Nearly Hamilton cycles in sublinear expanders, and applications

Shoham Letzter*   Abhishek Methuku†   Benny Sudakov‡

Abstract

We develop novel methods for constructing nearly Hamilton cycles in sublinear expanders with good regularity properties, as well as new techniques for finding such expanders in general graphs. These methods are of independent interest due to their potential for various applications to embedding problems in sparse graphs. In particular, using these tools, we make substantial progress towards a twenty-year-old conjecture of Verstraëte, which asserts that for any given graph F, nearly all vertices of every d-regular graph G can be covered by vertex-disjoint F-subdivisions. This significantly extends previous work on the conjecture by Kelmans, Mubayi and Sudakov, Alon, and Kühn and Osthus. Additionally, we present applications of our methods to two other problems.

1   Introduction

A *Hamilton cycle* in a graph G is a cycle passing through all vertices of G. A graph is called *Hamiltonian* if it admits a Hamilton cycle. Hamiltonicity is a central notion in graph theory. Since deciding whether a given graph contains a Hamilton cycle is known to be NP-complete, much effort has been devoted to obtaining sufficient conditions for the existence of a Hamilton cycle, for example see [2, 17, 19, 20, 23] and the surveys [31, 54]. Most existing Hamiltonicity conditions, such as Dirac's theorem [21], are typically applicable only to very dense graphs. Therefore, identifying Hamiltonicity conditions that can also be applied to sparse graphs is a topic of significant interest. Over the last 50 years, a major focus of research has been understanding Hamiltonicity in sparse random graphs. Erdős and Rényi [26] posed the foundational question of determining the threshold probability for Hamiltonicity in random graphs. After a series of efforts by various researchers, including Korshunov [46] and Pósa [66], the problem was ultimately resolved by Komlós and Szemerédi [43], and independently by Bollobás [9].

1.1   Long cycles in Expanders

Since Hamiltonicity in random graphs is well understood, a key area of research is to look for Hamilton cycles in deterministic graphs that satisfy 'pseudorandom' conditions which enable them to mimic the properties of random graphs. This line of inquiry is particularly valuable for various applications such as Hamiltonicity in random Cayley graphs and Alon and Bourgain's work on additive patterns in multiplicative subgroups [5]. A well-known class of pseudorandom graphs, introduced by Alon, is defined using spectral properties as follows. A graph G is an (n, d, λ)-*graph* if it is d-regular with n vertices and the second largest eigenvalue of G in absolute value is at most λ. In 2003, Krivelevich and Sudakov, in their influential paper [50], proved that if d is sufficiently larger than λ, then every (n, d, λ)-graph is Hamiltonian. In the same paper, they conjectured that there exists $C > 0$ such that if $\frac{d}{\lambda} \geq C$, then every (n, d, λ)-graph is Hamiltonian. Shortly after this conjecture was stated, several papers, such as [11], considered an even stronger conjecture, singling out the key properties of (n, d, λ)-graphs that were believed to be instrumental

in demonstrating their Hamiltonicity. This stronger conjecture asserts that there exists a constant $C > 0$ such that every 'C-expander' is Hamiltonian. A graph G is called a C-expander if it satisfies the following two properties: for every subset $X \subseteq V(G)$ with $|X| < \frac{n}{2C}$, the neighbourhood of X satisfies $|N(X)| \geq C|X|$, and there is an edge between any two disjoint sets of at least $\frac{n}{2C}$ vertices. Despite significant attention (see, e.g., [11, 30, 36, 47, 51]) and many motivating applications, these two conjectures were only recently resolved by Draganić, Montgomery, Munhá Correia, Pokrovskiy and Sudakov [23].

Graph expansion is a fundamental concept in graph theory and computer science, with a wide range of applications; see, for example, the comprehensive survey by Hoory, Linial, and Wigderson [37]. Most of the expanders studied in the literature are constant expanders, defined by their linear expansion (such as the C-expanders discussed earlier). More formally, such graphs G satisfy the property $|N_G(U)| \geq \lambda|U|$ for any subset $U \subseteq V(G)$ that is not too large and not too small, where the expansion factor λ is some strictly positive absolute constant independent of G. Sublinear expansion is a weaker notion of this classical expansion introduced by Komlós and Szemerédi in the '90s [44, 45], and characterised by taking a much smaller value of λ. More precisely, if G is a sublinear expander of order n, then λ can be taken to satisfy $\lambda = \Omega(\frac{1}{(\log n)^2})$. Although sublinear expanders exhibit weaker expansion properties, their key advantage is that they can be found in essentially any graph. This notion has played a central role in the recent resolution of several long-standing conjectures (see, e.g. [6, 13, 14, 59, 60, 64] for notable examples and the survey [57] for a rather comprehensive list).

The study of cycles in expanders is a key area of research; see, for example, [29, 34, 48]. Notably, a classic result by Krivelevich [49] establishes that every n-vertex expander with expansion factor λ contains a cycle of length $\Omega(\lambda n)$. Hence, every n-vertex sublinear expander with expansion factor, say $\lambda = \frac{1}{(\log n)^c}$, contains a cycle of length $\Omega(\frac{n}{(\log n)^c})$. In general, we cannot necessarily guarantee a linear-sized cycle in sublinear expanders, as shown by the imbalanced complete bipartite graph $K_{n,\lambda n}$. In this paper, we prove that, somewhat surprisingly, sublinear expanders with reasonably good regularity properties admit a nearly Hamilton cycle; see Lemma 2.2. We also show how to find such expanders in general graphs.

Using these techniques, we make significant progress towards resolving a long-standing conjecture of Verstraëte from 2002 on packing subdivisions in regular graphs, which we address in the next subsection. Our second application concerns the well-known conjecture of Magnant and Martin [63] from 2009, which asserts that any d-regular graph can be partitioned into $n/(d+1)$ paths. Recently, Montgomery, Müyesser, Pokrovskiy, and Sudakov [65] asymptotically confirmed this conjecture by showing that nearly all vertices of a d-regular graph can be partitioned into $\frac{n}{d+1}$ paths. As a simple consequence of our methods, we show that nearly all vertices of a d-regular graph with sufficiently large degree can actually be partitioned into $\frac{n}{d+1}$ *cycles* (see the discussion following Conjecture 1.1 for further details on this conjecture). Finally, our methods can also be used to find a cycle with many chords, addressing a question of Chen, Erdős, and Staton [16] from 1996, and recovering—up to a slightly weaker polylogarithmic factor—a recent result by Draganić, Methuku, Munhá Correia, and Sudakov [22] (see Section 8 for details on both applications). Given the prominence of sublinear expanders, these new tools are likely to find further applications in future research.

1.2 Packing subgraphs in regular graphs

Packings in graphs have been extensively studied. Given two graphs H and G, an H-packing in G is a collection of vertex-disjoint copies of H in G. An H-packing in G is called *perfect* if it covers all of the vertices of G. The celebrated Hajnal–Szemerédi theorem [33] from 1970 states that every graph whose order n is divisible by t and whose minimum degree is at least $(1 - \frac{1}{t})n$ contains a perfect K_t-packing. (The case $k = 3$ was proved earlier by Corrádi and Hajnal [18].)

This theorem is best possible in the sense that the bound on minimum degree cannot be lowered. For non-complete graphs H, a series of papers including Alon and Yuster [8], Komlós, Sárközy and Szemerédi [42], Komlós [41], determined the minimum degree thresholds which force a perfect H-packing in a graph, culminating in the work of Kühn and Osthus [53], who essentially settled the problem by giving the best possible such condition (up to an additive constant) for any graph H, in terms of the so-called *critical chromatic number*.

In view of the above results, rather surprisingly, Kühn and Osthus [52] showed that if we restrict our attention to packings in regular graphs, then *any* linear bound on the minimum degree guarantees an almost perfect H-packing.

More precisely, they showed that for every bipartite graph H and every $0 < c, \alpha \leq 1$, every cn-regular graph G of sufficiently large order n has an H-packing which covers all but at most αn vertices of G. Resolving a conjecture of Kühn and Osthus [52], in an upcoming paper [58] the authors show that the bound αn on the number of uncovered vertices can actually be significantly lowered to obtain an H-packing which covers all but a constant number of vertices of G, which is clearly best possible, in the sense that there is not always an H-packing covering all vertices of G, and in fact, there are examples where the number of uncovered vertices grows with $|V(H)|$ and $\frac{1}{c}$.

The notion of subdivisions has played an important role in topological graph theory since the seminal result of Kuratowski [55] from 1930 showing that a graph is planar if and only if it does not contain a K_5-subdivision or a $K_{3,3}$-subdivision. Here, for a given graph F, an *F-subdivision* (or a *subdivision of F*), denoted by TF, is a graph obtained by replacing each edge uv in F by a path with ends u and v, such that the internal vertex sets of these paths are pairwise vertex-disjoint and vertex-disjoint from the original vertices of F. These original vertices of F are called the *branch vertices* of TF. One of the most classical results in this area is due to Mader [62], who showed that there is some $d = d(t)$ such that every graph with an average degree at least d contains a subdivision of the complete graph K_t. Mader [62], and independently Erdős and Hajnal [24] conjectured that $d(t) = O(t^2)$. In the '90s, Komlós and Szemerédi [44, 45] (using sublinear expanders), and independently, Bollobás and Thomason [10] (using different methods) confirmed this conjecture. Since then, various extensions and strengthenings of this result have been studied. For instance, an old conjecture of Thomassen asks for finding a *balanced* subdivision of K_t and a conjecture of Verstraëte asks for finding vertex-disjoint isomorphic subdivisions of K_t. Recently, these two conjectures have been resolved in [28, 59].

In 2002, Verstraëte [73] made the bold conjecture that every d-regular graph contains an almost perfect packing of subdivisions. More precisely, given graphs F and G, a TF-packing in G is a collection of pairwise vertex-disjoint copies of subdivisions of F in G (which are not required to be isomorphic). Verstraëte [73] observed that in any d-regular graph G, one can find a TF-packing which covers about half of the vertices of G (by repeatedly removing subdivisions of F from G until we obtain a graph containing no subdivisions of F) and made the following conjecture.

Conjecture 1.1 (Verstraëte [73], 2002). *For any graph F and any $\eta > 0$, there exists an integer $d_0 = d_0(F, \eta)$ such that, for all $d \geq d_0$, every d-regular graph G of order n contains a TF-packing that covers all but at most ηn vertices of G.*

Note that when F is a complete graph of order two or three, Conjecture 1.1 becomes a problem of covering the vertices of a regular graph with vertex-disjoint paths or cycles, respectively. Hence, this conjecture is a far-reaching extension of Petersen's 2-factor theorem (see [61]) which states that for $k \geq 1$, every $2k$-regular graph contains a 2-factor. Problems involving covering the edges or vertices of a graph with paths/cycles have been extensively studied. Perhaps the most famous open problem in this area is the linear arboricity conjecture of Akiyama, Exoo, and Harary [3] from 1980, which states that every graph with maximum degree Δ can be decomposed into at most $\lceil (\Delta + 1)/2 \rceil$ path forests. This is related to another well-known conjecture, posed by Magnant and Martin [63] in 2009, which states that the vertices of any d-regular graph of order n can be covered by at most $n/(d + 1)$ vertex-disjoint paths. Indeed, the linear arboricity conjecture implies Magnant and Martin's conjecture for odd d by noting that the largest path forest in the conjectured decomposition yields the desired collection of paths. This latter conjecture is still open; see [27, 32, 65] for some interesting recent progress towards it. Even the much weaker conjecture by Feige and Fuchs [27] that every d-regular graph of order n can be covered by at most $O(n/(d + 1))$ vertex-disjoint paths (which follows from the linear arboricity conjecture for all d) remains wide open.

Conjecture 1.1 was motivated by an old result of Jørgensen and Pyber [39] on covering the *edges* of a graph with subdivisions which actually implies that Conjecture 1.1 holds if we do not require the subdivisions of F to be vertex-disjoint (as observed by Kühn and Osthus in [52]). In the last twenty years, there have been many results showing that Conjecture 1.1 holds in several natural special cases. A result of Kelmans, Mubayi and Sudakov [40] shows that Conjecture 1.1 holds when F is a tree. In 2003, Alon [4] proved that Conjecture 1.1 holds when F is a cycle, using careful estimates on permanents. Alon [4] also remarked that this result can be extended to the case when F is a unicyclic graph but that it does not extend to the case of more complicated graphs F.

In 2005, Kühn and Osthus [52] proved that Conjecture 1.1 holds when G is dense (i.e., $d = \Omega(n)$). However, a significant obstacle in the way of proving the conjecture in full generality is that this proof relies on Szemerédi's

regularity lemma and the Blow-up lemma which only apply to dense graphs G. In this paper, we substantially improve on their results by showing that Conjecture 1.1 holds in the following stronger form for all graphs G with at least polylogarithmic average degree.

Theorem 1.2. *For any graph F and large enough n, every d-regular graph G of order n with $d \geq (\log n)^{130}$ contains a TF-packing that covers all but at most $\frac{n}{(\log \log n)^{1/30}}$ vertices of G.*

As mentioned earlier, our proof of Theorem 1.2 involves novel methods for finding nearly Hamilton cycles in sublinear expanders with good regularity and techniques for finding such expanders in general graphs. We give a detailed outline of our methods in Section 2. In particular, in Lemma 4.1, we show that almost all vertices of every d-regular n-vertex graph with $d \geq 2 \log n$ can be covered by nearly-regular sublinear expanders (see Section 3.2 for a formal definition of these expanders). A key feature of this lemma is that it allows us to control the regularity properties of the expanders we obtain. As a consequence, Corollary 4.2 shows that every n-vertex graph with average degree at least $\Omega(d \log n)$ contains a sublinear expander with maximum degree at most d and average degree extremely close to d. This result and our techniques for finding nearly Hamilton cycles in sublinear expanders have significant potential for further applications (see Section 8 for some examples).

In the dense case, where $d = \Omega(n)$, Kühn and Osthus [52] showed that one can even find a perfect TK_t-packing in G when $t = 4$ and $t = 5$. It follows from the work of Gruslys and Letzter [32] towards the aforementioned conjecture of Magnant and Martin [63] that this holds for $t = 2$ and $t = 3$. Kühn and Osthus posed the question of whether this result holds for $t \geq 6$. Recently, the authors [58] answered this question positively using techniques very different from those used in this paper. It is known that for all $t \geq 3$, we need $d \geq \sqrt{n/2}$ to have a perfect TK_t-packing in G; see [52]. It would be interesting to determine the exact degree threshold at which a perfect TK_t-packing can be guaranteed in regular graphs.

1.3 Organization of the paper

The rest of the paper is organized as follows. In Section 2 we give a detailed sketch of our proof of Theorem 1.2. In Section 3, we introduce the notation and the two notions of expansion used throughout the paper, along with the required probabilistic tools. In Section 4, we develop methods for finding expanders with good regularity properties and prove our first key lemma, Lemma 4.1, which shows that one can cover nearly all vertices of a regular graph with such expanders. In Section 5, we prove Lemma 5.1 which shows that a collection of vertex-disjoint pairs of vertices (satisfying a certain expansion property) can be joined using vertex-disjoint paths through a random vertex subset of a sublinear expander. In Section 6, we develop methods for constructing nearly Hamilton cycles in sublinear expanders and use them to prove our second key lemma, Lemma 6.1, which shows that any sufficiently regular sublinear expander contains an almost-spanning F-subdivision. In Section 7, we put everything together to prove Theorem 1.2. In Section 8, we present two additional applications of our methods: one addressing the conjecture of Magnant and Martin [63], and the other concerning the existence of a cycle with many chords [16].

2 Proof sketch

In this section, we sketch the main ideas in our proof of Theorem 1.2. Let F be a given graph and let G be a d-regular graph of sufficiently large order n. Our strategy involves covering nearly all vertices of G with vertex-disjoint sublinear expanders that are close to regular and finding an almost-spanning subdivision of F within each such expander. Specifically, we proceed as follows.

Step 1. Find a collection $\mathcal{H}$ of vertex-disjoint sublinear expanders with good regularity properties (that is, with average degree very close to the maximum degree), covering nearly all vertices of G.

Step 2. Show that each expander $H \in \mathcal{H}$ contains a nearly Hamilton path P_H.

Step 3. Find a subdivision of F in each expander $H \in \mathcal{H}$ containing the nearly Hamilton path P_H.

It is easy to see that the F-subdivisions given by **Step 3** together cover nearly all vertices of G. A key contribution of our paper is the development of novel techniques for finding nearly Hamilton paths and cycles in sublinear expanders with good regularity properties. Such expanders are also useful for various other applications. The meta-problem of finding sublinear expanders with good regularity properties was therefore raised in [14] by Chakraborti, Janzer, Methuku and Montgomery. Our work here also makes progress towards this problem.

A natural strategy for constructing a nearly Hamilton path in a sublinear expander $H \in \mathcal{H}$ is to start with a small collection of vertex-disjoint paths $P_1, \ldots P_r$ in H, that together cover almost all vertices in H, and connect these paths through a small random set $V_0 \in V(H)$ of reserved vertices. Here we need V_0 to be small enough because the vertices of V_0 that are not used in the connecting paths are left uncovered by the nearly Hamilton path that we aim to find.

More precisely, we start by partitioning $V(H)$ into random sets $V_0, X_1, \ldots, X_t$ such that $V_0 = o(|V(H)|)$ and the sets $X_1, \ldots, X_t$ have roughly the same size. Then we obtain the paths $P_1, \ldots, P_r$, by finding a largest matching M_i between each pair of sets X_i, X_{i+1}, $1 \le i \le t-1$, taking the union of these matchings, and letting $P_1, \ldots, P_r$ be the connected components in the union that intersect all sets X_i (so that each of the paths $P_1, \ldots, P_r$ contains exactly one vertex from each X_i). In order for the matchings M_i to be large enough so that the paths $P_1, \ldots, P_r$ cover nearly all of the vertices of H, it is crucial that H has very good regularity properties. In particular, we need the following property.

> **(P)** If $H \in \mathcal{H}$ is a sublinear expander with average degree at least $d(1 - \varepsilon)$ and maximum degree d, then $\varepsilon \ll \frac{1}{t}$.

Indeed, using Vizing's theorem and standard concentration inequalities, it is easy to show that with high probability $|M_i| \ge |X_i|(1 - 2\varepsilon)$ for each $1 \le i \le t-1$. This implies that the paths $P_1, \ldots, P_r$ cover all but at most $2\varepsilon|V(H) - V_0|$ vertices from each X_i, so that up to $2\varepsilon t$ proportion of the vertices in $V(H) - V_0$ may be left uncovered by the paths $P_1, \ldots, P_r$. To make this a small enough proportion of the vertices of H, we need $\varepsilon \ll \frac{1}{t}$ as stated in **(P)**.

For connecting the paths $P_1, \ldots, P_r$ using paths through the random set V_0, we use some ideas from recent work of Bucić and Montgomery [13] and Tomon [71], showing that random vertex subsets in sublinear expanders with polylogarithmic average degree are likely to inherit some expansion properties. These ideas were slightly refined in [14] (see, e.g., [14, Lemma 8]) to show that a collection of vertex-disjoint pairs of vertices can be connected through a random subset of vertices of a sublinear expander (using vertex-disjoint paths) provided that the size of the random subset is sufficiently large compared to the number of pairs of vertices. Crucially, this means that for connecting the paths $P_1, \ldots, P_r$ through the random set V_0, we need the sets $X_1, \ldots, X_t$ to be much smaller than V_0. Since $|X_i| \approx |V(H)|/t$ and V_0 must be small, this implies that $1/t$ must be very small, which, in turn, requires ε to be small enough for **(P)** to hold. This explains why we need the sublinear expanders $H \in \mathcal{H}$ to have extremely good regularity properties.

In any d-regular graph G on n vertices, it is easy to find a sublinear expander with an average degree at least $d(1 - 2\lambda \log n)$ and an expansion factor of $\lambda = O\left(\frac{1}{\log n}\right)$, using standard methods such as iteratively removing sparse cuts. In fact, one can cover nearly all vertices of our d-regular graph G with such sublinear expanders. By choosing λ sufficiently small, these expanders indeed exhibit the desired strong regularity properties. However, these expanders may contain only a small number of vertices of G, so the expansion characterized by this λ might be too weak for following the aforementioned strategy to complete **Step 2**. To address this, we introduce a *refining procedure*, detailed at the end of this section, that begins with these expanders and gradually enhances their expansion while essentially preserving their strong regularity properties. Roughly speaking, this enables us to prove the following lemma; for a precise statement, see Lemma 4.1.

Lemma 2.1. *For any $c_1 \ge 2$, there is a constant $c_2 > 0$ such that the following holds. Nearly all vertices of every d-regular graph G with sufficiently large degree can be covered by vertex-disjoint (robust) sublinear expanders H with an average degree of at least $d(1 - \frac{1}{(\log|V(H)|)^{c_1}})$ and an expansion factor of $\frac{1}{(\log|V(H)|)^{c_2}}$.*

This lemma shows that by taking c_1 sufficiently large, we can obtain sublinear expanders H with sufficiently strong regularity (although at the expense of slightly weaker expansion). Unfortunately, this improved regularity of our

expanders H is still insufficient to satisfy (**P**) because the methods of [13, 14] require each of the sets $X_1, \ldots, X_t$ to be significantly smaller than V_0 (depending on the parameter c_1) – see, e.g., the proof of [14, Lemma 8]. Therefore, Lemma 2.1 still does not allow us to directly connect the paths $P_1, \ldots, P_r$ through the random set V_0, making the construction of a nearly Hamilton path challenging with the strategy described above. To overcome this difficulty, our main idea is to iteratively connect the paths $P_1, \ldots, P_r$ through V_0 using a procedure that, in each iteration, either directly connects a good proportion of the paths or identifies a well-expanding subset of their leaves (see Figure 1). More precisely, in each iteration of the procedure, we first greedily connect the paths through V_0 using as many vertex-disjoint paths of length two as possible (a similar idea for connecting paths, in a different context, was recently used in [65]). Crucially, when it is no longer possible to connect using vertex-disjoint paths of length two, we can identify a small subset S of the leaves of the paths $P_1, \ldots, P_r$ that expands very well into V_0. The improved regularity of expanders H (provided by Lemma 2.1) together with a variant of a lemma from [56] is now sufficient to connect the paths whose leaves lie in the subset S (through V_0). This allows us to join a good proportion of the paths $P_1, \ldots, P_r$ through V_0 in each iteration. By iterating this process $\Theta(\log|V(H)|)$ times, we eventually obtain a nearly Hamilton path P_H in H, completing **Step 2**.
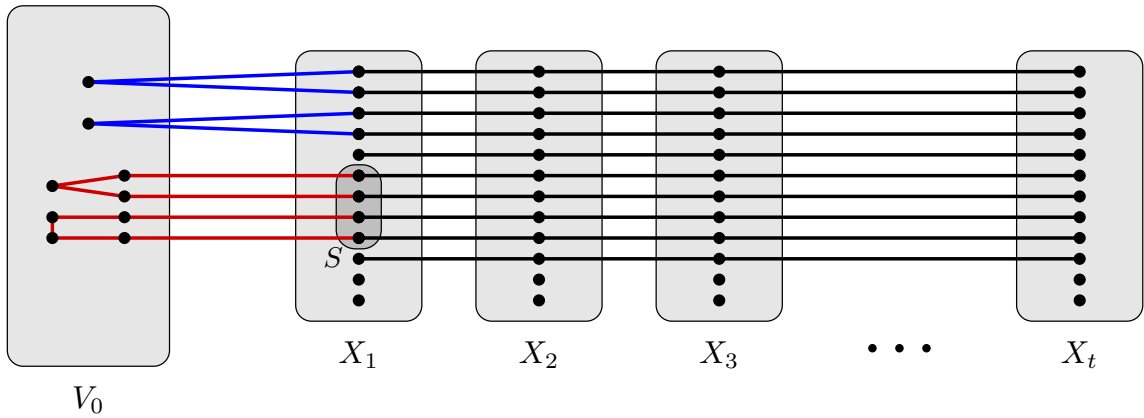
Figure 1: If we cannot find sufficiently many connecting paths of length two (shown in blue), we identify a set S of leaves that expands very well into V_0 and connect vertices in S using paths (shown in red) through V_0.

Since the procedure requires $\Theta(\log|V(H)|)$ iterations to connect the paths $P_1, \ldots, P_r$ through V_0, the sets $X_1, \ldots, X_t$ must be at least $\Theta(\log|V(H)|)$ times smaller than V_0. This constraint demands an even smaller choice of ε to satisfy (**P**). Nevertheless, by selecting c_1 sufficiently large in Lemma 2.1, we can still apply the iterative procedure to connect the paths and construct a nearly Hamilton path, even when the expanders H possess slightly weaker expansion properties. This yields the following lemma (see Lemma 6.1 for the precise formulation).

Lemma 2.2. *Let $c > 0$ be fixed, let $0 < \varepsilon < \frac{1}{(\log n)^4}$, let n be sufficiently large, and let $d \geq (\log n)^{10c+51}$. Then, every n-vertex (robust) sublinear expander H with an expansion factor of $\frac{1}{(\log n)^c}$, average degree at least $d(1 - \varepsilon)$ and maximum degree at most d, contains a cycle (and thus a path P_H) of length at least $n - \frac{n}{\log n}$.*

Next, we *absorb* the vertices of the path P_H into a subdivision of F that is found within a small random subset R of vertices in each of our expanders H (as illustrated in Figure 2). (Here, we use the aforementioned techniques for connecting vertex pairs, together with a classical result for finding subdivisions [10, 45].) This yields the desired almost-spanning subdivision of F in each of the expanders $H \in \mathcal{H}$, thus completing **Step 3** and the proof of Theorem 1.2.

As discussed earlier, sublinear expanders with strong regularity properties hold potential for a wide range of applications (some of which are discussed in Section 8). Given this independent interest, we conclude this section with a sketch of the proof of Lemma 2.1, showing that almost all vertices of a nearly regular graph can be covered with vertex-disjoint nearly regular expanders. We note that this lemma, together with known methods for 'regularising'

a graph, imply the existence of a nearly regular expander in every graph with sufficiently large average degree (see Corollary 4.2), a potentially very useful tool for applications.

A refining procedure. Let G be a graph with an average degree at least $d(1 - \varepsilon)$ and a maximum degree d. By repeatedly removing sparse cuts, it is easy to find an expander in G with an expansion factor of $\lambda = O(\frac{1}{\log n})$ and an average degree of at least $d(1 - 2\lambda \log n)$. By iteratively finding such expanders, removing their vertices and continuing with the remaining graph, we can cover nearly all vertices of the graph G using a collection $\mathcal{H}_1$ of vertex-disjoint expanders with an average degree of at least $d(1 - \varepsilon \log n)$ and an expansion factor of $\frac{\varepsilon}{\log n}$ (see Lemma 4.4). However, since these expanders may contain very few vertices compared to $n = |V(G)|$, the expansion factor $\frac{\varepsilon}{\log n}$ may represent only very weak expansion. Such a weak expansion is insufficient for most applications; in particular, we require a much stronger expansion to construct a nearly Hamilton cycle within these expanders. Let C be a large constant. If $\varepsilon = \frac{1}{(\log n)^{C-1}}$, then the expanders in $\mathcal{H}_1$ have an average degree of at least $d\left(1 - \frac{1}{(\log n)^{C-2}}\right)$ and an expansion factor of $\frac{1}{(\log n)^C}$. This means that, although the expanders in $\mathcal{H}_1$ may exhibit very weak expansion, they possess excellent regularity properties. We introduce a refining procedure that begins with the expanders in $\mathcal{H}_1$ and iteratively improves their expansion while largely preserving their strong regularity properties, ultimately producing expanders with both good expansion and regularity properties.

The main idea of this refining procedure is as follows. We carefully choose certain 'thresholds' $n_t \leq \cdots \leq n_1 = n$ for vertex sizes, where $\log n_{i+1} = (\log n_i)^{\frac{C-2}{C-1}}$, and regularity thresholds $\epsilon_1 \leq \cdots \leq \epsilon_t$, where $\epsilon_i = \frac{1}{(\log n_i)^{C-2}} = \frac{1}{(\log n_{i+1})^{C-1}}$. We iteratively construct a sequence of collections $\mathcal{H}_i$, $1 \leq i \leq t$, of expanders where each collection covers nearly all vertices of G, such that for each i, the expanders in $\mathcal{H}_{i+1}$ have slightly better expansion and only slightly weaker regularity properties compared to the expanders in the collection $\mathcal{H}_i$. We achieve this by 'refining' any expander $H \in \mathcal{H}_i$ that has too few vertices (using the aforementioned fact that any graph with average degree at least $d(1 - \varepsilon)$ and maximum degree d can be covered with vertex-disjoint expanders having an average degree of at least $d(1 - \varepsilon \log n)$ and an expansion factor of $\frac{\varepsilon}{\log n}$). More precisely, we replace any expander $H \in \mathcal{H}_i$ that has fewer than n_{i+1} vertices with a new collection $\mathcal{H}_{i+1}(H)$ of vertex-disjoint expanders that have improved expansion (and only slightly weaker regularity properties) covering nearly all vertices of H, and we let $\mathcal{H}_{i+1} = \bigcup_{H \in \mathcal{H}_i} \mathcal{H}_{i+1}(H)$ be the resulting collection of expanders. Crucially, the parameters are set up so that, if $H \in \mathcal{H}_i$ and $|V(H)| < n_{i+1}$, then the expanders in $\mathcal{H}_{i+1}(H)$ have an *improved* expansion factor of $\frac{1}{(\log n_{i+1})^C}$ while maintaining an average degree at least $d(1 - \varepsilon_{i+1})$. By repeating this refining procedure we eventually obtain a collection $\mathcal{H}_t = \mathcal{H}$ of expanders that cannot be refined further. This means that for every expander $H \in \mathcal{H}$, there is a step j at which it was last refined, where $n_j > |V(H)| \geq n_{j+1}$ and $H \in \mathcal{H}_j - \mathcal{H}_{j-1}$. It is then easy to see that H has an expansion factor of $\frac{1}{(\log n_j)^C} = \frac{1}{(\log n_{j+1})^{C(C-1)/(C-2)}} \geq \frac{1}{(\log |V(H)|)^{C(C-1)/(C-2)}}$, and an average degree of at least $d(1 - \varepsilon_j) = d(1 - \frac{1}{(\log n_j)^{C-2}}) \geq d(1 - \frac{1}{(\log |V(H)|)^{C-2}})$, as desired, proving Lemma 2.1.

3 Preliminaries

3.1 Notation

We write $c = a \pm b$ if $a - b \leq c \leq a + b$. For a set $U \subseteq V(G)$, let $\overline{U}$ denote the set $V(G) - U$. For a set $S \subseteq V(G)$, let $G - S$ denote the subgraph of G induced by $V(G) - S$. By $e(G)$, we denote the number of edges of G, and for $S \subseteq V(G)$, we denote by $e_G(S)$ the number of edges of G induced by S. For two disjoint sets $A, B \subseteq V(G)$, let $e_G(A, B)$ denote the number of edges of G which are incident to both A and B.

For a graph G, we denote by $d(G)$ its average degree, by $\delta(G)$ its minimum degree and by $\Delta(G)$ its maximum degree. For a vertex $v \in V(G)$, we denote its degree by $d_G(v)$ and the set of its neighbours by $N_G(v)$. For a set of vertices $X \subseteq V(G)$, and an integer $i \geq 0$, let $N_G(X)$ denote the set of vertices outside of X that are adjacent to at least one vertex of X, and we write $B_G^i(X)$ to be the set of vertices at distance at most i from X in G. We often omit subscripts and write, e.g. $N(X)$ instead of $N_G(X)$, if it is clear from the context which graph we are working with. Given vertices x, y, an (x, y)-path is a path from x to y.

All logarithms in this paper are base 2. When dealing with large numbers, we often omit floor and ceiling signs whenever they are not crucial.

3.2 Expansion

In this paper, we use two different notions of expansion that are closely related to each other. Here, we introduce these two notions and then prove a lemma that relates them. We will use the following standard notion of edge expansion.

Definition 3.1 (λ-expander)**.** Let $\lambda > 0$. We say that a graph H is a λ-expander if every set $U \subseteq V(H)$ with $|U| \leq \frac{1}{2}|V(H)|$ satisfies $e(U, \overline{U}) \geq \lambda d(H)|U|$.

We will mostly use the following notion of vertex expansion, which is a generalization of the notion of expansion introduced by Bucić and Montgomery [13]. Related notions of expansion were recently developed by Shapira and Sudakov [69], by Haslegrave, Kim, and Liu [35], and by Sudakov and Tomon [70].

Definition 3.2 ((ε, c, s)-expander)**.** An graph H is called an (ε, c, s)-expander if, for every $U \subseteq V(H)$ and $F \subseteq E(H)$ with $1 \leq |U| \leq \frac{2}{3}|V(H)|$ and $|F| \leq s|U|$, we have

$$|N_{H-F}(U)| \geq \frac{\varepsilon}{(\log |V(H)|)^c} \cdot |U|. \tag{1}$$

Expanders as in Definition 3.2 are sometimes called as *robust* sublinear expanders, referring to the graph F (Komlós and Szemerédi's definition of sublinear expanders did not include this feature). Notice that the expander becomes more 'robust' the larger s is.

The main difference between the notion of expansion we use (as given in Definition 3.2) and the one used in [13] is that we use an additional parameter c for our expanders (which is set to 2 in [13]). This parameter measures the rate of expansion, where larger values of c correspond to weaker expansion. Our approach in this paper relies crucially on expanders with excellent regularity properties, specifically those with average degree very close to maximum degree. In Lemma 4.3, we construct such expanders, which allows us to enhance their regularity by increasing the parameter c which, however, leads to a slight weakening of the expansion.

The following lemma connects the two notions of expansion.

Lemma 3.3. *Let $0 < \varepsilon \leq 1/4$, $c > 0$, let n be large and let $\lambda = \frac{1}{(\log n)^c}$. Let H be an n-vertex λ-expander with $\Delta(H) \leq d$ and $d(H) \geq d(1 - \varepsilon)$. Then H is a $(\frac{1}{8}, c, \frac{\lambda d}{4})$-expander.*

Proof. Let $U \subseteq V(H)$ with $1 \leq |U| \leq \frac{2}{3}n$, and let $F \subseteq E(H)$ with $|F| \leq \frac{\lambda d}{4}|U|$.

First consider the case when $1 \leq |U| \leq n/2$. Then, since H is a λ-expander, we have $e_H(U, \overline{U}) \geq \lambda d(H)|U|$. Therefore, $e_{H-F}(U, \overline{U}) \geq (\lambda d(H) - \frac{\lambda d}{4})|U| \geq (\lambda d(1 - \varepsilon) - \frac{\lambda d}{4})|U| \geq \frac{\lambda d}{4}|U|$. Since $\Delta(H) \leq d$, this implies that

$$|N_{H-F}(U)| \geq \frac{\lambda}{4}|U| \geq \frac{|U|}{8(\log n)^c}.$$

Now consider the case when $n/2 \leq |U| \leq 2n/3$. So $|U|/2 \leq n/3 \leq |\overline{U}| \leq n/2$. Then, $e_{H-F}(U, \overline{U}) \geq \lambda d(H)|\overline{U}| - \frac{\lambda d}{4}|U| \geq \frac{\lambda d(H)}{2}|U| - \frac{\lambda d}{4}|U| \geq (\frac{\lambda d(1 - \varepsilon)}{2} - \frac{\lambda d}{4})|U| \geq \frac{\lambda d}{8}|U|$. Since $\Delta(H) \leq d$, we have

$$|N_{H-F}(U)| \geq \frac{\lambda}{8}|U| = \frac{|U|}{8(\log n)^c}.$$

Therefore, H is a $(\frac{1}{8}, c, \frac{\lambda d}{4})$-expander, proving the lemma. $\qquad\square$

3.3 Probabilistic tools

We will often use a basic version of Chernoff's inequality for the binomial random variable (see, for example, [7, 72]).

Theorem 3.4 (Chernoff's bound)**.** *Let n be an integer, let $0 \leq p \leq 1$, let $X \sim \text{Bin}(n, p)$, and let $\mu = \mathbb{E}X = np$. Then the following hold.*

(i) *If $0 < \delta < 1$, then*

$$\mathbb{P}(X \le (1 - \delta)\mu) \le e^{-\frac{\delta^2 \mu}{2}}.$$

(ii) *If $\delta \ge 0$, then*

$$\mathbb{P}(X \ge (1 + \delta)\mu) \le e^{-\frac{\delta^2 \mu}{(2+\delta)}}.$$

Additionally, we use the following martingale concentration result (see Chapter 7 in [7]). We say that a function $f : \prod_{i=1}^n \Omega_i \to \mathbb{R}$, where Ω_i are arbitrary sets, is *k-Lipschitz* if $|f(u) - f(v)| \le k$ for every $u, v \in \prod_{i=1}^n \Omega_i$ that differ on at most one coordinate.

Lemma 3.5. *Let $X_1, \ldots, X_n$ be independent random variables, with X_i taking values in a set Ω_i for $i \in [n]$, and write $X = (X_1, \ldots, X_n)$. Suppose that $f : \prod_{i=1}^n \Omega_i \to \mathbb{R}$ is k-Lipschitz. Then,*

$$\mathbb{P}\big(|f(X) - \mathbb{E}f(X)| > t\big) \le 2\exp\left(\frac{-t^2}{2k^2 n}\right).$$

3.4 A regularisation theorem

Our main theorem (Theorem 1.2) is stated for regular graphs G with sufficiently large degree. However, for other applications of our methods (such as the one discussed in Section 8), the graph G does not need to be regular. In such cases, it is useful to extract a regular subgraph whose degree remains close to the average degree of G. We achieve this using a result by Chakraborti, Janzer, Methuku, and Montgomery [15].

Theorem 3.6. *There exists a constant γ such for all positive integers r, n with $r \le n/2$, every n-vertex graph with average degree at least $\gamma r \log(n/r)$ contains an r-regular subgraph.*

In [15], Theorem 3.6 and a related result for the case when r is small relative to n were used to resolve a problem of Rödl and Wysocka [68] from 1997 for almost all r and to obtain tight bounds for the Erdős-Sauer problem [25] (up to an absolute constant factor), improving the recent breakthrough of Janzer and Sudakov [38], who had resolved the problem up to a constant depending on r. For our application in this paper, it suffices to find an almost-regular subgraph with a large degree rather than a fully regular subgraph. This can be achieved, at the cost of an additional polylogarithmic factor, using a simple lemma by Bucić, Kwan, Pokrovskiy, Sudakov, Tran, and Wagner [12], which builds on methods developed by Pyber [67].

4 Packing a regular graph with nearly regular expanders

In this section, we prove our first key lemma, Lemma 4.1, which shows that one can find sufficiently regular, vertex-disjoint expanders covering almost all vertices of any regular graph with sufficiently large degree. An important feature of this lemma is that it allows us to control the regularity properties of the expanders we obtain (via the parameter C), which is crucial for constructing a nearly Hamilton cycle in each of these expanders (as explained in the proof sketch, see Section 2). As mentioned earlier, expanders with good regularity properties are useful for various other applications, and the question of finding such expanders was raised by Chakraborti, Janzer, Methuku, and Montgomery in [14]. Our lemma below makes progress on this problem.

Lemma 4.1. *Let $\alpha > 0$ be a fixed real number, let ε, C, n, d satisfy $d \ge 2\log n$, $0 \le \varepsilon \le (\log n)^{-(C-1)}$, $C \ge \max\{28\alpha + 3, 56\alpha + 1\}$ and let n be large enough. Let $c = \frac{C(C-1)}{C-28\alpha-1}$.*

Suppose that G is an n-vertex graph with $\Delta(G) \le d$ and $d(G) \ge d(1 - \varepsilon)$. Then there is a collection $\mathcal{H}$ of vertex-disjoint subgraphs of G such that every $H \in \mathcal{H}$ is a $(\frac{1}{8}, c, s_H)$-expander satisfying $d(H) \ge d(1 - \varepsilon_H)$ and $\delta(H) \ge d(H)/2$, where $s_H := \frac{d}{4(\log |V(H)|)^c}$, and $\varepsilon_H := (\log |V(H)|)^{-(C-28\alpha-1)}$. Moreover, $\sum_{H \in \mathcal{H}} |V(H)| \ge (1 - \frac{(\log \log \log n)^2}{(\log \log n)^\alpha})n.$

Note that increasing the parameter C in Lemma 4.1 improves the regularity properties of the expanders we obtain at the cost of slightly weakening their expansion. Indeed, in our proof we let $\alpha = \frac{1}{28}$ (though this choice is somewhat arbitrary), so that $c = \frac{C(C-1)}{C-2}$. Hence, as C increases, it is easy to see that $\varepsilon_H = (\log|V(H)|)^{-(C-2)}$ decreases (improving the regularity properties of our expanders) and c increases (making their expansion weaker).

In certain cases, we do not require the full strength of Lemma 4.1 and instead seek a single expander that is nearly regular. In such situations, we can actually remove the near-regularity assumption on G at the cost of a slightly worse lower bound on the average degree of G, as shown below.

Corollary 4.2. *There is a constant γ such that the following holds for all sufficiently large n and for C, d satisfying $C \geq 4$, $d \geq 2\log n$. Suppose that G is an n-vertex graph with $d(G) \geq \gamma d \log n$. Then there is a subgraph $H \subseteq G$ which is a $(\frac{1}{8}, c, s)$-expander satisfying $\Delta(H) \leq d$, $d(H) \geq d(1-\mu)$ and $\delta(H) \geq d(H)/2$, where $c := \frac{C(C-1)}{C-2}$, $s := \frac{d}{4(\log|V(H)|)^c}$, and $\mu := (\log|V(H)|)^{-(C-2)}$.*

Proof of Corollary 4.2 using Lemma 4.1. Apply the regularisation theorem, Theorem 3.6, to G to obtain a d-regular subgraph G'. Now apply Lemma 4.1 to the graph G' (with $\alpha = \frac{1}{28}$ and $\varepsilon = 0$), and let H be any graph in the collection $\mathcal{H}$ guaranteed by the lemma. Then, it is easy to see that H has the desired properties. $\qquad\square$

Remark. Notice that in the proof of Corollary 4.2 we used Theorem 3.6, due to Chakraborti, Janzer, Methuku and Montgomery [15], which gives a tight bound on the number of edges in an n-vertex graph needed to guarantee the existence of a d-regular subgraph. It is possible to use more elementary tools to prove a slightly weaker version of Corollary 4.2 (where $d(G)$ is required to be at least $5d(\log n)^C$ instead of $\gamma d \log n$). Indeed, using [12, Lemma 2.2] due to Bucić, Kwan, Pokrovskiy, Sudakov, Tran and Wagner, one can prove that every n-vertex graph with sufficiently large average degree contains a subgraph which is close to being d-regular, similarly to the proof of [12, Lemma 2.1].

We prove Lemma 4.1 by combining Lemma 3.3 with the following variant of Lemma 4.1 for edge expanders (rather than vertex expanders).

Lemma 4.3. *Let $\alpha > 0$ be a fixed real number, let ε, C, n, d satisfy $d \geq 2\log n$, $0 \leq \varepsilon \leq (\log n)^{-(C-1)}$, $C \geq \max\{28\alpha + 3, 56\alpha + 1\}$, and let n be large enough. Let $c = \frac{C(C-1)}{C-28\alpha-1}$.*

Let G be an n-vertex graph with $\Delta(G) \leq d$ and $d(G) \geq d(1-\varepsilon)$. Then there is a collection $\mathcal{H}$ of vertex-disjoint subgraphs of G such that every $H \in \mathcal{H}$ is a λ_H-expander satisfying $d(H) \geq d(1-\varepsilon_H)$ and $\delta(H) \geq d(H)/2$, where $\lambda_H := (\log|V(H)|)^{-c}$ and $\varepsilon_H := (\log|V(H)|)^{-(C-28\alpha-1)}$. Moreover, $\sum_{H\in\mathcal{H}} |V(H)| \geq (1 - \frac{(\log\log\log n)^2}{(\log\log n)^\alpha})n$.

The rest of this section is dedicated to proving Lemma 4.3. To that end, in Section 4.1, we first construct an almost-perfect packing using expanders that may exhibit very weak expansion but possess excellent regularity properties. Then, in Section 4.2, we introduce a refining procedure that begins with these expanders and iteratively improves their expansion while largely preserving their strong regularity properties, ultimately producing expanders with both good expansion and regularity properties, thereby proving Lemma 4.3.

4.1 Packing with very weak expanders

In this subsection, we prove the following lemma which shows that one can cover almost all vertices of an n-vertex graph G (with average degree at least $d(1-\varepsilon)$ and maximum degree d) using vertex-disjoint λ-expanders having expansion factor $\lambda = O(\frac{\varepsilon}{\log n})$ and having average degree very close to that of G. Since these expanders could have very few vertices compared to n, this λ may indicate very weak expansion (which is not sufficient for our purpose), but crucially, these expanders have good regularity properties if we pick ε small enough. This is crucial to our refining procedure in Section 4.2 which starts with these expanders and iteratively improves their expansion while only slightly weakening their regularity properties.

Lemma 4.4. *Let $\alpha > 0$ be a fixed real number and let ε, λ, n satisfy $\lambda \log n \leq \min\{\frac{1}{10}, \varepsilon\}$ and let n be sufficiently large. Let G be an n-vertex graph with $\Delta(G) \leq d$ and $d(G) \geq d(1-\varepsilon)$. Then there exists a collection $\mathcal{H}$ of vertex-disjoint λ-expanders in G such that $d(H) \geq d(1-\varepsilon(\log n)^{28\alpha})$ and $\delta(H) \geq d(H)/2$ for every $H \in \mathcal{H}$ and $\sum_{H\in\mathcal{H}} |V(H)| \geq (1 - \frac{1}{(\log n)^\alpha})n$.*

We build up to the proof of Lemma 4.4 using a sequence of lemmas as follows. First, in Lemma 4.5, we show how to find one λ-expander in a sufficiently regular graph G using standard methods (see, e.g., [22, 69]). Then, by iteratively applying Lemma 4.5, we show how to cover a good proportion of vertices of G with vertex-disjoint λ-expanders in Lemma 4.6. Finally, by iteratively applying Lemma 4.6 we show how to cover almost all vertices of G with vertex-disjoint λ-expanders (with average degree very close to that of G), proving Lemma 4.4.

Lemma 4.5. *Let G be an n-vertex graph with $d(G) = d$, and let $\lambda \log n \leq \frac{1}{10}$. Then G contains a λ-expander H with $d(H) \geq d(1 - 2\lambda \log n)$ and $\delta(H) \geq d(H)/2$.*

Proof. We perform a procedure which finds the desired λ-expander H in G. Before we can describe the procedure, we need the following claim.

Claim 4.5.1. *Let F be a subgraph of G that is not a λ-expander. Then there is a non-empty set $U \subseteq V(F)$ with $|U| \leq \frac{|V(F)|}{2}$ such that either $d(F[\overline{U}]) \geq d(F)$ or $d(F[U]) \geq (1 - 2\lambda)d(F)$.*

Proof of claim. Since F is not a λ-expander, there is a non-empty set $U \subseteq V(F)$ with $|U| \leq \frac{|V(F)|}{2}$ such that $e(U, \bar{U}) < \lambda d(F)|U|$. We will show that U is the desired subset. Suppose for a contradiction that $d(F[\overline{U}]) < d(F)$ and $d(F[U]) < (1 - 2\lambda)d(F)$. Then, using the bound on $e(U, \bar{U})$, we have $e(F) = e_F(U) + e_F(U, \overline{U}) + e_F(\overline{U}) < |U|d(F)\left(\frac{1-2\lambda}{2} + \lambda\right) + |\overline{U}|\frac{d(F)}{2} = |V(F)|\frac{d(F)}{2} = e(F)$, which is a contradiction. $\square$

Let us now describe the procedure. We start the procedure with $F = G$. At every step of the procedure, we consider a subgraph F of G and do the following.

- If F is a λ-expander with $\delta(F) \geq d(F)/2$, then we let $H := F$ be the desired λ-expander and stop the procedure.

- If F has a vertex v with degree less than $d(F)/2$, we remove it and define $F' := F - v$. Note that in this case $d(F') \geq d(F)$. Now we repeat the procedure with F' playing the role of F.

- Otherwise, by Claim 4.5.1, there is a non-empty set $U \subseteq V(F)$ with $|U| \leq \frac{|V(F)|}{2}$ such that either $d(F[\overline{U}]) \geq d(F)$ or $d(F[U]) \geq (1 - 2\lambda)d(F)$. In the former case, let $F' := F[\overline{U}]$, and in the latter case, let $F' := F[U]$. Then, we repeat the procedure with F' playing the role of F.

Note that at any step of our procedure, we can have $d(F') < d(F)$ only if $|V(F')| \leq |V(F)|/2$ and $d(F') \geq (1 - 2\lambda)d(F)$. Furthermore, since $|V(F')| \leq |V(F)|/2$ can only occur for at most $\log n$ steps, at any step of the procedure the subgraph F we consider satisfies $d(F) \geq (1-2\lambda)^{\log n}d(G) \geq (1-2\lambda \log n)d(G) \geq 4d(G)/5$. Combining this with the fact that the number of vertices of F strictly decreases after each step, it follows that the procedure eventually stops with a non-empty subgraph H which is a λ-expander satisfying $d(H) \geq (1 - 2\lambda \log n)d(G)$ and $\delta(H) \geq d(H)/2$. This proves the lemma. $\square$

By repeatedly applying Lemma 4.5, we obtain the following lemma which shows that one can cover a good proportion of the vertices of any (sufficiently regular) graph G with vertex-disjoint λ-expanders (whose average degree is very close to that of G).

Lemma 4.6. *Let ε, λ, n satisfy $\lambda \log n \leq \min\{\frac{1}{10}, \varepsilon\}$. Let G be an n-vertex graph with $\Delta(G) \leq d$ and $d(G) \geq d(1-\varepsilon)$. Then there exists a collection $\mathcal{H}$ of vertex-disjoint λ-expanders in G such that $d(H) \geq d(1-10\varepsilon)$ and $\delta(H) \geq d(H)/2$ for every $H \in \mathcal{H}$ and $\sum_{H \in \mathcal{H}} |V(H)| \geq n/4$.*

Proof. Let $\mathcal{H}$ be a maximal collection of vertex-disjoint λ-expanders in G with average degree at least $d(1 - 10\varepsilon)$. Let $U := \bigcup_{H \in \mathcal{H}} V(H)$. If $|U| \geq n/4$, then the λ-expanders in $\mathcal{H}$ satisfy the desired properties, proving the lemma. So suppose $|U| < n/4$. Then $e(U) + e(U, \overline{U}) \leq d|U| - e(U) \leq d|U| - \frac{d(1-10\varepsilon)}{2}|U| \leq \frac{d(1+10\varepsilon)}{2}|U|$. Hence,

$$\begin{aligned}
e(\overline{U}) &\geq e(G) - \left(e(U) + e(U, \overline{U})\right) \geq \frac{d(1-\varepsilon)}{2}n - \frac{d(1+10\varepsilon)}{2}|U| \\
&= \frac{d}{2}\left((1-\varepsilon)n - (1+10\varepsilon)|U|\right) = \frac{d}{2}\left(|\overline{U}| - \varepsilon n - 10\varepsilon|U|\right) \\
&\geq \frac{d}{2}\left(|\overline{U}| - \varepsilon n - 10\varepsilon\frac{n}{4}\right) \geq \frac{d}{2}\left(|\overline{U}| - 8\varepsilon|\overline{U}|\right) = \frac{d}{2}(1 - 8\varepsilon)|\overline{U}|.
\end{aligned}$$

Therefore, $d(\overline{U}) \geq d(1 - 8\varepsilon)$. Hence, by applying Lemma 4.5 to $G[\overline{U}]$, we obtain a λ-expander H with $d(H) \geq d(\overline{U})(1 - 2\lambda \log n) \geq d(1 - 8\varepsilon)(1 - 2\lambda \log n) \geq d(1 - 8\varepsilon - 2\lambda \log n) \geq d(1 - 10\varepsilon)$ and $\delta(H) \geq d(H)/2$. This contradicts the maximality of the collection $\mathcal{H}$ of vertex-disjoint λ-expanders. Therefore, $|U| \geq n/4$, proving the lemma. $\qquad\square$

Finally, we prove Lemma 4.4 by repeatedly applying Lemma 4.6.

Proof of Lemma 4.4. Let $C = 100$. For every $i \geq 1$, let $\varepsilon_i = C^i \varepsilon$, and let $\mathcal{H}_i$ be a collection of vertex-disjoint λ-expanders with average degree at least $d(1 - \varepsilon_i)$ in $G - (\bigcup_{H \in \mathcal{H}_1 \cup \ldots \cup \mathcal{H}_{i-1}} V(H))$ that maximises the number of vertices covered by expanders in $\mathcal{H}_i$ among all such collections. For all $i \geq 1$, let $V_i := \bigcup_{H \in \mathcal{H}_i} V(H)$.

Claim 4.6.1. *For all $i \geq 1$, $\sum_{k=1}^{i} |V_k| \geq (1 - (\frac{3}{4})^i)n$.*

Proof of claim. We prove the claim by induction. For $i = 1$, the claim immediately follows by applying Lemma 4.6 to G to obtain a collection $\mathcal{H}_1$ of vertex-disjoint λ-expanders covering at least $n/4$ vertices and satisfying $d(H) \geq d(1 - 10\varepsilon) \geq d(1 - C\varepsilon) = d(1 - \varepsilon_1)$ and $\delta(H) \geq d(H)/2$ for every $H \in \mathcal{H}_1$. This shows that $|V_1| \geq n/4$, as desired.

Now suppose that for all $1 \leq j \leq i$, we have $\sum_{k=1}^{j} |V_k| \geq (1 - (\frac{3}{4})^j)n$. We will show that $\sum_{k=1}^{i+1} |V_k| \geq (1 - (\frac{3}{4})^{i+1})n$.

Let $U := V_1 \cup \ldots \cup V_i$. Since for each $1 \leq k \leq i$, the average degree of the λ-expanders in $\mathcal{H}_k$ (whose union spans the vertex set V_k) is at least $d(1 - \varepsilon_k)$, we have $e(U) \geq \sum_{k=1}^{i} \frac{d}{2}(1 - \varepsilon_k)|V_k| = \frac{d}{2}|U| - \frac{d}{2}\sum_{k=1}^{i} \varepsilon_k |V_k|$. Hence,

$$e(U) + e(U, \overline{U}) \leq d|U| - e(U) \leq \frac{d}{2}|U| + \frac{d}{2}\sum_{k=1}^{i} \varepsilon_k |V_k|. \tag{2}$$

Note that under the constraints $\varepsilon_1 \leq \ldots \leq \varepsilon_i$ and $\sum_{k=1}^{j} |V_k| \geq (1 - (\frac{3}{4})^j)n$ for $1 \leq j \leq i$, the sum $\sum_{k=1}^{i} \varepsilon_k |V_k|$ is maximised if, for each $1 \leq k \leq i$, the size of V_k is chosen to be as large as possible, assuming the sizes of $V_{k+1}, \ldots, V_i$ were already chosen to be as large as possible. Since $\sum_{k=1}^{i-1} |V_k| \geq (1 - (\frac{3}{4})^{i-1})n$, the largest possible size of V_i is $(\frac{3}{4})^{i-1}n$. Given this, the largest possible size of V_{i-1} is $(1 - (\frac{3}{4})^{i-1})n - (1 - (\frac{3}{4})^{i-2})n = ((\frac{3}{4})^{i-2} - (\frac{3}{4})^{i-1})n = (\frac{3}{4})^{i-2}\frac{n}{4}$, using $\sum_{j=1}^{i-2} |V_j| \geq (1 - (\frac{3}{4})^{i-2})n$. Similarly, for each $1 \leq k \leq i-1$, the largest possible size of V_k is $(\frac{3}{4})^{k-1}\frac{n}{4}$. Hence,

$$\sum_{k=1}^{i} \varepsilon_k |V_k| \leq \sum_{k=1}^{i-1} \varepsilon_k |V_k| + \varepsilon_i |V_i| \leq \sum_{k=1}^{i-1} \varepsilon_k \left(\frac{3}{4}\right)^{k-1} \frac{n}{4} + \varepsilon_i \left(\frac{3}{4}\right)^{i-1} n$$

$$\leq \sum_{k=1}^{i} \frac{4}{3}\varepsilon \left(\frac{3C}{4}\right)^{k} n \leq \frac{\frac{4}{3}\varepsilon}{\frac{3C}{4} - 1}\left(\frac{3C}{4}\right)^{i+1} n. \tag{3}$$

Combining (2) and (3), we have

$$e(U) + e(U, \overline{U}) \leq \frac{d}{2}|U| + \frac{\frac{2}{3}\varepsilon d}{\frac{3C}{4} - 1}\left(\frac{3C}{4}\right)^{i+1} n.$$

Therefore,

$$e(\overline{U}) \geq e(G) - (e(U) + e(U, \overline{U})) \geq \frac{1}{2}nd(1 - \varepsilon) - \frac{d}{2}|U| - \frac{\frac{2}{3}\varepsilon d}{\frac{3C}{4} - 1}\left(\frac{3C}{4}\right)^{i+1} n$$

$$\geq \frac{d}{2}|\overline{U}| - \frac{1}{2}nd\varepsilon - \frac{8\varepsilon d}{9C - 12}\left(\frac{3C}{4}\right)^{i+1} n. \tag{4}$$

Since $\sum_{k=1}^{i}|V_k| \geq (1-(\frac{3}{4})^i)n$, we have $|\overline{U}| \leq (\frac{3}{4})^i n$. Thus, by (4),

$$d(\overline{U}) = \frac{2e(\overline{U})}{|\overline{U}|} \geq d\left(1 - \left(\frac{4}{3}\right)^i \varepsilon \left(1 + \frac{16}{9C-12}\left(\frac{3C}{4}\right)^{i+1}\right)\right)$$
$$= d\left(1 - \varepsilon\left(\left(\frac{4}{3}\right)^i + \frac{48C^{i+1}}{36C-48}\right)\right).$$

Note that since $C = 100$, we have $\left(\frac{4}{3}\right)^i + \frac{48C^{i+1}}{36C-48} \leq \frac{C^{i+1}}{10}$, so

$$d(\overline{U}) \geq d\left(1 - \frac{\varepsilon C^{i+1}}{10}\right) = d\left(1 - \frac{\varepsilon_{i+1}}{10}\right).$$

Now, note that $\lambda \log|\overline{U}| \leq \lambda \log n \leq \min\{\frac{1}{10}, \varepsilon\} \leq \min\{\frac{1}{10}, \frac{C^{i+1}\varepsilon}{10}\} = \min\{\frac{1}{10}, \frac{\varepsilon_{i+1}}{10}\}$, where in the last inequality we used $C = 100$. Hence, by Lemma 4.6, there exists a collection $\mathcal{H}_{i+1}$ of vertex-disjoint λ-expanders in $G[\overline{U}]$ with $d(H) \geq d(1-\varepsilon_{i+1})$, and $\delta(H) \geq d(H)/2$ for every $H \in \mathcal{H}_{i+1}$ such that, if $V_{i+1} = \bigcup_{H \in \mathcal{H}_{i+1}} V(H)$, then $|V_{i+1}| = |\bigcup_{H \in \mathcal{H}_{i+1}} V(H)| \geq |\overline{U}|/4$. Therefore, $n - \sum_{k=1}^{i+1}|V_k| \leq \frac{3}{4}|\overline{U}|$. Since $|\overline{U}| \leq (\frac{3}{4})^i n$, we have $n - \sum_{k=1}^{i+1}|V_k| \leq \frac{3}{4}|\overline{U}| \leq (\frac{3}{4})^{i+1}n$. This shows that $\sum_{k=1}^{i+1}|V_k| \geq (1-(\frac{3}{4})^{i+1})n$, proving the claim. $\qquad\square$

Let $t = \min\{i \mid (\frac{3}{4})^{i+1} \leq \frac{1}{(\log n)^\alpha}\}$. Then $(\frac{3}{4})^t \geq \frac{1}{(\log n)^\alpha}$, so $(\frac{4}{3})^t \leq (\log n)^\alpha$ which implies that $t \leq \frac{\alpha}{\log(4/3)}\log\log n \leq 3\alpha\log\log n$. Thus, $C^{t+1} \leq C^{3\alpha\log\log n+1} \leq C^{4\alpha\log\log n} \leq 2^{28\alpha\log\log n} = (\log n)^{28\alpha}$.

Now consider the collection of λ-expanders $\mathcal{H} := \mathcal{H}_1 \cup \ldots \cup \mathcal{H}_{t+1}$. Every $H \in \mathcal{H}$ is a λ-expander with $d(H) \geq d(1-\varepsilon_{t+1}) = d(1-\varepsilon C^{t+1}) \geq d(1-\varepsilon(\log n)^{28\alpha})$, $\delta(H) \geq d(H)/2$, and by Claim 4.6.1, these expanders cover at least $\sum_{k=1}^{t+1}|V_k| \geq (1-(\frac{3}{4})^{t+1})n \geq (1 - \frac{1}{(\log n)^\alpha})n$ vertices of G, proving the lemma. $\qquad\square$

4.2 Refining procedure

To prove Lemma 4.3 we use a refining procedure that gradually improves the expansion of the expanders given by Lemma 4.4 without significantly weakening their regularity. This is achieved through repeated applications of Lemma 4.4. The main idea of the refining procedure is described in the proof sketch at the end of Section 2.

Proof of Lemma 4.3. For convenience, let $\beta := 28\alpha$ and $\gamma := \frac{C-\beta-1}{C-1}$. Note that by the assumptions of the lemma, $\gamma \geq 1/2$. Let $n_1 = n$, and for every $i \geq 1$, let n_{i+1} satisfy $\log n_{i+1} = (\log n_i)^\gamma$ (so that $\log n_{i+1} = (\log n)^{\gamma^i}$). Let $\varepsilon_0 = (\log n_1)^{-(C-1)}$, and for every $i \geq 1$, let $\varepsilon_i = (\log n_i)^{-(C-\beta-1)}$ and $\lambda_i = (\log n_i)^{-C}$.

Let $\mathcal{H}_0 := \{G\}$. Since $d(G) \geq d(1-\varepsilon_0)$, and $\lambda_1 \log n_1 = (\log n_1)^{-(C-1)} = \varepsilon_0 \leq \frac{1}{10}$, applying Lemma 4.4 with G, λ_1, α and ε_0 playing the roles of G, λ, α and ε respectively, we obtain a collection $\mathcal{H}_1$ of λ_1-expanders with $d(H) \geq d(1-\varepsilon_0(\log n_1)^{28\alpha}) = d(1-\varepsilon_1)$ and $\delta(H) \geq d(H)/2$ for every $H \in \mathcal{H}_1$ such that $\sum_{H \in \mathcal{H}_1}|V(H)| \geq (1-\frac{1}{(\log n)^\alpha})n$.

Note that $\mathcal{H}_1$ is defined above. Now, we define a sequence $\mathcal{H}_2, \mathcal{H}_3, \ldots$, as follows. For every $i \geq 1$ such that $\varepsilon_i \leq \frac{1}{10}$, assuming that $\mathcal{H}_i$ was already defined, we do the following for each $H \in \mathcal{H}_i$.

(a)$_i$ If $|V(H)| \leq n_{i+1}$ and $d(H) \geq d(1-\varepsilon_i)$, then let $\mathcal{H}_{i+1}(H)$ be a collection of vertex-disjoint λ_{i+1}-expanders in H satisfying $d(F) \geq d(1-\varepsilon_i(\log|V(H)|)^{28\alpha})$, and $\delta(F) \geq d(F)/2$ for every $F \in \mathcal{H}_{i+1}(H)$, and $\sum_{F \in \mathcal{H}_{i+1}(H)}|V(F)| \geq (1 - \frac{1}{(\log|V(H)|)^\alpha})|V(H)|$, obtained by applying Lemma 4.4 with H, λ_{i+1}, α and ε_i playing the roles of G, λ, α and ε respectively. Note that Lemma 4.4 is indeed applicable because $\lambda_{i+1}\log|V(H)| \leq \lambda_{i+1}\log n_{i+1} = (\log n_{i+1})^{-(C-1)} = (\log n_i)^{-\gamma(C-1)} = (\log n_i)^{-(C-\beta-1)} = \varepsilon_i \leq \frac{1}{10}$.

(b)$_i$ Otherwise, let $\mathcal{H}_{i+1}(H) = \{H\}$.

Then we define $\mathcal{H}_{i+1} := \bigcup_{H \in \mathcal{H}_i} \mathcal{H}_{i+1}(H)$. We will now show that the following claim holds.

Claim 4.6.2. *For every $i \geq 1$, if $\varepsilon_{i-1} \leq \frac{1}{10}$, then the following holds.*

$(A)_i$ *For every $H \in \mathcal{H}_i$, we have $d(H) \geq d(1 - \varepsilon_i)$ and $\delta(H) \geq d(H)/2$.*

$(B)_i$ $\sum_{H \in \mathcal{H}_i} |V(H)| \geq (1 - \frac{i}{(\log \log n)^\alpha})n.$

Proof of claim. We prove the claim by induction. By our choice of $\mathcal{H}_1$ as discussed above, we know that $(A)_1$ holds, and that $\sum_{H \in \mathcal{H}_1} |V(H)| \geq (1 - \frac{1}{(\log n)^\alpha})n \geq (1 - \frac{1}{(\log \log n)^\alpha})n$, so $(B)_1$ holds as well. Now suppose $(A)_i$ and $(B)_i$ hold for $i = k$ for some $k \geq 1$, and we will show that they hold for $i = k + 1$. Suppose that $\varepsilon_k \leq \frac{1}{10}$ (otherwise there is nothing to prove).

Indeed, for every $H \in \mathcal{H}_k$ with $|V(H)| > n_{k+1}$, it follows from $(b)_k$ that $\mathcal{H}_{k+1}(H) = \{H\}$. So we have $d(H) \geq d(1 - \varepsilon_k) \geq d(1 - \varepsilon_{k+1})$ and $\delta(H) \geq d(H)/2$ by $(A)_k$. For every $H \in \mathcal{H}_k$ with $|V(H)| \leq n_{k+1}$, it follows from $(a)_k$ and $(A)_k$ that $d(F) \geq d(1 - \varepsilon_k(\log |V(H)|)^{28\alpha})$ and $\delta(F) \geq d(F)/2$ for every $F \in \mathcal{H}_{k+1}(H)$. Moreover, notice that we have

$$\varepsilon_k(\log |V(H)|)^{28\alpha} \leq \varepsilon_k(\log n_{k+1})^{28\alpha}$$
$$= (\log n_k)^{-(C-\beta-1)}(\log n_{k+1})^\beta$$
$$= (\log n_{k+1})^{-\frac{(C-\beta-1)}{\gamma}}(\log n_{k+1})^\beta = (\log n_{k+1})^{-(C-\beta-1)} = \varepsilon_{k+1}.$$

Therefore, for every $F \in \mathcal{H}_{k+1}(H)$, we have $d(F) \geq d(1 - \varepsilon_{k+1})$ in both cases. This proves that $(A)_{k+1}$ holds. Now our goal is to prove $(B)_{k+1}$. Note that

$$\sum_{H \in \mathcal{H}_{k+1}} |V(H)| = \sum_{H \in \mathcal{H}_k} \sum_{F \in \mathcal{H}_{k+1}(H)} |V(F)| \geq \sum_{H \in \mathcal{H}_k} \left(1 - \frac{1}{(\log |V(H)|)^\alpha}\right) |V(H)|, \tag{5}$$

where, in the last inequality we used $(a)_k$ in the case when $|V(H)| \leq n_{k+1}$, and in the case when $|V(H)| > n_{k+1}$, $(b)_k$ trivially implies that $\sum_{F \in \mathcal{H}_{k+1}(H)} |V(F)| = |V(H)|$. Now note that, for every $H \in \mathcal{H}_k$ we have $|V(H)| \geq d(H) \geq d(1 - \varepsilon_k) \geq \frac{9d}{10} \geq \log n$. Hence, $1 - \frac{1}{(\log |V(H)|)^\alpha} \geq 1 - \frac{1}{(\log \log n)^\alpha}$. Therefore, by (5), and our assumption that $(B)_k$ holds, we have

$$\sum_{H \in \mathcal{H}_{k+1}} |V(H)| \geq \left(1 - \frac{1}{(\log \log n)^\alpha}\right) \left(1 - \frac{k}{(\log \log n)^\alpha}\right) n \geq \left(1 - \frac{k+1}{(\log \log n)^\alpha}\right) n.$$

This proves that $(B)_{k+1}$ holds, and completes the proof of the claim. $\qquad \square$

Let $t = \max\{i \mid \log n_i \geq 4\}$. Then $4 \geq \log n_{t+1} = (\log n_t)^\gamma \geq (\log n_t)^{1/2} \geq 2$. In particular, $n_{t+1} \leq 16$. Since, by the assumptions of Lemma 4.3, $C \geq \beta + 3$, we have $\varepsilon_t = (\log n_t)^{-(C-\beta-1)} \leq 4^{-(C-\beta-1)} \leq \frac{1}{10}$. Moreover, since $\varepsilon_{t-1} \leq \varepsilon_t$, we also have $\varepsilon_{t-1} \leq \frac{1}{10}$, so by $(A)_t$ of Claim 4.6.2, for every $H \in \mathcal{H}_t$, we have $d(H) \geq d(1 - \varepsilon_t) \geq d(1 - \frac{1}{10}) > 16 \geq n_{t+1}$ and $\delta(H) \geq d(H)/2$. Therefore, for every $H \in \mathcal{H}_t$, we have $|V(H)| \geq d(H) > n_{t+1}$ and thus $\mathcal{H}_{t+1} = \mathcal{H}_t$.

Now let $H \in \mathcal{H}_t$, and let j be the smallest index such that $H \in \mathcal{H}_j$. Then H is a λ_j-expander with $d(H) \geq d(1 - \varepsilon_j)$. We claim that $n_j \geq |V(H)| > n_{j+1}$. Indeed, the upper bound clearly holds if $j = 1$. Otherwise, since $H \notin \mathcal{H}_{j-1}$, there exists $H' \in \mathcal{H}_{j-1}$ such that $|V(H')| \leq n_j$ and $H \in \mathcal{H}_j(H')$. Since H is a subgraph of H', it follows that $|V(H)| \leq n_j$, as desired. For the lower bound, notice that $H \in \mathcal{H}_j \cap \mathcal{H}_{j+1}$ (otherwise, H would not be in $\mathcal{H}_t$), showing that $|V(H)| > n_{j+1}$. Recall that, as defined in the statement of Lemma 4.3, $\lambda_H = (\log |V(H)|)^{-\frac{C(C-1)}{C-28\alpha-1}}$, and $\varepsilon_H = (\log |V(H)|)^{-(C-28\alpha-1)}$. Hence,

$$\lambda_j = (\log n_j)^{-C} = (\log n_{j+1})^{-C/\gamma} = (\log n_{j+1})^{\frac{-C(C-1)}{C-\beta-1}} \geq (\log |V(H)|)^{\frac{-C(C-1)}{C-\beta-1}} = \lambda_H,$$

and that

$$\varepsilon_j = (\log n_j)^{-(C-\beta-1)} \leq (\log |V(H)|)^{-(C-\beta-1)} = \varepsilon_H.$$

14

Hence, every $H \in \mathcal{H}_t$ is a λ_H-expander with $d(H) \geq d(1 - \varepsilon_H)$ and $\delta(H) \geq d(H)/2$. Finally, since $(\log n)^{\gamma^{t-1}} = \log n_t \geq 4$, we have $t \leq \frac{\log \log \log n}{\log(1/\gamma)} + 1 \leq (\log \log \log n)^2$. Thus, by $(B)_t$ of Claim 4.6.2, we have $\sum_{H \in \mathcal{H}_t} |V(H)| \geq (1 - \frac{t}{(\log \log n)^\alpha})n \geq (1 - \frac{(\log \log \log n)^2}{(\log \log n)^\alpha})n$. This shows that $\mathcal{H}_t$ is a collection of vertex-disjoint subgraphs of G satisfying the desired properties, completing the proof of Lemma 4.3. □

5 Connecting vertex pairs through a random vertex subset of an expander

In this section, we prove the following generalization of a lemma from [56], which allows for connecting pairs of vertices $(x_1, y_1), \ldots, (x_r, y_r)$ in a robust sublinear expander through a random set of vertices using vertex-disjoint paths, if all subsets of the set $\{x_1, \ldots, x_r, y_1, \ldots, y_r\}$ expand well. This is a variant of Theorem 16 in [13], which only requires the paths to be edge-disjoint. The proof follows [13] closely, which, in turn, uses some ideas from Tomon [71].

Lemma 5.1. *Let* $2^{-9} \leq \varepsilon < 1$, $c > 0$, $0 < q < 1$, $s \geq \frac{2(\log n)^{9c+21}}{q^{10}}$, *and let* n *be sufficiently large. Suppose that* G *is an* n*-vertex* (ε, c, s)*-expander, and let* V *be a random subset of* $V(G)$ *obtained by including each vertex independently with probability* q. *Then, with probability at least* $1 - \frac{1}{n}$, *the following holds. If* $x_1, \ldots, x_r, y_1, \ldots, y_r$ *are distinct vertices not contained in* V *such that* $|N(X)| \geq \frac{100(\log n)^{7c+19}}{q^6}|X|$ *for every* $X \subseteq \{x_1, \ldots, x_r, y_1, \ldots, y_r\}$, *then there is a sequence of paths* $P_1, \ldots, P_r$ *of length at most* $(\log n)^{c+4}$, *such that for each* $i \in [r]$, P_i *is a path from* x_i *to* y_i *with internal vertices in* V, *and the paths* $P_1, \ldots, P_r$ *are pairwise vertex-disjoint.*

We will introduce the relevant preliminaries in Section 5.1, and then prove Lemma 5.1 in three steps, as detailed in Sections 5.2 to 5.4.

5.1 Preliminaries

We will use the following two propositions, both of which show that every vertex set U (which is not too large) in an expander either 'expands well', meaning that it has a large neighbourhood, or 'expands robustly', meaning that there are many vertices in $N(U)$ with many neighbours in U. These two propositions are slight variations of propositions from [13]. The main difference is that here we do not have a set of forbidden edges (because we will not need this feature) and we have a new parameter c that controls the rate of expansion of our expanders.

First, we adapt Proposition 12 from [13]. For any graph G, parameter d, and $U \subseteq V(G)$, let $N_{G,d}(U) := \{v \in V(G) - U : |N_G(v) \cap U| \geq d\}$ i.e., $N_{G,d}(U)$ is the set of vertices in G outside U which have degree at least d in U.

Proposition 5.2. *Let* $\varepsilon, c > 0$, *and let* $0 < d \leq s$. *Suppose that* G *is an* n*-vertex* (ε, c, s)*-expander. Then for every set* $U \subseteq V(G)$ *with* $|U| \leq \frac{2n}{3}$, *either*

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

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

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

and therefore **b)** holds, as required. □

Now we adapt Proposition 13 from [13] which shows that more structure can be found in both outcomes of the above proposition.

15

Lemma 5.3. *There is an n_0 such that the following holds for all $n \geq n_0$, $2^{-9} < \varepsilon < 1$, $c > 0$, $r, t \geq (\log n)^2$ and $s \geq 20rt$. Let G be an n-vertex (ε, c, s)-expander, let $U \subseteq V(G)$ satisfy $|U| \leq 2n/3$. Then, in G we can find either*

(a) *$\frac{|U|}{10r}$ pairwise vertex-disjoint stars of size t, whose centers are in U and whose leaves are in $V(G) - U$, or*

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

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

- *every vertex in X has degree at least r in H and every vertex in U has degree at most $2t$ in H.*

Proof. Take a maximal collection $\mathcal{C}$ of pairwise vertex-disjoint stars in G with t leaves, centres in U and leaves outside of U. Let $C \subseteq U$ be the set of centres of these stars and $L \subseteq V(G) - U$ be the set consisting of all their leaves. Suppose **a)** does not hold. Then we can assume that $|C| \leq \frac{|U|}{10r}$ and thus $|L| = |C| \cdot t \leq \frac{|U|}{10r} \cdot t$, and, by the maximality of $\mathcal{C}$, that there is no vertex in $U - C$ with at least t neighbours in G in $V(G) - (U \cup L)$. Thus,

$$|N_G(U - C)| \leq |C| + |L| + |U - C| \cdot t \leq \frac{|U|}{10r} + |C| \cdot t + |U - C| \cdot t < 2|U| \cdot t. \tag{6}$$

We now construct a set $X \subseteq V(G) - U$ and a bipartite subgraph H with vertex classes U and X using 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 $k = |V(G) - U|$ and label the vertices of $V(G) - U$ arbitrarily as $v_1, \ldots, v_k$. For each $i \geq 1$, if possible, pick a star S_i in G with centre v_i and r leaves in U such that the vertices in U in the graph $H_{i-1} \cup S_i$ have degree at most $2t$, 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 $H = H_k$ and $X = X_k = V(H_k) - U$. We will now show that **b)** holds for this choice of H (with vertex classes U and X).

Firstly, observe that every vertex of U has degree at most $2t$ in H_i for each $i \in [k]$ by construction, and that every vertex v_i in X has degree exactly r in H, so the second condition in **b)** holds. Thus, we only need to show that $|X| \geq \frac{\varepsilon |U|}{2(\log n)^c}$ holds.

To see this, let U' be the set of vertices in $U - C$ with degree exactly $2t$ in H. As each vertex in $U - C$ has fewer than t neighbours in G in $X - L$ (due to the maximality of the collection of stars $\mathcal{C}$), the vertices in U' must have at least t neighbours in H in $X \cap L$. As each vertex in $X \cap L$ has r neighbours in H, we have

$$|U'| \leq \frac{r|X \cap L|}{t} \leq \frac{r}{t} \cdot |L| \leq \frac{r}{t} \cdot \frac{|U| \cdot t}{10r} = \frac{|U|}{10}.$$

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

$$|B| \leq \frac{|U|}{10r} + \frac{|U|}{10} \leq \frac{|U|}{2},$$

and, thus, $|U - B| \geq \frac{|U|}{2}$.

Then, by Proposition 5.2 applied to $U - B$ with $d = r$, we have either $|N_G(U - B)| \geq \frac{s|U-B|}{2r}$ or $|N_{G,r}(U - B)| \geq \frac{\varepsilon |U-B|}{(\log n)^c}$. As

$$\frac{s|U - B|}{2r} \geq \frac{s|U|}{4r} \geq 5t|U|,$$

the former inequality contradicts (6), so we have that $|N_{G,r}(U - B)| \geq \frac{\varepsilon |U-B|}{(\log n)^c}$. Every vertex v_i in $N_{G,r}(U - B)$ has at least r neighbours in G in $U - B$, and vertices of $U - B$ must all have degree strictly less than $2t$ in H (as they are not in U'). This implies that every v_i in $N_{G,r}(U - B)$, satisfies $v_i \in X$, since we could add it along with some r of its neighbours while constructing H. Hence, $N_{G,r}(U - B) \subseteq X$, and

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

as required. $\qquad \square$

As in [13], we also need the following hypergraph version of Hall's theorem, due to Aharoni and Haxell [1].

Theorem 5.4 (Aharoni–Haxell [1]). *Let $\ell, r \geq 1$ be integers, and let $\mathcal{H}_1, \ldots, \mathcal{H}_r$ be hypergraphs on the same vertex set with edges of size at most ℓ. Suppose that, for every $I \subseteq [r]$, there is a matching in $\bigcup_{i \in I} \mathcal{H}_i$ of size at least $\ell(|I| - 1)$. Then, there is an injective function $f : [r] \mapsto \bigcup_{i \in [r]} E(\mathcal{H}_i)$ such that $f(i) \in E(\mathcal{H}_i)$ for each $i \in [r]$ and $\{f(i) : i \in [r]\}$ is a matching with r edges.*

5.2 Expansion into a random vertex set

The following lemma shows that, given a sufficiently robust expander G and a random set of vertices $V \subseteq V(G)$, any (sufficiently large) set U expands well in V, while avoiding a given small set of vertices Z, with high probability. This lemma is a variant of Lemma 17 in [13]. The main difference is that instead of a forbidden set of edges we have a forbidden set of vertices. Moreover, we have an additional parameter c that controls the rate of expansion of our expanders and another parameter q that controls the sampling probability for the random set V.

Lemma 5.5. *Let $2^{-9} \leq \varepsilon < 1$, $0 < q < 1$, $c > 0$, $s \geq \frac{100(\log n)^{2c+6}}{q^3}$ and let n be large. Suppose that G is an n-vertex (ε, c, s)-expander and let $U, Z \subseteq V(G)$ be sets satisfying $|U| \geq \frac{(\log n)^{4c+9}}{q^5}$ and $|Z| \leq \frac{\varepsilon q |U|}{10^7 (\log n)^c}$. Let V be a random subset of $V(G)$, obtained by including each vertex independently with probability q. Then, with probability at least $1 - \exp\left(-\Omega\left(\frac{q^5 |U|}{(\log n)^{4c+8}}\right)\right)$,*

$$\left| B_{G[V']}^{(\log n)^{c+2}} (U \cap V') \right| > \frac{|V|}{2},$$

where $V' := V - Z$.

Proof. Let $\ell = (\log n)^{c+2}$ and let p be such that $1 - (1-p)^\ell (1 - \frac{q}{20})(1 - 0.9q) = q$, i.e., that $(1-p)^\ell = \frac{1-q}{(1-\frac{q}{20})(1-0.9q)}$, so that

$$p \geq \frac{q}{100\ell}. \tag{7}$$

Let $V_1, \ldots, V_\ell, V^*, V^{**}$ be random sets, where for each $1 \leq i \leq \ell$, V_i is obtained by including each vertex with probability p, independently, V^* is obtained by including each vertex with probability $\frac{q}{20}$, independently, and V^{**} is obtained by including each vertex with probability $0.9q$, independently. Notice that each vertex is in $V_1 \cup \ldots \cup V_\ell \cup V^* \cup V^{**}$ with probability q, and so we may think of the random set V as the union $V_1 \cup \ldots \cup V_\ell \cup V^* \cup V^{**}$.

Define $U^* = (U \cap V^*) - Z$. Then, by the Chernoff bound, with probability $1 - \exp(-\Omega(q|U|))$, we have that $|U^*| \geq \frac{q}{25}|U| - |Z| \geq \frac{q}{50}|U|$; condition on this being this case. Define $B_0 := U^*$ and, for $i \geq 1$, let B_i be the set of vertices in G that can be reached by a path in $G - Z$ that starts in U^*, has length at most i, and its internal vertices are in $V_1 \cup \ldots \cup V_i$. Let us emphasise that B_i is only required to be disjoint from Z, and need not be a subset of V_i. Notice that $B_i \subseteq B_{i+1}$ for every $i \geq 0$, implying that for all $i \geq 0$, we have

$$|B_i| \geq |U^*| \geq \frac{q}{50}|U|.$$

Claim 5.5.1. *For each $1 \leq i \leq \ell - 1$, if $|U^*| \geq \frac{q}{50}|U|$ and $|B_i| \leq \frac{2}{3}n$, then, with probability at least $1 - \exp\left(-\Omega\left(\frac{q^5}{\ell^4}|U|\right)\right)$,*

$$|B_{i+1} - B_i| \geq \frac{\varepsilon |B_i|}{10^5 (\log n)^c}.$$

Proof. Notice that a vertex in $N(B_i) - Z$ is in $B_{i+1} - B_i$ if at least one of its neighbours in B_i is sampled into V_{i+1}.

Consider the two possible outcomes obtained by applying Lemma 5.3 with B_i playing the role of U and with $r = \frac{\ell}{q}$ and $t = \left(\frac{\ell}{q}\right)^2 \cdot \frac{5}{(\log n)^c}$. (Note that the lemma indeed applies because $r, t \geq \ell \geq (\log n)^2$ and $s \geq \frac{100(\log n)^{2c+6}}{q^3} = 20rt$.)

Suppose that the first outcome of Lemma 5.3 holds, so there are $\frac{|B_i|}{10r}$ pairwise vertex-disjoint stars of size t with centres in B_i and leaves in $N(B_i)$. By the Chernoff bound, with probability at least $1 - \exp\left(-\Omega\left(\frac{p|B_i|}{10r}\right)\right) =$

17

$1 - \exp\left(-\Omega\left(\frac{q^3}{\ell^2}|U|\right)\right)$, at least $\frac{p|B_i|}{20r}$ centres of these stars are included in V_{i+1}, implying that

$$|B_{i+1} - B_i| \geq \frac{pt|B_i|}{20r} - |Z| \geq \frac{|B_i|}{20} \cdot \frac{q}{100\ell} \cdot \frac{5\ell^2}{q^2(\log n)^c} \cdot \frac{q}{\ell} - |Z|$$

$$\geq \frac{|B_i|}{400(\log n)^c} - |Z| \geq \frac{\varepsilon|B_i|}{10^5(\log n)^c},$$

using that $|Z| \leq \frac{\varepsilon q|U|}{10^7(\log n)^c} \leq \frac{|B_i|}{10^5(\log n)^c}$ (since $|B_i| \geq \frac{q}{50}|U|$ and $\varepsilon \leq 1$).

Now suppose that the second outcome of Lemma 5.3 holds, so there is a bipartite subgraph $H \subseteq G$ with parts B_i and $X \subseteq V(G) - B_i$, with $|X| \geq \frac{\varepsilon|B_i|}{2(\log n)^c}$, such that vertices in X have degree at least r in H while vertices in B_i have degree at most $2t$ in H. Let Y be the set of vertices in X that do not have an H-neighbour in V_{i+1}. Note that $\mathbb{E}|Y| \leq |X|(1-p)^r \leq |X|e^{-pr} \leq 0.999|X|$. Note also that $|Y|$ is $2t$-Lipschitz, since the outcome of the sampling of any single vertex in B_i affects the outcome of at most $2t$ vertices in X. Thus, by Lemma 3.5,

$$\mathbb{P}\left(|Y| > 0.9991|X|\right) \leq \mathbb{P}\left(|Y| > \mathbb{E}|Y| + \frac{|X|}{10^4}\right) \leq 2\exp\left(-\frac{|X|^2}{2 \cdot 10^8(2t)^2|B_i|}\right)$$

$$= \exp\left(-\Omega\left(\frac{q|U|}{t^2(\log n)^{2c}}\right)\right)$$

$$= \exp\left(-\Omega\left(\frac{q^5}{\ell^4}|U|\right)\right).$$

Hence, in this case, with probability at least $1 - \exp\left(-\Omega\left(\frac{q^5}{\ell^4}|U|\right)\right)$, we have $|B_{i+1} - B_i| \geq 0.0009|X| - |Z| \geq \frac{\varepsilon|B_i|}{10^4(\log n)^c} - |Z| \geq \frac{\varepsilon|B_i|}{10^5(\log n)^c}$ since $|Z| \leq \frac{\varepsilon q|U|}{10^7(\log n)^c} \leq \frac{\varepsilon|B_i|}{10^5(\log n)^c}$. $\qquad\square$

By repeatedly applying Claim 5.5.1, we obtain that with probability at least

$$1 - \exp\left(-\Omega(q|U|)\right) - \ell \cdot \exp\left(-\Omega\left(\frac{q^5}{\ell^4}|U|\right)\right) = 1 - \exp\left(-\Omega\left(\frac{q^5}{\ell^4}|U|\right)\right),$$

we have $|U^*| \geq \frac{q}{50}|U|$ and for all $i \in [\ell]$ such that $|B_i| \leq \frac{2}{3}n$, the following holds.

$$|B_i| \geq \left(1 + \frac{\varepsilon}{10^5(\log n)^c}\right)^i |U^*|.$$

This implies that $|B_\ell| \geq \frac{2}{3}n$ with probability at least $1 - \exp\left(-\Omega\left(\frac{q^5}{\ell^4}|U|\right)\right)$.

To complete the proof, note that any vertex in B_ℓ that gets sampled into V^{**} (or that is already in $V_1 \cup \ldots \cup V_\ell \cup V^*$) is in the set $B' := B_{G[V']}^{(\log n)^{c+2}}(U \cap V')$. Hence, conditioning on the event that $|B_\ell| \geq \frac{2}{3}n$, by the Chernoff bound, at least $\frac{17}{30}qn$ vertices of B_ℓ are in V, and $|V| \leq \frac{31}{30}qn$ with probability at least $1 - \exp(-\Omega(qn))$. This implies that $|B'| > \frac{1}{2}|V|$ with probability at least $1 - \exp\left(-\Omega\left(\frac{q^5}{\ell^4}|U|\right)\right) - \exp(-\Omega(qn)) = 1 - \exp\left(-\Omega\left(\frac{q^5}{\ell^4}|U|\right)\right)$, as required. $\qquad\square$

Next, we prove the following corollary by boosting the probability that a given set U expands into V (as compared to the probability in Lemma 5.5), so that we can apply the union bound to show that all sets U (of size at most $2n/3$) expand into V at the expense of a slightly smaller upper bound on $|Z|$. This is a variant of Lemma 19 in [13].

Corollary 5.6. *Let* $2^{-9} \leq \varepsilon < 1$, $c > 0$, $0 < q < 1$, $s \geq \frac{2(\log n)^{9c+21}}{q^{10}}$, *and let n be large. Suppose that G is an n-vertex (ε, c, s)-expander, and let V be a random subset of $V(G)$, obtained by including each vertex independently with probability at least q. Then, with probability at least $1 - \frac{1}{n^2}$, for every $U, Z \subseteq V(G)$ with $|U| \leq \frac{2}{3}n$ and $|Z| \leq \frac{q^5|U|}{(\log n)^{5c+11}}$, we have*

$$\left|B_{G[V']}^{(\log n)^{c+2}}(N(U) \cap V')\right| > \frac{|V|}{2}, \tag{8}$$

where $V' := V - Z$.

18

Proof. We say that a set $U \subseteq V(G)$ *expands well* if $|N(U)| \geq \frac{|U|(\log n)^{4c+10}}{q^5}$. Given a (non-empty) set U which expands well, and a set Z with $|Z| \leq |U|$, Lemma 5.5 (applied with $N(U)$ playing the role of U) shows that (8) holds, with probability at least $1 - \exp\left(-\Omega\left(|U|(\log n)^2\right)\right)$.

By the union bound, it follows that the probability that (8) fails for some pair (U, Z), where U expands well and $|Z| \leq |U|$, is at most

$$\sum_{u=1}^{n} (u+1) n^{2u} \exp\left(-\Omega\left(u(\log n)^2\right)\right) \leq \sum_{u=1}^{n} \exp\left(3u \log n - \Omega\left(u(\log n)^2\right)\right)$$
$$= \sum_{u=1}^{n} \exp\left(-\Omega\left(u(\log n)^2\right)\right) \leq \frac{1}{n^2}.$$

For the rest of the proof of Corollary 5.6 we condition on the event that (8) holds for all pairs (U, Z) where U expands well and $|Z| \leq |U|$. From this, we will deduce that (8) holds for all pairs (U, Z), where U need not expand well, $|U| \leq \frac{2}{3}n$, and $|Z| \leq \frac{q^5|U|}{(\log n)^{5c+11}}$. Fix such a pair of sets (U, Z).

Write $d := \frac{(\log n)^{5c+11}}{q^5}$. By Proposition 5.2, either $|N(U)| \geq \frac{s|U|}{2d}$, or $|N_d(U)| \geq \frac{\varepsilon|U|}{(\log n)^c}$ (using the notation $N_d(U)$ introduced before the proposition). Notice that the first outcome implies that U expands well (using the assumption on s and the definition of d) and so we already know that (8) holds. Hence, we may assume that the second outcome holds, and write $W := N_d(U)$. Let U' be a subset of U of size $\frac{|U|}{d}$, chosen uniformly at random. For a fixed $w \in W$, the probability that w has no neighbours in U' is at most

$$\frac{\binom{|U|-d}{|U|/d}}{\binom{|U|}{|U|/d}} \leq \left(\frac{|U|-d}{|U|}\right)^{|U|/d} \leq e^{-1}.$$

It follows that $\mathbb{E}[|W \cap N(U')|] \geq (1 - e^{-1})|W| \geq \frac{\varepsilon|U|}{2(\log n)^c} \geq \frac{|U'|(\log n)^{4c+10}}{q^5}$. In particular, there is a subset $U' \subseteq U$ of size $\frac{|U|}{d}$ with $|N(U')| \geq \frac{|U'|(\log n)^{4c+10}}{q^5}$. Since U' expands well, and $|Z| \leq \frac{q^5|U|}{(\log n)^{5c+11}} = |U'|$, (8) holds for (U', Z) and thus it also holds for (U, Z), completing the proof of the corollary. $\qquad\square$

5.3   A path connection through a random set

For proving Lemma 5.1, our end goal is to be able to connect r pairs of vertices through a random set V (using internally vertex-disjoint paths), provided that any subset of these vertices expands well. The following lemma shows that at least one pair of vertices can be joined while avoiding a small set of forbidden vertices. This is a variant of Proposition 8 in [13], where in addition to forbidding vertices instead of edges, we impose an expansion property on the vertices we wish to connect, and we use additional parameters c and q that control the rate of expansion of our expanders and the sampling probability of the random set V, respectively.

Lemma 5.7. *Let* $2^{-9} \leq \varepsilon < 1$, $c > 0$, $0 < q < 1$, $s \geq \frac{2(\log n)^{9c+21}}{q^{10}}$, *and let n be large. Suppose that G is an n-vertex (ε, c, s)-expander, and let V be a random subset of $V(G)$, obtained by including each vertex independently with probability q. Then, with probability at least $1 - \frac{1}{n}$, the following holds for every r: If $x_1, \ldots, x_r, y_1, \ldots, y_r$ are distinct vertices, satisfying $|N(X)| \geq \frac{100(\log n)^{7c+19}}{q^6}|X|$ for every subset $X \subseteq \{x_1, \ldots, x_r, y_1, \ldots, y_r\}$, and Z is a set of size at most $2r(\log n)^{2c+8}$ which is disjoint from $\{x_1, \ldots, x_r, y_1, \ldots, y_r\}$, then for some $i \in [r]$ there is an (x_i, y_i)-path in G whose internal vertices are in $V - Z$ and whose length is at most $(\log n)^{c+4}$.*

Proof. Let $V' := V - Z$. It is easy to see that by Corollary 5.6 and the Chernoff bound (Theorem 3.4) together with a union bound, the following three properties hold simultaneously with probability at least $1 - \frac{1}{n}$.

(T1) For every $U \subseteq V(G)$ satisfying $|U| \leq \frac{2n}{3}$ and $|Z| \leq \frac{q^5|U|}{(\log n)^{5c+11}}$, (8) holds.

(T2) We have $|V| \geq \frac{qn}{2}$.

(T3) For every $X \subseteq \{x_1, \ldots, x_r, y_1, \ldots, y_r\}$ with $|X| \geq \frac{r}{2}$, we have $|N(X) \cap V'| \geq \frac{48(\log n)^{7c+19}}{q^5}|X|$.

Indeed, by Corollary 5.6, (T1) holds with probability at least $1 - \frac{1}{n^2}$. Moreover, (T2) holds with probability at least $1 - \exp(-\Omega(qn)) \geq 1 - \frac{1}{n^2}$ by the Chernoff bound. To see why (T3) holds, note that for a given X, the probability that $|N(X) \cap V| \geq q|N(X)|/2 \geq \frac{50(\log n)^{7c+19}}{q^5}|X|$ is at least $1 - \exp(-\Omega(q|N(X)|)) \geq 1 - \exp(-\Omega((\log n)^{7c+19}/q^5)|X|)$. Hence by a union bound, for all $X \subseteq \{x_1, \ldots, x_r, y_1, \ldots, y_r\}$ with $|X| \geq r/2$, we have $|N(X) \cap V'| \geq \frac{q|N(X)|}{2} - |Z| \geq \frac{50(\log n)^{7c+19}}{q^5}|X| - |Z| \geq \frac{48(\log n)^{7c+19}}{q^5}|X|$ with probability at least $1 - \frac{1}{n^2}$.

Fix an outcome of V such that the properties (T1)–(T3) hold simultaneously. Write $\ell := (\log n)^{c+2} + 1$. For a set of vertices X and integer $d \geq 1$, define $R^d(X) := B_{G[V']}^d(N(X) \cap V')$. Let X_1 be the set of vertices x in $\{x_1, \ldots, x_r\}$ satisfying $|R^{\ell \log n}(x)| \leq \frac{|V|}{2}$.

Claim 5.7.1. $|X_1| < \frac{r}{2}$.

Proof. Suppose that $|X_1| \geq \frac{r}{2}$. We will first show that there is a sequence $(X_i)_{i \geq 1}$, such that for every $i \geq 1$, $X_{i+1} \subseteq X_i$, $|X_{i+1}| \leq \max\{1, \frac{|X_i|}{2}\}$, and

$$\left| R^{i\ell}(X_i) \right| > \frac{|V|}{2}. \tag{9}$$

Indeed, notice that $|N(X_1) \cap V'| \geq \frac{48(\log n)^{7c+19}}{q^5}|X_1| \geq \frac{r}{2} \cdot \frac{4(\log n)^{7c+19}}{q^5} \geq \frac{(\log n)^{5c+11}}{q^5}|Z|$, by the third item above and since $|X_1| \geq \frac{r}{2}$ and $|Z| \leq 2r(\log n)^{2c+8}$. Thus, by the first item, $|R^{(\log n)^{c+2}}(N(X_1) \cap V')| > \frac{|V|}{2}$, showing that $|R^\ell(X_1)| > \frac{|V|}{2}$, proving that (9) holds for $i = 1$. (Notice that (8) is only applicable here if $|N(X_1) \cap V'| \leq \frac{2n}{3}$, but if this fails then (9) holds trivially for $i = 1$.)

Now suppose that $X_1 \supseteq \ldots \supseteq X_i$ is a sequence, such that for all $j \in [i]$, $|X_j| \leq \max\{1, \frac{|X_{j-1}|}{2}\}$ and (9) holds. If $|X_i| = 1$ we take $X_{i+1} = X_i$ (which clearly satisfies the required properties). Otherwise, by partitioning X_i into at most three sets of size at most $\frac{|X_i|}{2}$, there is a subset $X_{i+1} \subseteq X_i$ of size at most $\frac{|X_i|}{2}$ satisfying $|R^{i\ell}(X_{i+1})| \geq \frac{|V|}{6}$. We will show that X_{i+1} satisfies (9). Indeed, consider the set $U := R^{i\ell}(X_{i+1})$. Then (9) trivially holds if $|U| \geq \frac{2n}{3}$, so suppose otherwise. We have $|U| \geq \frac{|V|}{6} \geq \frac{qn}{12} \geq \frac{q}{12} \cdot r \cdot \frac{100(\log n)^{7c+19}}{q^6} \geq \frac{(\log n)^{5c+11}}{q^5}|Z|$, where the third inequality follows from the assumption that $|N(\{x_1, \ldots, x_r\})| \geq r \cdot \frac{100(\log n)^{7c+19}}{q^6}$ (and by the trivial inequality $n \geq |N(\{x_1, \ldots, x_r\})|$). Thus, by the first item,

$$\left| B_{G[V']}^{(\log n)^{c+2}} (N(U) \cap V') \right| > \frac{|V|}{2}.$$

Notice that

$$B_{G[V']}^{(\log n)^{c+2}} (N(U) \cap V') \subseteq B_{G[V']}^\ell(U) = R^{(i+1)\ell}(X_{i+1}),$$

so X_{i+1} has the required properties. This completes the proof of the existence of a sequence $(X_i)_{i \geq 1}$ with the desired properties.

Fix such a sequence, and take $i := \log n$. Then $|X_i| \leq \max\{1, 2^{-\log n}|X_1|\} = 1$. This means $|R^{\ell \log n}(x)| > \frac{|V|}{2}$ for the single vertex x in X_i, contradicting the choice of X_1, and completing the proof of the claim. $\square$

Take Y_1 to be the set of vertices $y \in \{y_1, \ldots, y_r\}$ satisfying $|R^{\ell \log n}(y)| \leq \frac{|V|}{2}$. Then, analogously to Claim 5.7.1, $|Y_1| < \frac{r}{2}$. Hence, there exists $i \in [r]$ such that $|R^{\ell \log n}(x_i)|, |R^{\ell \log n}(y_i)| > \frac{|V|}{2}$, showing that, with probability at least $1 - \frac{1}{n}$, there is an (x_i, y_i)-path of length at most $2\ell \log n + 2 \leq (\log n)^{c+4}$, with internal vertices in V', completing the proof of Lemma 5.7. $\square$

5.4 Many vertex-disjoint path connections through a random set

Finally, we are ready to prove Lemma 5.1, which shows that given a sufficiently robust expander G and a random subset of vertices V, then with high probability for any set $\{x_1, \ldots, x_r, y_1, \ldots, y_r\}$ of vertices outside of V, whose subsets all expand well, the pairs (x_i, y_i), $i \in [r]$, can be connected using vertex-disjoint paths through V.

Proof of Lemma 5.1. Fix an outcome of V for which the conclusion of Lemma 5.7 holds. For each $i \in [r]$, let $\mathcal{H}_i$ be the hypergraph on the vertex set V where each edge is the set of internal vertices of P, for all (x_i, y_i)-paths P of length at most $\ell := (\log n)^{c+4}$ with internal vertices in V.

We will apply Theorem 5.4. Fix a subset $I \subseteq [r]$. We wish to show that there is a matching of size at least $\ell(|I| - 1)$ in $\mathcal{H}' := \bigcup_{i \in I} \mathcal{H}_i$. Suppose that no such matching exists. Let $\mathcal{M}'$ be a maximal matching in $\mathcal{H}'$, and let Z be the set of vertices in $\mathcal{M}'$. Then $|\mathcal{M}'| < \ell(|I| - 1)$, every edge in $\mathcal{H}'$ intersects Z, and $|Z| \leq \ell^2(|I| - 1) \leq |I|(\log n)^{2c+8}$. Hence, by applying Lemma 5.7 (with I playing the role of $[r]$) we obtain that for some $i \in I$ there is an (x_i, y_i)-path P of length at most ℓ whose internal vertices are in $V - Z$. But this means that $V(P) - \{x_i, y_i\}$ is an edge in $\mathcal{H}'$ that does not intersect Z, a contradiction.

Therefore, the assumptions in Theorem 5.4 are satisfied, showing that there is a matching $\mathcal{M}$ of size r in $\bigcup_{i \in [r]} \mathcal{H}_i$, such that the i-th edge of the matching is in $\mathcal{H}_i$. Let P_i be the path corresponding to the i-th edge in $\mathcal{M}$. Then, for each $i \in [r]$, P_i is an (x_i, y_i)-path with internal vertices in V. The internal vertex sets of the paths $P_1, \ldots, P_r$ are pairwise disjoint and are disjoint from $\{x_1, \ldots, x_r, y_1, \ldots, y_r\}$, as it is assumed that the vertices in this set are not contained in V, thereby proving the lemma. $\qquad\square$

6 Finding an almost-spanning F-subdivision in a nearly regular expander

In this section we prove our second key lemma (Lemma 6.1) which shows that any sufficiently regular expander contains an *almost-spanning* subdivision of any given (non-empty) graph F.

Lemma 6.1. *Let c, ε, d, s, n be parameters such that $c > 0$ is fixed, n is sufficiently large, $0 < \varepsilon \leq \frac{1}{(\log n)^4}$, $d \geq (\log n)^{10c+51}$ and $s \geq \frac{d}{4(\log n)^c}$. Let F be a non-empty graph on at most $\frac{\sqrt{d}}{(\log n)^2}$ vertices, and let H be an n-vertex $(\frac{1}{8}, c, s)$-expander such that $\Delta(H) \leq d$, $d(H) \geq d(1 - \varepsilon)$ and $\delta(H) \geq \frac{d(H)}{2}$. Then H contains a subdivision of F covering all but at most $\frac{n}{\log n}$ vertices of H.*

Lemma 6.1 implies that every sufficiently regular expander contains a nearly Hamilton cycle. Indeed, by letting F be a triangle in the lemma, we obtain that H contains a cycle C covering all but at most $\frac{n}{\log n}$ vertices of H. This statement is of independent interest and is likely to have other applications.

For instance, we can use it to prove that nearly all vertices of any d-regular graph of order n, with $d \geq (\log n)^{130}$, can be covered by at most $\frac{n}{d+1}$ vertex-disjoint cycles. Recently, Montgomery, Müyesser, Pokrovskiy and Sudakov [65] proved a similar result, showing that nearly all vertices of every d-regular graph of order n can be covered by at most $\frac{n}{d+1}$ vertex-disjoint paths, thus making progress on a conjecture of Magnant and Martin [63] (which states that all vertices can be covered by the same number of vertex-disjoint paths). Their result is stronger than ours in that it applies for all d, but ours is also slightly stronger in that it finds vertex-disjoint cycles rather than paths.

Additionally, combining Lemma 6.1 with Corollary 4.2, we derive a recent result of Draganić, Methuku, Munhá Correia, and Sudakov [22] on finding a cycle with many chords, albeit with a slightly larger logarithmic factor. We elaborate on both applications in Section 8.

In the rest of this section, we prove Lemma 6.1. Let us first provide a brief outline of the structure of the proof. We start by taking a random partition of the vertex set of our $(\frac{1}{8}, c, s)$-expander H into sets $V_0, X_1, \ldots, X_t, R$, where the sizes of these random sets are chosen carefully in Section 6.1. In particular, we ensure that $X := \bigcup_{i=1}^t X_i$ contains nearly all vertices of H and the size of V_0 is sufficiently larger than the sizes of the sets $X_1, \ldots, X_t, R$. In Section 6.2, we show that, with high probability, there exists a small collection of paths $P_1, \ldots, P_r$ contained in X, with leaves in $X_1 \cup X_t$, that covers nearly all vertices of H. Then, in Section 6.3, we iteratively connect these paths through V_0 to construct a nearly Hamilton path P. More precisely, in each iteration, we join a constant proportion of the paths $P_1, \ldots, P_r$ via V_0 using the following strategy: first, we greedily connect as many paths as possible through V_0 using vertex-disjoint paths of length two. When this is no longer possible, we show that there is a small subset S of the leaves of $P_1, \ldots, P_r$ that expands well into V_0, allowing us to apply Lemma 5.1 to connect the paths whose

leaves lie in S via V_0 (see Claim 6.1.2). Finally, in Section 6.4, we find a subdivision of F in H that contains the nearly Hamilton path P, yielding the desired almost spanning F-subdivision in H (see Figure 2).
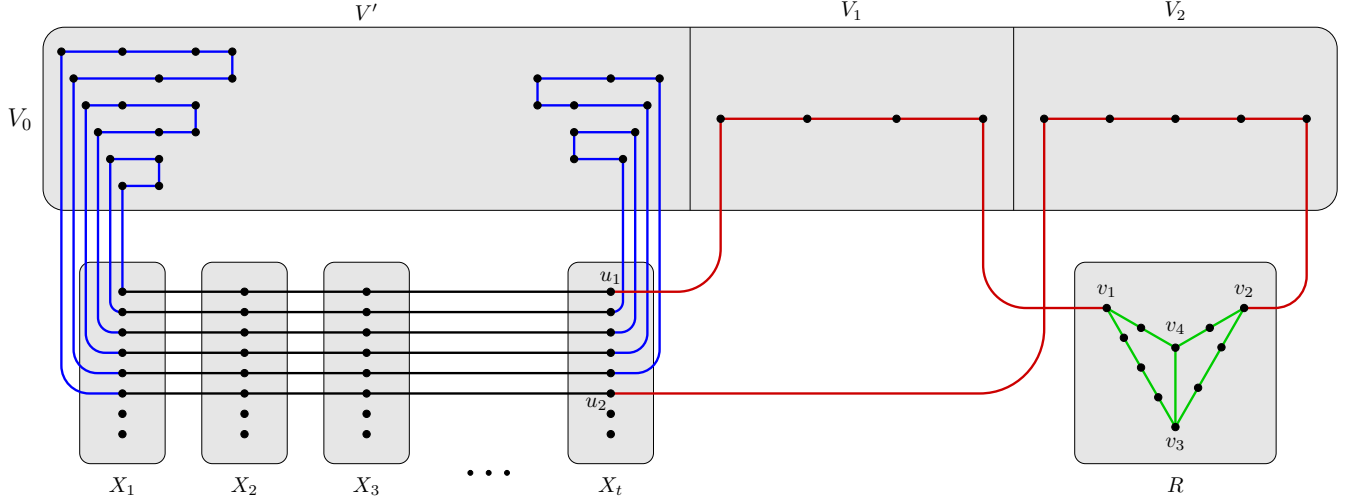
Figure 2: The figure shows how to construct an almost spanning F-subdivision in H (when $F = K_4$).

6.1 Setting up the parameters

Throughout the section, we let c, ε, d, s, n be as in the statement of Lemma 6.1, and we let H be an n-vertex $\left(\frac{1}{8}, c, s\right)$-expander with $s \geq \frac{d}{4(\log n)^c}$ such that $\Delta(H) \leq d$, $d(H) \geq d(1-\varepsilon)$ and $\delta(H) \geq \frac{d(H)}{2}$. Moreover, throughout this section, we let

$$q_1 := \frac{1}{3 \log n}, \quad q_2 := \frac{6}{(\log n)^3}, \quad \varepsilon_0 := \frac{1}{(\log n)^4}, \quad t := \frac{1}{6}(\log n)^3 - \frac{1}{18}(\log n)^2 - 1.$$

We make some quick observations that will be used in the rest of the proof. Note that $d(H) \geq d(1-\varepsilon) \geq d(1-\varepsilon_0)$ since $\varepsilon \leq \frac{1}{(\log n)^4} = \varepsilon_0$. Also note that $q_1 + (t+1)q_2 = 1$, that $q_2 \leq \frac{q_1}{4000 \log n}$ and that $t \leq \frac{1}{q_2}$. Moreover, $d \geq (\log n)^{10c+51} \geq \frac{8(\log n)^{10c+21}}{(2q_2)^{10}}$, $\varepsilon_0 \geq 2d^{-1/4}$, and $\frac{\varepsilon_0}{q_2} = \frac{1}{6 \log n}$.

Let us define a random partition $\{V_0, X_1, \ldots, X_t, R\}$ of $V(H)$ as follows. Independently, for each vertex $v \in V(H)$, the probability that v is included in V_0 is q_1, the probability that v is included in X_i is q_2 for each $i \in [t]$, and the probability that v is included in R is also q_2. (Notice that these probabilities sum to $q_1 + (t+1)q_2 = 1$.) For convenience, let $X := \bigcup_{i=1}^{t} X_i$, so that $V(H) = V_0 \cup X \cup R$.

6.2 Finding vertex-disjoint paths $P_1, \ldots, P_r$ covering most of the vertices of X

In the next claim, we show that, with high probability, there is a small number of paths in X, with leaves in $X_1 \cup X_t$, that cover nearly all vertices in H. In Section 6.3, we will join these paths to form a nearly Hamilton path in H.

Claim 6.1.1. *With probability at least $1 - \exp\left(-\Omega(\log n)^2\right)$, there exists a collection of vertex-disjoint paths $P_1, \ldots, P_r$ satisfying the following properties.*

(P1) $\bigcup_{i \in [r]} V(P_i) \subseteq X$.

(P2) *For $i \in [r]$, the leaves of P_i are in $X_1 \cup X_t$.*

(P3) $|X - \bigcup_{i=1}^{r} V(P_i)| \leq \frac{1}{2\log n}|X|$.

Proof of claim. Let us first show that the following three properties hold simultaneously with probability $1 - \exp\left(-\Omega\left((\log n)^2\right)\right)$.

(X1) $e_H(X_i, X_{i+1}) \geq d(1-\varepsilon)nq_2^2 - dq_2^2 n^{2/3}$ for all $i \in [t-1]$.

(X2) $|X_i| = (1 \pm n^{-1/3})q_2 n$ for each $i \in [t]$.

(X3) $\Delta(H[X_i, X_{i+1}]) \leq q_2 d(1 + d^{-1/3})$ for all $i \in [t-1]$.

To show this, it suffices to prove that each of the properties (X1)–(X3) holds with probability $1 - \exp\left(-\Omega\left((\log n)^2\right)\right)$.

For (X1), note that for any $i \in [t-1]$, we have $\mathbb{E}[e_H(X_i, X_{i+1})] = e(H) \cdot 2q_2^2 \geq d(1-\varepsilon)n \cdot q_2^2$ since for any edge $uv \in E(H)$, the probability that $u \in X_i, v \in X_{i+1}$ or $v \in X_i, u \in X_{i+1}$ is $2q_2^2$. Now note that changing the outcome of whether a certain vertex $v \in V(H)$ belongs to X_i, X_{i+1}, or neither, changes $e_H(X_i, X_{i+1})$ by at most $\Delta(H) \leq d$, so $e_H(X_i, X_{i+1})$ is d-Lipschitz. Hence, by Lemma 3.5,

$$\mathbb{P}\left[e_H(X_i, X_{i+1}) < d(1-\varepsilon)nq_2^2 - dq_2^2 n^{2/3}\right] \leq 2\exp\left(-\Omega(q_2^4 n^{1/3})\right).$$

Therefore, by the union bound, since $t \leq n$, (X1) holds with probability at least $1 - 2t\exp\left(-\Omega(q_2^4 n^{1/3})\right) \geq 1 - \exp\left(-\Omega(\log n)^2\right)$, as desired.

For (X2), note that we have $\mathbb{E}[|X_i|] = q_2 n$ for each $i \in [t]$. Then, by the Chernoff bound (Theorem 3.4), the probability that $|X_i| = (1 \pm n^{-1/3})q_2 n$ is at least $1 - \exp\left(-\Omega(q_2 n^{1/3})\right)$. Again, by the union bound, since $t \leq n$, (X2) holds with probability at least $1 - t\exp\left(-\Omega(q_2 n^{1/3})\right) \geq 1 - \exp\left(-\Omega(\log n)^2\right)$, as desired.

Finally, for (X3), note that since $|N_H(v)| \leq d$ for any $v \in V(H)$, we have $\mathbb{E}[|N_H(v) \cap X_i|] \leq q_2 d$. Hence, by the Chernoff bound (Theorem 3.4), for any $v \in V(H)$, $|N_H(v) \cap X_i| \leq q_2 d(1 + d^{-1/3})$ with probability at least $1 - \exp\left(-\Omega(\log n)^2\right)$. Therefore, by the union bound, for every $v \in V(H)$ and $i \in [t]$, $|N_H(v) \cap X_i| \leq q_2 d(1 + d^{-1/3})$ with probability at least $1 - tn\exp\left(-\Omega(\log n)^2\right) \geq 1 - \exp\left(-\Omega(\log n)^2\right)$, implying that (X3) holds with the required probability. This shows that (X1)–(X3) hold simultaneously with probability $1 - \exp\left(-\Omega\left((\log n)^2\right)\right)$, as desired.

We now assume that (X1)–(X3) hold, and use this to deduce that the paths $P_1, \ldots, P_r$ satisfying (P1)–(P3) exist with the required probability. By Vizing's theorem, for every $i \in [t-1]$, there is a matching M_i in $H[X_i, X_{i+1}]$ satisfying

$$|M_i| \geq \frac{e_H(X_i, X_{i+1})}{\Delta(H[X_i, X_{i+1}]) + 1} \geq \frac{d(1-\varepsilon)nq_2^2 - dq_2^2 n^{2/3}}{q_2 d(1 + d^{-1/4})}$$

$$\geq q_2 n(1 - \varepsilon - n^{-1/3})(1 - d^{-1/4})$$

$$\geq \frac{|X_i|}{(1 + n^{-1/3})} \cdot (1 - \varepsilon - n^{-1/3})(1 - d^{-1/4})$$

$$\geq |X_i|(1 - n^{-1/3})(1 - \varepsilon - n^{-1/3})(1 - d^{-1/4})$$

$$\geq |X_i|(1 - \varepsilon - 2n^{-1/3} - d^{-1/4})$$

$$\geq |X_i|(1 - \varepsilon_0 - 2d^{-1/4}) \geq |X_i|(1 - 2\varepsilon_0).$$

Let $M_1' := M_1$, and for each $2 \leq i \leq t-1$, let $M_i' \subseteq M_i$ be the subset of edges of M_i which are incident to an edge of M_{i-1}'. Then, for each $2 \leq i \leq t-1$, since $|M_i| \geq |X_i|(1 - 2\varepsilon_0)$, we have $|M_i'| \geq |M_{i-1}'| - 2\varepsilon_0|X_i|$. This implies that $|M_{t-1}'| \geq |M_1| - 2\varepsilon_0(\sum_{j=2}^{t-1}|X_j|) \geq |X_1| - 2\varepsilon_0(\sum_{j=1}^{t-1}|X_j|) \geq |X_1| - 2\varepsilon_0|X|$. Now, since every edge of M_i' is incident to an edge of M_{i-1}' for every $2 \leq i \leq t-1$, it follows that each edge of M_{t-1}' lies on a path from X_1 to X_t in $M_1 \cup \ldots \cup M_{t-1}$ which contains exactly one vertex from each X_j (for $j \in [t]$), such that the paths are vertex-disjoint. Let us denote these paths by $P_1, \ldots, P_r$, where $r := |M_{t-1}'|$. Then, it is easy to see that (P1) and (P2) hold. For (P3), notice

that since (X2) holds, $t|X_1| \geq t(1 - n^{-1/3})q_2 n = t(1 + n^{-1/3})q_2 n \cdot \frac{1 - n^{-1/3}}{1 + n^{-1/3}} \geq |X|(1 - 2n^{-1/3}) \geq |X|(1 - \varepsilon_0)$. Thus, $\sum_{i=1}^{r} |V(P_i)| = rt \geq (|X_1| - 2\varepsilon_0|X|)t \geq |X|(1 - 3\varepsilon_0 t)$. Hence, $|X - \bigcup_{i=1}^{r} V(P_i)| \leq 3\varepsilon_0 t|X| \leq \frac{3\varepsilon_0}{q_2}|X| = \frac{1}{2\log n}|X|$, as desired. This proves the claim. $\qquad\square$

6.3 Iteratively connecting the paths $P_1, \ldots, P_r$ through V_0 to construct a nearly Hamilton path

In this subsection, our aim is to construct a nearly Hamilton path in H with high probability. We do this in Claim 6.1.3 by repeatedly applying the following claim which shows that if V and W are two random sets, with W only slightly smaller than V, then with high probability one can find vertex-disjoint paths (with internal vertices in V) that connect a positive proportion of the vertices in any given subset of W. Note that, as discussed in Section 2, although the task becomes easier when W is significantly smaller than V (depending on c), it is crucial for our argument that W is allowed to be only slightly smaller than V.

Claim 6.1.2. *Let $0 < p_1, p_2 < 1$ such that $p_2 \leq \frac{p_1}{100}$, let $d \geq \frac{8(\log n)^{10c+21}}{p_2^{10}}$. Let $V, W \subseteq V(H)$ be disjoint random sets such that, independently, for each vertex $v \in V(H)$, the probability that v is included in V is p_1, and the probability that v is included in W is p_2. Then, with probability at least $1 - \frac{2}{n}$, for every $Y \subseteq W$ of size $k \geq 2$, there are at least $\frac{k}{10}$ pairwise vertex-disjoint paths of length at most $(\log n)^{c+4}$ whose internal vertices belong to V and whose endpoints lie in Y.*

Proof of claim. We claim that the following three properties hold simultaneously with probability at least $1 - \frac{2}{n}$.

(SP1) For any vertex $v \in V$, we have $|N_H(v) \cap W| \leq 2p_2 d$.

(SP2) For any vertex $v \in W$, we have $|N_H(v) \cap V| \geq \frac{p_1 d}{5}$.

(SP3) If $x_1, \ldots, x_r, y_1, \ldots, y_r \in V(H) - V$ are distinct vertices such that every subset $X \subseteq \{x_1, \ldots, x_r, y_1, \ldots, y_r\}$ satisfies $|N_H(X)| \geq \frac{100(\log n)^{7c+19}}{p_1^6}|X|$, then there are pairwise vertex-disjoint paths $Q_1, \ldots, Q_r$ of length at most $(\log n)^{c+4}$ with internal vertices in V, such that Q_i is a path joining x_i and y_i.

Indeed, since $\Delta(H) \leq d$, for any vertex $v \in V$, we have $\mathbb{E}[|N_H(v) \cap W|] \leq p_2 d$, so by the Chernoff bound (Theorem 3.4), we have $|N_H(v) \cap W| \leq 2p_2 d$ with probability at least $1 - \exp(-\Omega(p_2 d)) \geq 1 - \exp(-\Omega(\log n)^2)$. Therefore, by the union bound, (SP1) holds with probability at least $1 - \exp(-\Omega(\log n)^2)$. Since $\delta(H) \geq \frac{d(H)}{2} \geq \frac{d}{4}$, for any $v \in W$, we have $\mathbb{E}[|N_H(v) \cap V|] \geq \frac{p_1 d}{4}$, so by the Chernoff bound, we have $|N_H(v) \cap V| \geq \frac{p_1 d}{5}$ with probability at least $1 - \exp(-\Omega(p_1 d)) \geq 1 - \exp(-\Omega(\log n)^2)$. Therefore, by the union bound, (SP2) holds with probability at least $1 - \exp(-\Omega(\log n)^2)$. For (SP3), notice that H is a $(\frac{1}{8}, c, s)$-expander with $s \geq \frac{d}{4(\log n)^c} \geq \frac{2(\log n)^{9c+21}}{p_2^{10}} \geq \frac{2(\log n)^{9c+21}}{p_1^{10}}$, so we can apply Lemma 5.1 with H, p_1 playing the roles of G, q respectively. Thus, (SP3) holds with probability at least $1 - \frac{1}{n}$. By the union bound, (SP1)–(SP3) hold simultaneously, with probability at least $1 - \frac{2}{n}$.

In the rest of the proof of this claim, we assume that (SP1)–(SP3) hold and prove that for every $Y \subseteq W$ of size $k \geq 2$, there are at least $\frac{k}{10}$ pairwise vertex-disjoint paths of length at most $(\log n)^{c+4}$ whose internal vertices are in V and whose leaves are in Y. Indeed, let $Y \subseteq W$ be a set of size $k \geq 2$. Let $\mathcal{C}$ be a maximal collection of vertex-disjoint paths in H of the form abc with $b \in V$ and $a, c \in Y$. If $|\mathcal{C}| \geq k/10$, then $\mathcal{C}$ is the collection of vertex-disjoint paths required by the claim. So we may suppose $|\mathcal{C}| < k/10$.

Let $Y' \subseteq Y$ be the set of vertices in Y which are not contained in any of the paths in $\mathcal{C}$. Then $|Y'| \geq |Y| - 2|\mathcal{C}| \geq |Y| - \frac{k}{5} = \frac{4k}{5}$. Let Y'' be the set of vertices in Y' which have at least $\frac{p_1 d}{30}$ neighbours (in H) in the set $S := \{b \mid abc \in \mathcal{C}\}$. Note that $S \subseteq V$. We claim that $|Y''| \leq \frac{k}{10}$. Indeed, for every $v \in V$, $|N_H(v) \cap W| \leq 2p_2 d$ by (SP1) and $Y'' \subseteq W$, so we have $|N_H(v) \cap Y''| \leq 2p_2 d$. Thus,

$$|Y''| \cdot \frac{p_1 d}{30} \leq e_H(Y'', S) \leq |S| \cdot 2p_2 d = |\mathcal{C}| \cdot 2p_2 d \leq \frac{k}{10} \cdot 2p_2 d.$$

Hence, using that $p_2 \leq p_1/100$, we have $|Y''| \leq 6k \cdot \frac{p_2}{p_1} \leq \frac{k}{10}$, as desired.

Let $Y^* := Y' - Y''$. Then, since $|Y'| \geq \frac{4k}{5}$ and $|Y''| \leq \frac{k}{10}$, we have $|Y^*| \geq \frac{4k}{5} - \frac{k}{10} \geq \frac{7k}{10}$, and for every vertex $v \in Y^*$, we have $|N_H(v) \cap S| < \frac{p_1 d}{30}$ by the choice of Y^*. Thus, for every $v \in Y^* \subseteq W$, we have $|N_H(v) \cap (V-S)| \geq \frac{p_1 d}{5} - \frac{p_1 d}{30} = \frac{p_1 d}{6}$ since $|N_H(v) \cap V| \geq \frac{p_1 d}{5}$ by (SP2). Moreover, by the maximality of the collection $\mathcal{C}$, for any two distinct vertices $u, v \in Y^*$, we have $(N_H(u) \cap (V-S)) \cap (N_H(v) \cap (V-S)) = \emptyset$. In particular, this implies that for any $Z \subseteq Y^*$, we have $|N_H(Z)| \geq |Z| \frac{p_1 d}{6}$.

Let $Y^*_{\text{even}} \subseteq Y^*$ be a subset of size at least $|Y^*| - 1$ such that $|Y^*_{\text{even}}|$ is even. Note that for every $Z \subseteq Y^*_{\text{even}}$, we have $|N_H(Z)| \geq |Z| \frac{p_1 d}{6} \geq |Z| \frac{100(\log n)^{7c+19}}{p_1^6}$. Thus, using (SP3), with Y^*_{even} playing the role of $\{x_1, \ldots, x_r, y_1, \ldots, y_r\}$, we obtain $r = \frac{|Y^*_{\text{even}}|}{2} \geq \frac{|Y^*|-1}{2} \geq \frac{k}{10}$ pairwise vertex-disjoint paths of length at most $(\log n)^{c+4}$ whose internal vertices are in V, and whose leaves are in $Y^* \subseteq Y$. This proves that the claim holds with probability at least $1 - \frac{2}{n}$, as desired. $\qquad\square$

Recall that $\{V_0, X_1, \ldots, X_t, R\}$ is a random partition of $V(H)$, and $X = \bigcup_{i=1}^t X_i$. In the next claim we repeatedly apply Claim 6.1.2 to join the paths $P_1, \ldots, P_r$ (guaranteed by Claim 6.1.1) through the set V_0 to obtain a nearly Hamilton path in H with leaves in $X_1 \cup X_t$.

Claim 6.1.3. *Let V' be a random subset of V_0 obtained by including each vertex of V_0 in V' with probability $1/2$. Then, with probability at least $1 - \frac{21 \log n}{n}$, there is a path P in H such that $V(P) \subseteq V' \cup X$, the leaves of P are contained in $X_1 \cup X_t$, and $|X - V(P)| \leq \frac{1}{2 \log n}|X|$.*

Proof. Since every vertex $v \in V(H)$ belongs to V_0 with probability q_1, every vertex $v \in V(H)$ is included in V' with probability $\frac{q_1}{2}$. Let $\ell := 10 \log n$ and let $q_3 := \frac{q_1}{2\ell}$. Note that since $q_2 \leq \frac{q_1}{4000 \log n} = \frac{q_1}{400\ell}$, we have $q_2 \leq \frac{q_3}{200}$. Let $V' := \bigcup_{i=1}^{\ell} V_i'$ be a random partition of V' so that every vertex $v \in V(H)$ is included in V_i' with probability q_3, for every $i \in [\ell]$. In the rest of the proof of the claim, we condition on the existence of paths $P_1, \ldots, P_r$ satisfying (P1)–(P3) of Claim 6.1.1, and on the conclusion of Claim 6.1.2 holding with $V_i', X_1 \cup X_t, q_3, 2q_2$, playing the roles of V, W, p_1, p_2, respectively, for each $i \in [\ell]$. (Note that Claim 6.1.2 is applied ℓ times in the latter statement using that $2q_2 \leq \frac{q_3}{100}$ and $d \geq \frac{8(\log n)^{10c+21}}{(2q_2)^{10}}$.) Indeed, the former statement holds with probability at least $1 - \exp\left(-\Omega\left((\log n)^2\right)\right)$, and the latter statement holds with probability at least $1 - \frac{2\ell}{n}$, so both statements hold simultaneously with probability at least $1 - \frac{21 \log n}{n}$.

Let $P_1, \ldots, P_r$ be paths satisfying (P1)–(P3) of Claim 6.1.1, and let T_0 be the linear forest whose components are $P_1, \ldots, P_r$. We will iteratively construct linear forests $T_1, \ldots, T_\ell$ such that the number of leaves, say t_i, in T_i decreases quickly as i grows. More precisely, we claim that there is a sequence of linear forests $T_1, \ldots, T_\ell$ such that for every $i \in \{1, \ldots, \ell\}$, T_i satisfies the following properties.

(F1) $V(T_i) \subseteq V(T_0) \cup V_1' \cup \ldots \cup V_i'$,

(F2) The number of leaves t_i in T_i satisfies $2 \leq t_i \leq \max\{2, \frac{9}{10} t_{i-1}\}$, and

(F3) the leaves of T_i are contained in $X_1 \cup X_t$.

To prove this, we use induction on i. For $i = 0$, note that (F1) and (F2) are vacuously true, and (F3) holds due to the choice of the paths $P_1, \ldots, P_r$. Now, let us assume that there is a sequence of linear forests $T_0, \ldots, T_j$ such that T_j satisfies (F1)–(F3), and show that there is a linear forest T_{j+1} satisfying (F1)–(F3).

If $t_j = 2$, define $T_{j+1} = T_j$; it is then easy to check that T_{j+1} satisfies (F1)–(F3). So we may assume that $t_j > 2$. Then, in fact, $t_j \geq 4$ as the number of leaves in a linear forest is always even. Now, let Y be a set of leaves in T_j obtained by taking exactly one leaf from each path of T_j, so that $|Y| = t_j/2 \geq 2$. Note that since T_j satisfies (F3), $Y \subseteq X_1 \cup X_t$. Thus, by the assumption that the conclusion of Claim 6.1.2 holds (with $V_{j+1}', X_1 \cup X_t, q_3, 2q_2$ playing the roles of V, W, p_1, p_2, respectively), there is a collection $\mathcal{P}$ of at least $\frac{t_j/2}{10} = \frac{t_j}{20}$ vertex-disjoint paths of length at most $(\log n)^{c+4}$ whose internal vertices are contained in V_{j+1}', and whose leaves are contained in Y. Let T_{j+1} be the linear forest obtained by adding the edges of the paths in $\mathcal{P}$ to T_j. Note that since each path in $\mathcal{P}$ joins two different paths in T_j, it reduces the number of leaves in T_j by exactly 2, so we have

$t_{j+1} = t_j - 2|\mathcal{P}| \leq \max\{2, t_j - 2(t_j/20)\} = \max\{2, 9t_j/10\}$, showing that T_{j+1} satisfies (F2). Since the set of leaves of T_{j+1} is a subset of the set of leaves in T_j and T_j satisfies (F3), it follows that T_{j+1} satisfies (F3). Finally, since $V(T_{j+1}) \subseteq V(T_j) \cup V'_{j+1} \subseteq V(T_0) \cup V'_1 \cup \ldots \cup V'_{j+1}$, T_{j+1} satisfies (F1). This shows that there is a linear forest T_{j+1} satisfying (F1)–(F3), as desired.

Hence, by repeatedly using (F2) for $1 \leq i \leq \ell$, we obtain that $2 \leq t_\ell \leq \max\{2, (\frac{9}{10})^\ell t_0\}$. Since $\ell = 10 \log n$, we have $(\frac{9}{10})^\ell t_0 \leq e^{-\frac{\ell}{10}} t_0 \leq e^{-\log n} n = 1$. So $t_\ell = 2$. Therefore, the linear forest T_ℓ is actually a path. Moreover, $V(T_\ell) \subseteq V(T_0) \cup V'_1 \cup \ldots \cup V'_\ell \subseteq V' \cup X$ by (F1), and the leaves of T_ℓ are contained in $X_1 \cup X_t$ by (F3). Finally, $|X - V(T_\ell)| \leq |X - \bigcup_{i=1}^r V(P_i)| \leq \frac{1}{2 \log n}|X|$ since $\bigcup_{i=1}^r V(P_i) = V(T_0) \subseteq V(T_\ell)$ and $P_1, \ldots, P_r$ satisfy (P3). This shows that T_ℓ is a path P as required by the claim. $\qquad \square$

6.4 Finding an F-subdivision in H containing the nearly Hamilton path P

In this subsection we complete the proof of Lemma 6.1 by finding a subdivision of F that contains the nearly Hamilton path P guaranteed by Claim 6.1.3 with high probability. In fact, we will find a copy of a subdivision of the complete subgraph K_f with $f := |V(F)|$ such that a path P' containing the nearly Hamilton path P is one of the $\binom{f}{2}$ paths defining this copy. This is indeed sufficient to prove Lemma 6.1 because F is contained in K_f and we may assume that P' joins two vertices which are adjacent in F. To that end, we will need the following well-known result.

Theorem 6.2 (Bollobás-Thomason [10], Komlós-Szemerédi [45])**.** *Let p be a positive integer. Then every graph with average degree at least $512p^2$ contains a subdivision of the complete graph K_p.*

We are now ready to put everything together to complete the proof of Lemma 6.1.

Let $V_0 := V' \cup V_1 \cup V_2$ be a random partition of V_0 such that each $v \in V_0$ is independently included in V' with probability $1/2$, in V_1 with probability $1/4$, and in V_2 with probability $1/4$. We claim that with probability at least $1 - \frac{22 \log n}{n}$ the following five properties hold simultaneously.

(S1) For every $v \in R$, we have $|N_H(v) \cap R| \geq \frac{q_2 d}{5}$.

(S2) $|V_0| \leq 1.01 q_1 n$.

(S3) $|R| \leq 1.01 q_2 n$.

(S4) There is a path P in H such that $V(P) \subseteq V' \cup X$, the leaves of P, denoted u_1, u_2, are contained in $X_1 \cup X_t$, and $|X - V(P)| \leq \frac{1}{2 \log n}|X| \leq \frac{n}{2 \log n}$.

(S5) For every pair of distinct vertices $u, v \subseteq X_1 \cup X_t \cup R$, there is a path in H joining the vertices u and v whose internal vertices are in V_1, and there is another path in H joining the vertices u and v whose internal vertices are in V_2.

Indeed, for (S1), recall that every vertex $v \in V(H)$ is included in R with probability q_2, and that $\delta(H) \geq \frac{d(H)}{2} \geq \frac{d}{4}$. Therefore, for any given vertex $v \in R$, we have $\mathbb{E}[|N_H(v) \cap R|] \geq \frac{q_2 d}{4}$. Thus, (S1) holds with probability at least $1 - \exp\left(-\Omega\left((\log n)^2\right)\right)$, by the Chernoff bound and the union bound. Since each vertex $v \in V(H)$ is included in in V_0 with probability q_1, and in R with probability q_2, (S2) and (S3) hold with probability at least $1 - \exp\left(-\Omega\left((\log n)^2\right)\right)$, by the Chernoff bound. Moreover, by Claim 6.1.3, (S4) holds with probability at least $1 - \frac{21 \log n}{n}$. Finally, (S5) holds with probability at least $1 - \frac{4}{n}$, by applying Claim 6.1.2 twice (with $k = 2$ and $V_i, X_1 \cup X_t \cup R, \frac{q_1}{4}, 3q_2$ playing the roles of V, W, p_1, p_2, respectively, for $i \in [2]$) using that $3q_2 \leq \frac{q_1}{400}$ and $d \geq \frac{8(\log n)^{10c+21}}{(3q_2)^{10}}$. This shows that the properties (S1)–(S5) hold simultaneously with probability at least at least $1 - \frac{22 \log n}{n}$, as desired.

To prove Lemma 6.1, we now condition on the properties (S1)–(S5) holding, and show how to find a subdivision of F covering all but at most $\frac{n}{\log n}$ vertices of H.

Let f denote the number of vertices of F. By (S1), the average degree of $H[R]$ is at least $\frac{q_2 d}{5} \geq \frac{d}{(\log n)^3} \geq \frac{512 d}{(\log n)^4}$. Therefore, by Theorem 6.2, $H[R]$ contains a subdivision K of the complete graph of order $\frac{\sqrt{d}}{(\log n)^2} \geq f$. In other words, there are f vertices $v_1, \ldots, v_f \in V(K) \subseteq R$, and $\binom{f}{2}$ paths $P_{i,j}$ in $H[R]$, for $1 \leq i < j \leq f$, such that $P_{i,j}$ joins v_i and v_j, and the interiors of the paths $P_{i,j}$ are pairwise vertex-disjoint and disjoint of $\{v_1, \ldots, v_f\}$.

Recall that, by (S4), there is a path P in H joining vertices $u_1, u_2 \in X_1 \cup X_t$ such that $V(P) \subseteq V' \cup X$ and $|X - V(P)| \leq \frac{n}{2 \log n}$. Our plan is to replace the path $P_{1,2}$ joining v_1 and v_2 in the subdivision K with a path that contains P as a subpath. By (S5), we know that there is a path $P'_{u_1 v_1}$ in H joining the vertices u_1 and v_1 whose internal vertices are in V_1, and there is a path $P'_{u_2 v_2}$ in H joining the vertices u_2 and v_2 whose internal vertices are in V_2. Now let $P'_{1,2} := P \cup P'_{u_1 v_1} \cup P'_{u_2 v_2}$. Observe that $P'_{1,2}$ is a path in H joining v_1 and v_2, and $V(P'_{1,2}) \cap R = \{v_1, v_2\}$ since $V(P) \subseteq V' \cup X$ by (S4). Hence, $(V(P'_{1,2}) - \{v_1, v_2\}) \cap V(P_{i,j}) = \emptyset$ for every $1 \leq i < j \leq f$.

Let K' be the subgraph of H obtained by replacing the path $P_{1,2}$ in the subdivision K with the path $P'_{1,2}$ (see Figure 2). Then K' is also a subdivision of the complete graph of order $\frac{\sqrt{d}}{(\log n)^2} \geq f$. It is easy to see that (using that F is non-empty) by omitting some of the paths $P_{i,j}$ if necessary, but keeping the path $P'_{1,2}$, we can obtain a subdivision F' of F, we have

$$\begin{aligned}
|V(H) - V(F')| &\leq |V_0| + |X - V(P)| + |R| \leq 1.01 q_1 n + \frac{n}{2 \log n} + 1.01 q_2 n \\
&= \frac{1.01 n}{3 \log n} + \frac{n}{2 \log n} + \frac{6 \cdot 1.01 n}{(\log n)^3} \leq \frac{n}{\log n},
\end{aligned} \tag{10}$$

where we used that the omitted paths are contained in R and (S2)–(S4) for the second inequality. This completes the proof of Lemma 6.1.

7 Packing a regular graph with F-subdivisions

In this section we prove our main result (Theorem 1.2) by combining Lemma 4.1 and Lemma 6.1. The former lemma shows that one can cover almost all vertices of any regular graph with sufficiently regular expanders, and the latter lemma guarantees an almost-spanning F-subdivision within each of these expanders.

Proof of Theorem 1.2. Let G be a d-regular graph of order n with $d \geq (\log n)^{130}$. First, we plan to apply Lemma 4.1 to G. To that end, let $\alpha = \frac{1}{28}$, $C = 6$, so that $c = \frac{C(C-1)}{C - 28\alpha - 1} = 7.5$ and $C - 28\alpha - 1 = 4$. Let $\varepsilon = (\log n)^{-5}$ so that $0 < \varepsilon = (\log n)^{-(C-1)}$. Since G is d-regular, we have $d(G) = d \geq d(1 - \varepsilon)$ and $\Delta(G) \leq d$. Moreover, note that $\frac{C - 28\alpha - 1}{C - 1} = \frac{4}{5} \geq \frac{1}{2}$, and $C \geq 28\alpha + 3 = 4$. Hence, by Lemma 4.1, there is a collection $\mathcal{H}$ of vertex-disjoint subgraphs of G such that every $H \in \mathcal{H}$ is a $(\frac{1}{8}, c, s_H)$-expander satisfying $d(H) \geq d(1 - \varepsilon_H)$ and $\delta(H) \geq \frac{d(H)}{2}$, where $s_H = \frac{d}{4(\log |V(H)|)^c}$, and $\varepsilon_H = (\log |V(H)|)^{-4}$. Moreover,

$$\sum_{H \in \mathcal{H}} |V(H)| \geq \left(1 - \frac{(\log \log \log n)^2}{(\log \log n)^{1/28}}\right) n. \tag{11}$$

Let $H \in \mathcal{H}$ be an arbitrary member of the collection $\mathcal{H}$. Since G is d-regular and H is a subgraph of G, we have $\Delta(H) \leq d$. Hence, by applying Lemma 6.1 to H with ε_H and s_H playing the roles of ε and s, respectively, we obtain a copy K_H of an F-subdivision in H such that

$$\begin{aligned}
|V(K_H)| &\geq \left(1 - \frac{1}{\log |V(H)|}\right)|V(H)| \geq \left(1 - \frac{1}{\log d(H)}\right)|V(H)| \\
&\geq \left(1 - \frac{1}{\log \log n}\right)|V(H)|.
\end{aligned} \tag{12}$$

Note that Lemma 6.1 is indeed applicable because $d \geq (\log n)^{130} \geq (\log |V(H)|)^{10c+51}$. Now consider the collection

27

$\{K_H \mid H \in \mathcal{H}\}$ of vertex-disjoint copies of F-subdivisions in G. By (11) and (12), we have

$$\sum_{H \in \mathcal{H}} |V(K_H)| \geq \left(1 - \frac{1}{\log\log n}\right)\left(1 - \frac{(\log\log\log n)^2}{(\log\log n)^{1/28}}\right)n \geq \left(1 - \frac{1}{(\log\log n)^{1/30}}\right)n.$$

Hence $\{K_H \mid H \in \mathcal{H}\}$ is the desired TF-packing in G, covering all but at most $\frac{n}{(\log\log n)^{1/30}}$ vertices of G. This completes the proof of Theorem 1.2. □

Remark. Note that the proof of Theorem 1.2 actually shows that it suffices for a graph to be nearly regular (rather than regular) to find a TF-packing covering almost all of its vertices. More precisely, it shows that any graph G such that $d(G) \geq d(1 - \frac{1}{(\log n)^5})$, $\Delta(G) \leq d$ and $d \geq (\log n)^{130}$ contains a TF-packing which covers all but at most $\frac{n}{(\log\log n)^{1/30}}$ vertices of G.

8   Applications

In this section, we present two applications of our methods to other problems, each described in a separate short subsection as follows. Both applications rely on the fact that every sufficiently regular expander with large enough average degree contains a nearly Hamilton cycle – a simple consequence of our methods (obtained by letting F be a triangle in Lemma 6.1).

8.1   Cycle partitions of regular graphs

Our first application concerns the well-known conjecture of Magnant and Martin [63], asserting that every d-regular graph contains a collection of at most $\frac{n}{d+1}$ pairwise vertex-disjoint paths that cover all vertices. We establish an asymptotic version of this conjecture in a stronger form for sufficiently large d, allowing for $o(n)$ uncovered vertices but covering the vertices with cycles instead of paths.

Theorem 8.1. *Let G be a d-regular graph of order n, where n is large enough and $d \geq (\log n)^{130}$. Then there is a collection of at most $\frac{n}{d+1}$ vertex-disjoint cycles covering all but at most $O\left(\frac{n}{(\log\log n)^{1/30}}\right)$ vertices of G.*

Recently, Montgomery, Müyesser, Pokrovskiy and Sudakov [65] proved a similar result to Theorem 8.1. Their result is stronger than Theorem 8.1 in that it does not impose a lower bound on d, but it yields a slightly weaker conclusion, giving a collection of paths rather than cycles.

Proof of Theorem 8.1. Write $C := 6$ and $c := \frac{C(C-1)}{C-2} = 7.5$. We apply Lemma 4.1 to the graph G with parameters $\alpha = \frac{1}{28}$, $\varepsilon = 0$, n, d, C and c. Let $\mathcal{H}$ be the resulting collection of expanders guaranteed by the lemma, so that every $H \in \mathcal{H}$ is a $(\frac{1}{8}, c, s_H)$-expander satisfying $d(H) \geq (1 - \varepsilon_H)d$ and $\delta(H) \geq d(H)/2$, where $s_H := \frac{d}{4(\log |V(H)|)^c}$ and $\varepsilon_H := (\log |V(H)|)^{-(C-2)} = (\log |V(H)|)^{-4}$, and, moreover, $\sum_{H \in \mathcal{H}} |V(H)| \geq (1 - \frac{(\log\log\log n)^2}{(\log\log n)^{1/28}})n$. Note that $d \geq (\log n)^{130} \geq (\log |V(H)|)^{10c+51}$.

Next, we apply Lemma 6.1 to each $H \in \mathcal{H}$, with c, ε_H, d, s_H, and $|V(H)|$ taking the roles of c, ε, d, s, and n, respectively (noting that these parameters satisfy the lemma's requirements), and with F being a triangle. For each $H \in \mathcal{H}$, this yields a cycle C_H in H that covers all but at most $\frac{|V(H)|}{\log |V(H)|}$ vertices of H, meaning that $|C_H| \geq (1 - \frac{1}{\log |V(H)|})|V(H)|$. Since $|V(H)| \geq d(H) \geq (1 - \varepsilon_H)d = (1 - \frac{1}{(\log |V(H)|)^4})d$, this implies that $|C_H| \geq (1 - \frac{1}{\log |V(H)|} - \frac{1}{(\log |V(H)|)^4})d \geq (1 - \frac{1}{\log\log n})d$ (where in the last inequality we used that $\log |V(H)| \geq \log(d/2) \geq 2\log\log n$). Then, $\mathcal{C} := \{C_H : H \in \mathcal{H}\}$ is a collection of vertex-disjoint cycles of length at least $(1 - \frac{1}{\log\log n})d$,

covering all but at most $O\left(\frac{n}{(\log\log n)^{1/30}}\right)$ vertices of G as shown by the following.

$$\sum_{H\in\mathcal{H}}\frac{|V(H)|}{\log|V(H)|}+\left(n-\sum_{H\in\mathcal{H}}|V(H)|\right)\leq\sum_{H\in\mathcal{H}}\frac{|V(H)|}{\log\log n}+\frac{n(\log\log\log n)^2}{(\log\log n)^{1/28}}$$
$$\leq n\left(\frac{1}{\log\log n}+\frac{(\log\log\log n)^2}{(\log\log n)^{1/28}}\right)=O\left(\frac{n}{(\log\log n)^{1/30}}\right). \tag{13}$$

Let $\mathcal{C}'\subseteq\mathcal{C}$ be a subcollection of cycles consisting of $\min\{\frac{n}{d+1},|\mathcal{C}|\}$ cycles in $\mathcal{C}$. We claim that $\mathcal{C}'$ covers all but $O(\frac{n}{(\log\log n)^{1/30}})$ vertices in G. Indeed, this is clear from the choice of $\mathcal{C}$ and (13) if $|\mathcal{C}'|=|\mathcal{C}|$ (i.e. if $\mathcal{C}'=\mathcal{C}$). Otherwise, $\mathcal{C}'$ is a collection of $\frac{n}{d+1}$ vertex-disjoint cycles of length at least $(1-\frac{1}{\log\log n})d$, implying that the cycles in $\mathcal{C}'$ cover at least $(1-\frac{1}{\log\log n})d\cdot\frac{n}{d+1}\geq(1-\frac{2}{\log\log n})n=n-O(\frac{n}{\log\log n})$ vertices of G, as desired. □

8.2 Cycles with many chords

Our next application is about finding a cycle with many chords, addressing an old question of Chen, Erdős and Staton [16].

Corollary 8.2. *If n is sufficiently large, then every n-vertex graph with at least $n(\log n)^{130}$ edges contains a cycle C with at least $|C|$ chords.*

This gives a different proof of a recent result of Draganić, Methuku, Munhá Correia, and Sudakov [22], who proved this result with a smaller polylogarithmic factor. Specifically, they showed that $cn(\log n)^8$ edges are sufficient to find a cycle C with at least $|C|$ chords.

We now prove Corollary 8.2 by combining Corollary 4.2, which shows the existence of a nearly regular expander in a graph with sufficiently large average degree, with Lemma 6.1, which finds a nearly Hamilton cycle within such an expander.

Proof of Corollary 8.2. Let G be an n-vertex graph with at least $n(\log n)^{130}$ edges. Applying Corollary 4.2, with $C=6$ and $d=(\log n)^{126}$, we obtain a subgraph $H\subseteq G$ which is a $(\frac{1}{8},c,s)$-expander satisfying $\Delta(H)\leq d$, $d(H)\geq(1-\mu)d$ and $\delta(H)\geq d(H)/2$, where $c=7.5$, $s:=\frac{d}{4(\log m)^c}$, $\mu=(\log m)^{-4}$, and $m:=|V(H)|$.

Applying Lemma 6.1 to H with K_3 playing the role of F, we obtain a cycle C in H such that

$$|V(H)-V(C)|\leq\frac{m}{\log m}.$$

(Note that Lemma 6.1 is indeed applicable here because $d=(\log n)^{126}=(\log n)^{10c+51}\geq(\log m)^{10c+51}$.) Hence, at most $\frac{dm}{\log m}$ edges of H are incident to vertices of H not contained in C. This implies that the number of edges spanned by C is at least

$$\frac{d(1-\mu)m}{2}-\frac{dm}{\log m}\geq\frac{dm}{4}\geq 2|C|.$$

(Here, we rely on the fact that m is large, which follows from $m\geq d(H)\geq(1-\mu)d=\Omega((\log n)^{126})$ and the assumption that n is sufficiently large.) This shows that C is a cycle with at least $|C|$ chords, completing the proof of Corollary 8.2. □

References

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

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

[3] J. Akiyama, G. Exoo, and F. Harary. Covering and packing in graphs. III: Cyclic and acyclic invariants. *Mathematica Slovaca*, 30(4):405–417, 1980.

[4] N. Alon. Problems and results in Extremal Combinatorics–I. *Discrete Mathematics*, 273(1-3):31–53, 2003.

[5] N. Alon and J. Bourgain. Additive patterns in multiplicative subgroups. *Geometric and Functional Analysis*, 24:721–739, 2014.

[6] N. Alon, M. Bucić, L. Sauermann, D. Zakharov, and O. Zamir. Essentially tight bounds for rainbow cycles in proper edge-colourings. *arXiv:2309.04460*, 2023.

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

[8] N. Alon and R. Yuster. *H*-factors in dense graphs. *Journal of Combinatorial Theory, Series B*, 66(2):269–282, 1996.

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

[10] B. Bollobás and A. Thomason. Proof of a conjecture of Mader, Erdős and Hajnal on topological complete subgraphs. *European Journal of Combinatorics*, 19(8):883–887, 1998.

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

[12] M. Bucić, M. Kwan, A. Pokrovskiy, B. Sudakov, T. Tran, and A. Z. Wagner. Nearly-linear monotone paths in edge-ordered graphs. *Israel Journal of Mathematics*, 238(2):663–685, 2020.

[13] M. Bucić and R. Montgomery. Towards the Erdős-Gallai cycle decomposition conjecture. *Advances in Mathematics*, 437:109434, 2024.

[14] D. Chakraborti, O. Janzer, A. Methuku, and R. Montgomery. Edge-disjoint cycles with the same vertex set. *arXiv:2404.07190*, 2024.

[15] D. Chakraborti, O. Janzer, A. Methuku, and R. Montgomery. Regular subgraphs at every density. *arXiv:2411.11785*, 2024.

[16] G. Chen, P. Erdős, and W. Staton. Proof of a conjecture of Bollobás on Nested Cycles. *Journal of Combinatorial Theory, Series B*, 66(1):38–43, 1996.

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

[18] K. Corrádi and A. Hajnal. On the maximal number of independent circuits in a graph. *Acta Mathematica Hungarica*, 14(3-4):423–439, 1963.

[19] B. Csaba, D. Kühn, A. Lo, D. Osthus, and A. Treglown. *Proof of the 1-factorization and Hamilton decomposition conjectures*, volume 244. American Mathematical Society, 2016.

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

[21] G. A. Dirac. Some theorems on abstract graphs. *Proceedings of the London Mathematical Society*, 3(1):69–81, 1952.

[22] N. Draganić, A. Methuku, D. Munhá Correia, and B. Sudakov. Cycles with many chords. *Random Structures & Algorithms*, 65:3–16, 2024.

[23] N. Draganić, R. Montgomery, D. M. Correia, A. Pokrovskiy, and B. Sudakov. Hamiltonicity of expanders: optimal bounds and applications. *arXiv:2402.06603*, 2024.

[24] P. Erdős and A. Hajnal. On complete topological subgraphs of certain graphs. *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.*, 7:143–149, 1964.

[25] P. Erdős. Some recent progress on extremal problems in graph theory. *Congr. Numer*, 14:3–14, 1975.

[26] P. Erdős and A. Renyi. On the evolution of random graphs. *Publ. Math. Inst. Hungary. Acad. Sci.*, 5:17–61, 1960.

[27] U. Feige and E. Fuchs. On the path partition number of 6-regular graphs. *Journal of Graph Theory*, 101(3):345–378, 2022.

[28] I. G. Fernández, J. Hyde, H. Liu, O. Pikhurko, and Z. Wu. Disjoint isomorphic balanced clique subdivisions. *Journal of Combinatorial Theory, Series B*, 161:417–436, 2023.

[29] L. Friedman and M. Krivelevich. Cycle lengths in expanding graphs. *Combinatorica*, 41(1):53–74, 2021.

[30] S. Glock, D. Munhá Correia, and B. Sudakov. Hamilton cycles in pseudorandom graphs. *Advances in Mathematics*, 458, Part B:Article 109984, 2024.

[31] R. J. Gould. Recent advances on the Hamiltonian problem: Survey III. *Graphs and Combinatorics*, 30:1–46, 2014.

[32] V. Gruslys and S. Letzter. Cycle partitions of regular graphs. *Combin. Probab. Comput.*, 30:526–549, 2021.

[33] A. Hajnal and E. Szemerédi. Proof of a conjecture of P. Erdős. *Combinatorial Theory and Its Applications (P.Erdős, A.Rényi, V.Sós, Eds)*, pages 601–623, 1970.

[34] J. Haslegrave, J. Hu, J. Kim, H. Liu, B. Luan, and G. Wang. Crux and long cycles in graphs. *SIAM Journal on Discrete Mathematics*, 36(4):2942–2958, 2022.

[35] J. Haslegrave, J. Kim, and H. Liu. Extremal density for sparse minors and subdivisions. *International Mathematics Research Notices*, 2022(20):15505–15548, 2022.

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

[37] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.

[38] O. Janzer and B. Sudakov. Resolution of the Erdős–Sauer problem on regular subgraphs. In *Forum of Mathematics, Pi*, volume 11, page e19, 2023.

[39] L. K. Jørgensen and L. Pyber. Covering a graph by topological complete subgraphs. *Graphs and Combinatorics*, 6:161–171, 1990.

[40] A. Kelmans, D. Mubayi, and B. Sudakov. Asymptotically optimal tree-packings in regular graphs. *The Electronic Journal of Combinatorics*, 8(1):R38, 2001.

[41] J. Komlós. Tiling Turán theorems. *Combinatorica*, 20:203–218, 2000.

[42] J. Komlós, G. Sárközy, and E. Szemerédi. Proof of the Alon–Yuster conjecture. *Discrete Mathematics*, 235(1-3):255–269, 2001.

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

[44] J. Komlós and E. Szemerédi. Topological cliques in graphs. *Combinatorics, Probability & Computing*, 3:247–256, 1994.

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

[46] A. Koršunov. Solution of a problem of P. Erdős and A. Rényi on Hamiltonian cycles in undirected graphs. *Dokl. Akad. Nauk SSSR*, 228(3):529–532, 1976.

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

[48] M. Krivelevich. Expanders—how to find them, and what to find in them. In *Surveys in combinatorics 2019*, volume 456 of *London Math. Soc. Lecture Note Ser.*, pages 115–142. Cambridge Univ. Press, Cambridge, 2019.

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

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

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

[52] D. Kühn and D. Osthus. Packings in dense regular graphs. *Combinatorics, Probability and Computing*, 14(3):325–337, 2005.

[53] D. Kühn and D. Osthus. The minimum degree threshold for perfect graph packings. *Combinatorica*, 29(1):65–107, 2009.

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

[55] C. Kuratowski. Sur le probleme des courbes gauches en topologie. *Fundamenta mathematicae*, 15(1):271–283, 1930.

[56] S. Letzter. Separating path systems of almost linear size. *Transactions of the American Mathematical Society*, 2024.

[57] S. Letzter. Sublinear expanders and their applications. *Surveys in Combinatorics*, to appear.

[58] S. Letzter, A. Methuku, and B. Sudakov. Packing subgraphs in regular graphs, 2025. Manuscript in preparation.

[59] 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.

[60] H. Liu and R. Montgomery. A solution to Erdős and Hajnal's odd cycle problem. *Journal of the American Mathematical Society*, 36:1191–1234, 2023.

[61] L. Lovász. *Combinatorial Problems and Exercises*. North-Holland, 1979.

[62] W. Mader. Homomorphieeigenschaften und mittlere Kantendichte von Graphen. *Math. Ann.*, 174:265–268, 1967.

[63] C. Magnant and D. M. Martin. A note on the path cover number of regular graphs. *Australas. J Comb.*, 43:211–218, 2009.

[64] R. Montgomery. A proof of the Ryser-Brualdi-Stein conjecture for large even n. *arXiv:2310.19779*, 2023.

[65] R. Montgomery, A. Müyesser, A. Pokrovskiy, and B. Sudakov. Approximate path decompositions of regular graphs. *arXiv:2406.02514*, 2024.

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

[67] L. Pyber. Regular subgraphs of dense graphs. *Combinatorica*, 5:347–349, 1985.

[68] V. Rödl and B. Wysocka. Note on regular subgraphs. *Journal of Graph Theory*, 24:139–154, 1997.

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

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

[71] I. Tomon. Robust (rainbow) subdivisions and simplicial cycles. *Advances in Combinatorics*, 2024.

[72] E. Upfal. *Probability and computing: randomized algorithms and probabilistic analysis*. Cambridge university press, 2005.

[73] J. Verstraëte. A note on vertex-disjoint cycles. *Combinatorics, Probability and Computing*, 11(1):97–102, 2002.