graphs – random graphs and graphs with linear minimum degree. The conjecture was first established for a typical binomial random graph $G(n,p)$ (for any $p = p(n)$) by Conlon, Fox and Sudakov [11]. Korándi, Krivelevich, and Sudakov [36] found the correct leading constant here, showing that $(\frac{1}{4} + \frac{p}{2} + o(1))n$ cycles and edges are typically sufficient to decompose $G(n,p)$. For constant edge probability p, Glock, Kühn, and Osthus [25] were even able to determine with high probability the exact minimum number of cycles and edges required to decompose a (quasi)random graph. On the other hand, the Erdős-Gallai conjecture was first shown to hold for graphs with linear minimum degree again by Conlon, Fox and Sudakov [11]. Very recently, the asymptotically correct bound of $(\frac{3}{2} + o(1))n$ cycles and edges has been proved by Girão, Granet, Kühn, and Osthus [23] for large graphs with linear minimum degree.

A fundamental challenge towards establishing the Erdős-Gallai conjecture is its generality, and indeed these previous results make progress only by imposing a fairly strong constraint on the structure or randomness of the graph. For almost 50 years, the best known bound in the general case of the Erdős-Gallai conjecture (as observed by Erdős and Gallai) came from a simple argument involving the iterative removal of a longest cycle, which shows that an n-vertex graph can always be decomposed into $O(n \log n)$ cycles and edges. In 2014, Fox, Conlon and Sudakov [11] made the first major breakthrough on this problem, showing that such a decomposition with only $O(n \log \log n)$ cycles and edges always exists. Here we will give the following improvement on this bound, where $\log^{\star} n$ is the iterated logarithm function.

Theorem 2. *Any n-vertex graph can be decomposed into $O(n \log^{\star} n)$ cycles and edges.*

Key to the decompositions used by Conlon, Fox and Sudakov [11] was to show that *a)* graphs H with certain expansion properties can be decomposed into $O(|H|)$ cycles and few edges and *b)* any n-vertex graph G can be decomposed nicely into such 'expanders' H and a small number of leftover edges. Combined, this gives a decomposition of G into $O(n)$ cycles and some leftover edges, and it can be shown that iterating this on the leftover edges while removing any particularly long cycles causes the average degree of the leftover edges to drop significantly each time, so that after $\log \log n$ iterations the decomposition given by [11] is achieved.

To prove Theorem 2, essentially we need the average degree of the leftover edges to drop much faster, and so at *b)* we have to take a much weaker condition on the 'expanders' H. Effectively we replace the strong expansion used in [11], with a very weak *sublinear* expansion (as introduced by Komlós and Szemerédi [34, 35]), in particular using a *robust* sublinear expansion where sets expand sublinearly despite the additional removal of a possibly-superlinear set of edges (see Section 3.1 for a discussion of these forms of expansion and their background).

Using this much weaker form of expansion introduces a raft of issues when we decompose an expander into few cycles and edges (for *a)* above), resulting in a very different approach to that used in [11]. In order to do this, we introduce a

range of new tools, which we hope will find further applications. In particular, we would highlight a new approach to robust sublinear expansion (see Section 3) and the (surprisingly difficult) result that randomly sampling the vertices of an expander is likely to induce a subgraph with a (somewhat weaker) expansion property (see Lemma 19). Additional new tools include a similar result but while randomly sampling edges, the (almost) decomposition of any graph into robust sublinear expanders, and the finding of a sparse 'connective skeleton' in expanders to connect vertex pairs with paths. These, and other tools, and how they come together to prove Theorem 2, are discussed in Section 2.2.

As discussed in Section 3.1, sublinear expansion has been useful in many different settings in which our tools may also be useful (see Section 6 for some examples). In particular, *robust* sublinear expansion (specifically considering the deletion of superlinearly many edges) is a very recent concept and we hope our new perspective and tools will contribute to its development and use. Several of our intermediate results and tools might also ultimately prove useful towards proving the Erdős-Gallai conjecture in full as they often decompose any n-vertex graph into $O(n)$ cycles and a graph with some other structure imposed. This is discussed further in our concluding remarks in Section 6.

2 Preliminaries

After we introduce our notation, we give a detailed sketch of our methods before outlining the rest of the paper.

2.1 Notation

Given a graph G we will denote by $V(G)$ and $E(G)$ its vertex and edge set, respectively. Given a vertex $v \in V(G)$, we denote its degree by $d_G(v)$ and the set of its neighbours by $N_G(v)$. We write $\Delta(G)$ for the maximum degree of a vertex in a graph G. Given a subset of vertices $U \subseteq V(G)$ we denote by $N_G(U)$ the set of vertices in $V(G) \setminus U$ which have a neighbour in U. Given $U \subseteq V(G)$ we define $B_G^i(U)$ as the set of vertices at distance at most i from a vertex of U in the graph G, i.e. the ball of radius i around U in G, and write simply $B_G(U) = B_G^1(U)$. Given $V \subseteq V(G)$ we write $G[V]$ for the subgraph of G induced by the vertex set V, and write $G \setminus V$ for $G[V(G) \setminus V]$. Given $F \subseteq E(G)$ we write $G - F$ for the subgraph of G obtained by deleting all the edges in F. Given multiple (hyper)graphs $H_1, \ldots, H_t$ we write $H_1 \cup \ldots \cup H_t$ for the (hyper)graph with vertex set $\bigcup_{i \in [t]} V(H_i)$ and edge set $\bigcup_{i \in [t]} E(H_i)$. Given vertices v and u, by a vu-path/walk we refer to a path/walk joining v and u.

We write $X \sim \text{Bin}(n, p)$ to mean that X is a random variable distributed according to the binomial distribution with parameters n and p. We denote by $\mathcal{G}(n, p)$ the binomial random graph defined as the graph with vertex set $[n]$ in which we sample every edge with probability p independently from all other edges. We write $G \sim \mathcal{G}(n, p)$ to mean that G is sampled according to $\mathcal{G}(n, p)$.

All our logarithms have base two. For each $k \geq 1$, let $\log^{[k]}(n) = \underbrace{\log \log \ldots \log n}_{k \text{ times}}$, and let $\log^{[0]} n = n$. The iterated logarithm function $\log^{\star} n$ is the minimum number of times we need to apply the logarithm function to n until it becomes at most one, that is, the least $k \geq 0$ such that $\log^{[k]} n \leq 1$.

Throughout the paper we make no attempt to optimise constants and logarithmic factors; often we are wasteful to improve readability. With the same goal, we also omit floor and ceiling signs wherever they are not crucial.

2.2 Proof sketch

Our methods to find an (edge) decomposition into cycles and edges is iterative, where a single iteration, applied to an n-vertex graph G with average degree d, performs the following steps for some appropriately large constant C.

- Repeatedly remove any cycle of length at least d and add it to the decomposition, giving at most $\frac{n}{2}$ new cycles.

- Decompose G into edge disjoint subgraphs R_i with a certain expansion property which, combined with the lack of long cycles, guarantees that $|R_i| = O(d \log^4 d)$. These subgraphs are almost vertex disjoint, so that $\sum_i |R_i| \leq 2n$.

- Decompose each of the subgraphs R_i into cycles and edges, using in total $O(n)$ cycles and $O(n \log^C d)$ edges.

3

This finds a decomposition of G into $O(n)$ cycles and a subgraph consisting of leftover edges of average degree at most $O(\log^C d)$. We now iterate by applying the same argument to our much sparser graph consisting of leftover edges. After at most $O(\log^\star d)$ iterations we will be left with a graph of constant average degree. We then simply make all its edges part of our decomposition, which together with the $O(n)$ cycles we found at each of the $O(\log^\star d)$ iterations gives our desired decomposition for Theorem 2 – in fact, using only $O(n \log^\star d)$ edges for any n-vertex graph with average degree d.

The majority of our work lies in carrying out the final step in this iteration, where the key part of this step is to show the following intermediate result.

Theorem 3. *There exists $C > 0$ such that any r-vertex graph decomposes into $O(r)$ cycles and $O(r \log^C r)$ edges.*

Note that this intermediate result is applied to graphs R_i, each of which has order at most $O(d \log^4 d)$, so that, as $\sum_i |R_i| \leq 2n$, we get in total $O(n)$ cycles and $O(\sum_i |R_i| \log^C |R_i|) = O(\sum_i |R_i| \log^C d) = O(n \log^C d)$ edges in total from the decomposition in the final step of the iteration.

Though we use a slightly modified iteration argument, till now our approach has the same structure as the one taken by Conlon, Fox and Sudakov in [11], where they prove a weaker version of Theorem 3 in which they allow, instead of $O(r \log^C r)$, up to $O(r^{2-1/10})$ edges in the decomposition. This much weaker bound leads them to iterate $O(\log \log d)$ times and thus use a decomposition using $O(n \log \log d)$ cycles and edges in total. The key difference is that we are able to replace the extremely strong expansion properties used in [11] with a very weak form of expansion. To make this change successfully, we need to carefully develop the weak expansion property we use (which originates with Komlós and Szemerédi [34, 35]) as well as solve the variety of problems caused by working with this weak expansion. For this development, the main insight is the new perspective we bring to robust sublinear expansion. That we can decompose an arbitrary graph into edges and expanders follows relatively naturally from the definition of robust sublinear expansion (see Section 3.3 for more details on the expander partitioning lemma), and the same decomposition result allows us to carry out the second step of the iteration mentioned above. This leaves the difficult task of decomposing an expander graph into few cycles and edges, which we now discuss.

Decomposing expanders into few cycles and edges. For some constant $C > 0$, we now assume that we wish to decompose an n-vertex graph G with the following (slightly simplified) expansion condition (see Section 3 for the full condition we use): for each $U \subseteq V(G)$ and $F \subseteq E(G)$ with $|U| \leq \frac{2}{3}n$ and $|F| \leq |U| \log^C n$, we have

$$|N_{G-F}(U)| \geq \frac{1}{\log^2 n} |U|. \tag{1}$$

As the size of the neighbourhood guaranteed in (1) is smaller than $|U|$, this type of expansion is known as *sublinear expansion*.

Roughly speaking, our main strategy is to set aside a 'sparse connecting skeleton' $H \subseteq G$, before initially decomposing the edges of $G - H$ into $O(n)$ paths and cycles using a result of Lovász [43] stated in the introduction. We then use short paths from the connecting skeleton H to connect up each initial path into a cycle, before simply taking each unused edge of H as part of our decomposition. In order for this to produce a correct decomposition we need H to be very sparse, in particular with at most $n \log^{O(1)} n$ edges.

To aid the connection of the paths from the initial path/cycle decomposition, we need that each vertex does not appear too often as an endvertex of these paths. This we ensure by proving a simple, but crucial, corollary of Lovász's result (see Corollary 22), which will allow us to decompose $G - H$ into a collection $\mathcal{P}$ of $O(n)$ paths in which each vertex appears as an endvertex at most twice, so that the endvertices of the paths are well spread across the graph.

More problematically, note that in order to get an actual cycle we need to connect the endvertices of each path $P \in \mathcal{P}$ using a path in H which is internally vertex disjoint from P. To deal with this, we change this outline slightly as follows. We partition $V(G) = V_1 \cup V_2 \cup V_3$ by placing each vertex independently into a set V_i uniformly at random, and show that G contains sparse subgraphs H_1, H_2, H_3 (each with $n \log^{O(1)} n$ edges) with the following property for each $i \in [3]$, where a *path through V_i* is one whose interior vertices are all in V_i. We place no restriction on the endvertices themselves, so in particular a single edge path is a path through any set since it contains no interior vertices.

4

P For any set $\mathcal{P} \subseteq \binom{V(G)}{2}$ such that each vertex appears in at most 2 pairs in $\mathcal{P}$, there are edge disjoint paths P_{xy}, $\{x, y\} \in \mathcal{P}$, such that, for each $\{x, y\} \in \mathcal{P}$, P_{xy} is an xy-path through V_i in H_i with length $O(\log^7 n)$.

We then split the edges of $G - H_1 - H_2 - H_3$ into three subgraphs G_1, G_2, G_3 in such a way that $V(G_i) = V_{i+1} \cup V_{i+2}$, for each $i \in [3]$, with indices taken modulo 3. Applying our path decomposition corollary to each G_i then gives a decomposition of all the edges outside of $H_1 \cup H_2 \cup H_3$ into paths whose endvertices are well spread across the graph. Note that the paths decomposing G_i completely avoid V_i so by using the property **P** we can connect each of these paths into actual cycles using edges of H_i. In total we find $O(n)$ edge disjoint cycles which use all the edges of $G - H_1 - H_2 - H_3$ so the total number of uncovered edges, which all belong to $H_1 \cup H_2 \cup H_3$, is small.

With such a weak expansion property as that at (1), whether we can do this is initially far from clear. Building up to this, we ask the following three questions.

i) Can we connect pairs of vertices with edge disjoint paths using the whole of G?
   I.e., does property **P** hold if $H_i = G$ and $V_i = V(G)$?

ii) If so, can we do this using only a random subset of vertices V_i for the interior vertices of the paths?
   I.e., does property **P** hold if $H_i = G$?

iii) If so, can we do this using only a sparse subgraph H_i of G?
   I.e., can we find a sparse $H_i \subseteq G$ so that property **P** holds for H_i?

i) Finding edge disjoint paths in G. The expansion condition on G at (1) is sufficient to imply that any pair of vertices in H are connected by a path of length $O(\log^3 n)$ by expanding the neighbourhoods around each of the two vertices until they become large enough that they must overlap. Moreover, the robustness of our condition at (1) (i.e., that this expansion can avoid using an arbitrary, but not too large, set of edges F) allows us, with only a bit more work, to show that, for any collection $\mathcal{P}$ of pairs of vertices as in property **P**, we could find at least $\Theta(\log^4 n) \cdot |\mathcal{P}|$ edge disjoint paths in G which each connect *some* vertex pair in $\mathcal{P}$ and have length $O(\log^3 n)$. As this holds in fact for any subset $\mathcal{P}' \subseteq \mathcal{P}$ in place of $\mathcal{P}$, this allows us to use the Aharoni-Haxell hypergraph matching theorem (see Theorem 6) to select, for each pair $\{x, y\} \in \mathcal{P}$, an xy-path in G, so that all these paths are edge disjoint. In total this allows us to answer question i) positively.

ii) Connecting through the random vertex subset V_i. With $i \in [3]$ and V_i a random subset of V with size approximately $\frac{n}{3}$ as chosen above, unfortunately it seems it does not follow easily that $H[V_i]$ likely satisfies a similar expansion property to (1). Indeed, firstly, if the inequality (1) is tight for a set U, then with probability $\exp(-\Theta(|U|/\log^2 n))$ we have that U has no neighbours selected into our random subset V_i, too high a probability to naively take a union bound to avoid this event over all sets U with any fixed size. Secondly, the robustness condition we need in order to find multiple edge disjoint paths (to then apply the Aharoni-Haxell hypergraph matching theorem) requires us to avoid an arbitrary set of $|U| \log^{O(1)} n$ edges, and again there are too many choices to just take a union bound.

The first problem here is the most difficult to overcome. The second problem can be overcome by splitting the edges of G randomly into t subgraphs G_i, $i \in [t]$, for some $t = \log^{O(1)} n$, and showing then that (with high probability) each of these has some (slightly weaker) expansion property (see Lemma 15). When we look for edge disjoint collections of paths in each graph G_i separately for the application of the Aharoni-Haxell hypergraph matching theorem, by finding $\frac{1}{t}$ fraction of the required paths in each graph G_i we need to avoid fewer edges by expanding in each G_i separately rather than G. Thus, we have fewer sets of edges over which to take a union bound, solving the second problem above.

Solving the first problem to get some weak expansion into V_i is the crux of this paper, and is where our new perspective on robust sublinear expansion is critical. We will show that it follows from this new perspective that, for such sets U and F, we have either

1. $N_{G-F}(U)$ is actually much larger than guaranteed by (1), or

2. there is a set $U' \subseteq U$ which is much smaller than U but whose neighbourhood alone contains at least $\frac{|U|}{\log^2 n}$ vertices in $N_{G-F}(U)$.

Given this, a natural approach is to take a union bound over all 'well-expanding' sets U, meaning that they fall under condition 1. above, to guarantee a constant fraction of their neighbourhood gets sampled into our random set V_i. Since the subset U' from condition 2. is well-expanding this will guarantee us that it expands inside V_i. We would now like to use the fact that U' expands inside V_i to conclude the same happens for the original set U from condition 2. above containing U'. However, a major issue here is that in order to achieve this we would need a bound on $|(N_{G-F}(U') \setminus U) \cap V_i|$ to get the expansion for *any* relevant set U for which we use the well-expanding subset U' – having to bound a random variable depending on U' and U spoils our union bound approach over the smaller sets U'. For the first expansion of U into V_i this is avoidable (by only considering such sets disjoint from V_i), but we need to expand multiple times to reach most of the vertices in V_i and, after the first expansion, avoiding expanding sets that contain vertices in V_i is unavoidable.

To get around this, when we identify the well-expanding set U', we look at its successive neighbourhoods in $G - F$ and show that enough of these vertices are chosen to be in V_i together with all the vertices along a path going back to U' so that U' expands via such short paths to reach more than half of the vertices of V_i. In an early draft of this work, this was shown by carefully analysing an intricate random process. Fortunately, however, we will instead give here a much easier proof by combining our new perspective on robust sublinear expansion with an adaptation of a clever application of the sprinkling method appearing in a very recent work of Tomon [52]. We defer a more detailed sketch for this part of the argument to Section 4, in particular until after we have introduced in full our new perspective on robust sublinear expansion, which remains crucial for this new approach.

In total, though, this will allow us to answer question ii) positively. I.e., property **P** holds if we are allowed to use all the edges of G to make connections through V_i. Let us also stress an important point, which already played a role at various points in the above arguments and that is that the paths P_{xy} we find will always be short, namely of length $\log^{O(1)} n$. This again plays an important role in answering the next question, namely finding an appropriate sparse subgraph $H_i \subseteq G$ with the same property, which we turn to next.

iii) Finding sparse connecting skeletons. Before we look for a subgraph $H_i \subseteq G$ with the property **P** and $n \log^{O(1)} n$ edges, can we even find any graph with these properties? A binomial random graph is a natural candidate for such a graph, and, indeed, if H is a binomial random graph with vertex set $V(G)$ and edge probability $p = \omega\left(\frac{\log n}{n}\right)$ then it will have, with high probability, the property **P** if we replace H_i with H and choose V_i to be any fixed set of linear size. (We prove this as Lemma 10, with a larger than optimal value of p for simplicity.) We then use H as a *template* to construct the sparse expanding skeleton H_i. We first sample our large random subset of vertices V_i, and use our answer to question ii) to guarantee that property **P** holds with high probability in a slightly stronger form where every vertex is allowed to appear in $O(\log^5 n)$ pairs in $\mathcal{P}$. In particular, we will use it with $\mathcal{P}$ being the set of pairs of vertices making an edge of our template graph H, which we choose to be sparse and well-connected through V_i, as well as have maximum degree $O(\log^5 n)$. For each edge $xy \in E(H)$, we find an xy-path P_{xy} through V_i with length $\log^{O(1)} n$ in G so that all these paths are edge disjoint. We then let H_i be the union of all these paths, noting that, as H is sparse and the paths P_{xy} are relatively short, H_i is also relatively sparse. Then, given an arbitrary collection $\mathcal{P}$ of pairs to connect, we first find edge disjoint paths connecting them through V_i in H, before replacing each edge xy on one of these paths in H with the corresponding path P_{xy} through V_i. This creates a set of edge disjoint xy-*walks* through V_i in H_i — as each such walk contains an xy-path, we can find the paths required by property **P**. Note that, when we do this for each $i \in [3]$, we need to ensure that the graphs H_i we find are edge disjoint, but this is easy to do by reusing some of our previous work, splitting G into a union of edge disjoint expanders G_1, G_2, G_3, before finding each subgraph H_i in the respective subgraph G_i.

2.3 Organisation of the paper

In the rest of this section we will introduce some general preliminary results, including some concentration results in Section 2.4 and the Aharoni-Haxell hypergraph matching theorem in Section 2.5, before showing strongly expanding graphs (namely $\mathcal{G}(n, p)$) satisfy a certain strong connectivity property in Section 2.6. In Section 3, we introduce robust sublinear expansion and prove a number of useful properties of this type of expansion. In Section 4, we establish that our weaker expansion implies a similar (though weaker) connectivity property as that used in Section 2.6. In Section 5, we use the machinery we developed to prove our main result, Theorem 2. Finally, in Section 6, we make some concluding remarks.

2.4 Concentration inequalities

We will often use a basic version of Chernoff's inequality for the binomial random variable (see, for example, [2]).

Theorem 4 (Chernoff's bound). *Let n be an integer and $0 \leq \delta, p \leq 1$. If $X \sim \mathrm{Bin}(n, p)$, then, setting $\mu = \mathbb{E}X = np$, we have*

$$\mathbb{P}(X > (1 + \delta)\mu) \leq e^{-\delta^2 \mu/2}, \qquad and \qquad \mathbb{P}(X < (1 - \delta)\mu) \leq e^{-\delta^2 \mu/3}.$$

We will also make use of the following well-known martingale concentration result (see Chapter 7 of [2]).

Lemma 5. *Suppose that $X : \prod_{i=1}^{N} \Omega_i \to \mathbb{R}$ is k-Lipschitz. Then, for each $t > 0$,*

$$\mathbb{P}(|X - \mathbb{E}X| > t) \leq 2 \exp\left(\frac{-2t^2}{k^2 N}\right).$$

2.5 The Aharoni-Haxell hypergraph matching theorem and edge disjoint paths

We will use the following hypergraph version of Hall's theorem due to Aharoni and Haxell, which is an immediate consequence of Corollary 1.2 in [1] (noting that we can add new, unique, vertices to each edge in the theorem to make the hypergraphs ℓ-uniform). A *matching* in a hypergraph is a collection of pairwise vertex disjoint edges.

Theorem 6. *Let $r \in \mathbb{N}$, and let $H_1, \ldots, H_r$ be a collection of hypergraphs with at most ℓ vertices in each edge. Suppose that, for each $I \subseteq [r]$, there is a matching in $\bigcup_{i \in I} H_i$ containing more than $\ell(|I| - 1)$ edges. Then, there is an injective function $f : [r] \to \bigcup_{i \in [r]} E(H_i)$ such that $f(i) \in E(H_i)$ for each $i \in [r]$ and $\{f(i) : i \in [r]\}$ is a matching of r edges.*

We will use Theorem 6 to find edge disjoint paths between vertex pairs, to show that a graph is well-connected under the following definition, recalling that a path *through* V is a path with all its internal vertices in V.

Definition 7. A graph G is (ℓ, t)-*path connected* through a vertex subset $V \subseteq V(G)$ if, for any $\mathcal{P} \subseteq \binom{V(G)}{2}$ in which every vertex appears in at most t pairs in $\mathcal{P}$, there are edge disjoint paths $P_{\{x,y\}}$, $\{x, y\} \in \mathcal{P}$, such that, for each $\{x, y\} \in \mathcal{P}$, $P_{\{x,y\}}$ is an xy-path through V with length at most ℓ.

We denote by $\binom{V(G)}{2}$ the multiset of pairs of distinct vertices of G, so in particular the same pair may appear multiple times in the collection $\mathcal{P}$. Typically, t will be a small constant and ℓ will be at most polylogarithmic in the number of vertices.

Given a graph G, a vertex set $V \subseteq V(G)$ and a collection $\mathcal{P} \subseteq \binom{V(G)}{2}$, we translate the pair connectivity property into a hypergraph matching problem as follows. For some $\ell \in \mathbb{N}$, and each $\{x, y\} \in \mathcal{P}$, let $H_{\{x,y\}}$ be the hypergraph with vertex set $E(G)$ and add as an edge the set $E(P)$ for each xy-path P in G with interior vertices in V and length at most ℓ. If there is an injective function $f : \mathcal{P} \to \bigcup_{\{x,y\} \in \mathcal{P}} E(H_{\{x,y\}})$ such that $f(\{x, y\}) \in E(H_{\{x,y\}})$ for each $\{x, y\} \in \mathcal{P}$ and $\{f(\{x, y\}) : \{x, y\} \in \mathcal{P}\}$ is a matching, then, for each $\{x, y\} \in \mathcal{P}$, let $P_{\{x,y\}}$ be the path in G with edge set $f(\{x, y\})$. By the definition of $H_{\{x,y\}}$, each path $P_{\{x,y\}}$ is an xy-path in G with length at most ℓ and interior vertices in V, and, as $\{f(\{x, y\}) : \{x, y\} \in \mathcal{P}\}$ is a matching, these paths are all edge disjoint. Therefore, in order to prove that G is (ℓ, t)-path connected through V, it suffices to take a general such collection $\mathcal{P}$, define the relevant hypergraphs $H_{\{x,y\}}$, and prove that the associated condition holds for an application of Theorem 6.

2.6 Existence of well-connected sparse graphs

Random graphs typically present a natural candidate for a sparse, well-connected, graph, and we use this to prove the existence of our template in Lemma 10 (using the connectivity property in Definition 7). Similar properties have been studied before in random graphs for various applications (see e.g. [7, 24, 29]), and our main lemma (Lemma 10) can be obtained as a (not quite immediate) corollary of Lemma 3.4. from [46] combined with a multi-round exposure argument like the one we use below. We include a different proof of Lemma 10 for completeness but also use this to

introduce two intermediate results (Proposition 8 and Lemma 9) that we later use in the same manner to prove our key technical result, Theorem 16.

We first remind the reader that a path through a subset of vertices V is a path whose internal vertices are all in V. Let us also introduce, given $U, V \subseteq V(G)$, the *ball of radius i around U within V* which we will denote as $B_G^i(U, V)$, namely it is the set of vertices in V which can be reached by a path through V of length at most i starting from a vertex in U. The starting vertex in U is *not* required to be in V itself, as is usual with our definition of paths through a set. We do, however, only consider reachable vertices within V, so that $B_G^i(U, V) \subseteq V$.

For Proposition 8, we take an expansion property of sets of size $t \in \mathbb{N}$ and use this to connect a pair of vertices from a set of $2t - 1$ pairs (c.f. the collection $\mathcal{P}$ in Definition 7).

Proposition 8. *Let $1 \leq \ell, t \leq n$. Let G be an n-vertex graph and let $V \subseteq V(G)$ be of size $|V| \geq 4t - 2$ such that, for every $U \subseteq V(G)$ with size $|U| = t$, we have $|B_G^\ell(U, V)| > \frac{|V|}{2}$. Let $x_1, \ldots, x_{2t-1}, y_1, \ldots, y_{2t-1}$ be distinct vertices of G.*

Then, for some $j \in [2t - 1]$, there is an $x_j y_j$-path in G through V with length at most $4\ell \log n$.

Proof. Let I_x be the set of $i \in [2t - 1]$ for which $|B_G^{2\ell \log n}(x_i, V)| \leq \frac{|V|}{2}$ and let I_y be the set of $i \in [2t - 1]$ for which $|B_G^{2\ell \log n}(y_i, V)| \leq \frac{|V|}{2}$. Note that the required path can be found if there exists some $j \in [2t - 1]$ with $j \notin I_x$ and $j \notin I_y$, for then $B_G^{2\ell \log n}(x_j, V)$ and $B_G^{2\ell \log n}(y_j, V)$ each have size larger than $\frac{|V|}{2}$ and must therefore intersect. Thus, we can assume that there is no such j, and, therefore, without loss of generality, that $|I_x| \geq t$.

Let $r \geq 0$ be the largest integer for which there is a set $X \subseteq \{x_i : i \in I_x\}$ for which $|X| \leq t \left(\frac{2}{3}\right)^r$ and $|B_G^{(r+1)\ell}(X, V)| > \frac{|V|}{2}$, and let X be any such set. Note that this is possible as $|I_x| \geq t$ and any subset of $\{x_i : i \in I_x\}$ with size t satisfies these conditions for $r = 0$ by the assumption of the proposition.

Now, as $X \neq \emptyset$, we have that $t \left(\frac{2}{3}\right)^r \geq 1$, so that $r \leq 2 \log t \leq 2 \log \frac{n}{2}$ (using $\left(\frac{2}{3}\right)^2 < \frac{1}{2}$ and $n \geq |V| \geq 4t - 2 \geq 2t$) and thus $r < (2 \log n) - 1$. Thus, by the definition of I_x, $|B_G^{(r+1)\ell}(x_i, V)| \leq |B_G^{2\ell \log n}(x_i, V)| \leq \frac{|V|}{2}$ for each $i \in I_x$, and hence $|X| \geq 2$. This allows us to partition $X = X_0 \cup X_1$ with $|X_0|, |X_1| \leq \frac{2}{3}|X| \leq t \left(\frac{2}{3}\right)^{r+1}$. As

$$|B_G^{(r+1)\ell}(X_0, V)| + |B_G^{(r+1)\ell}(X_1, V)| \geq |B_G^{(r+1)\ell}(X_0 \cup X_1, V)| = |B_G^{(r+1)\ell}(X, V)| > \frac{|V|}{2} \geq 2t - 1,$$

we can pick $j \in [2]$ such that $|B_G^{(r+1)\ell}(X_j, V)| \geq t$. Therefore, using the expansion of t-sets into V, we have $|B_G^{(r+2)\ell}(X_j, V)| > \frac{|V|}{2}$, contradicting the maximality of r as $|X_j| \leq t \left(\frac{2}{3}\right)^r$. $\square$

We now take a stronger expansion property and use Proposition 8 in combination with the Aharoni-Haxell hypergraph matching theorem (Theorem 6) to find edge disjoint paths connecting a set of vertex pairs (see also the discussion after Theorem 6), proving Lemma 9.

Lemma 9. *Let $n \geq 2$ and $1 \leq \ell, k \leq n$. Let G be an n-vertex graph and let $V \subseteq V(G)$ so that $|V| \geq \frac{n}{8} + 1$ and suppose, for each $U \subseteq V(G)$ and $F \subseteq E(G)$ with $U \neq \emptyset$ and $|F| \leq 2^9 k|U|(\ell \log n)^2$, we have $|B_{G-F}^\ell(U, V)| > \frac{|V|}{2}$.*

Then, G is $(4\ell \log n, k)$-path connected through V.

Proof. To show that G is $(4\ell \log n, k)$-path connected through V, let $\mathcal{P} \subseteq \binom{V(G)}{2}$ be an arbitrary collection of vertex pairs such that each vertex appears in at most k pairs in $\mathcal{P}$. Let $r = |\mathcal{P}|$, and order the pairs in $\mathcal{P}$. For each $i \in [r]$, let H_i be the hypergraph with vertex set $E(G)$ and edge set corresponding to the edge sets of paths through V with length at most $4\ell \log n$ connecting the i-th pair of vertices in $\mathcal{P}$, noting that the size of any edge is at most $4\ell \log n$.

Now, for each $I \subseteq [r]$, let M_I be a maximal matching in $\bigcup_{i \in I} H_i$. We will show that $|M_I| \geq 4\ell \log n \cdot |I|$ for each $I \subseteq [r]$. Towards a contradiction, suppose that, for some $I \subseteq [r]$, we have $|M_I| < 4\ell \log n \cdot |I|$, noting that we must have $I \neq \emptyset$. Let F be the set of edges of G in any path corresponding to an edge in M_I, so that $F = \bigcup_{e \in M_I} V(e)$. Note that $|F| < 4\ell \log n \cdot 4\ell \log n \cdot |I| = (4\ell \log n)^2 |I|$. Let I' be a maximal subset of I such that no vertex appears in a pair in I' more than once, so that $2k|I'| \geq |I|$. Note also that this ensures that $|I'| \leq \frac{n}{2}$. Now, let $t = \lceil |I'|/16 \rceil \geq 1$, so that $2t-1 \leq 16t-15 \leq |I'| \leq 16t$ and $4t-2 \leq \lceil \frac{n}{8} \rceil + 1 \leq |V|$. Note that $|F| < (4\ell \log n)^2|I| \leq k\frac{|I'|}{2}(8\ell \log n)^2 \leq 8kt(8\ell \log n)^2$, and therefore any set $U \subseteq V(G)$ with size t satisfies $|B_{G-F}^\ell(U, V)| > \frac{|V|}{2}$. Then, by Proposition 8 applied to $G - F$, for

some $j \in I'$ there is a path in $G - F$ between the j-th pair in $\mathcal{P}$ with interior vertices in V and length at most $4\ell \log n$. Such a path corresponds to an edge of H_j with no vertices in F, a contradiction to the maximality of M_I. Thus, we must have $|M_I| \geq 4\ell \log n \cdot |I|$.

Therefore, by Theorem 6, there is a set of paths P_i, $i \in [r]$, in G with $E(P_i) \in H_i$ for each $i \in [r]$, such that $E(P_i)$, $i \in [r]$, form edge disjoint sets. Thus, G is $(4\ell \log n, k)$-path connected through V. $\qquad\square$

We now prove the existence of our template graph, by showing an appropriate expansion condition is likely in a certain binomial random graph and applying Lemma 9.

Lemma 10. *For any large enough n, there exists an n-vertex graph G with $\Delta(G) \leq 2^8 \log^5 n$ and a set $V \subseteq V(G)$ with $|V| = \frac{n}{6}$ such that G is $\left(\frac{1}{4} \log^2 n, 2\right)$-path connected through V.*

Proof. Let $p = \frac{150 \log^5 n}{n}$, let $V \subseteq [n]$ be a set of size $\frac{n}{6}$ and let $G \sim \mathcal{G}(n, p)$. As $p = \omega\left(\frac{\log n}{n}\right)$, a standard application of Chernoff's inequality (Theorem 4) shows that, with high probability, $\Delta(G) \leq \frac{3}{2} pn = 225 \log^5 n$. Therefore, it is sufficient to show that, with high probability, G is $\left(\frac{1}{4} \log^2 n, 2\right)$-path connected through V.

Claim. *With high probability, for each $U \subseteq V(G)$ and each $F \subseteq E(G)$ with $U \neq \emptyset$ and $|F| \leq 4|U| \log^4 n$ we have*

$$|B_{G-F}(U, V)| > \min\left\{2^{16}|U|, \frac{|V|}{2}\right\}. \tag{2}$$

Proof. For large n, we will show for each $U \subseteq V(G)$ and $F \subseteq E(G)$ with $U \neq \emptyset$ and $|F| \leq 4|U| \log^4 n$ that (2) holds with probability at least $1 - 2^{-10|U| \log^5 n}$, so that (2) holds for all $U \subseteq V(G)$ and $F \subseteq E(G)$ with $|F| \leq 4|U| \log^4 n$ with probability at least

$$1 - \sum_{u=1}^{n} \binom{n}{u} \sum_{f=0}^{4u \log^4 n} \binom{\binom{n}{2}}{f} \cdot 2^{-10u \log^5 n} \geq 1 - \sum_{u=1}^{n} n^u \cdot n^2 \cdot n^{8u \log^4 n} \cdot 2^{-10u \log^5 n} \geq 1 - \sum_{u=1}^{n} n^{-2} = 1 - n^{-1},$$

and the claim holds.

Let then $U \subseteq V(G)$ with $u = |U| \geq 1$ and $F \subseteq E(G)$ with $f = |F| \leq 4u \log^4 n$. Note that if $|B_{G-F}(U, V)| \leq \min\left\{2^{16}u, \frac{|V|}{2}\right\}$, then there is some set $X = B_{G-F}(U, V)$ with size at most $2^{16}u$ such that $|V \setminus X| \geq \frac{|V|}{2}$ and there are no edges between U and $V \setminus X$ other than in F. The probability of such a set X existing is at most

$$\sum_{i=0}^{2^{16}u} \binom{n}{i} \cdot (1-p)^{u \cdot \frac{|V|}{2} - |F|} \leq n^{2^{16}u + 1} \cdot e^{-p(un/12 - f)} \leq e^{-pun/15} \leq 2^{-10u \log^5 n},$$

as required. $\qquad\square$

With high probability then, we have that the conclusion of the claim holds for G. For any $U \subseteq V(G)$ and $F \subseteq E(G)$ with $U \neq \emptyset$ and $|F| \leq 4|U| \log^4 n$, we have then by induction that, for each $i \geq 0$,

$$|B_{G-F}^i(U, V)| > \min\left\{2^{16i}|U|, \frac{|V|}{2}\right\}.$$

Setting $i = \ell = \frac{\log n}{16}$, we thus have $|B_{G-F}^\ell(U) \cap V| > \frac{|V|}{2}$. Thus, by Lemma 9 with $\ell = \frac{\log n}{16}$ and $k = 2$, we have that G is $\left(\frac{1}{4} \log^2 n, 2\right)$-connected through V. $\qquad\square$

3 Robust sublinear expansion

In this section we will explain the expansion we use and its background, our new perspective on this type of expansion, and prove some key results using the expansion. Before doing so, for convenience we state the definition of expansion that we use, as follows.

Definition 11. An n-vertex graph G is an (ε, s)-expander if, for every $U \subseteq V(G)$ and $F \subseteq E(G)$ with $1 \leq |U| \leq \frac{2}{3}n$ and $|F| \leq s|U|$, we have

$$|N_{G-F}(U)| \geq \frac{\varepsilon|U|}{\log^2 n}. \tag{3}$$

As the bound on the size of the neighbourhood guaranteed at (3) is $o(|U|)$ as $n \to \infty$, we consider this to be *sublinear* expansion. We often use Definition 11 when s is polylogarithmic in n, so that the set of edges F may be of size $\omega(|U|)$ as $n \to \infty$, and call this *robust* sublinear expansion. This is discussed in more detail with the relevant background in Section 3.1. In Section 3.2, we then introduce an alternative perspective of this expansion. In Section 3.3, we prove a lemma which almost decomposes an arbitrary graph into expanders. Finally, in Section 3.4 we prove sublinear expanders can be (edge) partitioned into expanders (with only slightly weaker expansion parameters).

3.1 Expansion

Classical graph expansion is an immensely powerful idea in graph theory and computer science that has seen a very wide variety of applications (see for example the survey [30]). A typical such property in a graph G says that $|N_G(U)| \geq \lambda|U|$ for any set $U \subseteq V(G)$ which is not too large, where $\lambda \geq 2$, though other notions have been considered instead of requiring a large $|N_G(U)|$, for example bounding the number of edges in G between U and $V(G) \setminus U$ (as indeed used by Conlon, Fox and Sudakov [11]). Sublinear expansion is a weaker notion of this classical expansion introduced by Komlós and Szemerédi [34, 35], where we take a much smaller value of λ, but which is significant as every graph contains a sublinear expander H with $\lambda = \Theta(1/\log^2|H|)$ (and even has a nice decomposition into sublinear expanders, as we will prove and use). Komlós and Szemerédi used sublinear expansion to find minors in sparse graphs, and more recently sublinear expansion has found a host of other applications (see, for example, [8, 21, 22, 26, 27, 33, 39, 40, 41, 42, 45, 50]).

Such sublinear expansion in a graph G has some very weak *robustness* properties, in that if $\lambda|U|/2$ vertices in $V(G) \setminus U$ are removed from the graph then the set U will still expand (with the neighbourhood of U still having at least $\lambda|U|/2$ vertices), and this property is used in many of the applications of sublinear expansion cited above. However, we will distinguish *robust sublinear expansion* to be that where U expands despite the removal of any set F of at most $s|U|$ edges in G, where s grows with $|G|$ so that this bound is superlinear in $|U|$, as in Definition 11. Such robust sublinear expansion has recently been developed essentially independently by groups of different authors, appearing in some form in work by Haslegrave, Kim, and Liu [28] and by Sudakov and Tomon [51] (with the parallel clearer in the expansion used in subsequent work by Jiang, Methuku and Yepremyan [31] and by Tomon [52]). Roughly speaking, the expansion we use, as given in Definition 11, is a slightly weaker version of that used by Haslegrave, Kim, and Liu [28] (so that we can find an almost-decomposition into such expanders) and a stronger version than subsequent developments of the expansion used by Sudakov and Tomon [51] (which makes it more powerful when we use it).

3.2 An alternative notion of robustness

In Definition 11, we consider the expansion to be robust as sets expand despite an arbitrary removal of a small number of edges. This can be alternatively encoded by recording that every vertex subset U either expands very well (by some factor greater than 1) or that its 'robust neighbourhood' of vertices with plenty of edges towards U expands well, though perhaps sublinearly (as in Proposition 12 below). It will be convenient to define this robust neighbourhood for any parameter d as

$$N_{G,d}(U) := \{v \in V(G) \setminus U : |N_G(v) \cap U| \geq d\},$$

that is, the set of vertices in a graph G, outside of a subset of vertices U, which have degree at least d towards U.

Proposition 12. *Let G be an n-vertex (ε, s)-expander, $U \subseteq V(G), |U| \leq \frac{2}{3}n$ and F a set of at most $s|U|/2$ edges. Then, for any $0 < d \leq s$, either*

$$\textbf{a)} \quad |N_{G-F}(U)| \geq \frac{s|U|}{2d}, \quad \textit{or} \quad \textbf{b)} \quad |N_{G-F,d}(U)| \geq \frac{\varepsilon|U|}{\log^2 n}.$$

Proof. Suppose **a)** is not satisfied, so that $|N_{G-F}(U)| < \frac{s|U|}{2d}$. Let $X = N_{G-F}(U) \setminus N_{G-F,d}(U)$, so that $|X| < \frac{s|U|}{2d}$. Let F' be the edges of $G-F$ between U and X, so that $|F'| < |X|d \leq s|U|/2$, and hence $|F| + |F'| \leq s|U|$. Note that, by the definition of F', we have $N_{G-F,d}(U) = N_{G-F-F'}(U)$. As G is an (ε, s)-expander, we thus have

$$|N_{G-F,d}(U)| = |N_{G-F-F'}(U)| \geq \frac{\varepsilon|U|}{\log^2 n},$$

and therefore **b)** holds, as required. $\qquad \square$

The following proposition shows more structure can be found in both outcomes of the above proposition. Though a more general variant follows easily, the parameters are tailored for our intended application.

Proposition 13. *There is an n_0 such that the following holds for each $n \geq n_0$, $\varepsilon \geq 2^{-9}$ and $s \geq 8\log^{13} n$. Let G be an n-vertex (ε, s)-expander, let $U \subseteq V(G)$ have size $|U| \leq \frac{2}{3}n$ and let F be a set of at most $s|U|/4$ edges. Then, in $G - F$ we can find either*

a) $\frac{|U|}{\log^7 n}$ *vertex disjoint stars, each with $\log^9 n$ leaves, centre in U and all its leaves in $V(G) \setminus U$, or*

b) *a bipartite subgraph H with vertex classes U and $X \subseteq V(G) \setminus U$ such that*

- $|X| \geq \frac{\varepsilon|U|}{2\log^2 n}$ *and*

- *every vertex in X has degree at least $\log^4 n$ in H and every vertex of U has degree at most $2\log^9 n$ in H.*

Proof. Take a maximal collection of vertex disjoint stars in $G - F$ with $\log^9 n$ leaves and centre in U and leaves outside of U. Let $C \subseteq U$ be the set of centres of these star and $L \subseteq V(G) \setminus U$ be the set consisting of all their leaves. Assuming **a)** does not hold, we can thus assume that $|C| \leq \frac{|U|}{\log^7 n}$, $|L| \leq |U|\log^2 n$, and, by the maximality, that there is no vertex in $U \setminus C$ with at least $\log^9 n$ neighbours in $G - F$ in $V(G) \setminus (U \cup L)$. Thus,

$$|N_{G-F}(U \setminus C)| \leq |C| + |L| + |U \setminus C| \cdot \log^9 n \leq \frac{|U|}{\log^7 n} + |U|\log^2 n + |U|\log^9 n < 2|U|\log^9 n. \qquad (4)$$

Let $d = \log^4 n$ and $\Delta = 2\log^9 n$. We now construct the set $X \subseteq V(G) \setminus U$ and the bipartite subgraph H through the following process, starting with $X_0 = \emptyset$ and setting H_0 to be the graph with vertex set $U \cup X_0$ and no edges. Let $r = |V(G) \setminus U|$ and label the vertices of $V(G) \setminus U$ arbitrarily as $v_1, \ldots, v_r$. For each $i \geq 1$, if possible pick a star S_i in $G - F$ with centre v_i and d leaves in U such that $H_{i-1} \cup S_i$ has maximum degree at most Δ, and let $H_i = H_{i-1} \cup S_i$ and $X_i = X_{i-1} \cup \{v_i\}$, while otherwise we set $H_i = H_{i-1}$ and $X_i = X_{i-1}$. Finally, let $H = H_r$ and $X = X_r = V(H_r) \setminus U$. We will show that **b)** holds for this choice of H with bipartition (U, X).

Firstly, observe that $\Delta(H_i) \leq \Delta$ for each $i \in [r]$ by construction, and that every vertex v_i in X has degree exactly d in H, so the second condition in **b)** holds. Thus, we only need to show that $|X| \geq \frac{\varepsilon|U|}{2\log^2 n}$ holds, which will follow as no vertex in $U \setminus C$ has $\frac{\Delta}{2} = \log^9 n$ neighbours in $G - F$ in $X \setminus L$ due to the maximality of our family of disjoint stars defining C and L.

Indeed, let U' be the set of vertices in $U \setminus C$ with degree exactly Δ in H. As each vertex in $U \setminus C$ has fewer than $\frac{\Delta}{2}$ neighbours in $G - F$ in $X \setminus L$, it must have at least $\frac{\Delta}{2}$ neighbours in H in $X \cap L$. As each vertex in $X \cap L$ has d neighbours in H, we have

$$|U'| \leq \frac{d|X \cap L|}{\Delta/2} \leq \frac{2d|L|}{\Delta} \leq \frac{2d \cdot |U|\log^2 n}{\Delta} = \frac{2\log^4 n \cdot |U|\log^2 n}{2\log^9 n} \leq \frac{\varepsilon|U|}{8\log^2 n},$$

where the last inequality follows for sufficiently large n.

Let $B = C \cup U'$, so that

$$|B| \leq \frac{|U|}{\log^7 n} + \frac{\varepsilon|U|}{8\log^2 n} \leq \frac{\varepsilon|U|}{4\log^2 n},$$

and, in particular, $|U \setminus B| \geq \frac{|U|}{2}$.

11

Then, by Proposition 12 applied to $U \setminus B$ and F with d, using that $|F| \leq s|U|/4 \leq s|U \setminus B|/2$, we have either $|N_{G-F}(U \setminus B)| \geq \frac{s|U \setminus B|}{2d}$ or $|N_{G-F,d}(U \setminus B)| \geq \frac{\varepsilon|U \setminus B|}{\log^2 n}$. As

$$\frac{s|U \setminus B|}{2d} \geq \frac{s|U|}{4d} \geq 2|U| \log^9 n,$$

the former contradicts (4), and therefore we must have that $|N_{G-F,d}(U \setminus B)| \geq \frac{\varepsilon|U \setminus B|}{\log^2 n}$. Every vertex v_i in $N_{G-F,d}(U \setminus B)$ has at least d neighbours in $G - F$ in $U \setminus B$ which, not being in $B = U' \cup C$, by definition of U' must all have degree strictly less than Δ in H. This implies $v_i \in X$, since we could add it together with some d of these neighbours. Hence, we must have $N_{G-F,d}(U \setminus B) \subseteq X$, and

$$|X| \geq |N_{G-F,d}(U \setminus B)| \geq \frac{\varepsilon|U \setminus B|}{\log^2 n} \geq \frac{\varepsilon|U|}{2\log^2 n},$$

as required. $\qquad\qquad\square$

3.3 Almost decomposing an arbitrary graph into expanders

The following lemma almost decomposes a graph into robust sublinear expanders with, on average, very little overlap between their vertex sets. Setting $s = 0$ in the below lemma (as we do in one application) obtains a full decomposition, although without any robustness.

Lemma 14. *Given an n-vertex graph G, a non-negative integer s and $\varepsilon \leq 2^{-5}$ we can delete up to $4sn \log n$ edges from G so that the remaining edges may be partitioned into graphs $G_1, \ldots, G_r$ such that $\sum_{i=1}^{r} |G_i| \leq 2n$ and each G_i is an (ε, s)-expander.*

Proof. We prove this by induction on n, under the stronger condition that the graphs $G_1, \ldots, G_r$ in the partition satisfy $\sum_{i=1}^{r} |G_i| \leq 2n - \frac{2n}{2+\log n}$. Since $2n - \frac{2n}{2+\log n} \geq n$, and any 1-vertex graph G is trivially an (ε, s)-expander, the lemma holds for $n = 1$ with $G_1 = G$. Let us then assume $n \geq 2$ and that the claim holds for all graphs with at most $n - 1$ vertices.

Letting G be an n-vertex graph, note that, as $2n - \frac{2n}{2+\log n} \geq n$, if G is an (ε, s)-expander then the trivial partition of $G_1 = G$ demonstrates the claim holds for G. Thus, we can assume G is not an (ε, s)-expander, and in particular that there exists a non-empty set of vertices $U \subseteq V(G)$ with $|U| \leq \frac{2}{3}n$ and a set F of at most $s|U|$ edges such that $|N_{G-F}(U)| < \frac{\varepsilon|U|}{\log^2 n}$. Let $G_1 = G[U \cup N_{G-F}(U)] - F$ and let $G_2 = G \setminus U - E(G_1) - F$, so that G_1 and G_2 form an edge partition of $G - F$ and, setting $n_1 = |G_1|$ and $n_2 = G_2$, we have

$$n_1 + n_2 = |G_1| + |G_2| = |G| + |N_{G-F}(U)| < n + \frac{\varepsilon|U|}{\log^2 n} \leq n + \frac{\varepsilon n_1}{\log^2 n}. \qquad (5)$$

Now, $n_2 = n - |U| < n$ and

$$n_1 \leq |U| + \frac{\varepsilon|U|}{\log^2 n} \leq \frac{2}{3}n + \varepsilon n \leq \frac{3}{4}n < n, \qquad (6)$$

so there exist sets $E_1 \subseteq E(G_1)$ and $E_2 \subseteq E(G_2)$ and partitions $G_{1,1}, \ldots, G_{1,r_1}$ and $G_{2,1}, \ldots, G_{2,r_2}$ of $G_1 - E_1$ and $G_2 - E_2$ into edge disjoint (ε, s)-expanders so that, for each $i \in [2]$, $|E_i| \leq 4sn_i \log n_i$, and

$$\sum_{j=1}^{r_i} |G_{i,j}| \leq 2n_i - \frac{2n_i}{2 + \log n_i}.$$

Therefore, we can remove $F \cup E_1 \cup E_2$ from G and decompose the remaining edges into (ε, s)-expanders $G_{1,1}, \ldots, G_{1,r_1}$, $G_{2,1}, \ldots, G_{2,r_2}$. We need then only check that $|F \cup E_1 \cup E_2| \leq 4sn \log n$ and that the sum of the vertices of the expanders in this decomposition is at most $2n - \frac{2n}{2+\log n}$.

Firstly, note that, from (6), we have $\log n_1 \leq \log \frac{3}{4}n < \log n - \frac{2}{5}$, so that

$$\frac{1}{s}(|F| + |E_1| + |E_2|) \leq |U| + 4n_1 \log n_1 + 4n_2 \log n_2 \leq n_1 + 4n_1 \left(\log n - \frac{2}{5}\right) + 4n_2 \log n$$

$$= 4(n_1 + n_2)\log n - \frac{3}{5}n_1 \overset{(5)}{\leq} 4\left(n + \frac{\varepsilon n_1}{\log^2 n}\right)\log n - \frac{3}{5}n_1$$

$$\leq 4n \log n. \tag{7}$$

Secondly, again as $\log \frac{3}{4}n < \log n - \frac{2}{5}$, we have

$$\frac{2n_1}{2 + \log n_1} \overset{(6)}{\geq} \frac{2n_1}{2 + \log(3n/4)} \geq \frac{2n_1}{8/5 + \log n} = \frac{2n_1}{2 + \log n} + \frac{2n_1 \cdot 2/5}{(8/5 + \log n)(2 + \log n)} > \frac{2n_1}{2 + \log n} + \frac{n_1}{10\log^2 n}, \tag{8}$$

so that

$$\sum_{i=1}^{2}\sum_{j=1}^{r_i}|G_{i,j}| \leq 2n_1 + 2n_2 - \frac{2n_1}{2 + \log n_1} - \frac{2n_2}{2 + \log n_2} \overset{(8)}{<} (n_1 + n_2)\left(2 - \frac{2}{2 + \log n}\right) - \frac{n_1}{10\log^2 n}$$

$$\overset{(5)}{\leq} \left(n + \frac{\varepsilon n_1}{\log^2 n}\right)\left(2 - \frac{2}{2 + \log n}\right) - \frac{n_1}{10\log^2 n} \leq 2n - \frac{2n}{2 + \log n}.$$

In combination with (7), this shows that G has the required decomposition, completing the inductive step and hence the proof. $\qquad \square$

3.4 Decomposing an expander into many expanders

The following lemma partitions the edges of an expander into a chosen number of expanders with the same vertex set (and a slightly weaker expansion condition).

Lemma 15. *Let $n, k, s \in \mathbb{N}$ and $0 < \varepsilon \leq 1$. Suppose that G is an n-vertex (ε, s)-expander and $s \geq 2^{12}\varepsilon^{-1}k^2\log^4 n$. Then, there are edge disjoint graphs $G_1, \ldots, G_k$ such that $E(G) = \bigcup_{i \in [k]} E(G_i)$ and, for each $i \in [k]$, G_i is an $\left(\frac{\varepsilon}{4}, \frac{\sqrt{s\varepsilon}}{8k\log n}\right)$-expander with vertex set $V(G)$.*

Proof. If $n = 1$ then the claim is trivially true so let us assume that $n \geq 2$. Furthermore, observe that as $\varepsilon > 0$, we must have $s \leq \delta(G)$ for otherwise we can remove all the neighbours of a vertex with minimum degree by removing at most s edges, contradicting that G is an (ε, s)-expander, so certainly $s \leq n$.

Let H be a random subgraph of G with vertex set $V(G)$ which contains every edge independently with probability $\frac{1}{k}$. Then, assign every edge of G to one of the graphs $G_1, \ldots, G_k$ uniformly and independently at random, so that each G_i is a random subgraph with the same distribution as H. Letting $s' = \frac{\sqrt{s\varepsilon}}{8k\log n} \geq 8\log n$, we will show that the probability H is not an $\left(\frac{\varepsilon}{4}, s'\right)$-expander is strictly less than $\frac{1}{k}$. Thus, by a union bound, the probability that each G_i is an $\left(\frac{\varepsilon}{4}, s'\right)$-expander is strictly positive, so some decomposition as required by the lemma must exist.

To show that H is not an $\left(\frac{\varepsilon}{4}, s'\right)$-expander with probability less then $\frac{1}{k}$, we will take a union bound over all subsets U of $V(G)$ for the event that U fails the conditions of $\left(\frac{\varepsilon}{4}, s'\right)$-expansion in H. For this, set $d = \sqrt{s/\varepsilon}\log n$ and note that $\frac{s}{d} = 8ks'$, $s' = \frac{\varepsilon d}{8k\log^2 n}$ and $s \geq \sqrt{s} \cdot \sqrt{\varepsilon^{-1}\log^2 n} = d \geq 64k$.

Let $U \subseteq V(G)$ and $u = |U| \leq \frac{2n}{3}$. By Proposition 12 with d, U and $F = \emptyset$, we have either **a)** $|N_G(U)| \geq \frac{su}{2d} = 4ks'u$ or **b)** $|N_{G,d}(U)| \geq \frac{\varepsilon u}{\log^2 n}$.

If **a)** holds, then $|N_H(U)|$ is dominated by $\text{Bin}(4ks'u, 1/k)$, so that $\mathbb{P}(|N_H(U)| \geq 2s'u) \geq 1 - e^{-s'u/2}$ by a Chernoff bound (Theorem 4). Note that if $|N_H(U)| \geq 2s'u$ then for any $F \subseteq E(H)$ with $|F| \leq s'u$ we have $|N_{H-F}(U)| \geq s'u \geq \frac{\varepsilon u}{4\log^2 n}$. Thus, when **a)** holds, the probability U fails the $\left(\frac{\varepsilon}{4}, s'\right)$-expansion condition is at most $e^{-s'u/2} \leq e^{-4u\log n}$ as $s' \geq 8\log n$.

If **b)** holds, then note first that the probability that any vertex $v \in N_{G,d}(U)$ is not in $N_{H,d/2k}(U)$ is at most $p := \mathbb{P}\left(\text{Bin}\left(d, \frac{1}{k}\right) < \frac{d}{2k}\right)$, where we have once again by a Chernoff bound that

$$p \leq e^{-d/8k}. \tag{9}$$

Therefore, as $|N_{G,d}(U)| \geq \frac{\varepsilon u}{\log^2 n}$, we have

$$\mathbb{P}\left(|N_{H,d/2k}(U)| < \frac{\varepsilon u}{2\log^2 n}\right) \leq \binom{\frac{\varepsilon u}{\log^2 n}}{\frac{\varepsilon u}{2\log^2 n}} \cdot p^{\frac{\varepsilon u}{2\log^2 n}} \leq 2^{\frac{\varepsilon u}{\log^2 n}} \cdot p^{\frac{\varepsilon u}{2\log^2 n}} \overset{(9)}{\leq} e^{\frac{-\varepsilon u d}{32k\log^2 n}} = e^{-s'u/4} \leq e^{-2u\log n}. \tag{10}$$

Note that, if $|N_{H,d/2k}(U)| \geq \frac{\varepsilon u}{2\log^2 n}$, then, for any $F \subseteq E(H)$ with $|F| \leq s'u$, we have

$$|N_{H-F}(U)| \geq |N_{H,d/2k}(U)| - \frac{|F|}{d/2k} \geq \frac{\varepsilon u}{2\log^2 n} - \frac{2ks'u}{d} = \frac{\varepsilon u}{2\log^2 n} - \frac{su}{4d^2} = \frac{\varepsilon u}{4\log^2 n}.$$

Thus, (10) implies that the probability U does not satisfy the $(\varepsilon/4, s')$-expansion condition is at most $e^{-2u\log n}$.

Therefore, whichever of **a)** or **b)** holds, the probability U does not satisfy the $(\varepsilon/4, s')$-expansion condition is at most $e^{-2u\log n}$. Hence, the probability that H is not an $(\varepsilon/4, s')$-expander is at most

$$\sum_{u=1}^{2n/3} \binom{n}{u} e^{-2u\log n} \leq \sum_{u=1}^{2n/3} n^u \cdot n^{-2u} \leq n^{-1} + \sum_{u=2}^{2n/3} n^{-2} \leq \frac{2}{n} < \frac{1}{k},$$

as required, where in the last inequality we make use of the observation that $n \geq s \geq 2^{12}\varepsilon^{-1}k^2\log^4 n > 2k$. $\qquad\square$

4 Finding edge disjoint paths through random vertex sets

We will now show that a robust sublinear expander is not only well-connected (in the sense of Definition 7), but is likely to be well-connected through any large random vertex subset. That is, we prove the following result.

Theorem 16. *Let G be an n-vertex (ε, s)-expander with $1 \geq \varepsilon \geq 2^{-7}$ and $s \geq \log^{135} n$. Let $V \subseteq V(G)$ be a random subset chosen by including each vertex independently at random with probability $\frac{1}{3}$. Then, with high probability, G is $(4\log^5 n, 2^8\log^5 n)$-path connected through V.*

The challenge of proving Theorem 16 is discussed in Section 2.2, and in particular in the answer to question **ii)** there. Having since then proved Lemma 9, let us note that, for suitable polylogarithmic parameters $\bar{s}$ and ℓ, by this lemma, to prove Theorem 16 it is sufficient to show that, with high probability, $|B^\ell_{G-F}(U, V)| > \frac{|V|}{2}$ for each $U \subseteq V(G)$ and $F \subseteq E(G)$ with $|F| \leq \bar{s}|U|$.

We show this property in three stages. Firstly, in Section 4.1, we show that for any set $U \subseteq V(G)$ with $|N_G(U)| \geq |U|\log^{24} n$ and any $F \subseteq E(G)$ with $|F| \leq |U|$ we have $|B^\ell_{G-F}(U, V)| > \frac{|V|}{2}$ holds with probability $1 - \exp(-\Omega(|U|\log^2 n))$. That is, we show the desired property holds for a *well-expanding* set U with high enough probability that we can take a union bound over all well-expanding sets U (and any edge set F with say $|F| \leq |U|$) to get that this property holds for all well-expanding sets with probability $1 - o(1/n)$. As discussed in Section 2.2, if the set U does not expand well then we can not guarantee it has the property we want with high enough probability for a union bound over all sets U, so we only consider well-expanding sets U here. Then, in Section 4.2, we show that every set U in our expander contains a well-expanding set U' which is not that much smaller than U (see Proposition 18), indeed, we will find such a U' satisfying $|U'| \geq s'|U|$, where $s' \geq 1/\log^{27} n$. Thus, for any edge set F with $|F| \leq s'|U| \leq |U'|$ we have $|B^\ell_{G-F}(U, V)| \geq |B^\ell_{G-F}(U', V)| > \frac{|V|}{2}$. This is almost the condition we need to apply Lemma 9, however s' is too small, namely by a polylogarithmic factor smaller than the value of $\bar{s}$ that we need. Therefore, in Section 4.3, we find the stronger expansion property we need by first splitting an (ε, s)-expander into polylogarithmically many edge disjoint expanders via Lemma 15, before showing that with high probability each one of these has the above-mentioned weaker expansion property into V. Finally, we combine these properties to show that indeed G has the desired stronger expansion property into V. This allows us to apply Lemma 9, completing the proof of Theorem 16.

4.1 Expansion of well-expanding sets into a random vertex set

In an n-vertex (ε, s)-expander G, given $U \subseteq V(G)$ such that $|N_G(U)| \geq |U|\log^{24} n$ and $F \subseteq E(G)$ with $|F| \leq s|U|/4$, when $V \subseteq V(G)$ is chosen by selecting each vertex independently at random with probability $\frac{1}{3}$, we wish to show that,

with some large probability, we have $|B_{G-F}^\ell(U,V)| > \frac{|V|}{2}$, for some appropriate parameters ε, s and ℓ (thus proving Lemma 17 below). To prove this we will adapt a 'sprinkling' argument from a very recent work of Tomon [52], thus avoiding a much more complex argument from initial versions of this work. To prove this we will reveal the vertices in V in ℓ batches, using the so-called sprinkling method, by partitioning V randomly into sets $V_1 \cup \ldots \cup V_\ell$, weighted so that most of the vertices in V are likely to be in V_ℓ. A natural approach here would be to prove a likely bound on $B_{G-F}^i(U, V_1 \cup \ldots \cup V_i)$ for each $i \in [\ell]$, resulting in a bound on $B_{G-F}^\ell(U, V_1 \cup \ldots \cup V_\ell) = B_{G-F}^\ell(U,V)$, so let us emphasise that this is *not* what we do.

Instead, for $0 \le i \le \ell$ we track the size of sets B_i which are defined as the set of vertices $v \in V(G)$ which can be reached from U by a path all of whose internal vertices are in $V_1 \cup \ldots \cup V_{i-1}$ (i.e., a path through this set) of length at most i. It is *crucial* here that we do not insist v belongs to either U or $V_1 \cup \ldots \cup V_i$. One can think of vertices in B_i as having potential of making all their neighbours in $G - F$ reachable in the same way (so being in B_{i+1}) if they get sampled into our next random subset V_i. In particular, every vertex with a neighbour in B_i which gets sampled into V_i will join B_{i+1}. This combined with our notion of robust expansion (as discussed further below) allows us to show that it is likely that B_{i+1} will increase in size compared to B_i until for some $i \le \ell - 1$ its size is at least $\frac{2}{3}n$. In particular, we will have $B_\ell \ge \frac{2}{3}n$. The final stage is slightly different, here since B_ℓ is independent of our final random set V_ℓ we will likely have almost $\frac{2}{3}$ of the vertices of V_ℓ belonging to B_ℓ. As our random sets are weighted heavily towards V_ℓ, it is likely to contain more than $\frac{3}{4}$ of the vertices of V so that we will likely have $|V_\ell \cap B_\ell| > \frac{|V|}{2}$, so that, finally, we have

$$|B_{G-F}^\ell(U,V)| \ge |V_\ell \cap B_\ell| > \frac{|V|}{2},$$

as required.

That the sets B_i, $1 \le i \le \ell - 1$, are very likely to increase notably in size will follow from our notion of robust expansion (as proved in the claim below). In particular, at step i, Proposition 13 tells us that one of two cases **a)** or **b)** may occur.

a) B_i has many large vertex disjoint stars extending from B_i. In this case we use that, for each centre sampled into V_i, the (many) corresponding leaves are added to B_{i+1}. We will have that many more leaves are added for each successful centre than the sampling probability for V_i, so that this is a good increase in size.

b) B_i has a large robust neighbourhood whose vertices have many neighbours in $G - F$ in B_i. Each vertex in this robust neighbourhood is likely to have at least one of these neighbours in B_i sampled into V_i, whereupon it will then be in B_{i+1}. (In fact, we need a slightly stronger property to hold so that the sampling of each vertex in B_i does not have too strong an influence on the size of $B_{i+1} \setminus B_i$, which is why we use the subgraph H provided by case **b)** of Proposition 13.)

Thus, in either case $|B_i|$ is likely to increase.

Lemma 17. *Suppose that G is an n-vertex (ε, s)-expander with $2^{-9} \le \varepsilon \le 1$ and $s \ge 8 \log^{13} n$. Let $U \subseteq V(G)$ satisfy $|N_G(U)| \ge |U| \log^{24} n$ and let $F \subseteq E(G)$ satisfy $|F| \le |U|$. Let $V \subseteq V(G)$ be a random subset chosen by including each vertex independently at random with probability $\frac{1}{3}$.*

Then, with probability $1 - e^{-\Omega(|U|\log^2 n)}$,

$$|B_{G-F}^{\log^4 n}(U,V)| > \frac{|V|}{2}. \tag{11}$$

Proof. Let $\ell = \log^4 n$, $q = \frac{3}{11}$ and let p be such that $1 - (1-p)^{\ell-1}(1-q) = \frac{1}{3}$, i.e., that $(1-p)^{\ell-1} = \frac{11}{12}$, so that

$$p \ge \frac{1}{15 \log^4 n}. \tag{12}$$

Let G be an n-vertex (ε, s)-expander, $U \subseteq V(G)$ with $|N_G(U)| \ge |U| \log^{24} n$ and $F \subseteq E(G)$ with $|F| \le |U|$. Independently, for each $i \in \{1, \ldots, \ell\}$, let V_i be a random subset of $V(G)$ with each vertex included independently at random with probability p if $i \le \ell - 1$ and with probability q if $i = \ell$. Set $V = V_1 \cup \ldots \cup V_\ell$, and note that each vertex is included in V independently at random with probability $\frac{1}{3}$. Thus, we wish to show that, with probability at least $1 - e^{-\Omega(|U|\log^2 n)}$ we have $|B_{G-F}^\ell(U,V)| > \frac{|V|}{2}$.

For each $0 \leq i \leq \ell$, let B_i be the set of vertices of G which can be reached via a path in $G - F$ which starts in U and has length at most i and whose internal vertices (if there are any) are in $V_1 \cup \ldots \cup V_{i-1}$. In particular, then, we have $B_0 = U$ and $B_1 = B_{G-F}(U)$. Observe also that $B_0 \subseteq B_1 \subseteq \ldots \subseteq B_\ell$. We emphasise that the vertices of B_i do not have to themselves be inside $V_1 \cup \ldots \cup V_{i-1}$, only the interior vertices of some path from U to the vertex in B_i are required to be inside $V_1 \cup \ldots \cup V_{i-1}$. An important property of B_i is that it is completely determined by the sets $U, V_1, \ldots, V_{i-1}$, so is in particular independent of V_i. Note also that any vertex in $N_{G-F}(B_i)$ with a neighbour in B_i that gets sampled into V_i belongs to B_{i+1}. These two observations will be the key behind why the sets B_{i+1} will grow in size until they occupy most of the set $V(G)$. In particular, finally, observe that

$$B_\ell \cap V_\ell \subseteq B_{G-F}^\ell(U, V). \tag{13}$$

We now show that indeed, for each $1 \leq i \leq \ell - 1$, that, unless B_i is already very large, B_{i+1} is likely to be larger than B_i, as follows.

Claim. *For each $1 \leq i \leq \ell - 1$, with probability $1 - e^{-\Omega(|U|\log^2 n)}$, either $|B_i| \geq \frac{2}{3}n$, or*

$$|B_{i+1} \setminus B_i| \geq \frac{\varepsilon |B_i|}{2^6 \log^2 n}.$$

Proof. For each $v \in N_{G-F}(B_i)$, v is in B_{i+1} if at least one of its neighbours in $G - F$ in B_i gets sampled into V_i. That is,

$$\{v \in N_{G-F}(B_i) : N_{G-F}(v, B_i) \cap V_i \neq \emptyset\} \subseteq B_{i+1} \setminus B_i. \tag{14}$$

We will show that, for any set $W \subseteq V(G)$ with $|W| \leq \frac{2}{3}n$ and $B_1 \subseteq W$

$$\mathbb{P}\left(|\{v \in N_{G-F}(W) : N_{G-F}(v, W) \cap V_i \neq \emptyset\}| \geq \frac{\varepsilon |W|}{2^6 \log^2 n}\right) \geq 1 - e^{-\Omega(|B_1|/\log^{22} n)}. \tag{15}$$

Thus, we will have for all $1 \leq i \leq \ell - 1$

$$\mathbb{P}\left(|B_i| \geq \frac{2}{3}n \text{ or } |B_{i+1} \setminus B_i| \geq \frac{\varepsilon |B_i|}{2^6 \log^2 n}\right) \geq \mathbb{P}\left(|B_{i+1} \setminus B_i| \geq \frac{\varepsilon |B_i|}{2^6 \log^2 n} \,\Big|\, |B_i| \leq \frac{2}{3}n\right)$$

$$\stackrel{(14)}{\geq} \mathbb{P}\left(|\{v \in N_{G-F}(B_i) : N_{G-F}(v, B_i) \cap V_i \neq \emptyset\}| \geq \frac{\varepsilon |B_i|}{2^6 \log^2 n} \,\Big|\, |B_i| \leq \frac{2}{3}n\right)$$

$$\stackrel{(15)}{\geq} 1 - e^{-\Omega(|B_1|/\log^{22} n)} \geq 1 - e^{-\Omega(|U|\log^2 n)},$$

where in the last inequality we used $|B_1| = |B_{G-F}(U)| \geq |U|\log^{24} n - |F| \geq \frac{1}{2}|U|\log^{24} n$.

Let then $W \subseteq V(G)$ with $|W| \leq \frac{2}{3}n$ and $B_1 \subseteq W$. As $|W| \leq \frac{2}{3}n$, and $|F| \leq |U| \leq |B_1| \leq |W| \leq s|W|/4$ we can apply Proposition 13 to W and F to show one of two cases **a)** or **b)** holds and we will show that (15) holds in either case.

a) Suppose $G - F$ contains $\frac{|W|}{\log^7 n}$ vertex disjoint stars with $\log^9 n$ leaves, with the centre in W and all leaves in $N_{G-F}(W)$. Let $C \subseteq W$ be the set of centres of such a collection of stars, and note that

$$|\{v \in N_{G-F}(W) : N_{G-F}(v, W) \cap V_i \neq \emptyset\}| \geq |C \cap V_i|\log^9 n. \tag{16}$$

By a Chernoff bound (Theorem 4) and (12), with probability at least $1 - e^{-p|C|/8} = 1 - e^{-\Omega(|W|/\log^{11} n)}$, we have $|C \cap V_i| \geq \frac{p|C|}{2} \geq \frac{|W|}{2^6 \log^{11} n}$. Thus, in combination with (16), we have that (15) holds as $\varepsilon \leq 1$.

b) Suppose instead that there is a bipartite subgraph $H \subseteq G - F$ with vertex classes W and X such that

- $|X| \geq \frac{\varepsilon |W|}{2\log^2 n}$ and

- every vertex in X has degree at least $\log^4 n$ in H and every vertex of U has degree at most $\Delta := 2\log^9 n$ in H.

For each $v \in X$, the probability that v has no neighbours in H in V_i is at most

$$(1-p)^{\log^4 n} \leq e^{-p\log^4 n} \stackrel{(12)}{\leq} e^{1/15} \leq \frac{15}{16}.$$

16

Let Y be the random variable counting the number of vertices of X having a neighbour in V_i in H, so that $\mathbb{E}Y \geq \frac{|X|}{16}$. Observe also that Y is Δ-Lipschitz since for each $v \in W$ the event $\{v \in V_i\}$ affects Y by at most $d_H(v) \leq \Delta$. Hence, by Lemma 5 with $k = \Delta$, $t = \frac{|X|}{32}$ and $N = |W|$, we have

$$\mathbb{P}\left(Y < \frac{|X|}{32}\right) \leq \mathbb{P}\left(Y < \mathbb{E}Y - \frac{|X|}{32}\right) \leq 2\exp\left(-\frac{2^{-9}|X|^2}{\Delta^2|W|}\right) = e^{-\Omega\left(|W|/\log^{22} n\right)}.$$

Each vertex in X with a neighbour in V_i in H lies in $\{v \in N_{G-F}(W) : N_{G-F}(v, W) \cap V_i \neq \emptyset\}$, so therefore, with probability at least $1 - e^{-\Omega\left(|W|/\log^{22} n\right)}$, we have $|\{v \in N_{G-F}(W) : N_{G-F}(v, W) \cap V_i \neq \emptyset\}| \geq Y \geq \frac{|X|}{32} \geq \frac{\varepsilon|W|}{2^6 \log^2 n}$ and thus (15) holds as well in case **b)**, completing the proof. □

As B_ℓ and V_ℓ are independent and $q = \frac{3}{11}$ (so that $\frac{2q}{3} = \frac{2}{11} < \frac{4}{23}$), by Chernoff's bound (Theorem 4), we have that

$$\mathbb{P}\left(|B_\ell \cap V_\ell| \leq \frac{4}{23}n \,\Big|\, |B_\ell| \geq \frac{2}{3}n\right) \leq \mathbb{P}\left(\text{Bin}\left(\frac{2}{3}n, q\right) \leq \frac{4}{23}n\right) \leq e^{-\Theta(n)},$$

and, similarly, as $\frac{1}{3} < \frac{8}{23}$ we have $\mathbb{P}\left(|V| \geq \frac{8}{23}n\right) \leq e^{-\Theta(n)}$.

Thus, by the claim, we have in total that

i) for each $i \in [\ell - 1]$, $|B_i| \geq \frac{2}{3}n$ or $|B_{i+1} \setminus B_i| \geq \frac{\varepsilon|B_i|}{2^6 \log^2 n}$, and

ii) $|B_\ell| < \frac{2}{3}n$ or $|B_\ell \cap V_\ell| > \frac{4}{23}n$, and

iii) $|V| \leq \frac{8}{23}n$

with probability at least

$$1 - \log^4 n \cdot e^{-\Omega\left(|U|\log^2 n\right)} - 2^{-\Theta(n)} \geq 1 - e^{-\Omega\left(|U|\log^2 n\right)}.$$

However, if i)–iii) all hold, then, for each $i \in [\ell - 1]$, we have

$$|B_i| \geq \min\left\{\frac{2}{3}n, \left(1 + \frac{\varepsilon}{2^6 \log^2 n}\right)^i |U|\right\} \geq \min\left\{\frac{2}{3}n, \exp\left(\frac{\varepsilon i}{2^7 \log^2 n}\right)\right\},$$

so that, setting $i = \ell = \log^4 n$, we conclude $|B_\ell| \geq \frac{2}{3}n$, and hence, by ii) and iii), that $|B_\ell \cap V_\ell| > \frac{|V|}{2}$.

Thus, by (13), we have that $|B_{G-F}^\ell(U, V)| > \frac{|V|}{2}$ with probability at least $1 - e^{-\Omega\left(|U|\log^2 n\right)}$. □

4.2 Expansion into a random vertex set

Having picked $V \subseteq V(G)$ with vertex probability $\frac{1}{3}$ in an n-vertex (ε, s)-expander G, Lemma 17 tells us that for any *fixed*, well-expanding subset of vertices U and small set F of edges we can reach more than one half of the vertices of V by short paths through V in $G - F$ with pretty high probability. We now want to use this to show that a similar property holds *simultaneously* for *all* vertex subsets U. As we cannot directly take a union bound over all subsets U, we first show that any vertex subset U in an (ε, s)-expander contains a subset $U' \subseteq U$ which expands particularly well and which is not much smaller than U. This follows easily from the definition of expansion but is perhaps easier to immediately see why it is true from the perspective introduced in Proposition 12.

Proposition 18. *Let $n \geq 2$, $0 < \varepsilon \leq 1$ and $s \geq \log^{24} n$. Let G be an n-vertex (ε, s)-expander and let $U \subseteq V(G)$ have size $|U| \leq \frac{2}{3}n$.*

Then, there is a set $U' \subseteq U$ with $|N_G(U')| \geq |U'|\log^{24} n$ and $|U'| \geq \frac{\varepsilon|U|}{3\log^{26} n}$.

Proof. Let $U' \subseteq U$ be maximal subject to $|N_G(U')| \geq |U'|\log^{24} n$, noting this is possible as $U' = \emptyset$ satisfies these conditions. Suppose that $U \neq U'$, for otherwise U satisfies the conditions itself. Then $|N_G(U')| < (|U'|+1)\log^{24} n + 1$ or we could add an arbitrary vertex to U' and contradict maximality. Similarly we know that, for every vertex $v \in U \setminus U'$,

v has at most $\log^{24} n$ neighbours outside of $U' \cup N_G(U')$, for otherwise $U' \cup \{v\}$ contradicts the maximality. Let F be the set of edges between $U \setminus U'$ and $V(G) \setminus (U' \cup N_G(U'))$, so that $|F| \le |U \setminus U'| \log^{24} n \le s|U|$. Thus, we have by the definition of expansion that

$$\frac{\varepsilon |U|}{\log^2 n} \le |N_{G-F}(U)| \le |N_G(U')| \le (|U'| + 1) \log^{24} n + 1,$$

so that $|U'| \ge \frac{\varepsilon |U|}{3 \log^{26} n}$, as required. □

We now show that we can ensure the conclusion of Lemma 17 holds for *all* well-expanding sets simultaneously by taking a union bound, and then use Proposition 18 to deduce an expansion property for all sets, as follows.

Lemma 19. *Suppose that G is an n-vertex (ε, s)-expander with $2^{-9} \le \varepsilon \le 1$ and $s \ge 2 \log^{24} n$. Let $V \subseteq V(G)$ be a random subset chosen by including each vertex independently at random with probability $\frac{1}{3}$.*

Then, with probability at least $1 - o(1/n)$, for every $U \subseteq V(G)$ and every set $F \subseteq E(G)$ with $|F| \le \frac{|U|}{\log^{27} n}$

$$|B_{G-F}^{\log^4 n}(U, V)| > \frac{|V|}{2}. \tag{17}$$

Proof. Say a set $U' \subseteq V(G)$ *expands well* in G if $|N_G(U')| \ge |U'| \log^{24} n$. Given a non-empty well-expanding set $U' \subseteq V(G)$ and a set of edges F of size at most $|U'|$, Lemma 17 applied to U' implies that

$$|B_{G-F}^{\log^4 n}(U', V)| > \frac{|V|}{2} \tag{18}$$

fails with probability at most $e^{-\Omega(|U'| \log^2 n)}$.

Now a union bound over all pairs (U', F) such that U' is a well-expanding set in G and F is a set of at most $|U'|$ edges tells us that *some* such pair (U', F) fails (18) with probability at most

$$\sum_{(U', F)} e^{-\Omega(|U'| \log^2 n)} \le \sum_{u=1}^{n} \sum_{f=1}^{u} \binom{n}{u} \binom{n^2}{f} \cdot e^{-\Omega(u \log^2 n)}$$

$$\le \sum_{u=1}^{n} u \cdot n^{3u} \cdot e^{-\Omega(u \log^2 n)} \le \sum_{u=1}^{n} e^{-\Omega(u \log^2 n)} = o(1/n).$$

Thus, with probability $1 - o(1/n)$, we can assume that (18) holds for every well-expanding set U' and set $F \subseteq E(G)$ with $|F| \le |U'|$. We will now show that this implies (17) holds for all $U \subseteq V(G)$ and $F \subseteq E(G)$ with $|F| \le \frac{|U|}{\log^{27} n}$, completing the proof.

Let then $U \subseteq V(G)$ with $|U| \le \frac{2}{3}n$ and let $F \subseteq E(G)$ satisfy the (slightly weaker) condition $|F| \le \frac{2|U|}{\log^{27} n}$. Then, by Proposition 18, there is a set $U' \subseteq U$ which is well-expanding for which $|U'| \ge \frac{\varepsilon |U|}{3 \log^{26} n}$. Noting that $|F| \le |U'|$ (as we may assume n is large with probability $1 - o(1/n)$), we therefore have that

$$|B^{\log^4 n}(U, V)| \ge |B^{\log^4 n}(U', V)| > \frac{|V|}{2}.$$

Finally, consider $U \subseteq V(G)$ with $|U| > \frac{2}{3}n$ and let $F \subseteq E(G)$ satisfy $|F| \le \frac{|U|}{\log^{27} n}$. Let $\bar{U} \subseteq U$ be an arbitrary subset with $\frac{n}{2} \le |\bar{U}| \le \frac{2}{3}n$, so that we have $|F| \le \frac{2|\bar{U}|}{\log^{27} n}$, and hence, from what we have just shown,

$$|B^{\log^4 n}(U, V)| \ge |B^{\log^4 n}(\bar{U}, V)| > \frac{|V|}{2},$$

as required. □

4.3 Path connectedness through a random subset in expanders

We are now ready to prove Theorem 16. As discussed at the start of this section, we first split the edges of the graph G into expanders, before applying Lemma 19 to each of these, to get (with high probability), a strong enough expansion

condition to apply Lemma 9.

Proof of Theorem 16. To recap: we have an n-vertex (ε, s)-expander, G, with $2^{-7} \leq \varepsilon \leq 1$ and $s \geq \log^{135} n$, and a random subset $V \subseteq V(G)$ where each vertex is included independently at random with probability $\frac{1}{3}$. To prove Theorem 16, we need to show that, with high probability, G is $(4 \log^5 n, 2^8 \log^5 n)$-path connected through V.

Let $k = 2^{17} \log^{42} n$, so that $s \geq 2^{12} \varepsilon^{-1} k^2 \log^4 n$, and let $s' = \frac{\sqrt{s\varepsilon}}{8k \log n} \geq 2 \log^{24} n$. Using Lemma 15, take edge disjoint graphs $G_1, \ldots, G_k$ such that $E(G) = \bigcup_{i \in [k]} E(G_i)$ and, for each $i \in [k]$, G_i is an $\left(\frac{\varepsilon}{4}, s'\right)$-expander.

Then, by Lemma 19 and a union bound over the k graphs G_i, with high probability we can assume that, for each $i \in [k]$ and every $U \subseteq V(G_i)$ and $F \subseteq E(G_i)$ with $|F| \leq \frac{|U|}{\log^{27} n}$,

$$|B_{G_i - F}^{\log^4 n}(U, V)| > \frac{|V|}{2}.$$

Now, let $U \subseteq V(G)$ and $F \subseteq E(G)$ with $|F| \leq 2^{17} |U| \log^{15} n$. As the graphs G_i, $i \in [k]$, are edge disjoint, there must be some $i \in [k]$ with $|F \cap E(G_i)| \leq \frac{2^{17} |U| \log^{15} n}{k} = \frac{|U|}{\log^{27} n}$, and therefore

$$|B_{G-F}^{\log^4 n}(U, V)| \geq |B_{G_i - F}^{\log^4 n}(U, V)| > \frac{|V|}{2}.$$

Thus, by Lemma 9, applied with $k = 2^8 \log^5 n$ and $\ell = \log^4 n$ we conclude that G is $(4 \log^5 n, 2^8 \log^5 n)$-connected, as desired. $\qquad\square$

5 Cycle decompositions

In this section we will prove our main results, Theorems 2 and 3. Before doing this we need to put together a few final ingredients. In Section 4, we established a very robust connectivity property of expanders. In Section 5.1 we will show that in an expander one can find a subgraph with few edges and yet (effectively) the same connectivity property through a random subset V, with high probability. We will refer to this subgraph as a *skeleton* of our graph, divide the vertex set into three as $V(G) = V_1 \cup V_2 \cup V_3$, and find three matching skeletons. In Section 5.2 we show that any graph can be decomposed into few paths in such a way that no vertex is used as an endvertex many times. In Section 5.3 we combine these results to decompose any expander into linearly many cycles and a few leftover edges. As outlined in Section 2.2, we achieve this by setting aside the skeletons, then decomposing the remainder of the graph into three sets of paths and finally using the connection properties of the skeletons to join the endvertices of these sets of paths, where the sets of paths are matched to the skeletons to ensure this creates edge disjoint cycles. These cycles decompose all the edges in the graph which are not in the skeleton and since the skeleton is chosen to be sparse this gives us the result.

The final ingredient in the proof of Theorem 3, given in Section 5.4, is to decompose an arbitrary graph into expanders and a few leftover edges via Lemma 23, to each of which we can apply our expander decomposition result. All that will remain, then, is to iteratively apply Theorem 3 in Section 5.6, while removing some additional cycles, to deduce Theorem 2.

5.1 Finding the skeletons

To find the skeletons, we will use Theorem 16 to embed a sparse well-connected 'template' graph (from Lemma 10) with its edges replaced by relatively short edge disjoint paths, and show that the image of this embedding has the properties we need of a skeleton, as follows.

Lemma 20. *Let G be an n-vertex graph which is an (ε, s)-expander with $2^{-7} \leq \varepsilon \leq 1$ and $s \geq \log^{135} n$. Let $V \subseteq V(G)$ be chosen by including each vertex independently at random with probability $\frac{1}{3}$. Then, with high probability, there is a subgraph of G with at most $2^9 n \log^{10} n$ edges which is $(\log^7 n, 2)$-path connected through V.*

Proof. By Theorem 16 applied to G and V, G is with high probability $(4 \log^5 n, 2^8 \log^5 n)$-path connected through V. Note that, by Chernoff's inequality (Theorem 4), we can in addition ensure with high probability that $|V| \geq \frac{n}{6}$, and since our goal is to show a statement with high probability we may assume that n is large enough to apply Lemma 10. That is, by that lemma we may assume there is an auxiliary graph H with vertex set $V(H) = V(G)$ which is $\left(\frac{1}{4} \log^2 n, 2\right)$-path connected through V, and such that $\Delta(H) \leq 2^8 \log^5 n$.

Note that $E(H)$ is a collection of pairs of vertices in $V(G)$ with the property that every vertex appears in at most $\Delta(H) \leq 2^8 \log^5 n$ pairs. Hence, since G is $(4 \log^5 n, 2^8 \log^5 n)$-path connected through V, we can find for each $e \in E(H)$ a path P_e through V of length at most $4 \log^5 n$, such that all the paths P_e, $e \in E(H)$, are edge disjoint. Let G' have vertex set $E(G)$ and edge set $\bigcup_{e \in E(H)} E(P_e)$, noting that

$$|E(G')| \leq 4 \log^5 n \cdot |E(H)| \leq 2n \log^5 n \cdot \Delta(H) \leq 2n \log^5 n \cdot 2^8 \log^5 n = 2^9 n \log^{10} n.$$

Therefore, we need only show that G' is $(\log^7 n, 2)$-path connected. For this, let $\mathcal{P} \subseteq \binom{V(G)}{2}$ be a family of pairs of vertices from $V(G)$ with each vertex appearing in at most 2 different pairs. Since H is $\left(\frac{1}{4} \log^2 n, 2\right)$-path connected through V we can find edge disjoint paths in H through V, each of length at most $\frac{1}{4} \log^2 n$, through V, joining each pair in $\mathcal{P}$. If we now replace each edge e of H used by one of these paths with P_e we obtain a collection of edge disjoint *walks* in G' of length at most $\log^7 n$, through V, joining each pair in $\mathcal{P}$. Replacing each of the walks with its shortest subwalk joining the same endvertices, we obtain paths which are edge disjoint, each have length at most $\log^7 n$, and which connect the vertex pairs in $\mathcal{P}$. Thus, G' is $(\log^7 n, 2)$-path connected, as claimed. $\qquad \square$

5.2 Lovász path covering with well-spread endvertices

As quoted in the introduction, Lovász proved the following classical decomposition result in 1968.

Theorem 21 (Lovász [43])**.** *Every n-vertex graph can be decomposed into at most $\frac{n}{2}$ paths and cycles.*

Theorem 21 almost provides the path decompositions that we need. At the expense of using perhaps slightly more paths, we can ensure in addition that no vertex is used often as an endvertex of the paths in the decomposition through the following simple deduction.

Corollary 22. *Every n-vertex graph G can be decomposed into paths so that each vertex of G is an endvertex of at most two paths in the decomposition.*

Proof. Form a graph G' from G by adding a new vertex v_0 and an edge from v_0 to each vertex $v \in V(G)$ for which $d_G(v)$ is even. By Theorem 21, there is a collection $\mathcal{C}$ of at most $\frac{n+1}{2}$ cycles and paths which decomposes G'. Note that each vertex $v \in V(G)$ has odd degree in G', and therefore must be an endvertex of an odd number of paths in $\mathcal{C}$, and thus, in particular, must be an endvertex of some path in $\mathcal{C}$. As the paths in $\mathcal{C}$ together have at most $n+1$ endvertices, each vertex in G is the endvertex of at most 1 path in $\mathcal{C}$, as otherwise there would need to be $n - 1 + 3 > n + 1$ endvertices. Note that, furthermore, as there must be at least n endvertices together for the paths in $\mathcal{C}$, we must have that $\mathcal{C}$ contains at least $\frac{n}{2}$ paths. Thus, as $|\mathcal{C}| \leq \frac{n+1}{2}$, $\mathcal{C}$ is in fact a collection consisting only of paths, with no cycles.

Now, for each path $P \in \mathcal{C}$, if $v_0 \in V(P)$ then remove the vertex v_0 from P, and let $\mathcal{C}'$ be the collection of all the resulting paths. Then, $\mathcal{C}'$ is a decomposition of G into paths. Furthermore, observe that each vertex $v \in V(G)$ is an endvertex of a path $P \in \mathcal{C}$ only if v was an endvertex of some path in $\mathcal{C}$ which contained P or if P was created by removing the edge vv_0 from a path in $\mathcal{C}$. Thus, each vertex is an endvertex of at most two paths in $\mathcal{C}'$, so that $\mathcal{C}'$ decomposes G as required. $\qquad \square$

5.3 Decomposing expanders

In this section we will decompose an expander into linearly many cycles and a few leftover edges, as follows.

Lemma 23. *Any sufficiently large n-vertex graph G which is an (ε, s)-expander with $2^{-5} \leq \varepsilon \leq 1$ and $s \geq \log^{273} n$ can be decomposed into at most $3n$ cycles and at most $2^{11} n \log^{10} n$ edges.*

Proof. Let $s' = \frac{\sqrt{s\varepsilon}}{24 \log n} \geq \log^{135} n$. Using Lemma 15 with $k = 3$, take edge disjoint $\left(\frac{\varepsilon}{4}, s'\right)$-expander subgraphs G_1, G_2 and G_3 of G, each with vertex set $V(G)$, so that $E(G) = E(G_1) \cup E(G_2) \cup E(G_3)$. Next, partition $V(G) = V_1 \cup V_2 \cup V_3$ by assigning each vertex to a set in the partition uniformly and independently at random. Using Lemma 20, for each $i \in [3]$, find a subgraph $G_i' \subseteq G_i$ with at most $2^9 n \log^{10} n$ edges which is $(\log^7 n, 2)$-path connected through V_i.

For each $i \in [3]$, let H_i be the graph with vertex set $V(G) \setminus V_i$ whose edges are the edges of $G - G_1' - G_2' - G_3'$ lying within V_{i+1} or between V_{i+1} and V_{i+2}, with these indices taken appropriately modulo 3, noting that these graphs partition $G - G_1' - G_2' - G_3'$. For each $i \in [3]$, then, apply Corollary 22 to decompose H_i into a collection of paths $\mathcal{P}_i$ with the property that no vertex is an endvertex of more than two of the paths in $\mathcal{P}_i$, noting that, in particular, this implies that $|\mathcal{P}_i| \leq n$. Since G_i' is $(\log^7 n, 2)$-path connected through V_i and the paths in $\mathcal{P}_i$ have no vertices in V_i, for each path $P \in \mathcal{P}_i$, we can find a path Q_P through V_i in G_i' joining the endvertices of P, so that all these new paths are edge disjoint. As $V(H_i) \cap V_i = \emptyset$, for each $P \in \mathcal{P}_i$, $P \cup Q_P$ is a cycle, and furthermore, as H_i and G_i' are edge disjoint, the cycles form a collection, $\mathcal{C}_i$ say, of at most n edge disjoint cycles whose edges contain all of the edges of H_i.

As the subgraphs $H_i \cup G_i'$, $i \in [3]$, are edge disjoint, $\mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3$ is a collection of at most $3n$ cycles which contains every edge of G except for, perhaps, some edges in $G_1' \cup G_2' \cup G_3'$. Thus, these cycles cover all but at most $2^{11} n \log^{10} n$ edges, giving us the desired decomposition. $\qquad \square$

5.4 Decomposing a general graph

We are now ready to prove Theorem 3, which we do in the following more quantitative form for convenience.

Theorem 24. *Any n-vertex graph can be decomposed into at most $6n$ cycles and $O(n \log^{274} n)$ edges.*

Proof. Let $n_0 \geq 2^{12}$ be sufficiently large so that the statement of Lemma 23 holds for every graph with at least n_0 vertices, and let $C = 3n_0$. Let $s = \log^{273} n$ and $\varepsilon = 2^{-5}$. Using Lemma 14, decompose G into subgraphs $G_1, \ldots, G_r$ and at most $4sn \log n$ edges so that $|G_1| + \ldots + |G_r| \leq 2n$ and, for each $i \in [r]$, G_i is an (ε, s)-expander. Let $I \subseteq [r]$ be the set of $i \in [r]$ with $|G_i| \geq n_0$, so that

$$\left| E(G) \setminus \left(\bigcup_{i \in I} E(G_i) \right) \right| \leq n_0 \cdot \sum_{i=1}^{r} |G_i| + 4sn \log n \leq 2n_0 s \cdot n \log n.$$

For each $i \in I$, using Lemma 23, decompose G_i into at most $3|G_i|$ cycles and $2^{11} |G_i| \log^{10} n$ edges, giving in total at most $3|G_1| + \ldots + 3|G_r| \leq 6n$ cycles and at most $2^{11}(|G_1| + \ldots + |G_r|) \log^{10} n \leq n_0 s \cdot n \log n$ edges. In combination with the edges of $G - \bigcup_{i \in I} G_i$ this gives a decomposition of G into at most $6n$ cycles and $3n_0 sn \log n = Cn \log^{274} n$ edges. $\qquad \square$

5.5 Long cycles in expanders

A final ingredient we need before proving Theorem 2 is to show that, after removing long cycles and adding them to a decomposition, any expander in the resulting graph must be quite small. That is to say, any expander contains a long cycle even if it has no robustness at all. This follows from a result of Krivelevich [37] although since we do not need the full power or generality of that result, for completeness we include a short proof using the Depth First Search (DFS) algorithm (first used in this manner in [4]), as follows.

Lemma 25. *Any n-vertex $(\varepsilon, 0)$-expander, with $\varepsilon \geq 2^{-5}$ and $n \geq 2^{30}/\varepsilon^2$, contains a cycle of length $\Omega\left(\frac{n}{\log^4 n}\right)$.*

Proof. Let G be our $(\varepsilon, 0)$-expander. Observe first that the expansion condition guarantees G is connected, since otherwise we could find a connected component of G of size less than $\frac{|G|}{2}$, whose vertex set then does not expand at all.

We now run the DFS algorithm on G as follows. At any point during the process we have the set of unexplored vertices U, the path P with active endvertex $t(P)$, and the set of processed vertices R. Picking an arbitrary vertex $r \in V(G)$, We start with $U = V(G) \setminus \{r\}, R = \emptyset$ and with P being the path with vertex set $\{r\}$, and set $t(P) = r$. At each step,

if there is a neighbour v of $t(P)$ in U we add it to P with the edge $t(P)v$ and let $t(P) = v$. Otherwise, we move $t(P)$ from P to R and set its neighbour in P as the new $t(P)$.

In the above process, in each step we either move precisely one vertex from U to P or precisely one vertex from P to R. Note also that at any point in the process there are no edges between U and R since a vertex is only moved to R once it has no neighbours in U, and U only ever has vertices removed from it. Finally, as G is connected, note that the process finishes with all the vertices being in R and P, and U being empty.

Thus, we start with $|U| = n - 1$ and $|R| = 0$ and finish with $|U| = 0$ and $|R| = n$, at each step reducing $|U|$ by one or increasing $|R|$ by one. Therefore, at some point in the process we must have $|U| = |R|$. Since there are no edges between U and R we know that all the neighbours of U must belong to P, which therefore has size $|P| = n - 2|U| \geq |N_G(U)| \geq \frac{\varepsilon |U|}{\log^2 n}$, where the last inequality follows by the expansion property applied to U, which we can do since $|U| = \frac{n - |P|}{2} \leq \frac{n}{2}$. This in turn implies that $|P| \geq \frac{\varepsilon n}{3 \log^2 n}$ since either $|U| \geq \frac{n}{3}$ or $|P| = n - 2|U| \geq \frac{n}{3}$.

Now, let X, Y, Z be sets of consecutive vertices of P in that order which partition $V(P)$ so that $|X|, |Z| \geq \frac{|P|}{3}$ and $\frac{\varepsilon^2 n}{18 \log^4 n} \leq |Y| < \frac{\varepsilon^2 n}{9 \log^4 n}$. If X and Z are connected by some path in $G \setminus Y$, then take a shortest path, Q say, between X and Z in $G \setminus Y$ and note that, combined with the segment of P between the endvertices of Q, this gives a cycle containing each vertex in Y, which thus has size $\Omega\left(\frac{n}{\log^4 n}\right)$. If X and Z are not connected by a path in $G \setminus Y$, then we can take a partition $V(G) \setminus Y = X' \cup Z'$ with no edges between X' and Z' in G, and $X \subseteq X'$ and $Z \subseteq Z'$. Without loss of generality, suppose that $|X| \leq \frac{n}{2}$. By the expansion condition we have $|N_G(X)| \geq \frac{\varepsilon |X|}{\log^2 n} \geq \frac{\varepsilon |P|}{3 \log^2 n} \geq \frac{\varepsilon^2 n}{9 \log^4 n}$, yet we also have $|N_G(X)| \leq |Y| < \frac{\varepsilon n}{9 \log^4 n}$, a contradiction. $\square$

5.6 Proof of Theorem 2

To decompose a graph into cycles and edges and prove Theorem 2, we now repeatedly do the following:

- letting d be the average degree of the graph consisting of the remaining edges, we remove maximally many edge disjoint cycles with length at least d,

- we then decompose the remaining edges exactly into expanders (using Theorem 14 with $s = 0$) and show that these must be small subgraphs as they each have no cycle with length at least d (using Lemma 25),

- and finally we decompose each of these small subgraphs into cycles and edges using Theorem 24.

Together this comprises the iterative step we use, which decomposes the n-vertex graph G with average degree d into $O(n)$ cycles and $n \log^{O(1)} d$ edges. We then iterate on the graph of the edges in this decomposition, noting that its average degree is much smaller than d. We state and prove the outcome of one iterative step as the following lemma, for convenience and its own interest.

Lemma 26. *Any n-vertex graph G with average degree $d \geq 2$ can be decomposed into $O(n)$ cycles and a subgraph with average degree $O(\log^{274} d)$.*

Proof. Let $\mathcal{C}$ be a maximal collection of edge disjoint cycles with length at least d in G. As there are $\frac{nd}{2}$ edges in G, we have $|\mathcal{C}| \leq \frac{n}{2}$. Let G' be G with the edges of the cycles in $\mathcal{C}$ removed, so that G' has no cycles with length at least d. Apply Lemma 14 with $s = 0$ and $\varepsilon = 2^{-5}$ to obtain a full decomposition (since $s = 0$) of G' into subgraphs $G_1, \ldots, G_k$, such that $|G_1| + \ldots + |G_k| \leq 2n$ and, for each $i \in [k]$, G_i is a $(2^{-5}, 0)$-expander. For each $i \in [k]$, as G' and hence G_i has no cycle with length at least d, we have by Lemma 25 that $|G_i| = O(d \log^4 d)$. Using Theorem 24, we decompose each G_i into at most $6|G_i|$ cycles and $O(|G_i| \log^{274} |G_i|) = O(|G_i| \log^{274} d)$ edges. Collecting these cycles and edges over all $i \in [k]$, and including the cycles in $\mathcal{C}$, we get a decomposition of G into at most

$$\frac{n}{2} + \sum_{i=1}^{k} 6|G_i| \leq 13n$$

cycles and

$$O\left(\sum_{i=1}^{k} |G_i| \log^{274} d\right) = O(n \log^{274} d)$$

edges. Noting these edges form a subgraph with average degree $O(\log^{274} d)$ completes the proof. □

Finally, by iterating Lemma 26, we can prove Theorem 2, i.e., that any n-vertex graph can be decomposed into $O(n \log^* n)$ cycles and edges.

Proof of Theorem 2. Using Lemma 26, let $C \geq 1$ be large enough that any n-vertex graph with average degree $d \geq 2$ has a decomposition into at most Cn cycles and a subgraph with average degree at most $C \log^{274} n$. Let G_0 be any n-vertex graph, and, for each $i \geq 0$, let G_{i+1} be a graph with the fewest edges that can be formed by removing at most Cn edge disjoint cycles from G_i. For each $i \geq 0$, let d_i be the average degree of G_i, so that we have $d_{i+1} \leq C \log^{274} d_i$ for each $i \geq 0$ for which $d_i \geq 2$.

Let ℓ be the largest integer such that the log function applied iteratively ℓ times to n is still above $300C$, i.e., the largest integer such that $\log^{[\ell]} n \geq 300C$. Note that $\ell \leq \log^* n$ and $\log^{[\ell]} n < 2^{300C}$. We will show, for each $0 \leq i \leq \ell$, that $d_i \leq C(300 \log^{[i]} n)^{274}$. Note that $d_0 \leq n$, so that this easily holds with $i = 0$. Then, assuming it is true for some $0 \leq i \leq \ell - 1$ and that $d_i \geq 2$ (for otherwise $d_{i-1} \leq d_i \leq 2$), we have

$$
\begin{aligned}
d_{i+1} &\leq C \log^{274} d_i \leq C(\log(C(300 \log^{[i]} n)^{274}))^{274} \\
&= C(\log C + 274 \log 300 + 274 \log^{[i+1]} n)^{274} \leq C(300 \log^{[i]} n)^{274},
\end{aligned}
\tag{19}
$$

where in the last inequality we used $26 \log^{[i+1]} n \geq 26 \log^{[\ell]} n \geq 26 \cdot 300C \geq \log C + 274 \log 300$.

Thus, in particular, we have that the average degree of G_ℓ is at most $C(300 \log^{[\ell]} n)^{274} \leq C(300 \cdot 2^{300C})^{274} = O(1)$. Furthermore, to get from G_0 to G_ℓ we have removed at most $C\ell \leq Cn \log^* n$ cycles. Therefore, G_0 has a decomposition into $O(n \log^* n)$ cycles and $O(n)$ edges, which, as G_0 is an arbitrary n-vertex graph, completes the proof. □

6 Concluding remarks

Our results. In this paper, we gave new bounds on two of the most central open problems on cycle decompositions – the Erdős-Gallai conjecture from 1966 that any n-vertex graph can be decomposed into $O(n)$ cycles and edges, and Hajós's conjecture from 1968 asserting that any n-vertex Eulerian graph can be decomposed into at most $\frac{n}{2}$ cycles, where the bound in Hajós's conjecture follows easily from Theorem 2. Indeed, given any n-vertex Eulerian graph, we can apply this result first to remove $O(n \log^* n)$ cycles and leave only $O(n \log^* n)$ remaining edges. The remaining edges then still form an Eulerian graph, which has a cycle decomposition by the observation of Veblen quoted in the introduction. As this cycle decomposition has fewer cycles than the number of these edges, we get, altogether, a decomposition of the original Eulerian graph into $O(n \log^* n)$ cycles.

Lower bounds for the Erdős-Gallai conjecture. As noted in the introduction, Hajós's conjecture implies the Erdős-Gallai conjecture, as any n-vertex graph can be decomposed into an Eulerian graph and at most $n - 1$ edges — for example by taking the union of a maximal collection of disjoint cycles, and the edges in the remaining acyclic subgraph. Therefore, if Hajós's conjecture holds, then any n-vertex graph would have a decomposition into at most $\frac{n}{2}$ cycles and at most $n - 1$ edges, and thus at most $\frac{3}{2}n$ cycles and edges. It is known that only $(\frac{3}{2} + o(1))n$ cycles and edges are needed to decompose any n-vertex graph with linear minimum degree (due to Girão, Granet, Kühn, and Osthus [23]), and also that $\frac{3}{2}$ is best possible here. This latter fact was observed by Erdős [17] in 1983 who remarked that there are graphs requiring $(\frac{3}{2} - o(1))n$ cycles and edges, likely referring to the following generalisation of an example of Gallai (see [18]). Take $k \in \mathbb{N}$ and consider the complete bipartite graph G with disjoint vertex classes A and B with $|A| = 2k + 1$ and $|B| = n - 2k - 1$. Each vertex in $|B|$ has odd degree in G, and $G[B]$ contains no edges, so any decomposition of G into cycles and edges must contain at least $|B|$ edges. As each cycle in G has length at most $2|A|$, there is thus no decomposition of G into fewer than

$$|B| + \frac{|A||B| - |B|}{2|A|} = \left(\frac{3}{2} - \frac{1}{2|A|}\right)|B| = \left(\frac{3}{2} - \frac{1}{4k + 2} - o(1)\right)n$$

cycles and edges.

Our tools. Many of our tools decompose an n-vertex graph into $O(n)$ cycles and a 'leftover' subgraph, and might prove useful towards settling the conjecture in full. For example, the proof of Lemma 26 shows that any n-vertex graph with no cycles longer than t decomposes into $O(n)$ cycles and $O(n \log^{274} t)$ edges. Furthermore, for any constant k, running our iteration argument using Lemma 26 $k+1$ times shows that any n-vertex graph can be decomposed into $O(kn)$ cycles and $O(n \log^{[k]} n)$ edges (see (19)). This latter result reduces the Erdős-Gallai conjecture to the case of arbitrarily sparse graphs, although this seems only to focus on the most difficult case. The main bottleneck in our argument seems to be the number of edges left uncovered in the almost decomposition into robust expanders, i.e. when applying Lemma 14. It appears hard to reduce the number of edges enough to make an improvement on Theorem 2 while getting enough properties in the expansion (robust or otherwise) to aid any cycle decomposition.

Potential further applications. In addition to their application towards the Erdős-Gallai conjecture, we believe the tools developed here could be useful in other settings due to the variety of applications that have been found for sublinear expansion since its introduction in its original form by Komlós and Szemerédi, as well as the only very recent use of the robust sublinear expansion discussed in Section 3.1. In fact, one can use some of our ideas, specifically a more precise version of Proposition 12, to give a simpler proof of a result of Tomon [52] on finding rainbow cycles in properly coloured graphs, and give a different proof on a result of Wang [55] (itself an improvement of a result of Tomon [52]). In addition, Letzter [38] has very recently adapted some of our methods in order to find separating paths systems.

References

[1] R. Aharoni and P. Haxell. Hall's theorem for hypergraphs. *Journal of Graph Theory*, 35(2):83–88, 2000.

[2] N. Alon and J. H. Spencer. *The Probabilistic Method*. John Wiley & Sons, 2016.

[3] B. Awerbuch, M. Luby, A. Goldberg, and S. Plotkin. Network decomposition and locality in distributed computation. In *30th Annual Symposium on Foundations of Computer Science*, pages 364–369, 1989.

[4] I. Ben-Eliezer, M. Krivelevich, and B. Sudakov. Long cycles in subgraphs of (pseudo)random directed graphs. *J. Graph Theory*, 70(3):284–296, 2012.

[5] A. Blanché, M. Bonamy, and N. Bonichon. Gallai's path decomposition in planar graphs. *preprint arXiv:2110.08870*, 2021.

[6] J. A. Bondy. Small cycle double covers of graphs. In *Cycles and rays (Montreal, PQ, 1987)*, volume 301 of *NATO Adv. Sci. Inst. Ser. C: Math. Phys. Sci.*, pages 21–40. Kluwer Acad. Publ., Dordrecht, 1990.

[7] A. Z. Broder, A. M. Frieze, S. Suen, and E. Upfal. An efficient algorithm for the vertex-disjoint paths problem in random graphs. In *Proceedings of the seventh annual ACM-SIAM symposium on Discrete algorithms*, pages 261–268, 1996.

[8] D. Chakraborti, J. Kim, J. Kim, M. Kim, and H. Liu. Well-mixing vertices and almost expanders. *preprint arXiv:2108.12864*, 2021.

[9] F. R. K. Chung. On the coverings of graphs. *Discrete Math.*, 30(2):89–93, 1980.

[10] C. J. Colbourn and P. C. van Oorschot. Applications of combinatorial designs in computer science. *ACM Comput. Surv.*, 21(2):223–250, jun 1989.

[11] D. Conlon, J. Fox, and B. Sudakov. Cycle packing. *Random Structures & Algorithms*, 45(4):608–626, 2014.

[12] N. Dean. What is the smallest number of dicycles in a dicycle decomposition of an Eulerian digraph? *Journal of Graph Theory*, 10(3):299–308, 1986.

[13] N. Dean and M. Kouider. Gallai's conjecture for disconnected graphs. *Discrete Mathematics*, 213(1-3):43–54, 2000.

[14] P. Erdős. Some unsolved problems in graph theory and combinatorial analysis. In *Combinatorial Mathematics and its Applications (Proc. Conf., Oxford, 1969)*, pages 97–109. Academic Press, London, 1971.

[15] P. Erdős. Problems and results in combinatorial analysis. In *Colloq. Internat. Theor. Combin. Rome*, pages 3–17, 1973.

[16] P. Erdős. On the combinatorial problems which I would most like to see solved. *Combinatorica*, 1(1):25–42, 1981.

[17] P. Erdős. On some of my conjectures in number theory and combinatorics. In *Proceedings of the fourteenth Southeastern conference on combinatorics, graph theory and computing*, volume 39, pages 3–19, 1983.

[18] P. Erdős, A. W. Goodman, and L. Pósa. The representation of a graph by set intersections. *Canadian Journal of Mathematics*, 18:106–112, 1966.

[19] G. Fan. Subgraph coverings and edge switchings. *Journal of Combinatorial Theory, Series B*, 84(1):54–83, 2002.

[20] G. Fan. Covers of Eulerian graphs. *Journal of Combinatorial Theory, Series B*, 89(2):173–187, 2003.

[21] I. Gil Fernández, J. Kim, Y. Kim, and H. Liu. Nested cycles with no geometric crossings. *Proceedings of the American Mathematical Society, Series B*, 9(03):22–32, 2022.

[22] I. Gil Fernández and H. Liu. How to build a pillar: a proof of Thomassen's conjecture. *preprint arXiv:2201.07777*, 2022.

[23] A. Girão, B. Granet, D. Kühn, and D. Osthus. Path and cycle decompositions of dense graphs. *Journal of the London Mathematical Society*, 2021.

[24] R. Glebov. *On Hamilton cycles and other spanning structures*. PhD thesis, Freie Universität Berlin, 2013.

[25] S. Glock, D. Kühn, and D. Osthus. Optimal path and cycle decompositions of dense quasirandom graphs. *Journal of Combinatorial Theory, Series B*, 118:88–108, 2016.

[26] J. Haslegrave, J. Hu, J. Kim, H. Liu, B. Luan, and G. Wang. Crux and long cycles in graphs. *preprint arXiv:2107.02061*, 2021.

[27] J. Haslegrave, J. Hyde, J. Kim, and H. Liu. Ramsey numbers of cycles versus general graphs. *preprint arXiv:2112.03893*, 2021.

[28] J. Haslegrave, J. Kim, and H. Liu. Extremal density for sparse minors and subdivisions. *International Mathematics Research Notices*, 2021.

[29] P. E. Haxell. Tree embeddings. *Journal of Graph Theory*, 36(3):121–130, 2001.

[30] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561, 2006.

[31] T. Jiang, A. Methuku, and L. Yepremyan. Rainbow Turán number of clique subdivisions. *arXiv preprint arXiv:2106.13803*, 2021.

[32] S. Jukna. On graph complexity. *Combin. Probab. Comput.*, 15(6):855–876, 2006.

[33] J. Kim, H. Liu, M. Sharifzadeh, and K. Staden. Proof of Komlós's conjecture on Hamiltonian subsets. *Proc. Lond. Math. Soc. (3)*, 115(5):974–1013, 2017.

[34] J. Komlós and E. Szemerédi. Topological cliques in graphs. *Combinatorics, Probability and Computing*, 3(2):247–256, 1994.

[35] J. Komlós and E. Szemerédi. Topological cliques in graphs II. *Combinatorics, Probability and Computing*, 5(1):79–90, 1996.

[36] D. Korándi, M. Krivelevich, and B. Sudakov. Decomposing random graphs into few cycles and edges. *Combin. Probab. Comput.*, 24(6):857–872, 2015.

[37] M. Krivelevich. Long cycles in locally expanding graphs, with applications. *Combinatorica*, 39(1):135–151, 2019.

[38] S. Letzter. Separating paths systems of almost linear size. *manuscript*, 2022.

[39] S. Letzter and A. Girão. Immersion of complete digraphs in eulerian digraphs. *Israel Journal of Mathematics*, to appear.

[40] H. Liu and R. Montgomery. A proof of Mader's conjecture on large clique subdivisions in C_4-free graphs. *Journal of the London Mathematical Society*, 95(1):203–222, 2017.

[41] H. Liu and R. Montgomery. A solution to Erdős and Hajnal's odd cycle problem. *preprint arXiv:2010.15802*, 2020.

[42] H. Liu, G. Wang, and D. Yang. Clique immersion in graphs without fixed bipartite graph. *preprint arXiv:2011.10961*, 2020.

[43] L. Lovász. On covering of graphs. In *Theory of Graphs (Proc. Colloq., Tihany, 1966)*, pages 231–236. Academic Press New York, 1968.

[44] É. Lucas. *Récréations mathématiques*, volume 2. Gauthier-Villars, 1883.

[45] R. Montgomery. Logarithmically small minors and topological minors. *Journal of the London Mathematical Society*, 91(1):71–88, 2015.

[46] R. Montgomery. Spanning cycles in random directed graphs. *arXiv preprint arXiv:2103.06751*, 2021.

[47] L. Pyber. An Erdős—Gallai conjecture. *Combinatorica*, 5(1):67–79, 1985.

[48] L. Pyber. Covering the edges of a graph by …. In *Sets, graphs and numbers (Budapest, 1991)*, volume 60 of *Colloq. Math. Soc. János Bolyai*, pages 583–610. North-Holland, Amsterdam, 1992.

[49] L. Pyber. Covering the edges of a connected graph by paths. *Journal of Combinatorial Theory, Series B*, 66(1):152–159, 1996.

[50] A. Shapira and B. Sudakov. Small complete minors above the extremal edge density. *Combinatorica*, 35(1):75–94, 2015.

[51] B. Sudakov and I. Tomon. The extremal number of tight cycles. *International Mathematics Research Notices*, 2022(13):9663–9684, 2022.

[52] I. Tomon. Robust (rainbow) subdivisions and simplicial cycles. *preprint arXiv:2201.12309*, 2022.

[53] O. Veblen. An application of modular equations in analysis situs. *Ann. of Math. (2)*, 14(1-4):86–94, 1912/13.

[54] O. Veblen. The Cambridge Colloquium, 1916, part ii. analysis situs. *Bull. Amer. Math. Soc*, 30:357–358, 1924.

[55] Y. Wang. Rainbow clique subdivisions. *preprint arXiv:2204.08804*, 2022.

[56] L. Yan. *On path decompositions of graphs*. PhD thesis, Arizona State University, 1998.