Sublinear expanders

Matija Bucić

July 17, 2025

Introduction.

Expander graphs make one of the most widely useful class of graphs. They appear naturally across many different areas of mathematics and satisfy a variety of very useful properties.

There are many closely related definitions, and lots of their usefulness stems from translating between these definitions. At a rather high level one can think of them as "robustly well-connected" graphs. While many interpretations of both "robustly" and "well-connected" fit perfectly well here, for example one can think of well-connected as saying that there is a short path between any two vertices and robustness as saying that this property is preserved under minor modifications to the graph such as edge or vertex deletions.

With these properties in mind, and armed with the fact that one can find such graphs with very few edges one perhaps starts to see the utility. For example, both well-connectedness and robustness are very useful properties for a communication network, while having few edges translates to lower costs in building it.

Expander graphs have been extensively studied, both due to their applications, as well as the intrinsic interest. We point an interested reader to several wonderful surveys [11, 17, 21] on the topic. We make essentially no assumption on prior knowledge on the topic.

In this course we will focus on a certain weak notion of expansion, called sublinear expansion, which has found some spectacular applications in recent years. Our goal is to help develop some intuition of when this type of expanders are useful and how to use them. Due to this we will not include a comprehensive overview of the wide variety of applications and modifications that arose over the years but rather point an interested reader to an excellent recent survey by Letzter [18].

1 Linear expanders

Let us start with a definition of what is usually referred to as a vertex expander.

Definition 1.1. For any $\varepsilon > 0$, an n-vertex non-empty graph G is an ε-expander if for all $U \subseteq V(G)$ such that $|U| \leq n/2$ we have:

$$|N(U)| > \varepsilon |U|.$$

So a graph is an expander if by taking one step out from any set of vertices, which is not too large, we reach many new vertices. We note that here and throughout these notes $N(U) := \{v \in V(G) \setminus U \mid \exists u \in U : vu \in E(G)\}$ denotes the external neighbourhood of the set of vertices U, namely we do *not* include vertices already in U.[1]

We note also that when working with usual expanders one often restricts the set we are allowed to expand to be of size at most $\frac{n}{2(1+\varepsilon)}$ since expanding a set of this size one reaches $n/2$ vertices, so for larger sets one can simply expand a subset of this size. The benefit of this is that one can have a C-expander with $C > 1$ unlike in our definition, where a set of size $n/2$ can only ever expand by a factor of at most 1. Since our focus is on expanders with small expansion factors the above definition is a bit cleaner for us.

[1] In the context of usual, strong expanders, this point makes little difference since whether one has C or $C - 1$ factor expansion is rarely relevant, but our expanders are going to have ε very small and the distinction is important.

While there are by now plenty of interesting examples, arguably the simplest is simply K_n, the clique on n vertices. According to our definition it is easy to see it is a 1-expander. Perhaps the most important class of examples are random graphs. They also serve as an excellent source of intuition for a variety of arguments.

Exercise 1. How good of an expander is the binomial random graph $G(n, p)$ as p varies?

Let us now prove that our definition actually implies the well-connectedness property we mentioned earlier, namely any two vertices are not too far apart. Here, the diameter of a graph G, denoted $\operatorname{diam}(G)$, is the maximum distance[2] between two vertices in G.'

Lemma 1.2 (Diameter bound). *Let $\varepsilon > 0$ and G be an n-vertex ε-expander graph. Then,*

$$\operatorname{diam} G \le 2\left\lceil \log_{1+\varepsilon} \frac{n}{2} \right\rceil.$$

Proof. Let $\ell = \left\lceil \log_{1+\varepsilon} \frac{n}{2} \right\rceil$. Our goal is to show that any pair of vertices v, u is joined by a walk of length at most 2ℓ. Let $U_0 = \{u\}$, and $U_{i+1} = U_i \cup N(U_i)$, for each $0 \le i < \ell$. Then, we have $|U_i| > n/2$ or $|U_{i+1}| > (1 + \varepsilon)|U_i|$. Iterating this we get,

$$|U_{i+1}| \ge \min\{(1 + \varepsilon)^{i+1}, n/2\}.$$

By our choice of ℓ we have $|U_\ell| > n/2$. Similarly, if we define $V_0 = \{v\}$ and $V_{i+1} = V_i \cup N(V_i)$, then $|V_\ell| > n/2$.

So $|U_\ell| + |V_\ell| > n$, which implies that there exists a vertex $w \in U_\ell \cap V_\ell$. Since $w \in U_\ell$, there exists a uw walk of length at most ℓ, and since $w \in V_\ell$, there exists a vw walk of length at most ℓ. Concatenating these two walks gives a vu walk of length at most 2ℓ, as desired. □

Exercise 2. Prove that in an n-vertex ε-expander between any two sets of size $u \le n/2$ there is a path of length at most

$$2\left\lceil \log_{1+\varepsilon} \frac{n}{2u} \right\rceil.$$

2 Sublinear expanders.

Note that in Definition 1.1 nothing prevents us from letting ε depend on n. Informally, a *sublinear* expander is an ε-expander where $\varepsilon \to 0$ as $n \to \infty$. Thinking of $\varepsilon = \frac{1}{\log n}$ is a good concrete choice to keep in mind.

Exercise 3. Prove that the $N = 2^n$ vertex hypercube graph Q_n with vertex set $\{0, 1\}^n$ and two vertices adjacent iff they differ in exactly one coordinate is a $\Omega\left(\frac{1}{\sqrt{\log N}}\right)$-expander.

Hint. You may wish to use Harper's vertex isoperimetric inequality for the hypercube.

Now, a natural question is why would we consider a clearly weaker notion? The answer is that while usual expanders appear in many places, sublinear expanders appear essentially everywhere. We will see that every graph contains a sublinear expander with "essentially" the same degree, as well as that any graph can be decomposed into sublinear expanders (in both cases with expansion at least

[2] Defined as the minimum length of a path between the two vertices.

$1/\operatorname{polylog} n$). This means that for a wide variety of problems one can get a lot of mileage from understanding the problem well for sublinear expanders. On the other hand, they retain many of the useful properties of their stronger siblings (albeit usually slightly weakened), so it is often much more feasible to attack the problem on a sublinear expander than a completely arbitrary one.

For example, Lemma 1.2 applies equally as well for sublinear expanders, and for example, for $\frac{1}{\log n}$-expanders the diameter is guaranteed to be at most

$$2\left\lceil \log_{1+1/\log n} \frac{n}{2} \right\rceil = 2\left\lceil \frac{n/2}{\log(1+1/\log n)} \right\rceil \leq O(\log^2 n),$$

so still very small.

Finding long paths or cycles in graphs is a very classical problem in extremal combinatorics. Dirac's theorem is one of the most famous examples, which gives a criterion for a graph to contain the longest possible cycle (called Hamilton cycle), so one of length n. Namely, if a graph has minimum degree at least $\lfloor n/2 \rfloor$, and this is known to be tight. This, while tight is unfortunately quite dense. On the other hand, in a random graph $G(n,p)$ much less density is needed to guarantee a Hamilton cycle w.h.p., so most graphs do have one. This motivated a long line of work, with plenty of tantalizing open problems, of finding which conditions on a graph which guarantee Hamilton cycles (or at least very long ones). There has been very exciting recent progress in case of strong expander graphs [8], as well as for regular sublinear expanders [19]. We here will prove a simple, weaker lemma, which is nonetheless often very useful (as we will see later in the course). It also allows us to showcase the use of the so called DFS method of Ben-Eliezer, Krivelevich, and Sudakov [2], which has proven itself to be very useful in a variety of settings. See [17, Theorem 7.4] for a version with arbitrary expansion parameter and tighter bound for the cycle case.

Lemma 2.1. *Any n-vertex $\frac{1}{\log n}$-expander G, contains a path of length $\Omega(n/\log n)$ and a cycle of length at least $\Omega\big(n/\log^2 n\big)$.*

Proof. Observe first that the expansion condition guarantees G is connected, since otherwise we could find a connected component of G of size at most $\frac{|G|}{2}$, whose vertex set does not expand at all.

We run the DFS algorithm on G as follows. At any point during the process we have the set of unexplored vertices U, the path P with active endvertex $t(P)$ (called a stack), and the set of processed vertices R. Picking an arbitrary vertex $r \in V(G)$, We start with $U = V(G) \setminus \{r\}, R = \emptyset$ and with P being the path with vertex set $\{r\}$, and set $t(P) = r$. At each step, if there is a neighbor v of $t(P)$ in U we add it to P with the edge $t(P)v$ and let $t(P) = v$. Otherwise, we move $t(P)$ from P to R and set its neighbor in P as the new $t(P)$.

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

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

Let us turn to showing there is also a long cycle. let X, Y, Z be sets of consecutive vertices of P in that order which partition $V(P)$ so that $|X|, |Z| \geq \frac{|P|}{3}$ and $\frac{n}{6\log^2 n} \leq |Y| < \frac{n}{3\log^2 n}$. If X and Z are connected by some path in $G \setminus Y$, then take a shortest path, Q say, between X and Z in $G \setminus Y$ and note that, combined with the segment of P between the endvertices of Q, this gives a cycle containing each vertex in Y, which thus has size $\Omega\left(\frac{n}{\log^2 n}\right)$. If X and Z are not connected by a path in $G \setminus Y$, then we can take a partition $V(G) \setminus Y = X' \cup Z'$ with no edges between X' and Z' in G, and $X \subseteq X'$ and $Z \subseteq Z'$. Without loss of generality, suppose that $|X'| \leq \frac{n}{2}$. By the expansion condition we have $|N_G(X')| \geq \frac{|X|}{\log n} \geq \frac{|P|}{3\log n} \geq \frac{n}{3\log^2 n}$, yet we also have $|N_G(X')| \leq |Y| < \frac{n}{3\log^2 n}$, a contradiction. $\square$

We note that one can in fact find a cycle of length $\Omega(n/\log n)$ and that this lemma is tight up to a constant factor as can be seen by taking a complete bipartite graph with one side of size $n/\log n$. We opted to show this argument here since we will repeatedly use this type of "robustness" arguments (which say that if you remove some set of vertices from your expander, then the expansion properties of slightly larger sets remain essentially intact).

Exercise 4. Show that any n-vertex $\frac{1}{\log n}$-expander G, contains a cycle of length $\Omega(n/\log n)$.

2.1 Pass to expander lemma

In this section we will show how to find a sublinear expander inside an arbitrary graph while "essentially" preserving the average degree. We first introduce a definition of a more robust version of sublinear expanders which often give one a crucial extra bit of power when working with them.

Definition 2.2. A graph G on $n \geq 1$ vertices is called a *robust sublinear expander* if

- for every $0 \leq \varepsilon \leq 1$ and every non-empty subset $U \subseteq V(G)$ of size $|U| \leq n^{1-\varepsilon}$, and
- for every subset $F \subseteq E(G)$ of $|F| \leq \frac{\varepsilon}{4} \cdot d(G) \cdot |U|$ edges, we have

$$|N_{G-F}(U)| \geq \frac{\varepsilon}{4} \cdot |U|.$$

To build intuition let us for the moment ignore the second point. In this case, the definition becomes the usual vertex expansion one, albeit with a different level of expansion for sets of different sizes (the size is controlled by the ε parameter). For example if one picks $\varepsilon = 1/\log n$ the required condition simply becomes $\frac{1}{3\log n}$-expansion (since the upper bound on $|U|$ becomes $n/2$). On the other hand, if ε is an absolute constant then we require that sets of size up to $n^{1-\varepsilon}$ expand by an $\varepsilon/3$ factor. As ε transitions from an absolute constant to $1/\log n$ the size of the sets we require expand grows while the required expansion factor decreases smoothly. For comparison to simply pure $\frac{1}{\log n}$-expanders which we saw have diameter upper bounded by $O(\log^2 n)$ robust sublinear expanders have quite a bit better guarantee and come closer to the usual constant factor expander guarantee of $O(\log n)$.

Exercise 5. Show that an n-vertex robust sublinear expander H has $\mathrm{diam}(H) \leq O(\log n \log\log n)$.

Turning back to the second condition, it in addition allows us to remove a substantial number of edges from the graph and still preserve the expansion guarantees we discussed above. In most cases, it is used in order to be able to forbid usage of a number of "inconvenient" edges. We stress that the choice of F can be made after the subset U is specified, and its allowed size is controlled by the size of the set we are expanding.

The following lemma shows we can find a robust sublinear expander inside any graph while preserving a certain relative notion of the degree. We note that there are several different, pass to a sublinear expander lemmas in use depending on how large the average degree of the initial graph is. The following particular lemma is designed for graphs with average degree at least a large constant times $\log n$. The proofs of all these lemmas essentially follow the same strategy.

Lemma 2.3. *Every non-empty graph G contains a subgraph H which is a robust sublinear expander and has average degree*

$$d(H) \geq \frac{\log|V(H)|}{\log|V(G)|} \cdot d(G).$$

Proof. Among all subgraphs H of G with at least two vertices, let us choose a subgraph H maximizing the expression

$$\frac{d(H)}{\log|V(H)|}.$$

Then, we get the desired inequality since $H = G$ is itself a viable choice, namely

$$\frac{d(H)}{\log|V(H)|} \geq \frac{d(G)}{\log|V(G)|}.$$

It remains to show that H is a robust sublinear expander. Let $n = |V(H)| \geq 2$ and $d = d(H) > 0$. Consider $0 \leq \varepsilon \leq 1$, and let $U \subseteq V(H)$ be a non-empty subset of size $|U| = u \leq n^{1-\varepsilon}$. Furthermore, let $F \subseteq E(H)$ be a subset of $|F| \leq \frac{\varepsilon}{4} \cdot du$ edges. Our goal is to show that $|N_{H-F}(U)| \geq \frac{\varepsilon}{4} \cdot u$. Suppose towards a contradiction that $|N_{H-F}(U)| < \frac{\varepsilon}{4} \cdot u$.

Let $H_1 = H[U \cup N_{G-F}(U)]$ and let $H_2 = H[V(H) \setminus U]$, and denote by $n_1 = |H| \leq (1 + \varepsilon/4)u$ and $n_2 = |H_2| = n - u$. Then, every edge of H belongs to H_1, H_2, or F (with some edges within $H[N_{G-F}(U)]$ counted twice). This implies

$$e(H) \leq e(H_1) + e(H_2) + |F|.$$

On the other hand, the maximality[3] of H implies $d(H_i) \leq d \cdot \frac{\log n_i}{\log n}$ for both i. This, plugged into the above inequality implies

$$\begin{aligned} n \leq \frac{2e(H_1) + 2e(H_2) + 2|F|}{d} &\leq n_1 \cdot \frac{\log n_1}{\log n} + n_2 \cdot \frac{\log n_2}{\log n} + \frac{\varepsilon}{2} \cdot u \\ &\leq \left(1 + \frac{\varepsilon}{4}\right)u \cdot \frac{(1-\varepsilon)\log n + \log(1+\varepsilon/4)}{\log n} + n - u + \frac{\varepsilon}{2} \cdot u \\ &\leq \left(1 + \frac{\varepsilon}{4}\right)u \cdot \left(1 - \frac{3\varepsilon}{4}\right) + n - u + \frac{\varepsilon}{2} \cdot u < n, \end{aligned}$$

where in the penultimate inequality we used $\log(1 + \varepsilon/4) < \varepsilon/4 \leq \frac{\varepsilon}{4} \cdot \log n$. This is a contradiction completing the proof. $\square$

[3]Technically if $n_i = 1$ these inequalities do not follow from the maximality assumption but are immediate.

As we already mentioned, the above lemma is only useful if $d(G) \gg \log n$ since, e.g. if $d(G) \leq \log n$ then the density guarantee we get on H becomes vacuous and H being a single edge is a valid (albeit useless) output. When we have $d(G) \geq C \log n$ on the other hand we get the same guarantee of $d(H) \geq C \log |H|$ and can at least guarantee $|H| \geq C$. This in fact, is one of the main downsides of essentially all pass to expander lemmas, namely while one preserves the "relative" guarantee on the average degree, one might lose a lot in terms of the size of the graph. This however, can't be avoided, our original graph G might simply be a vertex disjoint union of cliques of size $d(G)+1$, so it is not possible to find a very large expander subgraph.

Note that the definition of a robust sublinear expander already guarantees that its minimum degree $\delta(G) \geq d(G)/4$, by simply applying the expansion property to the sets of size one. With this in mind one can view the pass to expander lemma as a significantly stronger version of the easy classical result that any graph of average degree d contains a subgraph with minimum degree at least $d/2$.

Exercise 6. Show that the above proof in fact ensures $\delta(H) \geq d(H)/2$.

One can obtain pass to expander lemmas even for graphs with only a large constant degree, and in fact these are precisely the first ones ever used due to Komlós and Szemerédi [16]. They are in a certain sense weaker than the robust sublinear expanders we introduced above, and more cumbersome, but more versatile and are behind an impressive number of recent results using the theory. We point an interested reader to [18] for a gentle description of their properties.

On the other hand, if one has at least polynomial average degree $d(H) \gg n^{\alpha}$ for some absolute constant $\alpha > 0$, then one can find stronger expander subgraphs by repeating the above proof with the potential function $\frac{d(H)}{n^{\alpha}}$. Note that as above, the lower bound on $d(H)$ is simply required to guarantee the expander we find is not too small.

Exercise 7. Show that

a) For any $1 \geq \alpha > 0$, every non-empty graph G contains an $\frac{\alpha}{1+\alpha}$-expander subgraph H with

$$d(H) \geq \frac{|V(H)|^{\alpha}}{|V(G)|^{\alpha}} \cdot d(G).$$

b) One can in fact get (significantly) better expansion for smaller sets.

c) One can sacrifice a tiny bit of vertex expansion to ensure robustness, namely that the expansion is guaranteed even upon deletion of a substantial number of edges.

3 Applications of pass to expander lemmas.

In this section we will illustrate the "pass to expander" method with some examples. We note that the examples are picked to be simple and instructive rather than as flashy as possible, we direct the reader to [18] for a collection of some of the most spectacular examples.

3.1 Finding topological cliques

How many edges does one need in a graph in order to guarantee K_k as a subgraph? This is precisely answered by the classical Turán's theorem and is one of the cornerstone results of extremal graph theory. Unfortunately, from the view of using this result, the answer is rather large, even for a triangle we require a quadratic number of edges. In some applications, it turns out a slightly relaxed notion of a clique, where we replace every edge by a path in such a way that the paths are internally vertex disjoint, suffices.

Definition 3.1. A *topological clique* is a clique in which we replaced every edge with a path in such a way that all the paths are internally vertex disjoint. The original vertices are called the *anchor points* and their number is the *order* of the topological clique.

We note that an alternative way of defining a topological clique is as a subdivision of a clique. For example, a topological clique of order three is simply a cycle. We know that it is much easier to guarantee a cycle in a graph (one needs average degree of at least 2) than it is to guarantee a triangle (one needs average degree at least $n/2$). A very natural question is what happens for larger orders. This is a classical question of Erdős and Hajnal, and Mader. It was resolved by Bollobás and Thomason [3] and independently Komlós and Szemerédi [16] who showed the following.

Theorem 3.2. *Any graph with average degree d contains a topological clique of order $\Omega(\sqrt{d})$.*

It is not hard to see that this theorem is tight.

Exercise 8. Show that for any integer d there exists a graph which does not contain a topological clique of order larger than $10\sqrt{d}$.

The proofs of Bollobás and Thomason and independently Komlós and Szemerédi are very different, and in fact the one from the latter paper is the origin of the theory of sublinear expanders[4].

We will prove a slightly weaker version of the result, with a significantly simpler proof.

Theorem 3.3. *Any n-vertex graph with average degree d contains a topological clique of order $\Omega(\sqrt{d}/\log^2 n)$.*

Proof. By Lemma 2.3 we can find a robust sublinear expander subgraph H with minimum degree

$$\delta(H) \geq d(H)/4 \geq \frac{\log |H|}{4 \log n} \cdot d \geq \frac{d}{4 \log n}.$$

Let $m = |H|$. Now fix an arbitrary set of distinct vertices $v_1, \ldots, v_t \in V(H)$ with $t = \sqrt{d}/(8 \log^2 n)$ to serve as anchor points.

Now take a maximal collection of internally vertex disjoint paths of length at most $4 \log^2 m$ between the anchor points, with at most one per pair. If each pair got a path we found the desired topological clique and are done so let us assume there are some $i < j$ such that we were not able to find a $v_i v_j$ path which would extend the collection.

Notice that the current collection of paths uses at most $t + \binom{t}{2} \cdot 4 \log^2 m \leq 2t^2 \log^2 m \leq d/(32 \log^2 n)$ vertices. Call this set of vertices V_F.

[4]Technically Komlós and Szemerédi introduced the method in a paper [15] which preceded this one by three weeks and proved only a slightly weaker bound.

Next observe that in $G \setminus V_F$ any set U of size $m/2 \geq |U| \geq 2|V_F| \log n$ still has

$$|N_{G \setminus F_v}(U)| \geq |N_G(U)| - |F_V| \geq |U|/\log n - |F_V| \geq \frac{|U|}{2 \log n}.$$

Finally, notice that

$$|N_{G \setminus F_v}(v_i)| \geq \delta(H) - |F_V| \geq \frac{d}{4 \log n} - |F_v| \geq \frac{d}{8 \log n} \geq 2|V_F| \log n$$

So, for any $\ell \geq 1$ we have

$$|B^{\ell}_{G \setminus F_v}(v_i)| \geq \frac{d}{4 \log n}(1 + 1/(2 \log n))^{\ell - 1}.$$

In particular, we get $|B^{2\log^2 m - 2}_{G \setminus F_v}(v_i)| > m/2$. Repeating for v_j we get $|B^{2\log^2 m - 2}_{G \setminus F_v}(v_j)| > m/2$ and conclude there is a path in $G \setminus F_v$ between v_i and v_j of length at most $4 \log^2 m$, contradicting maximality. $\square$

We note that we only used the vertex expansion property above to keep things simple. We were also rather loose with the logarithms.

Exercise 9.

a) Improve the power of the $\log n$ in the above result as much as you can.

b) Prove Theorem 3.2 under the assumption $d(G) \geq \varepsilon n$ for any $\varepsilon > 0$.

c)* Prove Theorem 3.2 in full.

For part b) it might help to work out Exercise 7. As a warning, part c) is still quite hard, there are several key ideas we did not touch upon. But part b) is useful and you want to develop a sparse variant of the pass to expander lemmas. If you need further hints, looking at section 2 of [18] contains a high level sketch which should help.

3.2 Rainbow cycles

Ramsey theory refers to a large body of results which all roughly speaking say that in any sufficiently large structure you can find a well-organized substructure. In other words, "complete chaos is impossible".

It is natural to ask if the opposite is true, namely can we find a very chaotic substructure.[5] In the classical Ramsey setting if we have an arbitrarily colored graph, Ramsey gives us a monochromatic clique. What would be the most chaotic coloring we can think of?

Definition 3.4. A coloring of a graph is *rainbow* if every edge gets a different color.

Now, of course there is no hope to find a rainbow copy of essentially any graph in an arbitrary colored graph since it might be monochromatic to start with. This leads us to the idea of imposing a slight amount of "chaos" on the coloring of our host graph captured by the following definition.

Definition 3.5. A proper edge coloring of a graph G is an assignment $\chi : E(G) \to C$ such that no two incident edges get the same color.

There are several natural ways of obtaining properly edge colored graphs which makes them useful in studying seemingly unrelated objects. For example every Cayley graph comes with a natural proper edge coloring (where you color by which element gave rise to the edge). Similarly, unit distance graphs in the plane can be (almost) properly edge colored by giving each edge the color corresponding to its direction.

Our original question, first codified by Keevash, Mubayi, Sudakov and Versträete, now becomes given a properly edge-colored graph can we find a rainbow copy of some structure we are interested in it. Note the local to global flavor of this question, namely proper edge coloring imposes a restriction that a star at every vertex should be rainbow and we are trying to see if this local condition guarantees some "global" rainbow copy.

Keevash, Mubayi, Sudakov and Versträete [14], were interested in the natural Turán instance of this question.

Question 3.6. Given a graph H, its n-vertex rainbow Turán number is the minimum number of edges $e = \mathrm{ex}^*(n, H)$ needed to guarantee that any properly edge colored graph G with at least e edges contains a rainbow copy of H.

In particular, motivated by certain applications to additive combinatorics they were interested in how many edges do we need in a properly edge colored graph in order to find a rainbow cycle.

Note that the non-rainbow variant of this question has a very easy answer, namely n edges guarantees a cycle. It turns out that in the rainbow setting if one considers the natural (Cayley) coloring of the hypercube graph, where every edge is colored according to its "direction" (the bit in which its vertices differ). It is easy to see that such a graph does not have a rainbow cycle since as soon as we use an edge of some color, we are "stuck" on the other side of the cut defined by the edges of these color. This shows there is a properly edge colored graph with $\frac{1}{2}n\lfloor \log_2 n \rfloor$ edges without a rainbow cycle. The best known bound improves the leading constant slightly.

Exercise 10. Show that there exists a properly edge-colored graph with average degree at least $(1 + \varepsilon) \log_2 n$, for some $\varepsilon > 0$ which does not contain a rainbow cycle.

[5] Note that this aligns slightly with our general theme of finding a sublinear expander subgraph (which is in some very weak sense "random-like") in any graph.

The question of whether this is tight attracted a lot of attention in the area in recent years and has become the de facto benchmark problem for using sublinear expansion (alongside the origin and development of Homomorphism counting method [13, 12] which has found many uses in its own right). The current state of the art is the following result due to [1].

Theorem 3.7. *There exists a constant $C > 0$ such that every properly edge-colored graph on $n \geq 3$ vertices with average degree at least $C \cdot \log n \cdot \log\log n$ contains a rainbow cycle.*

This theorem found surprising applications going besides combinatorics, namely to coding theory, additive number theory, discrete geometry.

We will prove the following weaker result using at a high level a similar approach.

Theorem 3.8. *Every properly edge-colored graph on $n \geq 3$ vertices with average degree at least $2^{13}(\log n \cdot \log\log n)^2$ contains a rainbow cycle.*

Proof. Let G be our properly edge-colored host graph. Let us pass to a robust sublinear expander H provided by Lemma 2.3. Let $m = |H|$, $d = d(H)$, and $\chi : E(H) \to C$ be the coloring. Lemma 2.3 guarantees

$$d \geq \frac{\log m}{\log n} d(G) \geq C(\log m \log\log m)^2.$$

Let us define $RN^i(v, C')$ to be the set of vertices in H which we can reach by rainbow paths of length at most i starting at v and only using the colors from $C' \subseteq C$.

Let $C(p)$ denote a random subset of C obtained by picking each color in C with probability p, independently between different colors.

Our goal is to show that if we set $r = 100 \log m \log\log m$, then $RN^r(v, C(1/2))$ has size more than $(m+1)/2$. If we can do this, then applying it once with with $C' \sim C(1/2)$ and once with its complement $\bar{C}' \sim C(1/2)$, by a union bound we conclude that there is a vertex $u \neq v$ which is in $RN^r(v, C') \cap RN^r(v, \bar{C}')$. This means there is a closed rainbow walk going from v to u using colors from C' and back from u to v. Passing to a minimum closed subwalk gives us the desired rainbow cycle. So the problem reduces to showing the following key claim.

Claim. For any $v \in V(H)$, if we set $r = 2^{11} \log m \log\log m$

$$\mathbb{P}(|RN^r(v, C(1/2))| > (m+1)/2) > 1/2.$$

Proof. Let $p = \frac{1}{4r} = \frac{1}{2^{13} \log m \log\log m}$.

We use the sprinkling method. Let $C'_i = C(p)$ for each $1 \leq i \leq r$. Then $C' = C'_1 \cup \ldots \cup C'_r \sim C(p')$ where $1 - p' = (1-p)^r \geq 1/2$, so showing $|RN^r(v, C')| > (m+1)/2$ with probability more than $1/2$ is sufficient.

Let us denote by $C_i = C'_1 \cup \ldots \cup C'_i$, and $U_i = RN^i(v, C_i)$ for every $i \leq r$. We will show that for any outcome of $C'_1, \ldots, C'_i$, (which determines C_i and U_i), we have with probability more than $1 - \frac{1}{2r}$ that

$$|U_{i+1}| > \min\{\varepsilon |U_i|/128, (m+1)/2\}, \tag{1}$$

where ε is defined via $|U_i| = m^{1-\varepsilon}$. Armed with this result for each i, we can ensure that in more than half of the outcomes of $C' = C_r$ we have (1) holds for each $1 \le i \le r$ (note that the ε varies with i). So the sets U_i grow essentially as if we looked at the growth of balls in our robust sublinear expander, where we know that in $10 \log n \log \log n$ steps we reach more than half of the graph (as done in Exercise 5). So similarly here, one can show that $|U_r| > (m+1)/2$, as desired.

So it remains to show (1) holds. Let us set $u = |U_i|$, and note that we may assume $u \le 3m/4$ or the desired claim is trivial since $U_i \subseteq U_{i+1}$. This implies $\varepsilon \ge \frac{1}{2 \log m}$. Now we wish to use our robust sublinear expansion property, which guarantees we have at least $\varepsilon u/4$ neighbors of U_i (even after a deletion of $\varepsilon du/4$ edges). However, in U_{i+1} we only count vertices which can be reached by a rainbow path from v and only using colors in C_{i+1}.

To deal with the first issue, for any $w \in U_i$ there is a rainbow path of length at most i joining it to v, let $C(w)$ denote the set of up to i colors used on this path. This means that any vertex a adjacent to w, along an edge with color in $C'_{i+1} \setminus C(w)$ implies $a \in U_{i+1}$. So let F consist of all edges incident to vertices w in U_i using the colors in $C(w)$. Observe that $|F| \le r|U| \le \varepsilon d|U|/8$. So this means we retain essentially the same robust sublinear expansion properties in $G - F$.

Turning to the second issue, we wish to retain the same level of vertex expansion when we sample the colors with probability p. Let us consider now the bipartite graph H' with parts $A = U_i$, $B = N_{G-F}(U_i)$ and inheriting all the edges from $G - F$. We will distinguish two cases here.

Case 1. Suppose there are at least $u\varepsilon/8$ vertices in B with degree at least d in H'. In this case, when we sample the colors with probability p the number of edges incident to any such vertex dominates $\mathrm{Bin}(d, p)$, so the chance none of the edges survives is at most $(1-p)^d \le e^{-pd} \le m^{-2}$. So a union bound allows us to ensure that with probability at least $1 - 1/m$ all such vertices survive sampling and hence get added to U_{i+1}. So we get $u\varepsilon/8$ new vertices with probability $1 - 1/m$, as desired.

Case 2. There are at most $u\varepsilon/8$ vertices in B with degree at least d in H'. In this case we simply delete them from B and are left with a subgraph H'' which still has the property that even deleting $\varepsilon du/8$ edges at least $\varepsilon u/8$ vertices in B retain an edge to A. Now, delete any vertices from A which have degree at most $\varepsilon d/16$. This deletes at most $\varepsilon du/16$ edges so our new bipartite subgraph H''' has the property that even after deleting $\varepsilon du/16$ edges at least $\varepsilon u/8$ vertices in B retain an edge to A. Let $d_1, \ldots, d_{u'}$ denote the degrees of vertices from A still in H'''. Now, we call a vertex in A bad if after sampling, it had less than $p\varepsilon d/32$ edges. By a Chernoff bound the probability that a vertex is bad is at most $p' \le e^{-\varepsilon dp/128} \le 1/\log^2 m$. Now let

$$Y := \sum_{i \text{ is bad}} d_i$$

be a random variable counting the sum of the degrees of bad vertices. Then,

$$\mathbb{E}Y = \sum_i \mathbb{P}(i \text{ is bad}) \cdot d_i \le \sum_i d_i / \log^2 n.$$

By Markov's inequality we have that the probability that $Y \ge \sum_i d_i/2$ is at most $\mathbb{E}Y/(\sum_i d_i/2) \le 2/\log^2 m \le 1/(4r)$. So with probability at least $1 - 1/4r$ we have $\sum_{i \text{ is not bad}} d_i \ge \frac{1}{2} \sum_i d_i \ge \varepsilon du/32$. This implies that the number of edges that survive the sampling is at least $p\varepsilon du/64$. On the other hand, we can ensure that every vertex on the right has maximum degree at most $2pd$ with probability at least $1 - m \cdot e^{-3pd} \ge 1 - 1/(4r)$. When both these outcomes occur we get at least $\varepsilon u/128$ new vertices and are done. $\square$

As described above, the claim implies the result and completes the proof. $\square$

A keen eyed reader might have noticed that the only place in the above argument where we really needed the additional $\log\log n$ factor was towards the end of Case 2, analysis where we needed it to guarantee we have few bad vertices with high enough probability. Without this extra factor we would only get a small constant upper bound on the probability that each of our r stages succeeds. With a bit more care one can avoid paying this additional factor in the above argument by adjusting the thresholds dyadically (and notice that we are only in trouble when ε is close to $1/\log n$).

Exercise 11. Remove one of the $\log\log n$ factors in the above argument.

Let us also mention, that there are multiple places where we really needed the stronger degree requirement compared to $\log n \log\log n$ of Theorem 3.7

Exercise 12. Show that for any fixed constant t, there is $C_t > 0$ such that in any properly edge colored graph with average degree at least $C_t(\log n \log\log n)^2$ one can find a rainbow topological clique of order t.

Note that $t = 3$ case is the above theorem. This is actually the best known result for this problem and whether one can remove the $\log n$ factor is an open problem. In fact, even the $\log^2\log n$ factor is slightly better than best known bounds form [22].

4 Decomposition

In this section, we will prove that not only can we find a single sublinear expander with similar relative degree but that actually we can decompose a graph into very few of them.

We will work with the following slightly looser definition, where we do not have variable expansion rate depending on the subset size. This is mostly for convenience and since we aren't able to decompose in quite as strong expanders as we have seen in Definition 2.2

Definition 4.1. An n-vertex graph G is an (ε, s)-expander if, for every $U \subseteq V(G)$ and $F \subseteq E(G)$ with $1 \le |U| \le \frac{2}{3}n$ and $|F| \le s|U|$, we have

$$|N_{G-F}(U)| \ge \varepsilon|U|. \tag{2}$$

We note that the ε we are going to use is going to be $\Omega(1/\log^2 n)$ while s is in effect going to be around $\Omega(d(G)/\log n)$. There is some room to do better if one allows more overlap between the expanders and one can in fact ensure better expansion for smaller sets even while keeping the same tradeoffs.

Lemma 4.2. *Given an n-vertex graph G, and a non-negative integer s we can delete up to $4sn\log n$ edges from G so that the remaining edges may be partitioned into graphs $G_1, \dots, G_r$ such that*

- $\sum_{i=1}^{r} |G_i| \le 2n$ *and*
- *each G_i is a $\left(\frac{1}{32\log^2|G_i|}, s\right)$-expander.*

Note that technically the above lemma is an almost decomposition lemma since we get a number of edges as leftover. However, we are allowed to take $s = 0$ in which case we do get a full decomposition into $\frac{1}{32\log^2 n}$-expanders, albeit with no robustness. Notice that the above lemma applies (with $s = 0$) for graphs with only constant degree, and in fact provides us with a sublinear expander subgraph with degree at least $d/2$.

Exercise 13. Show that any n-vertex graph G with average degree d with contains a $\frac{1}{32\log^2 n}$-expander with average degree at least $d/2$.

The proof follows the "sparse cutting" method (a close relative to potential maximization method we used for Lemma 2.3). The basic ideas is that so long as we can find a "sparse" cut, meaning it breaks the expansion condition, then we cut the graph along it into two pieces with only the few neighbors of one side added to the non-expanding side. We then repeat the argument on both subgraphs. The argument only stops once the graph under consideration is a desired expander (and we no longer have a suitable cut to continue). The analysis of this is a bit technical and should (in part) be reminiscent of the one we have done in the proof of Lemma 2.3.

Proof of Lemma 4.2. Let $\varepsilon_m = \frac{1}{32\log^2 m}$ for every m. The proof is by induction on n, under the stronger condition that the graphs $G_1, \ldots, G_r$ in the partition satisfy $\sum_{i=1}^r |G_i| \le 2n - \frac{2n}{2+\log n}$. Since $2n - \frac{2n}{2+\log n} \ge n$, and any 1-vertex graph G is trivially a (ε_n, s)-expander, the lemma holds for $n = 1$ with $G_1 = G$. Let us then assume $n \ge 2$ and that the claim holds for all graphs with at most $n - 1$ vertices.

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

$$n_1 + n_2 = |G_1| + |G_2| = |G| + |N_{G-F}(U)| < n + \varepsilon_n \cdot |U| \le n + \varepsilon_n \cdot n_1. \tag{3}$$

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

$$n_1 \le |U| + \varepsilon_n \cdot |U| \le 2n/3 + n/32 \le 3n/4 < n, \tag{4}$$

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

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

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

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

$$\begin{aligned}
\frac{1}{s}(|F| + |E_1| + |E_2|) \leq |U| + 4n_1 \log n_1 + 4n_2 \log n_2 &\leq n_1 + 4n_1\left(\log n - \frac{2}{5}\right) + 4n_2 \log n \\
&= 4(n_1+n_2)\log n - \frac{3}{5}n_1 \overset{(3)}{\leq} 4\left(n + \frac{n_1}{32\log^2 n}\right)\log n - \frac{3}{5}n_1 \\
&\leq 4n\log n. \qquad (5)
\end{aligned}$$

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

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

so that

$$\begin{aligned}
\sum_{i=1}^{2}\sum_{j=1}^{r_i} |G_{i,j}| &\leq 2n_1 + 2n_2 - \frac{2n_1}{2+\log n_1} - \frac{2n_2}{2+\log n_2} \overset{(6)}{<} (n_1+n_2)\left(2 - \frac{2}{2+\log n}\right) - \frac{n_1}{10\log^2 n} \\
&\overset{(3)}{\leq} \left(n + \frac{n_1}{32\log^2 n}\right)\left(2 - \frac{2}{2+\log n}\right) - \frac{n_1}{10\log^2 n} \leq 2n - \frac{2n}{2+\log n}.
\end{aligned}$$

In combination with (5), this shows that G has the required decomposition, completing the inductive step and hence the proof. $\square$

4.1 An application to Erdős-Gallai Cycle decomposition conjecture

The following beautiful conjecture was posed in the 1960's by Erdős and Gallai [10] and reiterated by him in a number of problem papers, as well as by multiple other authors over the years.

Conjecture 4.3. *Any n-vertex graph can be decomposed into $O(n)$ cycles and edges*[6].

In large part the interest stems from the fact that similar linear bounds are known for a number of similar path/cycle graph decomposition questions. One particularly nice and very powerful result, which we will use is due to Lovász [20] from 1968.

Theorem 4.4. *Every n-vertex graph can be decomposed into at most $n/2$ paths and cycles.*

The proof of Lovász, while characteristically beautiful is a delicate induction so unfortunately doesn't seem very amenable to modification. Nonetheless, this result has found plenty of applications in a black box manner. We will actually use it through the following corollary which gives us a tiny, but crucial, amount of control on the endpoints of the paths in the decomposition.

Corollary 4.5. *Every n-vertex graph G can be decomposed into paths so that each vertex of G is an endvertex of at most two paths in the decomposition.*

Proof. Form a graph G' from G by adding a new vertex v_0 and an edge from v_0 to each vertex $v \in V(G)$ for which $d_G(v)$ is even. By Theorem 4.4, there is a collection $\mathcal{C}$ of at most $\frac{n+1}{2}$ cycles and paths which decomposes G'. Note that each vertex $v \in V(G)$ has odd degree in G', and therefore must be an endvertex of an odd number of paths in $\mathcal{C}$, and thus, in particular, must be an endvertex of some path in $\mathcal{C}$. As the paths in $\mathcal{C}$ together have at most $n+1$ endvertices, each vertex in G is the endvertex of at most 1 path in $\mathcal{C}$, as otherwise there would need to be $n-1+3 > n+1$ endvertices. Note that, furthermore, as there must be at least n endvertices together for the paths in $\mathcal{C}$, we must have that $\mathcal{C}$ contains at least $\frac{n}{2}$ paths. Thus, as $|\mathcal{C}| \le \frac{n+1}{2}$, $\mathcal{C}$ is in fact a collection consisting only of paths, with no cycles.

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

I am not aware (and would be very interested in it!) of a significantly different proof of Theorem 4.4 compared to the original one of Lovász, even with $O(n)$ in place of $n/2$. There is however a lovely short proof of the covering version of this result. (The one I have in mind is due to [4].)

Exercise 14. Any n-vertex graph G contains $O(n)$ paths which in their union contain all the edges.

Hint. Consider a longest path, fix one of its endpoints v and consider the subset of its vertices which can be rotated into being the other endpoint (so fixing v as an endpoint throughout). Use induction to decompose various pieces of the picture and the long path to help you cover the edges in one of the parts.

Let us return to the Erdős-Gallai Conjecture. First note that it is tight since any spanning tree requires at least $n-1$ edges to be covered. The best known construction due to Erdős, in one of the papers in which he reiterated the problem [9], shows one needs at least $(3/2 - o(1))n$.

[6] In the original paper Erdős and Gallai viewed edges as cycles of length two.

Exercise 15. Show that there exists a graph in which any decomposition into cyles and edges uses at least $(3/2 - o(1))n$ cycles or edges.

The state of the art upper bound is $O(n \log^* n)$ cycles and edges due to [6] (see also [5] for a version with more focus on connections to computer science).

Theorem 4.6. *Any n-vertex graph can be decomposed into $O(n \log^* n)$ cycles and edges.*

We will show a weaker bound of $O(n \log \log n)$ due to Conlon, Fox, and Sudakov [7], with some of the ideas from [5] scattered in. This improved significantly upon the following result, dating back to the original paper.

Exercise 16. Any n-vertex graph can be decomposed into $O(n \log n)$ cycles and edges.

Theorem 4.6 is a corollary of the following asymmetric result.

Theorem 4.7. *Any n-vertex graph can be decomposed into $O(n)$ cycles and $n \log^{O(1)} n$ edges.*

Which is actually done via the following intermediate result.

Theorem 4.8. *Any n-vertex graph with average degree d can be decomposed into $O(n)$ cycles and $n \log^{O(1)} d$ edges.*

Proof that Theorem 4.8 $\implies$ Theorem 4.6. Suppose our initial graph G had n vertices and average degree d. Apply Theorem 4.8 to it to decompose it into $O(n)$ cycles and $n \log^{O(1)} d$ edges. Delete the cycles from the graph and notice that the leftover graph has average degree at most $\log^{O(1)} d$ (compared to d we started from). Now iterate $O(\log^* d)$ many times until the average degree becomes a constant. Since we took out $O(n)$ cycles in each of $O(\log^* d)$ many iterations we get a decomposition into $O(n \log^* d)$ cycles and $O(n)$ edges (belonging to the remainder subgraph of constant average degree). $\square$

Proof that Theorem 4.7 $\implies$ Theorem 4.8. Suppose our initial graph G had n vertices and average degree d. Repeatedly take out cycles of length at least d so long as one exists[7] Since G has $nd/2$ edges, by doing so we remove at most $n/2$ cycles. The remaining graph G' contains no cycle of length d or longer.

Now apply our decomposition lemma Lemma 4.2 with $s = 0$ to G' to split it into $G_1, \ldots, G_r$ with $|G_1| + \ldots + |G_r| \le 2n$ and G_i being a $\frac{1}{32 \log^2 |G_i|}$-expander. If $|G_i| \ge 10d \log^4 d$ Lemma 2.1 implies there is a cycle in G_i (and hence in G') of length at least d, which we ensured is not the case. So we have $|G_i| \le 10d \log^4 d$ for each i. Applying Theorem 4.7 to each G_i we obtain a decomposition into

$$\sum_i O(G_i) \le O\Big(\sum |G_i|\Big) \le O(n)$$

cycles and

$$\sum_i |G_i| \log^{O(1)} |G_i| \le \sum_i |G_i| \log^{O(1)}(10d \log^4 d) \le \log^{O(1)} d \cdot \sum_i |G_i| \le \log^{O(1)} d \cdot n,$$

edges, as desired. $\square$

[7] Without updating d to be the current average degree, d here is always the *original* average degree.

So the key part is to prove Theorem 4.7. The analogue for the $O(n \log\log n)$ result we aim to prove is to show

Theorem 4.9. *Any n-vertex graph can be decomposed into $O(n)$ cycles and $O(n^{2-\varepsilon})$ edges.*

Exercise 17. Verify that Theorem 4.9 implies $O(n \log\log n)$ bound in Conjecture 4.3.

Proof of Theorem 4.9. By applying Lemma 4.2 with $s = n^{2-\varepsilon}/\log n$ we get a decomposition into $O(n^{2-\varepsilon})$ edges and expanders $G_1, \ldots, G_t$ where each G_i is a $\left(\frac{1}{32\log^2|G_i|}, s\right)$-expander and $|G_1| + \ldots + |G_t| \leq 2n$. So provided we can prove that the theorem holds for each G_i we can conclude it for G itself. Let us hence w.l.o.g. assume $t = 1$ and that G itself was a $\left(\frac{1}{32\log^2 n}, s\right)$-expander to avoid the need to redefine the parameters.

Let $V = V(G)$ and let $U = V(p)$ where $p = n^{-\varepsilon}$. Note that with high probability $|U| \leq 2n^{1-\varepsilon}$, so any of the at most $O(n^{2-\varepsilon})$ unused edges touching U may simply be taken as part of the decomposition at the end.

We will wish to ensure a certain key "path-connectivity" property of our random subset U but let us first explain how it is going to be used. Once we sampled our "suitable" U we apply the corollary to Lovász' Theorem (Corollary 4.5) to decompose $G[V \setminus U]$ into paths in such a way that every vertex is an endpoint of at most two paths (note that this implies there are at most n paths in the decomposition). Note that these paths do not use any vertices from U. We will now use our carefully chosen property of U to connect the endpoints of all these paths by short paths with all internal vertices in U.

The ideal way of doing so would be to ensure U inherits some expansion properties. However, as in the rainbow cycles problem we simply have too many subsets we wish to expand compared to the strength of the expansion we are working with. There are two ways around this, one is to apply a similar sprinkling approach as we used in the rainbow cycles problem (here though we are sampling vertices and edges, rather than colors) to show that even if one samples edges with probability $1/n$ one can still with sufficiently high probability ensure that starting from any vertex of G we can reach more than half of U by short paths. While there are some differences and care needed in implementing this sprinkling technique, we leave this as an exercise after the proof and will follow the other approach of so called-robust neighborhoods.

Let $RoN^0(v,\delta) := \{v\}$, $RoN^1(v,\delta) := \{v\} \cup N_G(v)$ and for $\ell \geq 2$, let $RoN^\ell(v,\delta)$ be the set of vertices which have at least δ neighbors in $RoN^{\ell-1}(v)$. Notice that these are nested by definition, and that they all have size at least $|RoN^1(v,\delta)| = d(v)+1 \geq n^{1-\varepsilon}/\log n$, since this minimum degree bound in G is guaranteed by the expansion property (applied to single vertex sets). Let us now set $\delta = n^{1-2\varepsilon}/\log^2 n$. Furthermore, notice that if we let F be the set of edges incident to vertices with less than δ neighbors in $RoN^{\ell-1}(v,\delta)$, then $|F| \leq n\delta = n^{2-2\varepsilon}/\log^2 n \leq s \cdot |RoN^{\ell-1}(v,\delta)|$, so our expansion property guarantees that there are at least $\frac{|RoN^{\ell-1}(v,\delta)|}{32\log^2 n}$ new vertices in $RoN^\ell(v,\delta)$.

Notice that under vertex sampling by probability $p = n^{-\varepsilon}$. For every vertex u in $RoN^\ell(v,\delta) \setminus RoN^{\ell-1}(v,\delta)$, with $\ell \geq 2$, we expect at least $\delta p = n^{1-3\varepsilon}/\log^2 n$ of its neighbors in $RoN^{\ell-1}(v,\delta)$ to be sampled. So by a Chernoff bound, we can ensure that at least $\delta p/2$ of them do with probability at least $1 - e^{-p\delta/8} \geq 1 - 1/n^3$, since $\delta p/8 \geq 3\log n$. By a union bound over v and u we conclude that we can ensure this with high probability for any v and u. We fix such an outcome for U, in addition to having $|U| \leq 2np$, and every vertex $v \in G$ having degree at least $\delta p/2 = n^{1-2\varepsilon}/(2\log n)$ towards U.

We now claim that given any set F of up to $n\log^3 n$ edges inside $G[U]$, we can find a path with all internal vertices inside U, of length at most $\log^3 n$ in $G - F$, linking arbitrary two vertices of $V \setminus U$.

Let B be the set of vertices in U incident to more than $\sqrt{n}$ edges. To see this, let us fix $v, u \in V \setminus U$ and let $U_1 = N_G(u) \cap U$, and for $\ell \geq 2$ we define $U_\ell = N_{G-F}(U_{\ell-1}) \cap U$.

$\square$

References

[1] N. Alon, M. Bucić, L. Sauermann, D. Zakharov, and O. Zamir, *Essentially tight bounds for rainbow cycles in proper edge-colourings*, Proc. Lond. Math. Soc. (3) **130** (2025), no. 4, Paper No. e70044, 37.

[2] I. Ben-Eliezer, M. Krivelevich, and B. Sudakov, *Long cycles in subgraphs of (pseudo)random directed graphs*, J. Graph Theory **70** (2012), no. 3, 284–296.

[3] B. Bollobás and A. Thomason, *Proof of a conjecture of Mader, Erdős and Hajnal on topological complete subgraphs*, European J. Combin. **19** (1998), no. 8, 883–887.

[4] M. Bonamy, F. Botler, F. c. Dross, T. Naia, and J. Skokan, *Separating the edges of a graph by a linear number of paths*, Adv. Comb. (2023), Paper No. 6, 7.

[5] M. Bucić and R. Montgomery, *Towards the Erdős-Gallai Cycle Decomposition Conjecture*, Proceedings of the 55th Annual ACM Symposium on Theory of Computing (New York, NY, USA), STOC 2023, Association for Computing Machinery, 2023, p. 839–852.

[6] M. Bucić and R. Montgomery, *Towards the Erdos-Gallai cycle decomposition conjecture*, Adv. Math. **437** (2024), Paper No. 109434, 40.

[7] D. Conlon, J. Fox, and B. Sudakov, *Cycle packing*, Random Struct. Algorithms **45** (2014), no. 4, 608–626.

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

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

[10] P. Erdős, A. W. Goodman, and L. Pósa, *The representation of a graph by set intersections*, Can. J. Math. **18** (1966), 106–112.

[11] S. Hoory, N. Linial, and A. Wigderson, *Expander graphs and their applications*, Bull. Amer. Math. Soc. (N.S.) **43** (2006), no. 4, 439–561.

[12] O. Janzer, *Rainbow turán number of even cycles, repeated patterns and blow-ups of cycles*, Israel Journal of Mathematics (2022), 1–28.

[13] O. Janzer and B. Sudakov, *On the Turán number of the hypercube*, preprint arXiv:2211.02015 (2022).

[14] P. Keevash, D. Mubayi, B. Sudakov, and J. Verstraëte, *Rainbow Turán problems*, Combin. Probab. Comput. **16** (2007), no. 1, 109–126.

[15] J. Komlós and E. Szemerédi, *Topological cliques in graphs*, Combin. Probab. Comput. **3** (1994), no. 2, 247–256.

[16] J. Komlós and E. Szemerédi, *Topological cliques in graphs II*, Combinatorics, Probability and Computing **5** (1996), no. 1, 79–90.

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

[18] S. Letzter, *Sublinear expanders and their applications*, Surveys in combinatorics 2024, London Math. Soc. Lecture Note Ser., vol. 493, Cambridge Univ. Press, Cambridge, 2024, pp. 89–130.

[19] S. Letzter, A. Methuku, and B. Sudakov, *Nearly hamilton cycles in sublinear expanders, and applications*, preprint arXiv:2503.07147 (2025).

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

[21] A. Lubotzky, *Expander graphs in pure and applied mathematics*, Bull. Amer. Math. Soc. (N.S.) **49** (2012), no. 1, 113–162.

[22] Y. Wang, *Rainbow clique subdivisions*, preprint arXiv:2204.08804 (2022).