

# Separating path systems of almost linear size

Shoham Letzter\*

## Abstract

A *separating path system* for a graph  $G$  is a collection  $\mathcal{P}$  of paths in  $G$  such that for every two edges  $e$  and  $f$ , there is a path in  $\mathcal{P}$  that contains  $e$  but not  $f$ . We show that every  $n$ -vertex graph has a separating path system of size  $O(n \log^* n)$ . This improves upon the previous best upper bound of  $O(n \log n)$ , and makes progress towards a conjecture of Falgas-Ravry–Kittipassorn–Korándi–Letzter–Narayanan and Balogh–Csaba–Martin–Pluhár, according to which an  $O(n)$  bound should hold.

## 1 Introduction

Given a set  $A$  and a family  $\mathcal{F}$  of sets, we say that  $\mathcal{F}$  is a *strongly-separating system* for  $A$ , or that  $\mathcal{F}$  *strongly-separates*  $A$ , if for every two elements  $a, b \in A$  there is a set  $S \in \mathcal{F}$  satisfying  $a \in S$  and  $b \notin S$ . Similarly, we say that  $\mathcal{F}$  is a *weakly-separating system* for  $A$ , or that  $\mathcal{F}$  *weakly-separates*  $A$ , if for every two elements  $a, b \in A$  there is a set  $S \in \mathcal{F}$  that contains exactly one of  $a$  and  $b$ .<sup>1</sup> In particular, if  $\mathcal{F}$  strongly-separates  $A$  then every element of  $A$  is in some set in  $\mathcal{F}$ , and if it weakly-separates  $A$  then there is at most one element in  $A$  not contained in any set in  $\mathcal{F}$ .

The notion of separation was introduced by Rényi [36] in 1961. It is easy to see that the smallest weakly-separating system for a set of size  $n$  has size  $\lceil \log_2 n \rceil$ , and that the smallest strongly-separating system has size  $(1 + o(1)) \log_2 n$  (see [37]). The problem of determining the size of a smallest separating system becomes more interesting when one imposes restrictions on the members of the separating system. For example, separating systems with restrictions on the size of their members were studied in [25, 29, 35, 41], and different extremal questions about separating systems were studied in [6, 7, 19].

Another interesting direction is to consider separating systems where the ground set is a graph and the members of the separating system satisfy some graph theoretic properties; see [9, 17, 18, 24] for extremal examples [20, 38, 39] for algorithmic examples. Our focus will be on *separating path*

---

\*Department of Mathematics, University College London, Gower Street, London WC1E 6BT, UK. Email: [s.letzter@ucl.ac.uk](mailto:s.letzter@ucl.ac.uk). Research supported by the Royal Society.

<sup>1</sup>The notion of weak separation is sometimes referred to, simply, as *separation*. We keep the word ‘weak’ in the definition, as our focus will be on the strong version.

*systems*, where the ground set is the set of edges of a given graph  $G$  and the members of the separating system are paths in  $G$ . Separating path systems arise naturally in the study of network design (see [2, 20, 38]). In the context of extremal graph theory, this variant was proposed by Gyula Katona in the 5<sup>th</sup> Emléktábla Workshop (2013). Define  $\text{sep}^*(G)$  to be the size of a smallest weakly-separating path system for  $G$ , and let  $\text{sep}^*(n)$  be the maximum of  $\text{sep}^*(G)$  over all  $n$ -vertex graphs  $G$ ; define  $\text{sep}(G)$  and  $\text{sep}(n)$  analogously for strongly-separating path systems. Katona asked to determine  $\text{sep}^*(n)$ .

Separating path systems were first studied by Falgas-Ravry–Kittipassorn–Korándi–Narayanan and the author [13], who focused on the weak version, and, independently, by Balogh–Csaba–Martin–Pluhár [4], who focused on the strong version. Both groups considered path separation in various classes of graphs, including random graphs and trees. Moreover, both groups conjectured that the answer to Katona’s question is  $O(n)$ .<sup>2</sup>

**Conjecture 1.1** (Falgas-Ravry–Kittipassorn–Korándi–Letzter–Narayanan [13] and Balogh–Csaba–Martin–Pluhár [4]). *There exists a constant  $c$  such that every graph on  $n$  vertices has a (weakly- or strongly-)separating path system of size at most  $cn$ .*

In fact, Falgas-Ravry–Kittipassorn–Korándi–Letzter–Narayanan [13] make the bold suggestion that perhaps, for weak separation,  $c$  could be taken to be arbitrarily close to 1, for sufficiently large  $n$ . If true, that would be tight, as can be seen by considering the complete graph  $K_n$ . In fact, determining  $\text{sep}^*(K_n)$  and  $\text{sep}(K_n)$  seems hard; see Wickes [42] for the best-known upper bound on  $\text{sep}^*(K_n)$ .

The study of separating path systems of graphs is related to the well-researched topic of path decompositions. A famous conjecture of Gallai asserts that every connected graph on  $n$  vertices can be decomposed into at most  $\lfloor \frac{n+1}{2} \rfloor$  paths; here by *decomposition* we mean that each edge is covered exactly once. A fundamental result of Lovász [34] proves the following slightly weaker version of Gallai’s conjecture.

**Theorem 1.2** ([34]). *Every  $n$ -vertex graph can be decomposed into at most  $n/2$  paths and cycles. Consequently, every  $n$ -vertex graph can be decomposed into at most  $n$  paths.*

Theorem 1.2 will play an important role in our proof. We sometimes use instead its following easy corollary.

**Corollary 1.3.** *Every  $n$ -vertex graph can be decomposed into at most  $e(G)/d + n$  paths of length at most  $d$ .*

*Proof.* By Theorem 1.2, there is a collection  $\mathcal{P}$  of at most  $n$  paths that decomposes the edges of  $G$ . Decompose each path in  $\mathcal{P}$  into paths of length  $d$  and at most one shorter path. We thus obtain a decomposition  $G$  into paths of length at most  $d$ , at most  $n$  of which have length shorter than  $d$ . The corollary follows as there can be at most  $e(G)/d$  pairwise edge-disjoint paths of length  $d$ .  $\square$

---

<sup>2</sup>Balogh–Csaba–Martin–Pluhár’s version is slightly stronger because they use the stronger notion of separation; we suspect that the two variants are equally hard.

The currently best-known bounds towards Gallai’s conjecture are due to, independently, Dean–Kouider [10] and Yan [43], who proved that every  $n$ -vertex graph can be decomposed into at most  $\frac{2n}{3}$  paths. The conjecture was proved to hold for various families of graphs; see Blanché–Bonamy–Bonichon [5] for a proof for connected planar graphs and the citations therein for other families of graphs.

Theorem 1.2 implies an upper bound of  $\text{sep}(n) = O(n \log n)$ . Indeed, given an  $n$ -vertex graph  $G$ , let  $\mathcal{H}$  be a collection of  $O(\log n)$  subgraphs of  $G$  that separate the edges of  $G$ ; this exists by an aforementioned result regarding separating a set without underlying structure. Now decompose each  $H \in \mathcal{H}$  into at most  $n$  paths. This yields a collection of  $O(n \log n)$  paths separating  $G$ . Until now, this was the best-known general upper bound on  $\text{sep}(n)$ . In this paper we improve it significantly, showing that  $\text{sep}(n) = O(n \log^* n)$ . Here  $\log^* n$  is the iterated logarithm, namely the minimum number of times the base-2 logarithm has to be applied, iteratively, to yield a result which is smaller than 1.

**Theorem 1.4.** *Every  $n$ -vertex graph has a strongly-separating path system of size  $O(n \log^* n)$ . That is,  $\text{sep}(n) = O(n \log^* n)$ .*

A key component in this paper is the use of sublinear expanders. These were introduced by Komlós–Szemerédi [27, 28] and recently various variants of them were used to prove some interesting results [8, 14, 15, 16, 21, 22, 23, 26, 30, 31, 32, 33]. Specifically, we will use many ideas from a very recent paper by Bucić–Montgomery [8], who proved that every  $n$ -vertex graph can be decomposed into  $O(n \log^* n)$  cycles and edges. This is significant progress on the Erdős–Gallai conjecture [11], asserting that every  $n$ -vertex graph can be decomposed into  $O(n)$  cycles and edges. We will also use a tool from [32], where Liu and Montgomery solved the so-called odd cycle problem, due to Erdős–Hajnal [12].

In the following section, Section 2, we give an overview of the proof and an outline of the structure of the paper. We remark that, as our focus throughout the paper is on strong separation, from now on we will refer to it simply as *separation*.

## 2 Proof overview

The main purpose of this section is to give an overview of the proof of Theorem 1.4. We will first describe a general strategy for finding small separating path systems (see Section 2.1), we then show how sublinear expanders come into play (Section 2.2), and we give a more detailed proof overview in Section 2.3. At the end of the section we outline the structure of the paper (see Section 2.4) and define some notation (in Section 2.5).

### 2.1 A general strategy

We start by describing a general strategy for finding a separating path system for a graph  $G$  on  $n$  vertices, which was used in [13]. By Theorem 1.2, there is a collection  $\mathcal{P}$  of at most  $n$  paths that

decomposes  $G$ ; for an edge  $e$ , denote by  $P(e)$  the unique path in  $\mathcal{P}$  that contains  $e$ . We claim that there is a collection  $\mathcal{M}$  of at most  $3n$  matchings that decomposes  $G$ , such that  $\mathcal{P} \cup \mathcal{M}$  separates the edges of  $G$ .

Indeed, take  $\mathcal{M}$  to be a collection of  $3n$  pairwise edge-disjoint matchings, whose edges intersect each path in  $\mathcal{P}$  in at most one edge, and that covers a maximum number of edges under these conditions. Suppose that  $e$  is an edge in  $G$  which is not covered by any of the matchings in  $\mathcal{M}$ . Notice that there are at most  $n - 1$  matchings in  $\mathcal{M}$  that contain edges of  $P(e)$ , and at most  $2(n - 1)$  matchings in  $\mathcal{M}$  that contain an edge that intersects  $e$ . Hence, there is a matching  $M \in \mathcal{M}$  that contains neither edges of  $P(e)$  nor edges that intersect  $e$ . Replacing  $M$  by  $M \cup \{e\}$ , we reach a contradiction to the maximality of  $\mathcal{M}$ , thus showing that  $\mathcal{M}$  covers all edges in  $G$ . Observe that  $\mathcal{P} \cup \mathcal{M}$  separates the edges of  $G$ . Indeed, defining  $M(e)$  to be the unique matching in  $\mathcal{M}$  that contains the edge  $e$ , we see that for any two edges  $e$  and  $f$ , one of  $P(e)$  and  $M(e)$  contains  $e$  but not  $f$ .

Thus, if we could find a path  $P_M$ , for each  $M \in \mathcal{M}$ , that contains  $M$  but avoids all other edges of  $\bigcup_{e \in M} P(e)$ , then we would obtain a separating path system for  $G$  whose size is  $|\mathcal{P}| + |\mathcal{M}| \leq 4n$ .

Of course, this strategy does not always work; otherwise we would have been able to improve on our main result, Theorem 1.4. Nevertheless, if  $G$  has expansion properties and we impose additional restrictions on the matchings in  $\mathcal{M}$ , then this strategy often succeeds. Indeed, in [13] it was used to show that, with high probability, the random graph  $G(n, p)$  has a weakly-separating path system of size  $O(n)$  for every  $p = p(n)$  (similar ideas appear in [4], where the authors found much smaller strongly-separating path systems when  $p \geq \frac{1000}{\log n}$ ). In this paper we apply it to a more general class of expanders.

## 2.2 Expanders

The exact notions of expansion that we shall use will be introduced later (see Sections 4 and 5). We mention that they are variants of the ‘sublinear expanders’, which were introduced by Komlós and Szemerédi [27, 28], and have since been modified in various ways to prove many interesting results [8, 14, 15, 16, 21, 22, 23, 26, 30, 31, 32, 33]. As their name suggests, such expanders have a fairly weak expansion property. Their strength is in their omnipresence: every graph with a sufficiently large (but possibly still constant) average degree contains a sublinear expander. In fact, Bucić and Montgomery [8] recently showed that every graph can be decomposed (or almost decomposed, if we want the expanders to be ‘robust’) into sublinear expanders that, on average, cover each vertex a constant number of times. This is very useful for us; it shows that, essentially, it suffices to be able to find separating path systems of linear size in expanders.

With this decomposition-into-expanders result at hand, a rough plan would be to apply it to decompose a given graph into expanders, and then find a linear separating path system for each expander using the strategy above. If this plan could be realised as stated, then we would end up with a linear separating path system for the original graph. However, depending on the density of the expanders we find, we sometimes require them to be ‘robust’ (or ‘somewhat robust’), and then some edges might remain uncovered. Moreover, in some situations, our methods separate almost all, but not all,

edges of a given expander. Instead, we perform  $O(\log^* n)$  steps, each time significantly decreasing the number of unseparated edges, until remaining with  $O(n)$  edges that we can separate trivially by individual edges.

### 2.3 Proof sketch

We now give a more concrete outline of our proof. Our main task is to prove that, given a graph  $G$  on  $n$  vertices and with average degree  $d$ , there is a subgraph  $G_1$  on  $O(n \text{ polylog } d)$  edges and a collection  $\mathcal{P}$  of  $O(n)$  paths in  $G$  that separates the edges of  $G - G_1$ .

It now essentially suffices to separate the edges of  $G_1$ . Indeed, suppose that  $\mathcal{Q}$  is a collection of paths in  $G$  that separates  $G_1$ . By Theorem 1.2, there are collections  $\mathcal{P}'$  and  $\mathcal{Q}'$  of at most  $n$  paths that decompose  $G - G_1$  and  $G_1$ . We claim that  $\mathcal{P} \cup \mathcal{P}' \cup \mathcal{Q} \cup \mathcal{Q}'$  separates the edges of  $G$ . To see this, let  $e$  and  $f$  be distinct edges in  $G$ . If both are in  $G - G_1$ , then  $\mathcal{P}$  has a path containing  $e$  but not  $f$ ; if both are in  $G_1$  then  $\mathcal{Q}$  has such a path; if  $e$  is in  $G - G_1$  and  $f$  is not, then  $\mathcal{P}'$  has a path through  $e$  but not  $f$ ; and finally if  $e$  is in  $G_1$  but  $f$  is not then  $\mathcal{Q}'$  has such a path.

Iterating the argument in the paragraph before last thus yields a separating path system for  $G$  of size  $O(n \log^* n)$ . We complete this task in three steps.

**Step 1: dense expanders.** We show (in Lemma 3.1) that given a graph  $J$  on  $m$  vertices, it has a subgraph  $J_1$  of size  $O(m \text{ polylog } m)$  such that  $J - J_1$  can be separated using  $O(m)$  paths. To do so, we decompose  $J$  into ‘robust’ expanders (with total order  $O(m)$ ) and a remainder of size  $O(m \text{ polylog } m)$  and show that such expanders can be separated using a linear number of paths (in Corollary 6.4).

To separate a robust expander  $K$ , we show that if  $V$  is a set of vertices in  $K$  that includes each vertex with probability  $1/3$ , then, with high probability, the edges of  $K - V$  can be separated by a linear number of paths. To find these paths, we apply the strategy given in Section 2.1 to  $K - V$ , to obtain small collections  $\mathcal{P}$  of paths and  $\mathcal{M}$  of matchings that together separate the edges of  $K$ , while imposing a certain degree condition on the matchings in  $\mathcal{M}$ . We then partially connect the edges of each  $M \in \mathcal{M}$  using a subset  $V_1$  of  $V$ , chosen uniformly at random, thereby forming a path forest whose ends ‘expand well’ into  $V - V_1$ . Now, we adapt a result from [8] (which in turn uses ideas from a paper of Tomon [40]) that allows for connecting pairs of vertices, disjointly, via a random set of vertices.

**Step 2: separating edges touching large degrees in sparse expanders.** We again start by decomposing a graph  $G$  on  $n$  vertices and with average degree  $d$  into expanders (here the expanders are not required to be robust, and so we have no uncovered edges), whose sum of orders is  $O(n)$ . For each such expander  $H$ , we show (in Lemma 8.1) that its edges touching high degree vertices can be separated using  $O(n + \frac{e(H)}{d})$  paths.

Denoting by  $L$  the set of large degree vertices in  $H$ , let us explain how to separate the edges in  $H[L]$ . Writing  $k := n + \frac{e(H)}{d}$ , we apply the strategy in Section 2.1, while additionally requiring

that the paths in  $\mathcal{P}$  and matchings in  $\mathcal{M}$  have size at most  $d$ ; it is straightforward to show that there exist appropriate  $\mathcal{P}$  and  $\mathcal{M}$  of size  $O(k)$ . Now, given a matching  $M \in \mathcal{M}$ , let  $F$  be a path forest that extends  $M$ , avoids vertices in  $X := \bigcup_{e \in \mathcal{M}} V(P(e)) - V(M)$ , and under these conditions has as few components as possible, and as few edges as possible (prioritising the former). Suppose, towards a contradiction, that there are at least two components, and let  $u$  and  $v$  be leaves of distinct components of  $F$ . The minimality assumption allows us to show that, roughly speaking, balls around  $u$  in  $H - X$  have few neighbours in  $F$ . Adapting a lemma from Liu–Montgomery [32], it follows that more than  $|H|/2$  vertices in  $H$  can be reached from  $u$  in  $H - (X \cup V(F) - \{u\})$ , exploiting the fact that  $u$  has a large degree. Since the same holds for  $v$ , it follows that  $F$  can be extended by a path from  $u$  to  $v$ , a contradiction. Thus  $F = F_M$  is a path that extends  $M$  and avoids other edges from  $\bigcup_{e \in \mathcal{M}} P(e)$ , and  $\{F_M : M \in \mathcal{M}\} \cup \mathcal{P}$  is a collection of  $O(k)$  paths separating  $G[L]$ . In fact, similar reasoning can be applied to separate edges  $xy$  where  $x \in L$  and  $y$  has at least four neighbours in  $L$ , by allowing the members of  $\mathcal{M}$  to contain paths of length 2 whose ends are in  $L$ .

**Step 3: separating almost all remaining edges in  $H$ .** For each  $H$  as in the previous step, let  $F$  be its subgraph of ‘unseparated edges’; so  $F$  has small maximum degree. We now decompose  $F$  into ‘somewhat robust’ expanders (with total order  $O(|H|)$ ) and a remainder of  $O(\text{polylog } d \cdot |H|)$  edges. Consider one of these expanders  $J$ . If  $|J|$  is fairly large with respect to  $d$  (namely, if  $|J| \geq 2^{(\log d)^7}$ ) then we show (in Lemma 8.4) that its edges can be separated using  $O(|J|)$  edges.

The proof is similar to the one outlined in Step 2. Here, in addition to requiring the paths and matchings in  $\mathcal{P}$  and  $\mathcal{M}$  to have size at most  $d$ , we also require the edges in  $\mathcal{M}$  to be far from each other (the assumption that  $|J|$  is large with respect to  $d$  is used to prove that appropriate  $\mathcal{M}$  exists). Defining  $M$ ,  $X$  and  $F$  as above, it suffices to show that  $F$  has just one component. To this end, taking  $u$  to be a leaf in  $F$ , we first show that the ball of radius  $(\log d)^5$  around  $u$  in  $J' := J - (X \cup V(F) - \{u\})$  is large, exploiting the expansion property being ‘somewhat robust’ and the fact that the other edges of  $M$  are far from  $u$ . We then proceed like in the previous step to expand  $u$  further, and reach a contradiction if  $F$  has at least two components. As before, this implies the existence of a separating path system for  $J$  of size  $O(|J| + \frac{e(J)}{d})$ .

If instead  $|J| \leq 2^{(\log d)^7}$ , we apply the first step to show that all but  $O(|J| \text{polylog } |J|)$  edges of  $J$  can be separated using  $O(|J|)$  paths.

Altogether, we get a collection of  $O(n)$  paths that separate all but  $O(n \text{polylog } d)$  edges of the graph  $G$  that we started with at the beginning of the second step, concluding the iterative step.

## 2.4 Organisation of the paper

In the next section (Section 3) we state two key lemmas — Lemma 3.1, which covers Step 1 above, and Lemma 3.2, which covers Steps 2 and 3 — and show how our main theorem (Theorem 1.4) follows from these lemmas. In Section 4 we state and prove the decomposition-into-expanders lemma, Lemma 4.1, which will be used in the proofs of both key lemmas. Section 5 contains the proof of Lemma 5.10, which allows for connecting pairs of vertices in an expander disjointly through

a random set of vertices, and is a variant of a lemma from [8]. The first key lemma (Lemma 3.1) is then proved in Section 6. In Section 7 we prove a variant of a lemma from [32] regarding expansion with forbidden sets. Finally, in Section 8 we prove the second key lemma (Lemma 3.2).

## 2.5 Notation

Our notation is mostly standard. Given a graph  $G$ , a set of vertices  $X$ , and integer  $i \geq 0$ , we write  $N_G(X)$  to be the set of vertices *outside* of  $X$  that send at least one edge into  $X$ , and write  $B_G^i(X)$  to be the set of vertices at distance at most  $i$  from  $X$  in  $G$ . Given a vertex  $x$  and a set of vertices  $X$  in a graph  $G$ , we denote the number of neighbours of  $x$  in  $X$  by  $d_G(x, X)$ . As is customary, we omit the subscript  $G$  when it is clear from the context. Given a graph  $G$  and two disjoint sets of vertices  $A, B$  in it, we write  $G[A, B]$  for the bipartite subgraph of  $G$ , with bipartition  $(A, B)$ , whose edges are the edges of  $G$  with one end in  $A$  and the other in  $B$ . Given vertices  $x, y$ , an  $(x, y)$ -*path* is a path from  $x$  to  $y$ . We denote the length of a path  $P$  by  $\ell(P)$ . We say that a sequence of events  $(E_n)_{n \geq 1}$  holds *with high probability* if it holds with probability tending to 1 as  $n$  tends to infinity. All logarithms will be in base 2, unless specified otherwise, and  $\log^* n$  is *iterated logarithm*, namely the least number of times the base-2 logarithms has to be applied, iteratively and starting with  $n$ , to yield a number which is smaller than 1. When dealing with large numbers we often omit floor and ceiling signs when they are not crucial.

## 3 Main lemmas

In this section we state two key lemmas and then show how to use them to prove our main result (Theorem 1.4).

The first key lemma shows that given any graph on  $n$  vertices there is a linear collection of paths separating all but  $O(n \text{ polylog } n)$  of its edges.

**Lemma 3.1.** *Suppose that  $G$  is a graph on  $n$  vertices. Then there is a subgraph  $G_1 \subseteq G$ , with  $e(G_1) \leq n(\log n)^{55}$ , and a collection  $\mathcal{P}$  of at most  $1300n$  paths in  $G$  that separates the edges of  $G - G_1$ .*

The second key lemma has a more complicated statement; essentially, it shows that for every graph  $G$  there is a linear collection of paths that separates a subgraph of  $G$  whose complement in  $G$  can be decomposed into small subgraphs and a sparse remainder.

**Lemma 3.2.** *There exists  $d_0$  such that the following holds. Suppose that  $G$  is a graph on  $n$  vertices with average degree  $d$ , where  $d \geq d_0$ . Then there is a subgraph  $G_1 \subseteq G$ , a collection  $\mathcal{H}$  of subgraphs of  $G$  and a collection  $\mathcal{P}$  of paths in  $G$  with the following properties.*

- (1) *The graphs in  $\mathcal{H}$  are mutually edge-disjoint;  $|H| \leq 2^{(\log d)^7}$  for every  $H \in \mathcal{H}$ ; and  $\sum_{H \in \mathcal{H}} |H| \leq 4n$ .*

- (2) The collection  $\mathcal{P}$  separates the edges of  $G - G_1 - \bigcup_{H \in \mathcal{H}} H$ , and satisfies  $|\mathcal{P}| \leq 80n$ .
- (3) The graph  $G_1$  has average degree at most  $(\log d)^3$ .

The following corollary combines the two key lemmas to show that given a graph with average degree  $d$ , there is a linear collection of paths that separates all edges but a remainder whose average degree is  $O(\text{polylog } d)$ .

**Corollary 3.3.** *Suppose that  $G$  is a graph on  $n$  vertices with average degree  $d$ . Then there is a subgraph  $G_1 \subseteq G$  with average degree at most  $O((\log d)^{420})$  and a collection  $\mathcal{P}$  of  $O(n)$  paths in  $G$  that separates  $G - G_1$ .*

*Proof.* We may assume that  $d$  is large, because otherwise we could take  $G_1$  to be the empty graph and  $\mathcal{P} = E(G)$ . Apply Lemma 3.2 to obtain  $\mathcal{H}$  and  $\mathcal{P}$  that satisfy the conditions of the lemma. Namely, the graphs in  $\mathcal{H}$  are mutually edge-disjoint and consist of at most  $2^{(\log d)^7}$  vertices,  $\sum_{H \in \mathcal{H}} |H| \leq 4n$ ,  $|\mathcal{P}| \leq 80n$ , and the paths in  $\mathcal{P}$  separate the edges in  $G'$ , where  $G' := G - G'' - \bigcup_{H \in \mathcal{H}} H$  and  $G''$  is a subgraph of  $G'$  with average degree  $O((\log d)^3)$ . Let  $\mathcal{P}'$  be a collection of at most  $n$  paths in  $G'$  that decomposes the edges of  $G'$ ; such a collection exists by Theorem 1.2.

Apply Lemma 3.1 to each graph  $H$  in  $\mathcal{H}$ ; denote the resulting subgraph and collection of paths by  $G_H$  and  $\mathcal{P}_H$ . So  $e(G_H) \leq |H|(\log |H|)^{55}$ ,  $|\mathcal{P}_H| = O(|H|)$ , and  $\mathcal{P}_H$  is a collection of paths in  $H$  that separates the edges of  $H - G_H$ . Write  $G_1 := \bigcup_{H \in \mathcal{H}} G_H \cup G''$ . Then

$$\sum_{H \in \mathcal{H}} e(G_H) \leq \sum_{H \in \mathcal{H}} |H|(\log |H|)^{55} \leq \left( \sum_{H \in \mathcal{H}} |H| \right) \cdot (\log d)^{420} = O(n(\log d)^{420}).$$

It follows that  $e(G_1) = O(n(\log d)^{420})$ , or, equivalently,  $G_1$  has average degree  $O((\log d)^{420})$ . Let  $\mathcal{P}'_H$  be a collection of at most  $|H|$  paths in  $H$  that decomposes  $E(H)$ , and write  $\mathcal{Q} := \mathcal{P} \cup \mathcal{P}' \cup \bigcup_{H \in \mathcal{H}} (\mathcal{P}_H \cup \mathcal{P}'_H)$ . Then  $|\mathcal{Q}| = O(n)$ , using  $\sum_{H \in \mathcal{H}} |H| = O(n)$ .

We claim that  $\mathcal{Q}$  separates the edges of  $G - G_1$ . To see this, consider two edges  $e$  and  $f$  in  $G - G_1$ . We consider four cases.

- There is  $H \in \mathcal{H}$  such that  $e, f \in E(H)$ . Then, there is  $P \in \mathcal{P}_H$  such that  $e \in E(P)$  and  $f \notin E(P)$ .
- There is  $H \in \mathcal{H}$  such that  $e \in E(H)$  and  $f \notin E(H)$ . Then, there is a path  $P \in \mathcal{P}'_H$  in  $H$  that contains  $e$ , and so  $f \notin E(P)$ .
- Both  $e$  and  $f$  are in  $G'$ . Then, there is  $P \in \mathcal{P}$  such that  $e \in E(P)$  and  $f \notin E(P)$ .
- The edge  $e$  is in  $G'$  and  $f$  is not in  $G'$ . Then, there is  $P \in \mathcal{P}'$  such that  $e \in E(P)$  and  $f \notin E(P)$ .

The subgraph  $G_1$  and collection  $\mathcal{Q}$  of paths satisfy the requirements of the corollary.  $\square$



The next corollary, applies the previous one twice, to prove a very similar statement, with the average degree of the remainder now being at most  $\log d$ . We could have omitted this corollary and proved Theorem 1.4 directly from Corollary 3.3, but the counting is slightly simpler with Corollary 3.4.

**Corollary 3.4.** *Let  $G$  be a graph on  $n$  vertices with average degree  $d$ , where  $d$  is large. Then there is a subgraph  $G_1 \subseteq G$  with average degree at most  $\log d$  and a collection  $\mathcal{P}$  of  $O(n)$  paths in  $G$  that separates  $G - G_1$ .*

*Proof.* We apply Corollary 3.3 twice. Denote by  $G_1$  and  $\mathcal{P}_1$  the subgraph and collection of paths resulting from the first application, and let  $G_2$  and  $\mathcal{P}_2$  be the subgraph and collection of paths resulting from a second application with the graph  $G_1$ . Apply Theorem 1.2 to get a collection  $\mathcal{P}'_1$  of at most  $n$  paths that decomposes  $G - G_1$ , and define  $\mathcal{P}'_2$  similarly with respect to  $G_1 - G_2$ . Write  $\mathcal{Q} = \mathcal{P}_1 \cup \mathcal{P}_2 \cup \mathcal{P}'_1 \cup \mathcal{P}'_2$ . Then  $|\mathcal{Q}| = O(n)$  and  $\mathcal{Q}$  separates the edges of  $G - G_2$  (this is easy to verify using a similar analysis to the proof of Corollary 3.3).

By the properties of  $G_1$  and  $G_2$  given by Corollary 3.3 we have  $d(G_1) = O((\log d)^{420})$  and  $d(G_2) = O((\log d(G_1))^{420}) = O((\log \log d)^{420})$ . Since  $d$  is large, it follows that  $d(G_2) \leq \log d$ . The subgraph  $G_2$  and collection  $\mathcal{Q}$  satisfy the requirements of the corollary.  $\square$

Finally, we deduce our main result (Theorem 1.4) from Corollary 3.4.

*Proof of Theorem 1.4.* Write  $G_0 := G$  and  $d := d(G)$ . We define sequences  $(G_i)_{i \geq 1}$  and collections of paths  $(\mathcal{P}_i)_{i \geq 1}$  as follows. As long as  $d(G_i)$  is large enough to apply Corollary 3.4, let  $G_{i+1}$  and  $\mathcal{P}_{i+1}$  be the graph and collection of paths resulting from an application of Corollary 3.4 to  $G_i$ . Then  $d(G_{i+1}) \leq \log d(G_i)$ ,  $|\mathcal{P}_{i+1}| = O(n)$  and  $\mathcal{P}$  separates the edges of  $G_i - G_{i+1}$ . Suppose that the last graph we defined was  $G_k$ . Then  $k \leq \log^* n$  and  $e(G_k) = O(n)$ . For  $i < k$  apply Theorem 1.2 to obtain a collection of paths  $\mathcal{P}'_i$  that decomposes  $G_i - G_{i+1}$  and let  $\mathcal{P}'_k := E(G_k)$ .

Let  $\mathcal{Q} := \bigcup_{0 \leq i < k} \mathcal{P}_i \cup \bigcup_{0 \leq i \leq k} \mathcal{P}'_i$ . Then  $|\mathcal{Q}| = O(nk) = O(n \log^* n)$ . We claim that  $\mathcal{Q}$  separates  $E(G)$ . To see this, let  $e$  and  $f$  be two edges in  $G$ , let  $i$  be such that  $e \in G_i - G_{i+1}$  (with  $i = k$  if  $e \in G_k$ ), and define  $j$  similarly with respect to  $f$ . If  $i \neq j$  then there is a path  $P \in \mathcal{P}'_i$  that contains  $e$  but not  $f$ . If  $i = j \neq k$  then there is a path  $P \in \mathcal{P}_i$  that contains  $e$  but not  $f$ . Finally, if  $i = j = k$ , then  $e$  is a path in  $\mathcal{P}'_k$  which contains  $e$  but not  $f$ . Thus  $\mathcal{Q}$  is a collection of  $O(n \log^* n)$  paths that separates the edges of  $G$ .  $\square$

## 4 Decomposition into expanders

We say that a graph  $G$  on  $n$  vertices is an  $(\varepsilon, s, t)$ -*expander* if  $|N_{G-F}(X)| \geq \frac{\varepsilon|X|}{(\log|X|+1)^2}$  for every  $X \subseteq V(G)$  and  $F \subseteq E(G)$  satisfying  $1 \leq |X| \leq 2n/3$  and  $|F| \leq s \cdot \min\{|X|, t\}$ . An  $(\varepsilon, s)$ -*expander* is an  $(\varepsilon, s, 2n/3)$ -expander (namely,  $F$  is not restricted by  $t$ ). An  $\varepsilon$ -*expander* is an  $(\varepsilon, 0)$ -expander (namely,  $F$  is always empty). This definition builds on previously defined notions of expansion: Komlós and Szemerédi [27, 28] introduced a notion of sublinear expanders, which are (up to a

slight change in parameters)  $\varepsilon$ -expanders in our notation. Haslegrave, Kim and Liu [23] introduced a notion of robust sublinear expanders, which are  $(\varepsilon, s)$ -expanders in our notation. We add the parameter  $t$ , which provides a cut-off in the robustness: vertex sets of size at most  $t$  expand robustly, whereas larger vertex sets expand less robustly.

The following lemma shows that every graph  $G$  can be decomposed into expanders, or almost decomposed into robust expanders. This is a variant of Lemma 14 in [8] and the proof is similar but more technical. The main difference is the introduction of the parameter  $t$ , which allows us to obtain expanders where small sets of vertices expand robustly, while not leaving too many edges uncovered. As a sanity check, note that as  $s$  and  $t$  increase, the expansion property becomes stronger, and the potential number of uncovered edges grows.

We will apply this lemma with several regimes for  $s$  and  $t$ : in Section 6 we take  $s$  to be polylogarithmic in  $n$  (and take  $t = 2n/3$ ). In Section 8 we apply this lemma twice, once with  $s = 0$  (in which case the choice of  $t$  does not matter), and once with  $s$  a large constant and  $t$  being the average degree of the ground graph.

**Lemma 4.1.** *Let  $s \geq 0$ ,  $t \geq 1$  and let  $0 < \varepsilon \leq 1/48$ . Let  $G$  be a graph on  $n$  vertices. Then there is a collection  $\mathcal{H}$  of pairwise edge-disjoint  $(\varepsilon, s, t)$ -expanders that covers all but at most  $48sn(\log t + 1)^2$  edges of  $G$  and satisfies  $\sum_{H \in \mathcal{H}} |H| \leq 2n$ .*

*Proof.* We leave a few simple calculus claims (see Claims 4.3 to 4.5) until after this proof. Notice that we may assume  $t \leq \frac{2n}{3}$ , by our definition of expanders.

We prove the statement by induction on  $n$ , with the following slightly stronger upper bounds:  $24sn \cdot (\log t + 1)^2 (2 - \frac{1}{\log n + 1})$  on the number of edges removed, and  $n \cdot (2 - \frac{1}{\log n + 1})$  on the sum of orders. If  $G$  is itself an  $(\varepsilon, s, t)$ -expander then we may take  $\mathcal{H} = \{G\}$ , noting that  $n \cdot (2 - \frac{1}{\log n + 1}) \geq n$ .

Suppose then that  $G$  is not an  $(\varepsilon, s, t)$ -expander. In particular,  $n \geq 2$  because the singleton graph is an  $(\varepsilon, s, t)$ -expander. This means that there are subsets  $X \subseteq V(G)$  and  $F \subseteq E(G)$  such that  $1 \leq |X| \leq 2n/3$ ,  $|F| \leq s \cdot \min\{|X|, t\}$ , and  $|N_{G-F}(X)| \leq \frac{\varepsilon|X|}{(\log|X|+1)^2}$ . Let  $G_1 := G[X \cup N_{G-F}(X)]$  and let  $G_2$  be the subgraph obtained from  $G - X$  by removing the edges with both ends in  $N_{G-F}(X)$ . Then  $G_1$  and  $G_2$  are edge-disjoint, together they cover all edges in  $G - F$ , and they satisfy

$$\begin{aligned} n \leq |G_1| + |G_2| &= n + |N_{G-F}(X)| \leq n + \frac{\varepsilon|X|}{(\log|X|+1)^2} \\ &\leq n + \frac{3\varepsilon|G_1|}{(\log|G_1|+1)^2} \leq n + \frac{|G_1|}{16(\log|G_1|+1)^2}, \end{aligned} \tag{1}$$

where the penultimate inequality follows from Claim 4.3 stated below.

The bulk of the calculations needed for this lemma will be performed in the following claim.

**Claim 4.2.**

$$|G_1| \left( 2 - \frac{1}{\log|G_1|+1} \right) + |G_2| \left( 2 - \frac{1}{\log|G_2|+1} \right) + \frac{\min\{|G_1|, t\}}{24(\log t + 1)^2} \leq n \left( 2 - \frac{1}{\log n + 1} \right).$$

*Proof.* Write  $D$  for the difference between the left-hand side and the right-hand side.

$$\begin{aligned}
D &= |G_1| \left( 2 - \frac{1}{\log |G_1| + 1} \right) + |G_2| \left( 2 - \frac{1}{\log |G_2| + 1} \right) - n \left( 2 - \frac{1}{\log n + 1} \right) + \frac{\min\{|G_1|, t\}}{24(\log t + 1)^2} \\
&= 2(|G_1| + |G_2| - n) - \frac{|G_1|}{\log |G_1| + 1} - \frac{|G_2|}{\log |G_2| + 1} + \frac{n}{\log n + 1} + \frac{\min\{|G_1|, t\}}{24(\log t + 1)^2} \\
&\leq \frac{|G_1|}{8(\log |G_1| + 1)^2} - \frac{|G_1|}{\log |G_1| + 1} - \frac{|G_2|}{\log n + 1} + \frac{n}{\log n + 1} + \frac{\min\{|G_1|, t\}}{24(\log t + 1)^2} \\
&\leq |G_1| \left( \frac{1}{8(\log |G_1| + 1)^2} - \frac{1}{\log |G_1| + 1} + \frac{1}{\log n + 1} + \frac{\min\{1, t/|G_1|\}}{24(\log t + 1)^2} \right),
\end{aligned}$$

where for the first inequality we used (1). We consider two cases:  $|G_1| \geq t$  and  $|G_1| < t$ . In the first case, using Claim 4.3, we have

$$\frac{\min\{1, t/|G_1|\}}{24(\log t + 1)^2} = \frac{t}{24|G_1|(\log t + 1)^2} \leq \frac{3|G_1|}{24|G_1|(\log |G_1| + 1)^2} = \frac{1}{8(\log |G_1| + 1)^2}.$$

Thus, noting that  $|G_1| \leq |X| + \frac{\varepsilon|X|}{(\log |X| + 1)^2} \leq |X| \left( 1 + \frac{1}{24} \right) \leq \frac{2n}{3} \cdot \frac{25}{24} \leq \frac{3n}{4}$ , and using Claim 4.5,

$$D \leq |G_1| \left( \frac{1}{4(\log |G_1| + 1)^2} - \frac{1}{\log |G_1| + 1} + \frac{1}{\log n + 1} \right) \leq 0.$$

In the second case, by Claim 4.4, we have  $\frac{1}{8(\log |G_1| + 1)^2} - \frac{1}{\log |G_1| + 1} \leq \frac{1}{8(\log t + 1)^2} - \frac{1}{\log t + 1}$ . Thus,

$$\begin{aligned}
D &\leq |G_1| \left( \frac{1/8 + 1/24}{(\log t + 1)^2} - \frac{1}{\log t + 1} + \frac{1}{\log n + 1} \right) \\
&\leq |G_1| \left( \frac{1}{4(\log t + 1)^2} - \frac{1}{\log t + 1} + \frac{1}{\log n + 1} \right) \leq 0,
\end{aligned}$$

where the last inequality follows from Claim 4.5, using  $1 \leq t \leq 2n/3$ . Either way,  $D \leq 0$ , as required.  $\square$

For  $i \in [2]$ , apply the induction hypothesis to  $G_i$  to obtain a family  $\mathcal{H}_i$  of pairwise edge-disjoint subgraphs of  $G_i$  that cover all but at most  $24s(\log t + 1)^2|G_i|(2 - \frac{1}{\log |G_i| + 1})$  edges of  $G_i$  and satisfy  $\sum_{H \in \mathcal{H}_i} |H| \leq |G_i| \left( 2 - \frac{1}{\log |G_i| + 1} \right)$ . Take  $\mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2$ . By Claim 4.2, the graphs in  $\mathcal{H}$  cover all but at most the following number of edges in  $G$ :

$$\begin{aligned}
&24s(\log t + 1)^2 \left( |G_1| \left( 2 - \frac{1}{\log |G_1| + 1} \right) + |G_2| \left( 2 - \frac{1}{\log |G_2| + 1} \right) + \frac{\min\{|G_1|, t\}}{24(\log t + 1)^2} \right) \leq \\
&24sn(\log t + 1)^2 \left( 2 - \frac{1}{\log n + 1} \right).
\end{aligned}$$

Similarly,

$$\begin{aligned} \sum_{H \in \mathcal{H}} |H| &\leq \sum_{H \in \mathcal{H}_1} |H| + \sum_{H \in \mathcal{H}_2} |H| \leq |G_1| \left( 2 - \frac{1}{\log |G_1| + 1} \right) + |G_2| \left( 2 - \frac{1}{\log |G_2| + 1} \right) \\ &\leq n \left( 2 - \frac{1}{\log n + 1} \right). \end{aligned}$$

So  $\mathcal{H}$  satisfies the requirements, proving the induction hypothesis.  $\square$

We now state and prove the calculus claims used in the proof of Lemma 4.1.

**Claim 4.3.**  $\frac{x}{(\log x+1)^2} \leq \frac{3y}{(\log y+1)^2}$  for integers  $x, y$  with  $1 \leq x \leq y$ .

*Proof.* Consider the function  $f(x) = \frac{x}{(\log x+1)^2}$ . Then

$$f'(x) = \frac{(\log_e(2))^2 \log_e(2x/e^2)}{(\log_e(2x))^3}.$$

In particular,  $f'(x) \geq 0$  for  $x \geq 4$ , showing  $f(x) \leq f(y) \leq 3f(y)$  when  $4 \leq x \leq y$  (using  $f(y) \geq 0$  for  $y \geq 4$ ). A calculation shows  $1/3 \leq f(x) \leq 1$  for  $x \in [4]$ . It follows that  $f(x) \leq 3f(y)$  for  $x \in [3]$  and any integer  $y \geq 1$ .  $\square$

**Claim 4.4.**  $\frac{1}{8(\log x+1)^2} - \frac{1}{\log x+1} \leq \frac{1}{8(\log y+1)^2} - \frac{1}{\log y+1}$  when  $1 \leq x \leq y$ .

*Proof.* Write  $f(x) = \frac{1}{8(\log x+1)^2} - \frac{1}{\log x+1}$ . Then the derivative  $f'(x)$  satisfies the following for  $x \geq 1$ .

$$f'(x) = \frac{\log_e(2)(4 \log_e x + \log_e 8)}{4x(\log_e(2x))^3} \geq 0.$$

In particular,  $f$  is increasing in the range  $x \geq 1$ , as required.  $\square$

**Claim 4.5.** The following holds for  $1 \leq x \leq \log(3n/4) + 1$  and  $n \geq 2$ .

$$\frac{1}{4x^2} - \frac{1}{x} + \frac{1}{\log n + 1} \leq 0.$$

*Proof.* Write  $f(x) = \log n + 1 - 4x(\log n + 1) + 4x^2$ , and notice that it suffices to prove  $f(x) \leq 0$  for  $1 \leq x \leq \log(2n/3) + 1$ .

Since the set of real numbers  $x$  for which  $f(x) \leq 0$  is an interval, it suffices to show that  $f(x) \leq 0$

for  $x = 1$  and  $x = \log(2n/3) + 1$ .

$$f(1) = \log n + 1 - 4(\log n + 1) + 4 = 1 - 3 \log n \leq 0.$$

$$\begin{aligned} f(\log(3n/4) + 1) &= \log n + 1 - 4(\log(3n/4) + 1)(\log n + 1) + 4(\log(3n/4) + 1)^2 \\ &= \log n + 1 - 4(\log n + \log(3/2))(\log n + 1) + 4(\log n + \log(3/2))^2 \\ &= \log n \cdot (1 - 4 - 4 \log(3/2) + 8 \log(3/2)) + 1 - 4 \log(3/2) + 4(\log(3/2))^2 \\ &\leq -\frac{1}{2} \cdot \log n + \frac{1}{2} \leq 0. \end{aligned}$$

This completes the proof of the claim. □

## 5 Finding vertex-disjoint paths through a random vertex set

In this section we prove Lemma 5.10, which allows for connecting pairs of vertices  $(x_1, y_1), \dots, (x_r, y_r)$  in a robust expander through a random set of vertices via vertex-disjoint paths, provided that any subset of the vertices  $x_1, \dots, x_r, y_1, \dots, y_r$  expands well. This is a variant of Theorem 10 in [8], which does something very similar, but avoids the expansion assumption on  $x_1, \dots, x_r, y_1, \dots, y_r$  and only requires the paths to be edge-disjoint. Our proof follows [8] very closely. We note that, as mentioned also in [8], some of the ideas that appear in the proof are due to Tomon [40].

We will introduce the relevant preliminaries (including a different notion of expansion) in Section 5.1, we then prove Lemma 5.10 in three steps, given in Sections 5.2 to 5.4.

### 5.1 Preliminaries

Say that a graph  $G$  on  $n$  vertices is a *weak*  $(\varepsilon, s)$ -*expander* if  $|N_{G-F}(U)| \geq \frac{\varepsilon|U|}{(\log n)^2}$  for every  $U \subseteq V(G)$  and  $F \subseteq E(G)$  with  $1 \leq |U| \leq 2n/3$  and  $|F| \leq s|U|$ . In [8] such a graph is called simply an  $(\varepsilon, s)$ -expander; we add ‘weak’ to its name to distinguish it from  $(\varepsilon, s)$ -expanders, which have stronger expansion properties for smaller sets of vertices. Note that an  $(\varepsilon, s)$ -expander on at least two vertices is a weak  $(\frac{\varepsilon}{2}, s)$ -expander.

We will use the following two propositions from [8], both of which show that every not-too-large vertex set  $U$  in an expander either ‘expands well’ (i.e. has a large neighbourhood) or ‘expands robustly’ (i.e. there are many vertices in  $N(U)$  with many neighbours in  $U$ ). We state a slightly simplified version of the propositions, where there is no set of forbidden edges, because we will not use this feature.

**Proposition 5.1** (Proposition 12 in [8]). *Let  $\varepsilon > 0$ , let  $0 < d \leq s$ . Suppose that  $G$  is an  $n$ -vertex  $(\varepsilon, s)$ -weak expander. Then one of the following holds, for every set  $U \subseteq V(G)$  with  $|U| \leq \frac{2n}{3}$ .*

- $|N(U)| \geq \frac{s|U|}{2d}$ , or

- $|\{u \in V(G) : |N(u) \cap U| \geq d\}| \geq \frac{\varepsilon|U|}{(\log n)^2}$ .

**Proposition 5.2** (Proposition 13 in [8]). *Let  $\varepsilon \geq 2^{-9}$ ,  $s \geq 8(\log n)^{13}$ , and let  $n$  be large. Suppose that  $G$  is an  $(\varepsilon, s)$ -weak expander, and let  $W \subseteq V(G)$  satisfy  $|W| \leq 2n/3$ . Then  $G$  contains one of the following*

- a collection of at least  $\frac{|W|}{(\log n)^7}$  pairwise vertex-disjoint stars of size at least  $(\log n)^9$ , whose centre is in  $W$  and its leaves are in  $V(G) - W$ ,
- a bipartite graph  $H$  with parts  $W$  and  $X \subseteq V(G) - W$ , such that
  - $|X| \geq \frac{\varepsilon|W|}{2(\log n)^2}$ ,
  - every vertex in  $X$  has degree at least  $(\log n)^4$  and every vertex in  $W$  has degree at most  $2(\log n)^9$  in  $H$ .

As in [8], we also need the following hypergraph version of Hall’s theorem, due to Aharoni and Haxell.

**Theorem 5.3** (Aharoni–Haxell [1]). *Let  $s, r \geq 1$  be integers, and let  $\mathcal{H}_1, \dots, \mathcal{H}_r$  be hypergraphs on the same vertex set whose edges have at most  $s$  vertices. Suppose that, for every  $I \subseteq [r]$ , there is a matching in  $\bigcup_{i \in I} \mathcal{H}_i$  of size at least  $s(|I| - 1)$ . Then there is a matching of size  $r$  in  $\bigcup_{i \in [r]} \mathcal{H}_i$  whose  $i^{\text{th}}$  edge is in  $\mathcal{H}_i$ .*

Additionally, we use the following martingale concentration result (see Chapter 7 in [3]). We say that a function  $f : \prod_{i=1}^n \Omega_i \rightarrow \mathbb{R}$ , where  $\Omega_i$  are arbitrary sets, is  $k$ -Lipschitz if  $|f(u) - f(v)| \leq k$  for every  $u, v \in \prod_{i=1}^n \Omega_i$  that differ on at most one coordinate.

**Lemma 5.4.** *Let  $X_1, \dots, X_n$  be independent random variables, with  $X_i$  taking values in a set  $\Omega_i$  for  $i \in [n]$ , and write  $X = (X_1, \dots, X_n)$ . Suppose that  $f : \prod_{i=1}^n \Omega_i \rightarrow \mathbb{R}$  is  $k$ -Lipschitz. Then,*

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

## 5.2 Expansion into a random vertex set

The following lemma shows that, given a sufficiently robust weak expander  $G$  and a random set of vertices  $V$ , obtained by including each vertex independently with probability  $1/6$ , any set  $U$  expands well in  $V$ , while avoiding a given small set of vertices  $Z$ , with quite large probability. This lemma is a variant of Lemma 17 in [8], where instead of a forbidden set of edges we have a forbidden set of vertices, and the proof is essentially the same.

**Lemma 5.5.** *Let  $2^{-9} \leq \varepsilon < 1$ ,  $s \geq 8(\log n)^{13}$  and let  $n$  be large. Suppose that  $G$  is an  $n$ -vertex weak  $(\varepsilon, s)$ -expander and let  $U, Z \subseteq V(G)$  be sets satisfying  $|U| \geq (\log n)^{23}$  and  $|Z| \leq \frac{|U|}{(\log n)^3}$ . Let  $V$*

be a random subset of  $V(G)$ , obtained by including each vertex independently with probability  $1/6$ . Then, with probability at least  $1 - \exp\left(-\Omega\left(\frac{|U|}{(\log n)^{22}}\right)\right)$ ,

$$\left|B_{G[V']}^{(\log n)^4}(U \cap V')\right| > \frac{|V|}{2},$$

where  $V' := V - Z$ .

*Proof.* Take  $\ell := (\log n)^4$ , and let  $p$  satisfy  $1 - (1 - p)^\ell(1 - \frac{3}{20})(1 - \frac{1}{120}) = \frac{1}{6}$ , so that

$$1 - p\ell \leq (1 - p)^\ell = \frac{1 - 1/6}{(1 - 3/20) \cdot (1 - 1/120)} \leq \frac{99}{100},$$

which implies that  $p \geq \frac{1}{100(\log n)^4}$ . Let  $V_1, \dots, V_\ell, V^*, V^{**}$  be random sets, where  $V_i$  is obtained by including each vertex with probability  $p$ , independently,  $V^*$  is obtained by including each vertex with probability  $\frac{1}{120}$ , independently, and  $V^{**}$  is obtained by including each vertex with probability  $\frac{3}{20}$ , independently. Notice that each vertex is in  $V_1 \cup \dots \cup V_\ell \cup V^* \cup V^{**}$  with probability  $1/6$ , so we may think of  $V$  as the union  $V_1 \cup \dots \cup V_\ell \cup V^* \cup V^{**}$ .

Define  $U^* = (U \cap V^*) - Z$ . Then, with probability  $1 - \exp(-\Omega(|U|))$ , we have  $|U^*| \geq \frac{1}{200}|U| - |Z| \geq \frac{1}{400}|U|$ . 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 interior is in  $V_1 \cup \dots \cup V_i$ . We 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  $|B_i| \geq |B_0| = |U^*| \geq \frac{1}{400}|U|$  for  $i \geq 0$  when  $|U^*| \geq \frac{1}{400}|U|$ .

**Claim 5.6.** *If  $|B_i| \leq \frac{2}{3}n$ , then, with probability at least  $1 - \exp\left(-\Omega\left(\frac{|U|}{(\log n)^{22}}\right)\right)$ ,*

$$|B_{i+1} - B_i| \geq \frac{\varepsilon|B_i|}{10^4(\log n)^2}.$$

*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  $V_{i+1}$ . We consider the two possible outcomes of Proposition 5.2 for  $B_i$ .

Suppose that the first outcome holds, so there are  $\frac{|B_i|}{(\log n)^7}$  pairwise vertex-disjoint stars of size  $(\log n)^9$  with centres in  $B_i$  and leaves in  $N(B_i)$ . By a Chernoff bound, with probability at least  $1 - \exp\left(-\Omega\left(\frac{p|B_i|}{(\log n)^7}\right)\right) = 1 - \exp\left(-\Omega\left(\frac{|U|}{(\log n)^{11}}\right)\right)$ , at least  $\frac{p|B_i|}{2(\log n)^7}$  centres are added to  $V_{i+1}$ , implying that

$$|B_{i+1} - B_i| \geq (\log n)^9 \cdot \frac{p|B_i|}{2(\log n)^7} - |Z| \geq \frac{|B_i|}{200(\log n)^2} - \frac{|U|}{(\log n)^3} \geq \frac{|B_i|}{400(\log n)^2}.$$

Now suppose that the second outcome holds, so there is a bipartite subgraph  $H \subseteq G$  with parts  $B_i$  and  $X \subseteq V(G) - B_i$ , with  $|X| \geq \frac{\varepsilon|B_i|}{4(\log n)^2}$ , such that vertices in  $X$  have degree at least  $d := (\log n)^4$  in  $H$  while vertices in  $B_i$  have degree at most  $D := 2(\log n)^9$  in  $H$ . Let  $Y$  be the set of vertices in  $X$  that do not have an  $H$ -neighbour in  $V_{i+1}$ . Note that  $\mathbb{E}|Y| \leq |X|(1 - p)^d \leq |X|e^{-pd} \leq \frac{399}{400}|X|$ . Note

also that  $|Y|$  is  $D$ -Lipschitz, since the outcome of the sampling of any single vertex in  $B_i$  affects the outcome of at most  $D$  vertices in  $X$ . Thus, by Lemma 5.4,

$$\begin{aligned} \mathbb{P}\left(|Y| > \frac{799}{800}|X|\right) &\leq \mathbb{P}\left(|Y| > \mathbb{E}|Y| + \frac{|X|}{800}\right) \leq 2 \exp\left(-\frac{|X|^2}{2 \cdot 800^2 \cdot D^2 |B_i|}\right) \\ &= \exp\left(-\Omega\left(\frac{|U|}{(\log n)^{22}}\right)\right). \end{aligned}$$

So, with probability at least  $1 - \exp\left(-\Omega\left(\frac{|U|}{(\log n)^{22}}\right)\right)$ , we have  $|B_{i+1} - B_i| \geq \frac{|X|}{800} - |Z| \geq \frac{\varepsilon |B_i|}{10^4 (\log n)^2}$ .  $\square$

Iterating the claim, with probability at least

$$1 - (\log n)^4 \cdot \exp\left(-\Omega\left(\frac{|U|}{(\log n)^{22}}\right)\right) \geq 1 - \exp\left(-\Omega\left(\frac{|U|}{(\log n)^{22}}\right)\right),$$

if  $|B_i| \leq \frac{2}{3}n$  and  $i \leq (\log n)^4$ , then,

$$|B_i| \geq \left(1 + \frac{\varepsilon}{10^4 (\log n)^2}\right)^i |U|,$$

which implies  $|B_\ell| \geq \frac{2}{3}n$  with probability  $1 - \exp\left(-\Omega\left(\frac{|U|}{(\log n)^{22}}\right)\right)$ .

To complete the proof, note that any vertex in  $B_\ell$  that gets sampled into  $V^{**}$  (or that is already in  $V_1 \cup \dots \cup V_\ell \cup V^*$ ) is in the set  $B' := B_{G[V']}^{(\log n)^4}(U \cap V')$ . By Chernoff, with probability at least  $1 - e^{-\Omega(n)}$ , at least  $\frac{99}{100} \cdot \frac{3}{20} \cdot \frac{2}{3}n = \frac{99}{1000}n$  vertices of  $B_\ell$  are in  $V$ , and  $|V| \leq \frac{101}{100} \cdot \frac{1}{6}n = \frac{101}{600}n$ . If all this holds, then  $|B'| > \frac{1}{2}|V|$ , as required.  $\square$

The next corollary, which is a variant of Lemma 19 in [8], boosts the probability of expansion of  $U$  into  $V$ , so as to be amenable to an application of the union bound, with the expense of a slightly smaller upper bound on  $|Z|$ .

**Corollary 5.7.** *Let  $2^{-9} \leq \varepsilon < 1$ ,  $s \geq 2^{11}(\log n)^{50}$ , and let  $n$  be large. Suppose that  $G$  is an  $n$ -vertex weak  $(\varepsilon, s)$ -expander, and let  $V$  be a random subset of  $V(G)$ , obtained by including each vertex independently with probability at least  $1/6$ . Then, with high probability, if  $U, Z \subseteq V(G)$  are sets with  $|Z| \leq \frac{|U|}{(\log n)^{30}}$ , then*

$$\left|B_{G[V']}^{(\log n)^4}(N(U) \cap V')\right| > \frac{|V|}{2}, \quad (2)$$

where  $V' := V - Z$ .

*Proof.* Say that a set  $U \subseteq V(G)$  expands well if  $|N(U)| \geq |U|(\log n)^{24}$ . Given a (non-empty) set  $U$  which expands well, and a set  $Z$  with  $|Z| \leq |U|$ , Lemma 5.5 (applied with  $U_{5.5} = N(U)$ ) tells us that (2) holds, with probability at least  $1 - \exp\left(-\Omega(|U|(\log n)^2)\right)$ .

By a union bound, we conclude that the probability that (2) fails for  $(U, Z)$ , with  $U$  well-expanding



and  $|Z| \leq |U|$ , is at most

$$\begin{aligned} \sum_{u=1}^n \binom{n}{u}^2 \exp(-\Omega(u(\log n)^2)) &\leq \sum_{u=1}^n \exp(2u \log n - \Omega(u(\log n)^2)) \\ &= \sum_{u=1}^n \exp(-\Omega(u(\log n)^2)) = o(1/n). \end{aligned}$$

For the rest of the proof we assume that (2) holds for all such pairs  $(U, Z)$ . To complete the proof, we will deduce that (2) holds for all pairs  $(U, Z)$ , where  $U$  need not expand well,  $|U| \leq \frac{2}{3}n$ , and  $|Z| \leq \frac{|U|}{(\log n)^{30}}$ . Fix such  $U, Z$ .

Write  $d := \frac{2}{\varepsilon}(\log n)^{26}$ . By Proposition 5.1, either  $|N(U)| \geq \frac{s|U|}{2d}$ , or  $|\{u : |N(u) \cap U| \geq d\}| \geq \frac{\varepsilon|U|}{(\log n)^2}$ . Notice that the first outcome implies that  $U$  expands well and so we already know that (2) holds. We thus assume that the second one holds, and write  $W := \{u : |N(u) \cap U| \geq d\}$ . 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  $\mathbb{E}[|W \cap N(U')|] \geq (1 - e^{-1})|W| \geq \frac{\varepsilon|U|}{2(\log n)^2} = |U'|(\log n)^{24}$ . In particular, there is a subset  $U' \subseteq U$  of size  $\frac{|U|}{d}$  with  $|N(U')| \geq |U'|(\log n)^{24}$ . Since such  $U'$  expands well, and  $|Z| \leq \frac{|U|}{(\log n)^{30}} \leq |U'|$ , Equation (2) holds for  $(U', Z)$  and thus for  $(U, Z)$ .  $\square$

### 5.3 A path connection through a random set

Our end goal is to be able to join (disjointly)  $r$  pairs of vertices through a random set  $V$ , provided that they expand well. Here we show how to join one pair of vertices (while avoiding a small set of forbidden vertices). This is a variant of Proposition 8 in [8], where in addition to forbidding vertices instead of edges, we also impose the expansion property on the vertices we wish to connect. The proof is essentially the same.

**Lemma 5.8.** *Let  $2^{-9} \leq \varepsilon < 1$ ,  $s \geq 2^{11}(\log n)^{50}$ , and let  $n$  be large. Suppose that  $G$  is an  $n$ -vertex weak  $(\varepsilon, s)$ -expander, and let  $V$  be a random subset of  $V(G)$ , obtained by including each vertex independently with probability  $1/6$ . Then, with high probability, the following holds for every  $r$ : if  $x_1, \dots, x_r, y_1, \dots, y_r$  are distinct vertices, satisfying  $|N(X)| \geq |X|(\log n)^{50}$  for every subset  $X \subseteq \{x_1, \dots, x_r, y_1, \dots, y_r\}$ , and  $Z$  is a set of size at most  $2r(\log n)^{12}$  which is disjoint from  $\{x_1, \dots, x_r, y_1, \dots, y_r\}$ , then for some  $i \in [r]$  there is an  $(x_i, y_i)$ -path in  $G$  whose interior is in  $V - Z$  and whose length is at most  $(\log n)^6$ .*

*Proof.* Fix an outcome of  $V$  such that Equation (2) holds for every disjoint  $U, Z \subseteq V(G)$  with  $|Z| \leq \frac{|U|}{(\log n)^{30}}$  (here  $V' := V - Z$ ), and moreover  $|V| \geq \frac{n}{8}$ . By Corollary 5.7 and a Chernoff bound,

both these assumptions hold with high probability, and so we are justified in making them. Write  $\ell := (\log n)^4 + 1$ . 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  $\{x_1, \dots, x_r\}$  satisfying  $|R^{\ell \log n}(x)| \leq \frac{|V|}{2}$ .

**Claim 5.9.**  $|X_1| < \frac{r}{2}$ .

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

$$|R^{i\ell}(X_i)| > \frac{|V|}{2}, \quad (3)$$

for  $i \geq 1$ .

Notice that  $|N(X_1)| \geq \frac{r}{2}(\log n)^{50}$ , using  $|X_1| \geq \frac{r}{2}$  and the assumption on the  $x_i$ 's. Thus  $|N(X_1)| \geq |Z|(\log n)^{30}$ , and, by (2),  $|R^{(\log n)^4}(N(X_1) \cap V')| > \frac{|V|}{2}$ . It follows that  $|R^\ell(X_1)| \geq |R^{(\log n)^4}(X_1)| > \frac{|V|}{2}$ , proving (3) for  $i = 1$ .

Now suppose that  $X_1 \supseteq \dots \supseteq X_j$  and (3) holds for  $i \in [j]$ . If  $|X_i| = 1$  we take  $X_{i+1} = X_i$  (which clearly satisfies the requirements). Otherwise, by dividing  $X_i$  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  $|R^{i\ell}(X_{i+1})| \geq \frac{|V|}{6}$ . Consider the set  $U := R^{i\ell}(X_{i+1})$ . It satisfies  $|U| \geq \frac{|V|}{6} \geq \frac{n}{48} \geq \frac{r}{48}(\log n)^{50} \geq |Z|(\log n)^{30}$ , where the third inequality follows implicitly from the assumption that  $|N(\{x_1, \dots, x_r\})| \geq r(\log n)^{50}$ . Thus, by (2),

$$|B_{G[V']}^{(\log n)^4}(N(U) \cap V')| > \frac{|V|}{2}.$$

Notice that

$$B_{G[V']}^{(\log n)^4}(N(U) \cap V') = B_{G[V']}^\ell(U) = R^{(i+1)\ell}(X_{i+1}),$$

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 above 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  $X_1$ .  $\square$

Take  $Y_1$  to be the set of vertices  $y \in \{y_1, \dots, y_r\}$  with  $|R^{\ell \log n}(y)| \leq \frac{|V|}{2}$ . Then, analogously to the above claim,  $|Y_1| < \frac{r}{2}$ . 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 there is an  $(x, y)$ -path of length at most  $2\ell \log n + 2$ , with interior in  $V'$ , as required.  $\square$

## 5.4 Many disjoint path connections through a random set

Finally, we prove that given a sufficiently robust weak expander and pairs  $(x_1, y_1), \dots, (x_r, y_r)$ , such that the set  $\{x_1, \dots, x_r, y_1, \dots, y_r\}$  expands well, the pairs can be connected via vertex-disjoint paths through a random set  $V$ . This is a variant of Lemma 10 in [8] (though their statement looks quite different).

**Lemma 5.10.** *Let  $2^{-9} \leq \varepsilon < 1$ ,  $s \geq 2^{11}(\log n)^{50}$ , and let  $n$  be large. Suppose that  $G$  is an  $n$ -vertex weak  $(\varepsilon, s)$ -expander, and let  $V$  be a random subset of  $V(G)$ , that includes each vertex independently with probability  $\frac{1}{6}$ . Then, with high probability, for every sequence of vertices  $x_1, \dots, x_r, y_1, \dots, y_r$  satisfying  $|N(X)| \geq |X|(\log n)^{50}$  for every  $X \subseteq \{x_1, \dots, x_r, y_1, \dots, y_r\}$ , there is a sequence of paths  $P_1, \dots, P_r$ , whose interiors are pairwise vertex-disjoint and in  $V$ , such that  $P_i$  is a path from  $x_i$  to  $y_i$ .*

*Proof.* We assume that the conclusion of Lemma 5.8 holds (it holds with high probability, so we are justified in making this assumption). Let  $\mathcal{H}_i$  be the hypergraph on vertex set  $V$  with edges the interiors of  $P$ , for all  $(x_i, y_i)$ -paths  $P$  of length at most  $\ell := (\log n)^6$  with interior in  $V$ .

We will use Theorem 5.3. Fix a subset  $I \subseteq [r]$ . We wish to show that there is a matching of size  $\ell(|I| - 1) + 1$  in  $\mathcal{H}' := \bigcup_{i \in I} \mathcal{H}_i$ . Suppose no such matching exist, and let  $\mathcal{M}'$  be a maximal matching in  $\mathcal{H}'$ . Then  $|\mathcal{M}'| \leq \ell(|I| - 1)$  and every edge in  $\mathcal{H}'$  intersects  $Z := V(\mathcal{M}')$ , a set of size at most  $\ell^2(|I| - 1) \leq |I|(\log n)^{12}$ . The conclusion of Lemma 5.8 tells us that for some  $i \in I$  there is an  $(x_i, y_i)$ -path  $P$  of length at most  $\ell$  whose interior is 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.

So, the assumptions in Theorem 5.3 are satisfied, showing that there is a matching  $\mathcal{M}$  of size  $r$  in  $\bigcup_{i \in [r]} \mathcal{H}_i$ , whose  $i$ -th edge is in  $\mathcal{H}_i$ . Let  $P_i$  be the path corresponding to the  $i$ -th edge in  $\mathcal{M}$ . Then  $P_i$  is an  $(x_i, y_i)$ -path with interior in  $V$ , and the interiors of the  $P_i$ 's are pairwise vertex-disjoint, proving the lemma.  $\square$

## 6 Proof of Lemma 3.1: separating dense expanders

In this section we prove the first key lemma, Lemma 3.1, which shows that for any graph  $G$ , there is a linear (i.e. of size  $O(|G|)$ ) path system that separates a subgraph of  $G$ , whose complement in  $G$  has polylogarithmic average degree. This will follow quite easily from Lemma 4.1, that decomposes graphs into expanders, and Corollary 6.4 that shows that robust weak expanders can be separated by a linear path system.

To prove the latter corollary, given a robust weak expander  $G$ , we will take a random partition  $\{V_1, V_2, V_3\}$  of  $V(G)$  and show that, with high probability,  $G - V_i$  can be separated using a linear path system, for  $i \in [3]$ . Taking the union of these collections of paths, as well as a decomposition of  $G - V_i$  into a linear number of paths, we obtain the desired linear separating paths system for  $G$ .

It thus suffices to prove that, with high probability, if  $V$  is a random set, obtained by including each vertex independently with probability  $1/3$ , then  $G - V$  has a linear separating path system. This is done in the following lemma, whose proof proceeds as follows. First, we decompose  $G - V$  into at most  $n$  paths  $\mathcal{P}$  (using Theorem 1.2), and then we show that  $G - V$  can be further decomposed into  $O(n)$  matchings  $M$ , each of which intersects each path in  $\mathcal{P}$  in at most one edge, and moreover each matching satisfies a certain degree condition. Next, we take a random partition  $\{V_1, V_2\}$  of  $V$ , taken uniformly at random. We first extend each  $M$  through  $V_1$  to a path forest  $F$ , in such a way

that any set of leaves of  $F$  expands in  $G$ , and thus in  $V$ . This is exactly the property needed for an application of Lemma 5.10, so we can use it to connect the leaves of  $F$  via paths through  $V$ , vertex-disjointly. We thus get a path  $P_M$  that extends  $M$  and contains no other edges from  $G - V$ , and so  $\mathcal{P} \cup \{P_M : M \in \mathcal{M}\}$  is a linear collection of paths separating  $G - V$ .

**Lemma 6.1.** *Let  $2^{-9} \leq \varepsilon < 1$ ,  $s \geq 2^{11}(\log n)^{51}$ , and let  $n$  be large. Suppose that  $G$  is a weak  $(\varepsilon, s)$ -expander, and let  $V$  be a random subset of  $V(G)$ , that includes each vertex independently with probability  $\frac{1}{3}$ . Then, with high probability, there is a collection of  $211n$  paths that separates  $G - V$ .*

*Proof.* Let  $\{V_1, V_2\}$  be a partition of  $V$ , chosen uniformly at random. We will make the following assumptions.

- (a) if  $|N(U)| \geq |U|(\log n)^2$  then  $|N(U) \cap V_1| \geq \frac{1}{10}|N(U)|$ , for every set of vertices  $U$ ,
- (b) the conclusion of Lemma 5.10 holds for  $V_2$ .

Noting that each vertex is included in  $V_i$  with probability  $\frac{1}{6}$ , independently, a Chernoff bound and Lemma 5.10 allow us to make these assumptions.

Write  $G' := G - V$ . Let  $\mathcal{P}$  be a collection of at most  $n$  paths that decomposes  $G'$ . For an edge  $e = xy$  in  $G'$ , let  $P(e)$  be the unique path in  $\mathcal{P}$  that contains  $e$ , and define  $\bar{d}(e) := \min\{d_G(x), d_G(y)\}$ . Notice that, by expansion,  $\delta(G) \geq s \geq (\log n)^{51}$  and so  $\bar{d}(e) \geq (\log n)^{51}$  for every edge  $e$  in  $G$ .

**Claim 6.2.** *There is a collection  $\mathcal{M}$  of  $210n$  pairwise edge-disjoint matchings in  $G'$  that covers all edges in  $G'$  and every  $M \in \mathcal{M}$  satisfies: every path in  $\mathcal{P}$  intersects at most one edge in  $M$ ; and for every  $d$  there are at most  $\frac{d}{50}$  edges  $e$  in  $M$  with  $\bar{d}(e) \leq d$ .*

*Proof.* Let  $\mathcal{M}$  be a collection of  $210n$  pairwise edge-disjoint matchings such that every  $M \in \mathcal{M}$  satisfies the following properties,

- every path in  $\mathcal{P}$  contains at most one edge from  $M$ ,
- for every integer  $r \geq 1$ , there are at most  $\frac{1}{200}2^r$  edges  $e$  in  $M$  with  $\bar{d}(e) \in [2^{r-1}, 2^r)$ ,

and  $\mathcal{M}$  covers a maximum number of edges in  $G'$  under these conditions. We will show that the matchings in  $\mathcal{M}$  together cover all edges of  $G'$ . Suppose not, and take  $e$  to be an edge in  $G'$  such that  $e \notin \bigcup \mathcal{M}$ . Let  $r$  be the integer for which  $\bar{d}(e) \in [2^{r-1}, 2^r)$ .

There are at most  $n$  matchings in  $\mathcal{M}$  containing an edge of  $P(e)$ , and at most  $2n$  matchings containing an edge that intersects  $e$ . Notice that  $G$  has at most  $2^r n$  edges touching vertices of degree at most  $2^r$ . In particular,  $G'$  has at most  $2^r n$  edges  $f$  with  $\bar{d}(f) \in [2^{r-1}, 2^r)$ , so the number of matchings  $M \in \mathcal{M}$  that contain at least  $\frac{1}{200}2^r - 1$  such edges is at most

$$\frac{2^r n}{\frac{1}{200}2^r - 1} = 200n + \frac{40000n}{2^r - 200} \leq 201n,$$

using that  $r \geq \log(\bar{d}(e)) - 1 \geq 51 \log \log n - 1$  and that  $n$  is large. Hence, there exists  $M \in \mathcal{M}$  that has no edges in  $P(e)$  or edges that intersect  $e$ , and has at most  $\frac{1}{200}2^r - 1$  edges  $f$  with  $\bar{d}(f) \in [2^{r-1}, 2^r)$ . Replacing  $M$  with  $M \cup \{e\}$ , we reach a contradiction to the maximality of  $\mathcal{M}$ , showing that  $\mathcal{M}$  covers all edges in  $G'$ .

It remains to show that there are at most  $\frac{d}{50}$  edges  $e$  in  $M$  with  $\bar{d}(e) \leq d$ , for every  $d \geq 1$  and  $M \in \mathcal{M}$ . To see this, fix  $M$  and  $d$ , and let  $r$  be the integer such that  $d \in [2^{r-1}, 2^r)$ . Then, there are at most  $\frac{1}{200}(1 + 2 + \dots + 2^r) \leq \frac{1}{100}2^r \leq \frac{d}{50}$  edges  $e$  in  $M$  with  $\bar{d}(e) \leq 2^r$ , as required.  $\square$

Let  $\mathcal{M}$  be as in the above claim. For an edge  $e$  in  $G'$ , denote by  $M(e)$  the unique matching in  $\mathcal{M}$  containing  $e$ . Fix  $M \in \mathcal{M}$ . We will show that there is a path  $P_M$  that contains  $E(M)$  and its edges which are not in  $M$  have at least one end in  $V = V_1 \cup V_2$ . Write  $M = \{e_1, \dots, e_{|M|}\}$ , where  $\bar{d}(e_1) \leq \dots \leq \bar{d}(e_{|M|})$ . We will define a sequence  $F_0, \dots, F_{|M|}$  of path forests, as follows. Take  $F_0 := M$ . Having defined  $F_{i-1}$ , if a vertex  $v$  in  $e_i$  is a leaf in  $F_{i-1}$  and has a common neighbour  $u$  in  $V_1$  with a leaf  $v'$  of another component in  $F_{i-1}$ , add the edges  $uv, uv'$  to  $F_{i-1}$  (thereby obtaining a new path forest which has one component fewer). Repeat this with the other vertex in  $e_i$ , and denote the resulting graph by  $F_i$ . Write  $F := F_{|M|}$ , denote the components of  $F$  by  $P_1, \dots, P_r$ , and let  $x_i, y_i$  be the ends of  $P_i$ .

**Claim 6.3.** *Every  $X \subseteq \{x_1, \dots, x_r\}$  satisfies  $|N(X)| \geq 2|X|(\log n)^{50}$ .*

*Proof.* Let  $i_j$  be such that  $x_j \in e_{i_j}$ , for  $j \in [r]$ , and suppose that  $i_1 < \dots < i_r$ . We claim that  $|N(x_j) \cap (N(x_{j+1}) \cup \dots \cup N(x_r)) \cap V_1| \leq 2i_j$ , for  $j \in [r]$ . Indeed, the forest  $F_{i_j}$  uses at most  $2i_j$  vertices from  $V_1$ . Because  $x_j$  remains a leaf in  $F_{i_j}$ , the vertices  $x_j$  and  $x_\ell$ , for  $\ell \in [j+1, r]$ , have no common neighbours in  $V_1 - V(F_{i_j})$ , showing that  $x_j$  has at most  $2i_j$  neighbours in  $V_1$  that are also neighbours of a vertex in  $\{x_{j+1}, \dots, x_r\}$ , as claimed.

We additionally claim that, if  $v$  is a leaf of  $F$  and a vertex in  $e_i$ , then  $|N(v) \cap V_1| \geq 2i + 2(\log n)^{50}$ . Recall that  $M$  has at most  $\frac{d}{50}$  edges  $e$  with  $\bar{d}(e) \leq d$ , for every  $d$ . Since there are at least  $i$  edges  $e$  in  $M$  with  $\bar{d}(e) \leq \bar{d}(e_i)$ , it follows that  $\bar{d}(e_i) \geq 50i$ . Recalling Assumption (a) and that  $\bar{d}(e_i) \geq (\log n)^{51}$ , we find that  $|N(v) \cap V_1| \geq \frac{1}{10}|N(v)| \geq \frac{1}{10}\bar{d}(e_i) \geq \max\{5i, (\log n)^{51}\} \geq 2i + 2(\log n)^{50}$ .

Writing  $X = \{x_{i_1}, \dots, x_{i_\rho}\}$ , where  $i_1 < \dots < i_\rho$ , we have

$$\begin{aligned} |N(X)| &\geq |N(X) \cap V_1| \geq \sum_{j \in [\rho]} (|N(x_{i_j}) \cap V_1| - |N(x_{i_j}) \cap (N(x_{i_{j+1}}) \cup \dots \cup N(x_{i_\rho})) \cap V_1|) \\ &\geq \sum_{j \in [\rho]} (2i_j + 2(\log n)^{50} - 2i_j) = \rho \cdot 2(\log n)^{50} = 2|X|(\log n)^{50}, \end{aligned}$$

proving the claim.  $\square$

An analogous argument shows that  $|N(Y)| \geq 2|Y|(\log n)^{50}$  for every  $Y \subseteq \{y_1, \dots, y_r\}$ . Thus, if  $X \subseteq \{x_1, \dots, x_r, y_1, \dots, y_r\}$ , without loss of generality the set  $X' := X \cap \{x_1, \dots, x_r\}$  has size at least  $\frac{1}{2}|X|$ , and then by the claim  $|N(X)| \geq |N(X')| \geq 2|X'|(\log n)^{50} \geq |X|(\log n)^{50}$ . Assumption (b) thus shows that there exist paths  $Q_1, \dots, Q_{r-1}$ , where  $Q_i$  has ends  $y_i, x_{i+1}$ , and the interiors

of the  $Q_i$ 's are pairwise vertex-disjoint and contained in  $V_2$ . Because  $F$  was chosen so as not to contain vertices from  $V_2$ , the union  $Q_1 \cup \dots \cup Q_{r-1} \cup F$  is a path  $P_M$  that contains  $M$  and no other edges from  $G'$ .

Take  $\mathcal{Q} := \{P_M : M \in \mathcal{M}\}$ . We claim that  $\mathcal{P} \cup \mathcal{Q}$  separate the edges of  $G'$ . Indeed, given edges  $e, f$  in  $G'$ , it is easy to see that one of  $P(e)$  and  $P_{M(e)}$  contains  $e$  but not  $f$ .

To summarise  $\mathcal{P} \cup \mathcal{Q}$  is a collection of at most  $211n$  paths that separates the edges  $e$  of  $G'$ , as required.  $\square$

The following easy corollary of Lemma 6.1 shows that weak robust expanders have linear separating path systems.

**Corollary 6.4.** *Let  $2^{-9} \leq \varepsilon < 1$  and  $s \geq (\log n)^{51}$ . Suppose that  $G$  is a weak  $(\varepsilon, s)$ -expander. Then there is a collection  $\mathcal{P}$  of at most  $636n$  paths that separates the edges in  $G$ .*

*Proof.* Note that we may assume that  $n$  is large (otherwise take  $G_1 = G$  and  $\mathcal{P} = \emptyset$ ). Let  $\{V_1, V_2, V_3\}$  be a random partition of  $V(G)$ , chosen uniformly at random. By Lemma 6.1, with high probability, there is a collection of paths  $\mathcal{P}_i$  of size at most  $211n$  that separates  $G[V_{i+1} \cup V_{i+2}]$  (indices taken modulo 3); fix an outcome of  $V_1, V_2, V_3$  that has these properties. Let  $\mathcal{Q}_i$  be a collection of at most  $n$  paths that decomposes  $G[V_{i+1} \cup V_{i+2}]$ , for  $i \in [3]$ .

We claim that  $\mathcal{P} := \mathcal{P}_1 \cup \mathcal{P}_2 \cup \mathcal{P}_3 \cup \mathcal{Q}_1 \cup \mathcal{Q}_2 \cup \mathcal{Q}_3$  separates the edges of  $G$ . Indeed, consider two edges  $e, f$  in  $G$ , and pick  $i \in [3]$  so that  $e \in G[V_{i+1} \cup V_{i+2}]$  (notice that such  $i$  exists but need not be unique). If  $f \in G[V_{i+1} \cup V_{i+2}]$  then there is a path in  $\mathcal{P}_i$  that contains  $e$  but not  $f$ , and otherwise there is a path in  $\mathcal{Q}_i$  that contains  $e$  but not  $f$ .

To complete the proof it suffices to observe that  $|\mathcal{P}| \leq 3 \cdot 211n + 3n \leq 636n$ .  $\square$

Finally, we prove the first key lemma (Lemma 3.1), restated here.

**Lemma 3.1.** *Suppose that  $G$  is a graph on  $n$  vertices. Then there is a subgraph  $G_1 \subseteq G$ , with  $e(G_1) \leq n(\log n)^{55}$ , and a collection  $\mathcal{P}$  of at most  $1300n$  paths in  $G$  that separates the edges of  $G - G_1$ .*

*Proof of Lemma 3.1.* We assume that  $n$  is large (otherwise, we can take  $G_1 = G$  and  $\mathcal{P} = \emptyset$ ). Take  $s = (\log n)^{51}$ ,  $\varepsilon = 2^{-8}$  and  $t = \frac{2n}{3}$ . Apply Lemma 4.1 to obtain a collection  $\mathcal{H}$  of pairwise edge-disjoint  $(\varepsilon, s)$ -expanders that covers all but at most  $48sn(\log n)^3$  edges of  $G$  and satisfies  $\sum_{H \in \mathcal{H}} |H| \leq 2n$ . Let  $G_1$  be the subgraph of  $G$  spanned by edges not covered by  $\mathcal{H}$ . Then  $e(G_1) \leq n(\log n)^{55}$  and the graphs in  $\mathcal{H}$  are weak  $(\frac{\varepsilon}{2}, s)$ -expanders.

Apply Corollary 6.4 (with  $\frac{\varepsilon}{2}$  and  $s$ ) to each  $H \in \mathcal{H}$  to obtain a collection  $\mathcal{P}_H$  of at most  $636|H|$  paths that separates  $H$ . Let  $\mathcal{Q}_H$  be a collection of at most  $|H|$  paths that decomposes  $H$  (which exists by Theorem 1.2). Take  $\mathcal{P} := \bigcup_{H \in \mathcal{H}} (\mathcal{P}_H \cup \mathcal{Q}_H)$ .

We claim that  $\mathcal{P}$  separates  $G - G_1$ . To see this, consider two edges  $e, f$  in  $G - G_1$  and let  $H \in \mathcal{H}$  be a graph containing  $e$ . If  $f \in H$  then  $\mathcal{P}_H$  has a path containing  $e$  but not  $f$ , and otherwise  $\mathcal{Q}_H$  has such a path.

To finish, note the following bound.

$$|\mathcal{P}| \leq \sum_{H \in \mathcal{H}} (|\mathcal{P}_H| + |\mathcal{Q}_H|) \leq \sum_{H \in \mathcal{H}} 637|H| \leq 1300n.$$

So the graph  $G_1$  and the collection of paths  $\mathcal{P}$  satisfy the requirements of the lemma.  $\square$

## 7 Expanding with forbidden sets

As a preparation for the proof of the second key lemma, in this section we prove a variant of Lemma 3.2 from [32], which allows one to expand a set of vertices  $A$  while avoiding another set  $X$ , provided that  $X$  has a small intersection with each ball around  $A$  (in [32] the authors list two other properties of  $X$  that guarantee the expansion of  $A$  while avoiding  $X$ , but we shall not use them). In both cases, the conclusion is that the radius- $i$  ball around  $A$  in  $G - X$  is large. In [32] this was only shown for specific, fairly large values of  $i$ , whereas here we give a general lower bound that can be applied for any  $i$ , as per our needs later on.

Following [32], say that a vertex set  $A$  has  $k$ -limited contact with a vertex set  $X$  in a graph  $G$  if  $|N_G(B_{H-X}^{i-1}(A)) \cap X| \leq ki$  for every integer  $i \geq 1$ . The following lemma is similar to Lemma 3.2 in [32], but tailored to our needs.

**Lemma 7.1.** *For every  $\varepsilon > 0$  there exists  $k_0 = k_0(\varepsilon)$  such that the following holds for  $k \geq k_0(\varepsilon)$ . Suppose that  $G$  is an  $n$ -vertex  $\varepsilon$ -expander, and let  $A$  and  $X$  be disjoint subsets of  $V(G)$  where  $|A| \geq k^3$  and  $A$  has  $k$ -limited contact with  $X$ . Then  $|B_{G-X}^i(A)| > \min\{2^{i^{1/4}}, n/2\}$ , for every integer  $i \geq 1$ .*

*Proof.* Write  $H := G - X$  and  $B_i := B_H^i(A)$  for  $i \geq 0$ . Our task is thus to show  $|B_i| > \min\{2^{i^{1/4}}, n/2\}$  for  $i \geq 1$ .

**Claim 7.2.** *For every integer  $i \geq 0$ , either  $|B_i| > n/2$  or  $|N_H(B_i)| \geq \frac{\varepsilon|B_i|}{2(\log|B_i|)^2}$ .*

*Proof.* We prove the claim by induction on  $i$ . Let  $i \geq 0$  be such that  $|B_i| \leq n/2$  and suppose that the statement holds for  $j$  with  $0 \leq j < i$ . (We will prove the induction base and step simultaneously.)

For  $j \in [i]$ , since  $B_j$  is the disjoint union of  $B_{j-1}$  and  $N_H(B_{j-1})$ , we have

$$|B_j| \geq \left(1 + \frac{\varepsilon}{2(\log|B_{j-1}|)^2}\right) |B_{j-1}| \geq \left(1 + \frac{\varepsilon}{2(\log|B_i|)^2}\right) |B_{j-1}|,$$

using that  $B_{j-1} \subseteq B_i$ . Iterating this, we get

$$|B_i| \geq \left(1 + \frac{\varepsilon}{2(\log|B_i|)^2}\right)^i |A|. \tag{4}$$

(Notice that this inequality holds for  $i = 0$ , as  $B_0 = A$ , since  $A$  is assumed to be disjoint of  $X$ , allowing us to treat the induction base and step simultaneously.)

We conclude that  $|B_i| \geq \max\{k^3, i^3\}$ . Indeed, if  $i \leq k$  this holds because  $|B_i| \geq |A| \geq k^3$ . So suppose that  $i > k$  and  $|B_i| < i^3$ . Then (4) implies

$$|B_i| \geq \left(1 + \frac{\varepsilon}{2(\log i^3)^2}\right)^i |A| \geq \left(\exp\left(\frac{\varepsilon}{4(\log i^3)^2}\right)\right)^i = \exp\left(\frac{\varepsilon i}{36(\log i)^2}\right) \geq i^3,$$

using that  $i$  is large and thus  $\frac{\varepsilon}{2(\log i^3)^2} \leq 1$  (which follows from  $k$  being large and the assumption  $i > k$ ). This is a contradiction, proving that  $|B_i| \geq \max\{k^3, i^3\}$ . Write  $m := \max\{k, i\}$ . Then

$$\frac{\varepsilon|B_i|}{(\log |B_i|)^2} \geq \frac{\varepsilon m^3}{(\log m)^2} \geq 4m^2 \geq 2k(i+1), \quad (5)$$

using that  $m$  is large.

Recall that  $|X \cap N_G(B_i)| \leq k(i+1)$  (this is what it means for  $A$  to have  $k$ -limited contact with  $X$ ) and notice that  $N_G(B_i) - X \subseteq N_H(B_i)$ . Thus,

$$|N_H(B_i)| \geq |N_G(B_i)| - k(i+1) \geq \frac{\varepsilon|B_i|}{(\log |B_i|)^2} - k(i+1) \geq \frac{\varepsilon|B_i|}{2(\log |B_i|)^2},$$

where the second inequality follows from expansion (using  $|B_i| \leq n/2$ ), and the third inequality follows from (5). This proves the induction hypothesis for  $i$ .  $\square$

Suppose that  $i \geq 1$  satisfies  $|B_i| \leq \min\{2^{i^{1/4}}, n/2\}$ . Since  $|B_i| \geq |A| \geq k^3$ , it follows that  $2^{i^{1/4}} \geq k^3$ . In particular,  $i$  is large. By Claim 7.2, every  $j \in [i]$  satisfies

$$|B_j| = |B_{j-1}| + |N_H(B_{j-1})| \geq \left(1 + \frac{\varepsilon}{2(\log |B_{j-1}|)^2}\right) |B_{j-1}| \geq \left(1 + \frac{\varepsilon}{2\sqrt{i}}\right) |B_{j-1}|,$$

using that  $B_{j-1} \subseteq B_i$  and the assumption  $|B_i| \leq \min\{2^{i^{1/4}}, n/2\}$ . Iterating this, we get

$$|B_i| \geq \left(1 + \frac{\varepsilon}{2\sqrt{i}}\right)^i \geq \exp\left(\frac{\varepsilon i}{4\sqrt{i}}\right) > 2^{i^{1/4}},$$

using that  $i$  is large, and yielding a contradiction.  $\square$

## 8 Proof of Lemma 3.2: separating sparse expanders

In this section we prove the second key lemma, Lemma 3.2. The proof will be split into two parts: Lemma 8.1, proved in Section 8.1, and Lemma 8.4, proved in Section 8.2. In both parts, we are presented with an expander  $G$  on  $n$  vertices and a number  $d$ , and we wish to separate edges of  $G$  by  $O(n + e(G)/d)$  paths. In the first part, we separate edges that touch large degree vertices, and in



the second part we consider expanders with small maximum degree and many vertices. Lemma 3.2 itself is proved in Section 8.3.

## 8.1 Dealing with large degree vertices

The following lemma shows that given an expander  $G$  on  $n$  vertices and a number  $d$ , there is a collection of  $O(n + e(G)/d)$  paths that separate all edges of  $G$  except for a remainder with maximum degree at most  $d^7$ . Take  $L_1$  to be the set of vertices of degree at least  $d^7$  and  $L_2$  to be the set of vertices outside of  $L_1$  that send at least four edges into  $L_1$ . We will let the remainder graph be  $G - L_1$ . Noting that we can separate the edges of  $G[L_1, V(G) - (L_1 \cup L_2)]$  by  $O(n)$  single edges, it suffices to separate the edges in  $G' := G[L_1] \cup G[L_1, L_2]$ .

To do so, as usual we first decompose  $G'$  into  $O(n + e(G)/d)$  paths  $\mathcal{P}$  of length at most  $d$ . Then, for each  $u \in L_2$  we pick a family  $\mathcal{A}_u$  of  $d(u, L_1)$  paths of length 2 with the central vertex being  $u$  and the other two vertices in  $L_1$ , such that each edge between  $u$  and  $L_1$  is covered twice, and each such 2-path intersects each path in  $\mathcal{P}$  in at most one edge. Write  $\mathcal{A} = (\bigcup_{u \in L_2} \mathcal{A}_u) \cup E(G[L_1])$ , so that  $\mathcal{A}$  is a collection of paths of length at most 2. Then, we find a collection  $\mathcal{M}$  of  $O(n + e(G)/d)$  graphs  $M$ , which are disjoint unions of at most  $d$  paths of length at most 2 from  $\mathcal{A}$ , such that  $\mathcal{P} \cup \mathcal{M}$  separates  $G'$ . Now it suffices to show that each  $M \in \mathcal{M}$  can be extended to a path that avoids using other edges in  $E := \bigcup_{e \in E(M)} P(e)$  (where  $P(e)$  is the unique path in  $\mathcal{P}$  that contains  $e$ ). Take  $F$  to be a linear forest that extends  $M$ , avoids other edges in  $E$ , and minimises the number of components and then edges under these assumptions. We show that  $F$  must be a path, using Lemma 7.1 and that all leaves in  $F$  have degree at least  $d^7$  in  $G$ , and are thus easy to expand.

**Lemma 8.1.** *Let  $\varepsilon > 0$  and let  $d$  be sufficiently large. Suppose that  $G$  is an  $n$ -vertex  $\varepsilon$ -expander. Then there is a subgraph  $G_1 \subseteq G$  with maximum degree at most  $d^7$  and a collection  $\mathcal{P}$  of at most  $17 \max\{n, e(G)/d\}$  paths in  $G$  that separates the edges of  $G - G_1$ .*

*Proof.* Write  $m := e(G)$  and  $k := \max(m/d, n)$ .

Let  $L_1$  be the set of vertices in  $G$  with degree at least  $d^7$  and let  $L_2$  be the set of vertices outside of  $L_1$  that have at least four neighbours in  $L_1$ . Define  $G_1 = G - L_1$ . We will find a collection  $\mathcal{P}$  of at most  $17 \max\{n, e(G)/d\}$  paths in  $G$  that separate the edges of  $G - G_1$ .

Write  $H_1 := G[L_1] \cup G[L_1, L_2]$  and  $H_2 := G[L_1, V(G) - (L_1 \cup L_2)]$ . Let  $\mathcal{P}_1$  be a collection of at most  $n + m/d \leq 2k$  paths of length at most  $d$  that decomposes  $H_1$ ; such a collection exists by Corollary 1.3. For an edge  $e$  in  $H_1$ , let  $P(e)$  be the unique path in  $\mathcal{P}_1$  that contains  $e$ .

For each  $u \in L_2$ , let  $v_1, \dots, v_{d_u}$ , where  $d_u := d(u, L_1)$ , be an ordering of the neighbours of  $u$  in  $L_1$  such that  $uv_i$  and  $uv_{i+1}$  are not in the same path in  $\mathcal{P}_1$ , for  $i \in [d_u]$  (indices taken modulo  $d_u$ ). Notice that such an ordering exists because  $d_u \geq 4$  and the paths in  $\mathcal{P}_1$  are pairwise edge-disjoint. Define  $\mathcal{A}_u := \{v_i v_{i+1} : i \in [d_u]\}$  (with indices taken modulo  $d_u$ ) for  $u \in L_2$  and  $\mathcal{A} := (\bigcup_{u \in L_2} \mathcal{A}_u) \cup E(G[L_1])$ , so  $\mathcal{A}$  is a collection of edges and 2-paths.

**Claim 8.2.** *There is a collection  $\mathcal{M}$  of at most  $12k$  graphs which are disjoint unions of at most  $d$  paths of lengths at most 2 from  $\mathcal{A}$ , such that each path in  $\mathcal{A}$  is a component in exactly one graph  $M \in \mathcal{M}$ ; and every path in  $\mathcal{P}_1$  shares at most one edge with any  $M \in \mathcal{M}$ .*

*Proof.* Let  $\mathcal{M}$  be a collection of  $12k$  graphs that are disjoint unions of at most  $d$  paths from  $\mathcal{A}$ , such that: each  $A \in \mathcal{A}$  is a component in at most one  $M \in \mathcal{M}$ ; every path in  $\mathcal{P}_1$  shares at most one edge with any  $M \in \mathcal{M}$ ; and a maximum number of paths in  $\mathcal{A}$  are components in some graph in  $\mathcal{M}$ . We claim that every edge and 2-path in  $\mathcal{A}$  is a component in some  $M \in \mathcal{M}$ . Indeed, suppose not, and let  $A \in \mathcal{A}$  be an edge or 2-path which is not a component in any  $M \in \mathcal{M}$ .

Notice that  $|\mathcal{A}| \leq m$ , so there are at most  $m/d$  graphs  $M \in \mathcal{M}$  with  $d$  components.

We claim that every vertex  $v \in L_1 \cup L_2$  is in at most  $2n$  paths in  $\mathcal{A}$ . Indeed, if  $v \in L_1$  then there are at most  $|L_1|$  edges in  $\mathcal{A}$  that contain  $v$  and at most  $2|L_2|$  paths of length 2 in  $\mathcal{A}$  that contain  $v$ , because each edge in  $G[L_1, L_2]$  is used by exactly two such 2-paths. If  $v \in L_2$  then  $v$  is in at most  $2|L_1|$  paths of length 2 in  $\mathcal{A}$  (and in no edges in  $\mathcal{A}$ ). Either way, there are at most  $2|L_1| + 2|L_2| \leq 2n$  paths in  $\mathcal{A}$  that contain  $v$ . As  $|V(A)| \leq 3$ , it follows that  $A$  shares a vertex with at most  $6n$  paths in  $\mathcal{A}$ . In particular, there are most  $6n$  graphs  $M \in \mathcal{M}$  such that  $A$  and  $M$  share a vertex.

Notice also that each path in  $\mathcal{P}_1$  has at most  $n$  edges and thus intersects the edges of at most  $2n$  paths in  $\mathcal{A}$ . Thus, there are at most  $4n$  graphs  $M \in \mathcal{M}$  for which there is a path  $P \in \mathcal{P}_1$  that shares an edge with both  $A$  and  $M$ .

In summary, since  $m/d + 6n + 4n \leq 11k < 12k$ , there is  $M \in \mathcal{M}$  such that  $M$  has fewer than  $d$  components,  $A$  and  $M$  do not share a vertex, and there is no path  $P \in \mathcal{P}_1$  that shares an edge with  $A$  and  $M$ . Thus we can replace  $M$  by  $M \cup A$ , contradicting the maximality of  $\mathcal{M}$ .  $\square$

Let  $\mathcal{M}$  be a collection as guaranteed by the above claim. For a graph  $M \in \mathcal{M}$ , write  $E_M = \bigcup_{e \in E(M)} E(P(e))$  and  $X_M := V(E_M) - V(M)$  (so  $E_M$  is the set of edges that appear on paths from  $\mathcal{P}_1$  that share edges with  $M$ , and  $X_M$  is the set of vertices on such paths, minus the vertices of  $M$ ). Note that  $|X_M| \leq 2d^2$  (there are at most  $2d$  edges in  $M$ , each of which causes the insertion of at most  $d$  vertices into  $X_M$ ).

**Claim 8.3.** *There is a path in  $G$  that contains  $M$  and avoids the edges in  $E_M - E(M)$ , for  $M \in \mathcal{M}$ .*

*Proof.* Fix  $M \in \mathcal{M}$ . For convenience, write  $E := E_M$  and  $X := X_M$ . Let  $\mathcal{Q}$  be a collection of pairwise vertex-disjoint paths of length at least 2 in  $G - X$  with the following properties.

- (a) The union of  $M$  and the paths in  $\mathcal{Q}$  is a path forest whose components are paths that alternate between components of  $M$  and paths in  $\mathcal{Q}$ , starting and ending with a path in  $M$ . In particular,  $|\mathcal{Q}| \leq d$ .
- (b) Subject to (a),  $|\mathcal{Q}|$  is maximal.
- (c) Subject to (a) and (b),  $\sum_{Q \in \mathcal{Q}} \ell(Q)$  is minimal.

Denote the path forest, which is the union of  $M$  and paths in  $\mathcal{Q}$ , by  $F$ .

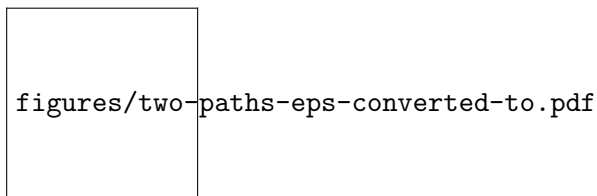
Let  $R$  be a component in  $F$ , let  $u$  be an end of  $R$ , and let  $X_u := (X \cup V(F)) - \{u\}$ . We claim that the following holds for  $r \geq 1$ .

$$|N_G(B_{G-X_u}^{r-1}(u) \cap V(F))| \leq (r+4)d. \quad (6)$$

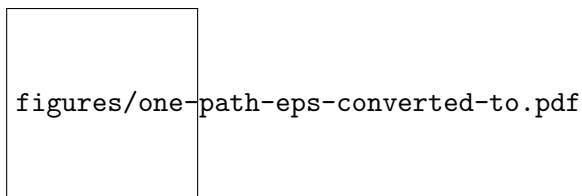
Suppose this is not the case. Because  $|V(M)| \leq 3d$  and  $|\mathcal{Q}| \leq d$ , there is a path  $Q \in \mathcal{Q}$  such that

$$|N_G(B_{G-X_u}^{r-1}(u) \cap V(Q^\circ))| > \frac{(r+4)d - 3d}{d} \geq r+1, \quad (7)$$

where  $Q^\circ$  is the interior of  $Q$ . Let  $R'$  be the component of  $F$  that contains  $Q$ . Let  $v$  be one of the ends of  $Q$ ; if  $R' = R$  take  $v$  to be the further end of  $Q$  from  $u$  in  $R$  (see Figures 1 and 2). Let  $w$  be the first vertex in  $Q$  (starting from  $v$ ) in  $N_G(B_{G-X_u}^{r-1}(u) \cap V(Q^\circ))$ , and let  $Q_1$  be the subpath of  $Q$  that starts at  $v$  and ends at  $w$ . Take  $Q_2$  to be a shortest path in  $G$  from  $u$  to  $w$  whose interior vertices are not in  $V(F) \cup X$ . By choice of  $w$  we have  $\ell(Q_2) \leq r$ . Finally, let  $Q'$  be the concatenation of  $Q_1$  and  $Q_2$  and define  $\mathcal{Q}' := (\mathcal{Q} - \{Q\}) \cup \{Q'\}$ .



**Figure 1:** The case  $R' \neq R$ . The crosses signify the vertices in  $N_G(B_{G-X_u}^{r-1}(u) \cap V(Q^\circ))$ , and the blue paths replace  $R$  and  $R'$  in the path forest  $M \cup \mathcal{Q}'$ .



**Figure 2:** The case  $R' = R$ . The crosses are as above, and the blue path replaces  $R$  in the path forest  $M \cup \mathcal{Q}'$ .

Notice that the paths in  $\mathcal{Q}'$  are pairwise vertex-disjoint paths of length at least 2 in  $G - X$  (in particular,  $Q'$  was indeed chosen to have length at least 2 and be in  $G - X$ ). Notice also that  $\mathcal{Q}'$  satisfies (a). Indeed, if  $R' = R$  then by choice of  $v$  we have that  $(R - Q) \cup Q'$  is a path, and if  $R' \neq R$  then  $((R \cup R') - Q) \cup Q'$  is a vertex-disjoint union of two paths. Other components of  $F$  are not affected by the switch from  $\mathcal{Q}$  to  $\mathcal{Q}'$ . Clearly,  $|\mathcal{Q}'| = |\mathcal{Q}|$ . Finally, notice that  $\ell(Q') = \ell(Q_1) + \ell(Q_2) \leq \ell(Q_1) + r < \ell(Q_1) + \ell(Q - Q_1)$ , where the last inequality holds because  $Q - Q_1$  is a path with at least  $r$  interior vertices, due to (7). This implies that  $\sum_{Q \in \mathcal{Q}'} \ell(Q) < \sum_{Q \in \mathcal{Q}} \ell(Q)$ , contradicting (c), and thus proving (6).

We now conclude that  $F$  is a path, namely it has exactly one component. Suppose not, and let  $R_1$  and  $R_2$  be two components in  $F$ . Let  $u_j$  be an end of  $R_j$ , for  $j \in [2]$ . By (a), both  $u_1$  and  $u_2$  are ends of paths in  $M$ , and thus have degree at least  $d^7$  in  $G$ . Let  $A_j := N_G(u_j) - (X \cup V(F))$ . Then

$$|A_j| \geq |N_G(u_j)| - |X| - |N_G(u_j) \cap V(F)| \geq d^7 - 2d^2 - 5d \geq (9d^2)^3,$$

using (6) (with  $r = 1$ ),  $|X| \leq 2d^2$  and that  $d$  is large. Also,

$$\left| N_G \left( B_{G-V(F)-X}^{r-1}(A_j) \right) \cap (V(F) \cup X) \right| \leq (r+5)d + 2d^2 + 1 \leq 9rd^2,$$

using (6) (with  $r+1$ ), where the  $2d^2+1$  term accounts for  $X \cup \{u\}$ . In particular,  $A_j$  has  $9d^2$ -limited contact with  $V(F)$  in  $G'$  (the notion of limited contact was defined at the beginning of Section 7). Take  $r = (\log n)^4$ . Then, by Lemma 7.1,  $|B_{G-V(F)-X}^r(A_j)| > n/2$  for  $j \in [2]$ . Hence, there is a path  $Q$  in  $G - V(F) - X$  joining  $A_1$  and  $A_2$ . Let  $Q'$  be the extension of  $Q$  by edges from  $u_j$  to  $A_j$ , for  $j \in [2]$ . Take  $\mathcal{Q}' := \mathcal{Q} \cup \{Q'\}$ . Then  $\mathcal{Q}'$  is a collection of pairwise vertex-disjoint paths of length at least 2 in  $G - X$  that satisfy (a) and  $|\mathcal{Q}'| > |\mathcal{Q}|$ , contradicting (b). This proves that  $F$  consists of a single component.

We claim that  $F$  satisfies the requirements of the claim. Clearly,  $F$  contains  $M$ . Moreover, any edge in  $F$  which is not in  $E(M)$  is incident with at least one vertex which is not in  $X_M \cup V(M) = V(E_M)$  and thus it is not in  $E_M$ .  $\square$

For  $M \in \mathcal{M}$ , denote by  $P_M$  the paths whose existence is guaranteed by Claim 8.3. To prove the theorem, take  $\mathcal{P}$  to be the union of  $\mathcal{P}_1, \mathcal{P}_2$  and the edges of  $H_2 = G[L_1, V(G) - (L_1 \cup L_2)]$  (each edge in  $H_2$  is considered as a separate path in  $\mathcal{P}$ ). Notice that  $|\mathcal{P}| = |\mathcal{P}_1| + |\mathcal{P}_2| + e(H_2) \leq 2k + 12k + 3n \leq 17k = 17 \max\{n, e(G)/d\}$ .

We claim that  $\mathcal{P}$  separates all edges in  $G$  that touch  $L_1$ . To see this, consider two distinct edges,  $e$  and  $f$ , that touch  $L_1$ . If  $e \in E(H_2)$  then  $e$  itself is a path in  $\mathcal{P}$  (which obviously does not contain  $f$ ). Otherwise, either the path  $P(e)$  contains  $e$  but not  $f$ , or there exists  $M \in \mathcal{M}$  that contains  $e$  but not  $f$  and then  $P_M$  contains  $e$  but not  $f$ .

Recall that  $G_1 = G - L_1$ . Then  $G_1$  has maximum degree at most  $d^7$ , and the previous paragraph shows that  $\mathcal{P}$  separates the edges of  $G - G_1$ . This proves Lemma 8.1.  $\square$

## 8.2 Sparse expanders with small maximum degree

The next lemma shows that given a ‘somewhat robust’ expander  $G$  on  $n$  vertices with maximum degree at most  $d^7$ , where  $n$  is large in terms of  $d$ , the edges of  $G$  can be separated by  $O(n + e(G)/d)$  paths. The proof is very similar in structure to the proof of Lemma 8.1. We first decompose  $G$  into paths  $\mathcal{P}$  and find matchings  $\mathcal{M}$ , such that  $|\mathcal{P}|, |\mathcal{M}| = O(n + e(G)/d)$ , the union  $\mathcal{P} \cup \mathcal{M}$  separates  $G$ , the paths in  $\mathcal{P}$  and matchings in  $\mathcal{M}$  have size at most  $d$ , and for any two edges  $e, f$  in the same matching, the paths  $P(e)$  and  $P(f)$  are far apart in  $G$  (we need  $n$  to be large with respect to  $d$  to be able to satisfy the last property).

Now it remains to show that for each  $M \in \mathcal{M}$  there is a path that contains  $M$  and avoids  $E := \bigcup_{e \in M} P(e) - E(M)$ . As before, we take  $F$  to be a path forest that extends  $M$ , avoids edges in  $E$ , and minimises the number of components and then edges. We again use Lemma 7.1 to deduce that  $F$  is a path. This time we do so by first showing that small-radius balls around any leaf of  $F$  expand well, using that such balls have few edges of  $E$  (by assumption on  $M$ ) and the moderate robustness property. We can then expand such balls further similarly to the proof of Lemma 8.1, using that we now start from sizeable sets.

**Lemma 8.4.** *Let  $\varepsilon > 0$ , let  $d$  be sufficiently large, let  $n \geq 2^{(\log d)^7}$ , and let  $s \geq (k_0(\varepsilon)_{7.1} + 3)^3$ . Suppose that  $G$  is an  $n$ -vertex  $(\varepsilon, s, d)$ -expander with maximum degree at most  $d^7$ . Then there is a collection  $\mathcal{P}$  of at most  $5 \max\{n, e(G)/d\}$  paths in  $G$  that separates the edges of  $G$ .*

*Proof.* As before, write  $m := e(G)$  and  $k := \max\{n, m/d\}$ . Let  $\mathcal{P}_1$  be a collection of at most  $n + m/d \leq 2k$  paths of length at most  $d$  that decompose the edges in  $G$ ; it exists by Corollary 1.3. For an edge  $e$  denote by  $P(e)$  the unique path in  $\mathcal{P}_1$  that contains  $e$ . Write  $r_0 = (\log d)^5$ .

**Claim 8.5.** *There is a decomposition  $\mathcal{M}$  of  $G$  of size  $3k$  such that every  $M \in \mathcal{M}$  satisfies:  $M$  is a matching of size at most  $d$ ; and the paths  $P(e)$  and  $P(f)$  are a distance at least  $2r_0$  apart for every distinct  $e, f \in M$ .*

*Proof.* Let  $\mathcal{M}$  be a collection of  $3k$  pairwise edge-disjoint matchings that satisfy the conditions of the claim, and together cover a maximum number of edges in  $G$ . We claim that  $\mathcal{M}$  covers all edges of  $G$ . Suppose not, and let  $e \in E(G) - \bigcup \mathcal{M}$ .

As  $e(G) = m$ , there are at most  $m/d$  matchings  $M \in \mathcal{M}$  of size  $d$ .

We have the following upper bound on the ball of radius  $2r_0$  around  $P(e)$ .

$$|B_G^{2r_0}(P(e))| \leq |P(e)| \cdot (1 + d^7 + \dots + (d^7)^{2r_0}) \leq d^{14r_0+2},$$

using the maximum degree assumption, that  $|P(e)| \leq d + 1$ , and that  $d$  is large. Notice that every  $P \in \mathcal{P}_1$  intersects the edges of at most  $d$  matchings  $M \in \mathcal{M}$ , implying that each edge  $f$  is in  $\bigcup_{h \in M} P(h)$  for at most  $d$  matchings  $M \in \mathcal{M}$ . It follows from the maximum degree assumption that every vertex  $u$  is in  $\bigcup_{h \in M} P(h)$  for at most  $d^8$  matchings  $M \in \mathcal{M}$ . Hence, the sets  $B_G^{2r_0}(P(e))$  and  $\bigcup_{h \in M} P(h)$  share a vertex, for at most  $d^8 \cdot d^{14r_0+2} \leq d^{14r_0+10} \leq 2^{(\log d)^7} \leq n$  matchings  $M$ , where for the last inequality we used the assumption on  $n$  and  $d$ . In other words,  $P(e)$  is at distance at least  $2r_0$  from  $\bigcup_{h \in M} P(h)$ , for all but at most  $n$  matchings  $M \in \mathcal{M}$ .

To summarise, since  $m/d + n < 3k$ , there is a matching  $M \in \mathcal{M}$  such that  $|M| < d$  and  $P(e)$  and  $\bigcup_{h \in M} P(h)$  are a distance at least  $2r_0$  apart. We can thus replace  $M$  by  $M \cup \{e\}$ , a contradiction to the maximality of  $\mathcal{M}$ .  $\square$

Let  $\mathcal{M}$  be a decomposition of  $G$  into matchings, as guaranteed by the above claim. For an edge  $e$ , denote by  $M(e)$  the unique matching in  $\mathcal{M}$  that contains  $e$ .

**Claim 8.6.** *There is a path in  $G$  that contains all edges in  $M$  and avoids the edges in  $\bigcup_{e \in M} (P(e) - \{e\})$ , for  $M \in \mathcal{M}$ .*

*Proof.* Fix  $M \in \mathcal{M}$ , and write  $E := \bigcup_{e \in M} P(e)$  and  $G' := G - E$ . Let  $\mathcal{Q}$  be a collection of pairwise vertex-disjoint paths in  $G'$  with the following properties.

- (a) The union of  $M$  with the paths in  $\mathcal{Q}$  is a path forest whose paths alternate between  $M$  and  $\mathcal{Q}$ , starting and ending with an edge from  $M$ .
- (b) Subject to (a),  $|\mathcal{Q}|$  is maximal.
- (c) Subject to (a) and (b),  $\sum_{Q \in \mathcal{Q}} \ell(Q)$  is minimal.

Denote the path forest which is the union of  $M$  and  $\mathcal{Q}$  by  $F$ .

Let  $u$  be an end of a path in  $F$ , let  $v$  be the neighbour of  $u$  in  $F$  (so  $uv \in M$ ), let  $R$  be the component in  $F$  that contains  $u$ , and write  $X_u := V(F) - \{u\}$ . We claim the following, for  $r < r_0$ ,

$$\left| N_G \left( B_{G'-X_u}^{r-1}(u) \right) \cap V(F) \right| \leq r + 1. \quad (8)$$

Suppose this is violated for some  $r \leq r_0$ . As edges in  $M$  are a distance at least  $2r_0$  apart in  $G$ , the set  $N_G(B_{G'-X_u}^{r-1}(u)) \cap V(M)$  is contained in  $\{v\}$ , so it has size at most 1. Hence there is a path  $Q \in \mathcal{Q}$  such that either  $v$  is not an end of  $Q$  and  $N_G(B_{G'-X_u}^{r-1}(u)) \cap V(Q^\circ) \neq \emptyset$  or  $v$  is an end of  $Q$  and  $|N_G(B_{G'-X_u}^{r-1}(u)) \cap V(Q^\circ)| \geq r + 1$ .

Suppose that the former holds, and let  $Q \in \mathcal{Q}$  be a path, none of whose ends is  $v$ , that satisfies  $N_G(B_{G'-X_u}^{r-1}(u)) \cap V(Q^\circ) \neq \emptyset$ . Let  $w_0$  be a vertex in this intersection, let  $Q_0$  be a shortest path in  $G'$  from  $u$  to  $w_0$  whose interior avoids  $X_u$ ; so  $\ell(Q_0) \leq r < r_0$ . Denote the ends of  $Q$  by  $w_1$  and  $w_2$ , where if  $Q$  is in  $R$  then  $w_1$  is further away from  $u$  in  $R$ . Let  $Q_i$  be the subpath of  $Q$  from  $w_0$  to  $w_i$ , for  $i \in [2]$ . Notice that  $\ell(Q_i) \geq r_0$ , because of the assumption that edges in  $M$  are a distance at least  $2r_0$  apart in  $G$ ; indeed, otherwise the concatenation  $Q_0Q_i$  of  $Q_0$  and  $Q_i$  is a path of length shorter than  $2r_0$  between two distinct edge in  $M$ . Write  $\mathcal{Q}' := \mathcal{Q} - \{Q\} \cup \{Q_0Q_1\}$ . It is easy to check that  $\mathcal{Q}'$  is a collection of vertex-disjoint paths in  $G'$  that satisfies (a),  $|\mathcal{Q}'| = |\mathcal{Q}|$ , and  $\sum_{Q \in \mathcal{Q}'} \ell(Q) < \sum_{Q \in \mathcal{Q}} \ell(Q)$ , contradicting (c).

It remains to consider the second case, namely that  $Q \in \mathcal{Q}$  has ends  $w_1$  and  $w_2$ , where  $w_2 = v$ , and  $|N_G(B_{G'-X_u}^{r-1}(u)) \cap V(Q^\circ)| \geq r + 1$ . Let  $w_0$  be the vertex in the intersection which is furthest away from  $w_2$  in  $Q$  and let  $Q_0$  be a shortest path from  $u$  to  $w_0$  in  $G'$  whose interior avoids  $X_u$ ; so  $\ell(Q_0) \leq r$ . Let  $Q_i$  be the subpath of  $Q$  from  $w_0$  to  $w_i$ , for  $i \in [2]$ . Then  $\ell(Q_2) \geq r + 1$  by assumption on the intersection size  $|N_G(B_{G'-X_u}^{r-1}(u)) \cap V(Q^\circ)|$ . Defining  $\mathcal{Q}' := \mathcal{Q} - \{Q\} \cup \{Q_0w_0Q_1\}$ , we reach a contradiction as in the previous paragraph. Thus (8) is proved.

We now conclude the following.

$$\left| B_{G'-X_u}^{r_0}(u) \right| \geq 2^{(\log d)^2} \geq d^8. \quad (9)$$

Write  $G'' := G - P(uv)$ . We claim that  $B_{G''-X_u}^r(u) = B_{G''-X_u}^r(u)$  for every  $r \leq r_0$ . Indeed, this holds because  $G'' = G' - \bigcup_{f \in M - \{uv\}} P(f)$ , and  $P(f)$  is a distance at least  $2r_0$  away from  $uv$  in  $G$  for every  $f \in M - \{uv\}$ . In particular, (8) implies

$$\left| N_{G''} \left( B_{G''-X_u}^{r-1}(u) \right) \cap X_u \right| \leq r + 1, \quad (10)$$

for  $r < r_0$ . We claim further that  $G''$  is an  $\varepsilon$ -expander. To see this, let  $X \subseteq V(G'')$  satisfy  $1 \leq |X| \leq 2n/3$ . Let  $E''$  be the set of edges from  $P(uv)$  that touch  $X$ . Then  $|E''| \leq \min\{2|X|, d\}$  and  $N_{G''}(X) = N_{G-E''}(X)$ . Thus, as  $G$  is an  $(\varepsilon, s, d)$ -expander and  $s \geq 2$ ,  $|N_{G''}(X)| = |N_{G-E''}(X)| \geq \frac{\varepsilon|X|}{(\log|X|+1)^2}$ , as required for  $\varepsilon$ -expansion. Note also that  $G$  has minimum degree at least  $s + 1$ , by  $(\varepsilon, s, d)$ -expansion, implying that  $G''$  has minimum degree at least  $s - 1$ . Write  $A_u := N_{G''-X_u}(u) \cup \{u\}$ . Then (10) implies  $|A_u| \geq |N_{G''}(u)| - 2 \geq s - 3$  and  $|N_{G''}(B_{G''-X_u}^{r-1}(A_u)) \cap X_u| \leq r + 2 \leq 3r$  for  $r \leq r_0 - 2$ . Write  $X'_u := X_u \cap N_{G''}(B_{G''-X_u}^{r_0-3}(u))$ . Then  $A_u$  has 3-limited contact with  $X'_u$ . Using that  $s$  is large, Lemma 7.1 shows  $|B_{G''-X'_u}^{r_0-2}(A_u)| \geq 2^{(r_0-2)^{1/4}} \geq 2^{(\log d)^2} \geq d^8$ . This proves (9), as  $B_{G''-X_u}^{r_0-1}(u) = B_{G''-X'_u}^{r_0-2}(A_u)$ .

Next, we follow the proof of Claim 8.3 quite closely to show that the set  $B_{G''-X_u}^{r_0}(u)$  expands. Define  $V := \bigcup_{e \in M} V(P(e))$ ,  $B_u := B_{G''-X_u}^{r_0}(u) - V$  and  $Y := V \cup V(F)$ . Notice that  $|B_u| \geq d^8 - d(d+1) \geq d^7$  by (9) and since  $|V| \leq d(d+1)$ . We claim the following for  $r \geq 1$ .

$$\left| N_G(B_{G-Y}^{r-1}(B_u)) \cap V(F) \right| \leq d(r + r_0 + 1). \quad (11)$$

The proof of this is very similar to that of (6) in Claim 8.3. Suppose this does not hold for some  $r \geq 1$ . Because  $|V(M)| \leq 2d$  and  $|\mathcal{Q}| \leq d$ , there is a path  $Q \in \mathcal{Q}$  such that

$$\left| N_G(B_{G-Y}^{r-1}(B_u)) \cap V(Q^\circ) \right| \geq \frac{d(r + r_0 + 3) - 2d}{d} \geq r + r_0 + 1.$$

Let  $R'$  be the component in  $F$  that contains  $Q$ . Denote the ends of  $Q$  by  $w_1$  and  $w_2$ , where  $w_1$  is further from  $u$  in  $R$  if  $R' = R$ . Let  $w_0$  be the vertex in  $N_G(B_{G-Y}^{r-1}(B_u)) \cap V(Q^\circ)$  which is closest to  $w_1$  in  $Q$ , and let  $Q_0$  be a shortest path in  $G'$  from  $u$  to  $w_0$  whose interior avoids  $Y$ ; so  $\ell(Q_0) \leq r + r_0$ . Let  $Q_i$  be the subpath of  $Q$  from  $w_0$  to  $w_i$ , for  $i \in [2]$ ; then  $\ell(Q_2) \geq r + r_0 + 1$ . Taking  $\mathcal{Q}' := \mathcal{Q} - \{Q\} \cup \{Q_0 w_0 Q_1\}$  yields the desired contradiction to (c). Thus (11) is proved.

Next, we deduce the following.

$$\left| B_{G-Y}^{(\log n)^4}(B_u) \right| > n/2. \quad (12)$$

Notice that (11) implies  $|N_G(B_{G-Y}^{r-1}(B_u)) \cap Y| \leq d(r + r_0 + d + 2) \leq 3d^2 r$  for  $r \geq r_0$  (using that  $d$  is large), i.e.  $B_u$  has  $3d^2$ -limited contact with  $Y$  in  $G$ . As  $|B_u| \geq d^7 \geq (3d^2)^3$ , Lemma 7.1 proves (12).

Finally, we conclude that  $F$  consists of a single component; namely, it is a path. Suppose not, and let  $u_1$  and  $u_2$  be ends of distinct components in  $F$ . By (12), the sets  $B_{G-Y}^{(\log n)^4}(B_{u_1})$  and  $B_{G-Y}^{(\log n)^4}(B_{u_2})$  intersect, showing that there is a path  $Q$  in  $G'$  from  $u_1$  to  $u_2$  whose interior avoids  $V(F)$ . Taking  $\mathcal{Q}' := \mathcal{Q} \cup \{Q'\}$ , we reach a contradiction to (b). We have thus proved that  $F$  is a path. As it contains  $M$  and has no other edges in  $\bigcup_{e \in M} P(e)$ , it satisfies the requirements of Claim 8.6.  $\square$

For  $M \in \mathcal{M}$ , let  $P_M$  be a path as guaranteed by Claim 8.6. Write  $\mathcal{P}_2 := \{P_M : M \in \mathcal{M}\}$ , and take  $\mathcal{P} := \mathcal{P}_1 \cup \mathcal{P}_2$ . Then  $|\mathcal{P}| \leq |\mathcal{P}_1| + |\mathcal{P}_2| \leq 5k$ . Notice that  $\mathcal{P}$  separates the edges of  $G$ . Indeed, given  $e, f \in E(G)$ , then one of  $P(e)$  and  $M(e)$  contains  $e$  but not  $f$ . Thus Lemma 8.4 is proved.  $\square$

### 8.3 Proof of Lemma 3.2

Finally, we put everything together to prove Lemma 3.2, restated here. The proof follows quite straightforwardly from Lemmas 8.1 and 8.4.

**Lemma 3.2.** *There exists  $d_0$  such that the following holds. Suppose that  $G$  is a graph on  $n$  vertices with average degree  $d$ , where  $d \geq d_0$ . Then there is a subgraph  $G_1 \subseteq G$ , a collection  $\mathcal{H}$  of subgraphs of  $G$  and a collection  $\mathcal{P}$  of paths in  $G$  with the following properties.*

- (1) *The graphs in  $\mathcal{H}$  are mutually edge-disjoint;  $|H| \leq 2^{(\log d)^7}$  for every  $H \in \mathcal{H}$ ; and  $\sum_{H \in \mathcal{H}} |H| \leq 4n$ .*
- (2) *The collection  $\mathcal{P}$  separates the edges of  $G - G_1 - \bigcup_{H \in \mathcal{H}} H$ , and satisfies  $|\mathcal{P}| \leq 80n$ .*
- (3) *The graph  $G_1$  has average degree at most  $(\log d)^3$ .*

*Proof of Lemma 3.2.* Let  $G$  be a graph on  $n$  vertices with average degree  $d$ , where  $d$  is large. Take  $\varepsilon := 1/48$  and apply Lemma 4.1 with  $\varepsilon$ ,  $s = 0$  and  $t = 2n/3$  (the last choice does not matter as  $t$  does not play a role when  $s = 0$ ) to obtain a collection  $\mathcal{H}$  of  $\varepsilon$ -expanders that decomposes  $G$  and satisfies  $\sum_{H \in \mathcal{H}} |H| \leq 2n$ .

Apply Lemma 8.1 to each  $H \in \mathcal{H}$ , to obtain a subgraph  $F_H \subseteq H$  and a collection  $\mathcal{P}_H$  of paths in  $H$  such that:  $F_H$  has maximum degree at most  $d^7$ ;  $|\mathcal{P}_H| \leq 17 \max\{|H|, e(H)/d\}$ ; and  $\mathcal{P}_H$  separates the edges of  $H - F_H$ . Let  $\mathcal{P}'_H$  be collection of at most  $|H|$  paths decomposing  $H - F_H$  (which exists by Theorem 1.2).

Now, for each  $H \in \mathcal{H}$ , apply Lemma 4.1 to  $F_H$  with  $\varepsilon$ ,  $s_{4.1} = ((k_0(\varepsilon))_{7.1} + 3)^3$  and  $t_{4.1} = d$ , to obtain a collection  $\mathcal{F}_H$  of pairwise edge-disjoint subgraphs of  $F_H$  that cover all but at most  $48s|F_H|(\log d + 1)^2$  edges of  $F_H$  and satisfy  $\sum_{F \in \mathcal{F}_H} |F| \leq 2|F_H|$ . Denote by  $E_H$  the set of edges in  $H$  uncovered by  $\mathcal{F}_H$ .

Let  $\mathcal{F}_1$  be the collection of graphs  $F \in \bigcup_{H \in \mathcal{H}} \mathcal{F}_H$  that satisfy  $|F| \geq 2^{(\log d)^7}$  and let  $\mathcal{F}_2 := (\bigcup_{H \in \mathcal{H}} F_H) - \mathcal{F}_1$ . Apply Lemma 8.4 to each  $F \in \mathcal{F}_1$  to obtain a collection  $\mathcal{Q}_F$  of paths in  $F$  such that  $|\mathcal{Q}_F| \leq 5 \max\{|F|, e(F)/d\}$  and  $\mathcal{Q}_F$  separates the edges of  $F$ .

Take  $\mathcal{P}$  to be the union of the collections  $\mathcal{P}_H, \mathcal{P}'_H$ , for  $H \in \mathcal{H}$ , and  $\mathcal{Q}_F$ , for  $F \in \mathcal{F}_1$ . Define  $G_1$  to be the graph on vertices  $V(G)$  and edges  $\bigcup_{H \in \mathcal{H}} E_H$ . We show that the statement of Lemma 3.2 holds (with  $\mathcal{H}_{3.2} = \mathcal{F}_2$ ). Notice that the graphs in  $\mathcal{F}_2$  indeed are pairwise edge-disjoint, because the graphs in  $\mathcal{H}$  are pairwise edge-disjoint and so are the graphs in  $\mathcal{F}_H$  for every  $H \in \mathcal{H}$ . Moreover,  $|F| \leq 2^{(\log d)^7}$  for every  $F \in \mathcal{F}_2$ , by choice of  $\mathcal{F}_2$ . Finally,

$$\sum_{F \in \mathcal{F}_1 \cup \mathcal{F}_2} |F| = \sum_{H \in \mathcal{H}} \sum_{F \in \mathcal{F}_H} |F| \leq \sum_{H \in \mathcal{H}} 2|F_H| \leq \sum_{H \in \mathcal{H}} 2|H| \leq 4n,$$



using the choice of  $\mathcal{H}$  and  $\mathcal{F}_H$ . In particular,  $\sum_{F \in \mathcal{F}_2} |F| \leq 4n$ . Thus (1) is proved.

Next, notice that

$$\begin{aligned} |\mathcal{P}| &\leq \sum_{H \in \mathcal{H}} (|\mathcal{P}_H| + |\mathcal{P}'_H|) + \sum_{F \in \mathcal{F}_1} |Q_F| \\ &\leq \sum_{H \in \mathcal{H}} 18(|H| + e(H)/d) + \sum_{F \in \mathcal{F}_1} 5(|F| + e(F)/d) \\ &\leq 36n + 18e(G)/d + 20n + 5e(G)/d \leq 80n. \end{aligned}$$

For the second inequality we used that the graphs in  $\mathcal{H}$  are pairwise edge-disjoint and similarly for  $\mathcal{F}_1$ , and for the last inequality we used that  $G$  has average degree  $d$ .

We claim that  $\mathcal{P}$  separates the edges of  $G' := G - G_1 - \bigcup_{F \in \mathcal{F}_2} F$ . Indeed, let  $e$  and  $f$  be distinct edges in  $G'$ , and let  $H \in \mathcal{H}$  satisfy  $e \in H$ . If  $e \in H - F_H$ , then if  $f \notin H - F_H$  there is a path in  $\mathcal{P}'_H$  that contains  $e$  but not  $f$ , and otherwise there is such a path in  $\mathcal{P}_H$ . Now suppose  $e \in F_H$  and let  $F \in \mathcal{F}_H$  be a graph that contains  $e$ . Then  $Q_F$  has a path that contains  $e$  but not  $f$ . This proves (2).

Finally, we have

$$2e(G_1) \leq 2 \sum_{H \in \mathcal{H}} |E_H| \leq \sum_{H \in \mathcal{H}} 96s|F_H|(\log d + 1)^2 \leq \frac{1}{2} \sum_{H \in \mathcal{H}} |H|(\log d)^3 \leq n(\log d)^3,$$

using that  $d$  is large and  $s$  is a constant (it depends only on  $\varepsilon$ , which itself is a constant). This completes the proof of (3) and Lemma 3.2.  $\square$

## 9 Conclusion

Recall that  $\text{sep}^*(G)$  is the size of a smallest weakly-separating path system for  $G$ , and  $\text{sep}^*(n)$  is the maximum of  $\text{sep}^*(G)$  over all  $n$ -vertex graphs  $G$ ;  $\text{sep}(G)$  and  $\text{sep}(n)$  are defined analogously for strong separation.

In this paper we proved that  $\text{sep}(n) = O(n \log^* n)$ . This is significant progress from the easy initial bound of  $O(n \log n)$ , towards a conjecture of, independently, Falgas-Ravry–Kittipassorn–Korándi–Narayanan and the author [13], and Balogh–Csaba–Martin–Pluhár [4], that  $\text{sep}(n) = O(n)$ .

It is plausible that our methods could be used to fully settle this conjecture. For example, one could try to prove a statement like the following, by induction on  $n$ . Suppose that  $G$  is an  $\varepsilon$ -expander and let  $H$  be a subgraph of  $G$  on  $n$  vertices. Then the edges of  $H$  can be separated using  $O(n)$  paths in  $G$ . To accomplish this, it would suffice to find  $O(n)$  paths in  $G$  that separated all edges of  $H$  touching a set  $V$  of  $\Omega(n)$  vertices in  $H$ . An obstruction to achieving this is a graph  $H$  as follows:  $H$  has average degree  $d$ , but almost all vertices in  $H$  have degree much lower than  $d$  (say, at most  $\log \log d$ ), and, moreover, there are relatively small sets of low degree vertices in  $H$  with few low degree neighbours.

In fact, the authors of [13] suggest that, perhaps,  $\text{sep}^*(n) = (1 + o(1))n$ ; it is easy to see that this bound would be tight by considering complete graphs. If true, this is likely to be very hard to prove. Indeed, even the restriction to complete graph seems hard. Wickes [42] proved that  $\text{sep}^*(K_n) \leq (\frac{21}{16} + o(1))n$  for any integer  $n$ , and that  $\text{sep}^*(K_n) \leq n$  when  $n = p$  or  $p+1$  for some prime  $p$ . It would be nice, and perhaps not completely out of reach, to prove that  $\text{sep}^*(K_n) = (1 + o(1))n$ . It also makes sense to study  $\text{sep}(K_n)$ .

## Acknowledgements

I would like to thank António Girão for reminding me of this problem and making the connection to sublinear expansion. I am grateful to Matija Bucić and Richard Montgomery for insightful discussions regarding their very nice paper [8] and for giving me a copy of an early draft. Some of the research leading up to this paper was done while the author attended the MFO workshop Combinatorics, Probability and Computing, held in Oberwolfach in April 2022, and the BIRS workshop Cross Community Collaborations in Combinatorics, held in Banff in June 2022. I am thankful to the organisers of both workshops for inviting me, and to MFO and BIRS for providing great infrastructure for research. I would also like to thank the kind anonymous referee for carefully reading a previous version of this paper, and for their various helpful suggestions.

## References

- [1] R. Aharoni and P. Haxell, *Hall's theorem for hypergraphs*, J. Graph Theory **35** (2000), 83–88. [14](#)
- [2] S. S. Ahuja, S. Ramasubramanian, and M. Krunz, *Single-link failure detection in all-optical networks using monitoring cycles and paths*, IEEE/ACM Transactions on Networking **17** (2009), 1080–1093. [2](#)
- [3] N. Alon and J. H. Spencer, *The probabilistic method*, 4 ed., John Wiley & Sons, 2016. [14](#)
- [4] J. Balogh, B. Csaba, R. R. Martin, and A. Pluhár, *On the path separation number of graphs*, Discr. Appl. Math. **213** (2016), 26–33. [2](#), [4](#), [33](#)
- [5] A. Blanché, M. Bonamy, and N. Bonichon, *Gallai's path decomposition in planar graphs*, arXiv:2110.08870 (2021). [3](#)
- [6] B. Bollobás and A. Scott, *On separating systems*, Europ. J. Combin. **28** (2007), 1068–1071. [1](#)
- [7] B. Bollobás and A. Scott, *Separating systems and oriented graphs of diameter two*, J. Combin. Theory Ser. B **97** (2007), 193–203. [1](#)
- [8] M. Bucić and R. Montgomery, *Towards the Erdős-Gallai Cycle Decomposition Conjecture*, arXiv:2211.07689 (2022). [3](#), [4](#), [5](#), [7](#), [10](#), [13](#), [14](#), [16](#), [17](#), [18](#), [34](#)
- [9] M.-C. Cai, *On separating systems of graphs*, Discr. Math. **49** (1984), 15–20. [1](#)
- [10] N. Dean and M. Kouider, *Gallai's conjecture for disconnected graphs*, Discr. Math. **213** (2000), 43–54. [3](#)
- [11] P. Erdős, A. W. Goodman, and L. Pósa, *The representation of a graph by set intersections*, Can. J. Math. **18** (1966), 106–112. [3](#)

- [12] P. Erdős and A. Hajnal, *On chromatic number of graphs and set-systems*, Acta Math. Acad. Sci. Hungar **17** (1966), 61–99. [3](#)
- [13] V. Falgas-Ravry, T. Kittipassorn, D. Korándi, S. Letzter, and B. Narayanan, *Separating path systems*, J. Combin. **5** (2014), 335–354. [2](#), [3](#), [4](#), [33](#), [34](#)
- [14] I. G. Fernández, J. Hyde, H. Liu, O. Pikhurko, and Z. Wu, *Disjoint isomorphic balanced clique subdivisions*, J. Combin. Theory Ser. B **161** (2023), 417–436. [3](#), [4](#)
- [15] I. G. Fernández, J. Kim, Y. Kim, and H. Liu, *Nested cycles with no geometric crossings*, Proc. Am. Math. Soc. Ser. B **9** (2022), 22–32. [3](#), [4](#)
- [16] I. G. Fernández and H. Liu, *How to build a pillar: A proof of Thomassen’s conjecture*, J. Combin. Theory Ser. B **162** (2023), 13–33. [3](#), [4](#)
- [17] F. Foucaud, S. Gravier, R. Naserasr, A. Parreau, and P. Valicov, *Identifying codes in line graphs*, J. Graph Theory **73** (2013), no. 4, 425–448. [1](#)
- [18] F. Foucaud and M. Kovše, *Identifying path covers in graphs*, J. Discrete Algorithms **23** (2013), 21–34. [1](#)
- [19] G. Hansel, *Nombre minimal de contacts de fermeture nécessaires pour réaliser une fonction booléenne symétrique de  $n$  variables*, C. R. Acad. Sci. Paris **258** (1964), 6037–6040. [1](#)
- [20] N. J. A. Harvey, M. Patrascu, Y. Wen, S. Yekhanin, and V. W. S. Chan, *Non-adaptive fault diagnosis for all-optical networks via combinatorial group testing on graphs*, IEEE INFOCOM 2007-26th IEEE International Conference on Computer Communications, IEEE, 2007, pp. 697–705. [1](#), [2](#)
- [21] J. Haslegrave, J. Hu, J. Kim, H. Liu, B. Luan, and G. Wang, *Crx and long cycles in graphs*, arXiv:2107.02061 (2021). [3](#), [4](#)
- [22] J. Haslegrave, J. Hyde, J. Kim, and H. Liu, *Ramsey numbers of cycles versus general graphs*, arXiv:2112.03893 (2021). [3](#), [4](#)
- [23] J. Haslegrave, J. Kim, and H. Liu, *Extremal density for sparse minors and subdivisions*, Int. Math. Res. Not. (2021), 15505–15548. [3](#), [4](#), [10](#)
- [24] I. Honkala, M. G. Karpovsky, and S. Litsyn, *Cycles identifying vertices and edges in binary hypercubes and 2-dimensional tori*, Discrete Appl. Math. **129** (2003), 409–419. [1](#)
- [25] G. Katona, *On separating systems of a finite set*, J. Combin. Theory **1** (1966), 174–194. [1](#)
- [26] J. Kim, H. Liu, M. Sharifzadeh, and K. Staden, *Proof of Komlós’s conjecture on Hamiltonian subsets*, Proc. Lond. Math. Soc. **115** (2017), 974–1013. [3](#), [4](#)
- [27] J. Komlós and E. Szemerédi, *Topological cliques in graphs*, Combin. Probab. Comput. **3** (1994), 247–256. [3](#), [4](#), [9](#)
- [28] ———, *Topological cliques in graphs II*, Combin. Probab. Comput. **5** (1996), 79–90. [3](#), [4](#), [9](#)
- [29] A. Kündgen, D. Mubayi, and P. Tetali, *Minimal completely separating systems of  $k$ -sets*, J. Combin. Theory Ser. A **93** (2001), 192–198. [1](#)
- [30] S. Letzter and A. Girão, *Immersion of complete digraphs in Eulerian digraphs*, Israel J. Math., to appear. [3](#), [4](#)
- [31] H. Liu and R. Montgomery, *A proof of Mader’s conjecture on large clique subdivisions in  $C_4$ -free graphs*, J. Lond. Math. Soc. **95** (2017), 203–222. [3](#), [4](#)

- [32] ———, *A solution to Erdős and Hajnal’s odd cycle problem*, J. Amer. Math. Soc. **36** (2020), 1191–1234. [3](#), [4](#), [6](#), [7](#), [23](#)
- [33] H. Liu, G. Wang, and D. Yang, *Clique immersion in graphs without a fixed bipartite graph*, J. Combin. Theory Ser. B **157** (2022), 346–365. [3](#), [4](#)
- [34] L. Lovász, *On covering of graphs*, Theory of Graphs (Proc. Colloq., Tihany, 1966), Academic Press, New York, 1968, pp. 231–236. [2](#)
- [35] C. Ramsay and Roberts I. T., *Minimal completely separating systems of sets*, Australas. J. Combin. **13** (1996), 129–150. [1](#)
- [36] A Rényi, *On random generating elements of a finite Boolean algebra*, Acta Sci. Math. Szeged **22** (1961), 75–81. [1](#)
- [37] J. Spencer, *Minimal completely separating systems*, J. Combin. Theory **8** (1970), 446–447. [1](#)
- [38] J. Tapolcai, L. Rónyai, and P.-H. Ho, *Link fault localization using bi-directional  $m$ -trails in all-optical mesh networks*, IEEE Transactions on Communications **61** (2012), 291–300. [1](#), [2](#)
- [39] J. Tapolcai, B. Wu, and P.-H. Ho, *On monitoring and failure localization in mesh all-optical networks*, IEEE INFOCOM 2009, IEEE, 2009, pp. 1008–1016. [1](#)
- [40] I. Tomon, *Robust (rainbow) subdivisions and simplicial cycles*, arXiv:2201.12309 (2022). [5](#), [13](#)
- [41] I. Wegener, *On separating systems whose elements are sets of at most  $k$  elements*, Discr. Math. **28** (1979), 219–222. [1](#)
- [42] B. Wickes, *Separating path systems for the complete graph*, arXiv:2209.04302 (2022). [2](#), [34](#)
- [43] L. Yan, *On path decompositions of graphs*, Ph.D. thesis, Arizona State University, 1998. [3](#)