Approximate path decompositions of regular graphs

Richard Montgomery*   Alp Müyesser†   Alexey Pokrovskiy†   Benny Sudakov‡

Abstract

We show that the edges of any d-regular graph can be almost decomposed into paths of length roughly d, giving an approximate solution to a problem of Kotzig from 1957. Along the way, we show that almost all of the vertices of a d-regular graph can be partitioned into $n/(d+1)$ paths, asymptotically confirming a conjecture of Magnant and Martin from 2009.

1 Introduction

A typical decomposition problem asks if the edges of some graph can be partitioned into copies of another graph. The origins of this area can be traced back to Euler who asked the following question in 1782: for which n does the balanced complete 4-partite graph $K_{n,n,n,n}$ decompose into copies of the complete graph on 4 vertices (i.e. copies of K_4)? Euler's problem is customarily phrased using the language of Latin squares and transversals, but the two formulations are equivalent. In 1847, Kirkman studied decompositions of a complete graph into triangles (copies of K_3). Such decompositions are also referred to as *Steiner triple systems*. In 1882, Walecki studied decompositions of complete graphs into Hamilton paths and cycles. All of this work, and other early approaches to graph decomposition problems, used exclusively algebraic or constructive techniques. In contrast, there have been many applications of the probabilistic method [2] to graph decomposition problems in the last few decades, and several major problems have recently been resolved. Highlights include, but are certainly not limited to, Keevash's resolution [25] of the Existence Conjecture for designs, which is a far reaching generalisation of the work of Kirkman (see also [20]), and the proof for large n of Ringel's conjecture [31, 25] which states that K_{2n+1} decomposes into copies of any n-edge tree. We refer the reader to [10, 21] for two recent surveys of the area.

The common denominator in these recent advances is that they concern decompositions of highly dense structures, such as complete graphs. There is also a large body of conjectures on the decomposition of sparse graphs, where there has been much less progress as existing proof techniques like those involving Szemerédi's regularity lemma only apply to dense graphs. Perhaps the most famous conjecture about decomposing sparse graphs is the linear arboricity conjecture of Akiyama, Exoo, and Harary [1] from 1980, which says that every graph with maximum degree Δ can be decomposed into $\lceil (\Delta + 1)/2 \rceil$ path forests. The best general bound on this is by Lang and Postle [28] who proved that such a graph always has a decomposition into $\Delta/2 + 3\sqrt{\Delta} \log^4 \Delta$ path forests. For decompositions into paths rather than path forests, there is a conjecture of Gallai which predicts that every connected n-vertex graph can be decomposed into at most $(n + 1)/2$ paths (see [29]). The best general result for this is that every such graph can be decomposed into at most $2n/3$ paths, proved in [14, 33]. What if we require all the paths to be of the same length d? Here some extra conditions are required on the host graph G, since at the very least we need d to divide $e(G)$. Botler, Mota, Oshiro, and Wakabayashi [7] showed that such a

P_d-decomposition exists if, furthermore, the graph is $f(d)$-edge-connected, for some function f, a result motivated by a now-solved conjecture of Barát and Thomassen (see [4]). A natural way to ensure d divides $e(G)$ when d is odd is to require G to be d-regular. In 1957, Kotzig [27] proved that a 3-regular graph has a P_3-decomposition if and only if G has a perfect matching, and raised the following general question for larger d.

Problem 1.1. *When d is odd, which d-regular graphs can be decomposed into paths of length d?*

In 1990, Bondy [5] proposed a further extension of this problem without the regularity assumption. There are very few general results known about these problems. Some progress has been made on a related problem of Favaron, Genest, and Koudier [16], who conjectured that a d-regular graph G (for odd d) decomposes into copies of P_d if G contains a perfect matching. See [9, 8, 6] for some partial progress on this conjecture concerning the cases when d is small, or when the host graph G is assumed to have high girth.

As obtaining a precise decomposition appears to be so difficult, it is of independent interest to show that the existing conjectures hold true approximately, in part as it may eventually be used as a stepping stone towards a full resolution. For example, even the approximate version of the Existence Conjecture for designs was a longstanding open problem of Erdős and Hanani [15] before it was resolved through the influential work of Rödl [32]. This approximate solution plays a key component in the work of Keevash [25] resolving the full conjecture.

In this paper, we address a relaxed version of Kotzig's problem by showing that almost all of the edges of a d-regular graph can be decomposed into paths of length roughly d.

Theorem 1.2. *For every $\varepsilon > 0$, there exists $d_0 \in \mathbb{N}$ such that the following holds for all $d \geq d_0$. If G is a d-regular n-vertex graph, then all but at most $\varepsilon n d$ edges of G can be decomposed into paths of length $\lceil (1 - \varepsilon)d \rceil$.*

Constructing large edge-disjoint path forests plays a large part in our methods. However, note that Theorem 1.2 cannot be strengthened to require that the paths in the decomposition can be arranged into $\lceil (d + 1)/2 \rceil$ path forests (as in the linear arboricity conjecture), as seen by considering the vertex-disjoint union of d-regular graphs of order $3d/2$. Our methods, though, can be used to make progress on a problem related to the linear arboricity conjecture, as follows. Note that, when d is odd, if the linear arboricity conjecture is true for d-regular graphs then one of the path forests in such a decomposition would have at most $\frac{nd}{2} / \frac{d+1}{2} = n - \frac{n}{d+1}$ edges, and therefore (adding isolated vertices if necessary) form a spanning path forest with at most $\frac{n}{d+1}$ paths. Though naturally much weaker than the linear arboricity conjecture, showing even the existence of such a path forest in a d-regular graph appears very difficult. This (for odd or even d) is the topic of the following conjecture by Magnant and Martin [30] from 2009 (which, as per [18], also has an interesting connection to tour length problems).

Conjecture 1.3. *The vertices of every d-regular n-vertex graph can be partitioned into $\lfloor n/(d+1) \rfloor$ paths.*

The disjoint union of complete graphs on $d + 1$ vertices shows that this conjecture would be optimal. It was confirmed for all $d \leq 5$ by Magnant and Martin [30], and for $d = 6$ by Feige and Fuchs [17]. It has also recently been confirmed in the dense case ($d = \Omega(n)$) by Gruslys and Letzter [22], but in general even the weaker conjecture by Feige and Fuchs [17] that every d-regular n-vertex graph can be partitioned into $O(n/(d+1))$ paths remains wide open. We prove the following approximate version of Conjecture 1.3.

Theorem 1.4. *For every $\varepsilon > 0$, there exists $d_0 \in \mathbb{N}$ such that the following holds for all $d \geq d_0$. Let G be a d-regular n-vertex graph. Then, all but at most εn of the vertices of G can be partitioned into at most $n/(d+1)$ paths.*

If we consider decomposing approximately the edges of a d-regular graph G into copies of P_ℓ starting with small ℓ, then the extremal examples for why we cannot increase ℓ beyond d all contain small dense subgraphs (e.g. the cliques in the disjoint union of copies of K_{d+1}), which we refer to as 'dense spots'. In d-regular graphs without these dense spots it would be relatively easy (using techniques developed for the linear arboricity conjecture) to decompose almost all of the edges of G into paths with length $\lceil (1 - \varepsilon)d \rceil$ by using the local lemma to partition $V(G)$ into sets between which we can find many edge-disjoint matchings which combine to form many paths,

and then iteratively connect some of these paths together using a set of reserved vertices to get longer and longer paths (as we do later in parts of our proof). The main idea in our proof is to set aside the dense spots in a d-regular graph and then show that if the preceding approach goes wrong then we will be able to connect paths which we cannot lengthen into dense spots, so that we can then decompose approximately the dense spots along with their attached paths into paths with length $\lceil (1-\varepsilon)d \rceil$. To do so, we will need the set of reserved vertices to 'sample' each dense spot appropriately in a precise manner, and how we manage this is the key technical novelty in our methods. However, there are many further challenges in developing these initial ideas into a proof of Theorem 1.2, and even approximately decomposing the dense spots with connected paths requires new ideas. Thus, in the next section, after outlining our notation, we give a detailed sketch of our methods. Following this, we outline the rest of the paper. After our proofs, we make some concluding remarks in Section 6, including on generalisations of path decomposition conjectures to other trees.

2 Preliminaries

Following an overview of our notation in Section 2.1, in Section 2.2 we give a sketch of the proof before outlining the rest of the paper. For ease of notation, we will assume throughout that d is even, for example constructing paths forests $\mathcal{P}_i$, $i \in [d/2]$. The case where d is odd follows almost identically as the proofs apply to $(d-1)$-regular graphs without any significant alteration.

2.1 Notation

For a graph G, let $V(G)$ and $E(G)$ denote the vertex set and edge set of G, respectively, and let $|G| = |V(G)|$. For a vertex $v \in V(G)$, let $N_G(v)$ denote the neighbourhood of v in G and, for a subset $U \subset V(G)$, let $N_G(v, U) = N_G(v) \cap U$. We let $d_G(v) = |N_G(v)|$ denote the degree of v and write $d_G(v, U) = |N_G(v, U)|$ for the degree of v into a subset $U \subset V(G)$. As we do elsewhere, we omit G from the subscript whenever there is no risk of confusion. The minimum and maximum degree of G is denoted by $\delta(G)$ and $\Delta(G)$ respectively. For a subset $U \subseteq V(G)$, let $G[U]$ denote the graph induced by U and, given a subset $U' \subset V(G) \setminus U$, let $G[U, U']$ denote the bipartite graph with bipartition $U \cup U'$ and edges of the form $uu' \in E(G)$ with $u \in U$ and $u' \in U'$. Given a subset $U \subset V(G)$, we let $G - U$ be the graph $G[V(G) \setminus U]$, which is G with the vertices in U removed, and use similar natural and common other notation. When $\mathcal{F}$ is a collection of graphs, $V(\mathcal{F})$ and $E(\mathcal{F})$ will be the set of vertices and edges used in some graph in $\mathcal{F}$ respectively.

For each positive integer n, we let $[n] = \{1, \ldots, n\}$. Given real numbers a, b, c, we write $a = b \pm c$ to denote that $b - c \leq a \leq b + c$. We say a graph G is $(1 \pm \gamma)d$-regular if $d_G(v) = (1 \pm \gamma)d$ for each $v \in V(G)$. We use standard notation for "hierarchies" of constants, writing $x \ll y$ to mean that there is a non-decreasing function $f : (0, 1] \to (0, 1]$ such that all the relevant subsequent statements hold for $x \leq f(y)$. Hierarchies with multiple constants are defined similarly, with the functions chosen from right to left, for example, for $x \ll y \ll z$. We omit rounding signs where they are not crucial.

In this paper, we will often add paths to connect paths in a path forest. Given two path forests $\mathcal{P}$ and $\mathcal{Q}$ we use $\mathcal{P} + \mathcal{Q}$ to denote the graph with vertex set $V(\mathcal{P}) \cup V(\mathcal{Q})$ and edge set $E(\mathcal{P}) \cup E(\mathcal{Q})$. Furthermore, we exclusively use this when the resulting graph is also a path forest, and each path in $\mathcal{Q}$ has at least one endpoint among the endpoints of $\mathcal{P}$.

2.2 Proof sketch

Let G be an n-vertex d-regular graph and let $1/d \ll \varepsilon$ and $\ell \leq (1-\varepsilon)d$. Suppose we wish to find edge-disjoint paths of length ℓ in G covering all but at most εnd edges (noting that if we can do this with $\ell = (1-\varepsilon)d$ then we will have proved Theorem 1.2). The following natural strategy, developed in part for attacking the

linear arboricity conjecture (see, for example, Section 1.1 in [19] which builds on [3, 26]), can accomplish this if $\ell \leq d^{1/5-o(1)}$. Take a random partition of $V(G)$ into roughly equal sets $A_1 \cup \ldots \cup A_{\ell+1}$. Using the local lemma (see Section 2.4.2), show that, with positive probability, each bipartite subgraph $G[A_i, A_j]$ with $i \neq j$ is almost $(d/(\ell+1))$-regular. Taking such a partition, then, decompose each $G[A_i, A_j]$ into k_0 matchings $M_{ij,k}$, $k \in [k_0]$, where k_0 is only slightly more than $d/(\ell+1)$ (using, say, Vizing's theorem). Decompose the complete graph $K_{\ell+1}$ into paths with length ℓ (it is well-known that this is possible whenever $\ell+1$ is even), and then, for each path and each k, take the corresponding matchings $M_{ij,k}$ for the edges ij of the path, and combine them to find many paths of length ℓ in G (along with some shorter paths where the matchings do not align exactly). Carried out carefully, this approach will give paths of length ℓ covering all but at most $\varepsilon n d$ edges, as long as $\ell \leq d^{1/5-o(1)}$ so that not too many edges are lost from the potentially misaligned matchings. The constant $1/5$ here is the natural barrier for our implementation of this method later, but more generally such an approach encounters a strong natural barrier for $\ell \approx d^{1/2}$ (see [28]).

Of course, we wish to have an approach that works for ℓ up to $(1-\varepsilon)d$. A tempting route forward is to first take aside a small 'sample' set X of $\approx pn$ vertices (likely chosen via the local lemma), with $p \ll \varepsilon$, where $G - X$ is close to regular, and then decompose almost all of $G - X$ into paths with length $d^{1/5-o(1)}$. Then, we could use vertices from X to iteratively join these paths into longer paths. The above partitioning and matching argument naturally produces edge-disjoint path forests $\mathcal{P}_1, \ldots, \mathcal{P}_{d/2}$ covering most of the edges of $G - X$ using paths with length $d^{1/5-o(1)}$. Given any set of $(1+\varepsilon)n/d$ paths in one of these path forests, if we take one endpoint from each path to form the set Y, then, as the neighbours of the vertices in Y must overlap, we could hope that some of the overlap will be sampled into X so that we can join up two of the paths using a single vertex in X, and, perhaps, even to do this iteratively until the path forest contains at most $(1+\varepsilon)n/d$ paths.

By making sure that the initial path forests do not together use any vertex as an endpoint too much, and that no path forest counts too many neighbours of a vertex among its endpoints, and by further dividing X into several subsets to exhaust in turn while connecting paths, this can be made to work (see Sections 3.1 and 3.2), and, furthermore, even, used to prove Theorem 1.4 (see Section 3.3). However, this approach will approximately decompose the n-vertex d-regular graph G into paths with *average* length $(1-\varepsilon)d$ (see Lemma 2.3), rather than paths of length $(1-\varepsilon)d$, so really is only the starting point for our proof of Theorem 1.2.

To go further, we consider what might stop us from joining up more paths in our path forests via X. After all, if we could continue to connect paths together until our paths forests mostly had paths with length at least Kd, for some $1/K \ll \varepsilon$, then each such path can be further decomposed into paths with length $(1-\varepsilon)d$ and at most $d-1$ other edges, where these other edges will then in total be a small portion of the edges of the graph G so that we will not need to decompose them. The disjoint union of copies of K_{d+1} demonstrates that hoping to always get paths of this length, Kd, is fanciful. However, the presence of small dense subgraphs (which we shall call 'dense spots') turns out to be the only thing that can stop this from working. Indeed, suppose we had $r := n/Kd$ vertex-disjoint paths $P_1, \ldots, P_r$ in $G - X$, and could not find a short path (say of length at most k, with $1/d \ll 1/k \ll 1/K$) in G between two endpoints of different paths, so that the short path has all its interior vertices in X. Then, if, for each $i \in [r]$, Y_i was the set of vertices in X which can be reached by a path in G of length at most $k/2$ from an endpoint of P_i while using otherwise only vertices from X, then the sets Y_i, $i \in [r]$, must all be disjoint. Then, one of these sets Y_i will be small (with $\leq Kd$ vertices), which will imply that $G[Y_i]$ must contain some dense subgraph. It is reasonable (if challenging), then, to hope that, starting with some initial paths forests, we could join the endpoints of these paths together via short connecting paths using new vertices in X until most of the paths are either **a)** long (with length $\geq Kd$) or **b)** connected into a 'dense spot' in $G[X]$.

In the case **b)**, we cannot hope to decompose these paths along with the dense spot in $G[X]$ they are connected to, for, as $|X| \approx pn$ and X will be chosen randomly, such a dense spot (say a sample of a copy of K_{d+1}) may only have around pd vertices, and thus only have paths of length up to pd. Therefore, we will have to argue that we can take the sample X accurately enough that any dense spot in $G[X]$ must lie within some dense spot in G, so that we can mostly decompose the original dense spot along with any paths we have attached to it into paths with length $(1-\varepsilon)d$. Reordering, slightly, then, we will do the following (see Figure 1):

    i) Take a maximal collection $\mathcal{F} = \{G_1, \ldots, G_t\}$ of dense spots in G and a small 'sample set' X.
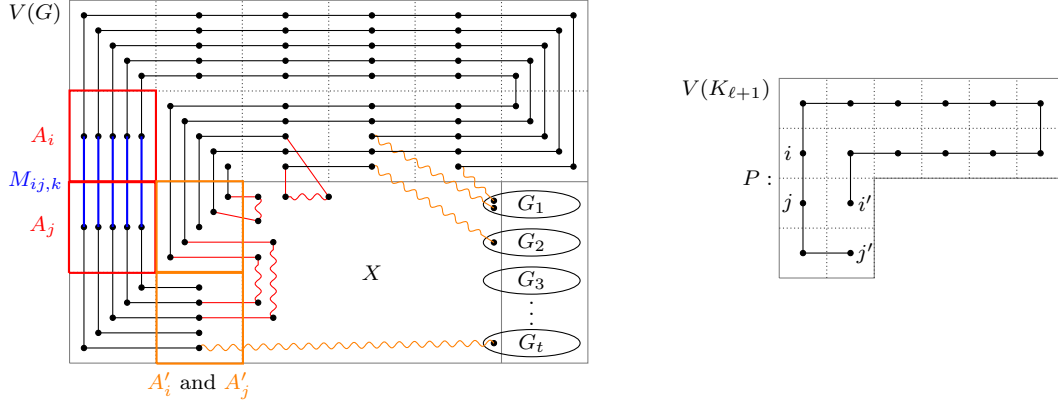
Figure 1: As shown on the left, having found a maximal collection of vertex-disjoint dense spots $G_1, \ldots, G_t$ in a graph G, we take a 'sample set' X which will intersect with each of these dense spots. Partitioning $V(G) \setminus (X \cup V(G_1) \cup \ldots \cup V(G_t))$ into sets $A_1 \cup \ldots \cup A_{\ell+1}$, we then cover most of the edges between each A_i and A_j with matchings $M_{ij,k}$, for $k \in [d']$ with $d' \approx d/s$. For a path P with length ℓ in $K_{\ell+1}$ (as on the right, with endvertices $i', j' \in V(K_{\ell+1})$ and each k we put together the matchings $M_{ij,k}$ for each $ij \in E(P)$ to get a path forest of few paths, most of which end in $A_{i'}$ and $A_{j'}$. We then connect up as many as possible of these endvertices with short paths using internal vertices in X (shown in red) and connect more of them into the dense spots with short paths using vertices in X (shown in orange).

ii) Mostly decompose $G - V(\mathcal{F}) - X$ into $d/2$ path forests consisting of medium-length paths.

iii) Iteratively join paths within each path forest together using short paths with internal vertices in X.

iv) Having exhausted the possible connections, we argue that, except for some small number of paths, most of the paths must be connectable into a dense spot in $G[X]$ and therefore, somehow, connectable into one of the dense spots G_i, $i \in [t]$.

v) Mostly decompose the long paths into paths of length $\ell = (1 - \varepsilon)d$ and, similarly, mostly decompose the dense spots along with the paths connected into them.

By this scheme, we will find the required decomposition of all but at most $\varepsilon n d$ edges of G into paths of length $\ell = (1 - \varepsilon)d$, but in setting aside the collection $\mathcal{F}$ we have introduced a further issue. While we can choose X so that $G - X$ is nearly regular, we now have the problem that the graph $G - V(\mathcal{F}) - X$ may not be nearly regular (which we require for partitioning into vertex sets and finding many edge-disjoint matchings between them). Critically, here, we will use two results shown by very recent techniques of Chakraborti, Janzer, Methuku, and Montgomery [12, 13] to efficiently find nearly-regular subgraphs. Moreover, we will develop these results non-trivially, using them to show that we can mostly decompose $G - V(\mathcal{F}) - X$ into a small number of edge-disjoint nearly-regular subgraphs (see Section 3.4). We also apply one of the results of [12, 13] while decomposing the dense spots along with the attached paths.

This completes an overall sketch of our methods for Theorem 1.2. In Section 2.3, we will show that Theorem 1.2 follows from three key lemmas, Lemmas 2.3, 2.4, and 2.5, which roughly correspond to steps ii), iv) and v) above, respectively. The proofs of each of these three key lemmas will need more new ideas than could be included in the sketch above, and they are discussed in more detail later where appropriate. In Section 2.4, we cover some tools we will need, including some comments on our use of the local lemma in Section 2.4.2. In Section 3, we give the decomposition into path forests we will use for $G - V(\mathcal{F}) - X$, thus proving Lemma 2.3. In Section 4, we will show the existence of a good 'sample set' X and show how to connect some paths to small dense spots within $G[X]$ and hence to some dense spot in $\mathcal{F}$, thus proving Lemma 2.4. In Section 5, we will decompose the dense spots along with some attached paths, proving Lemma 2.5 and hence completing the proof of Theorem 1.2. Finally, in Section 6, we will make some concluding remarks.

5

2.3 Proof of Theorem 1.2 subject to three key lemmas

We now introduce two key definitions to formalise some of the notions in Section 2.2. The first we will use to quantify what we want from our 'dense spots'.

Definition 2.1. *A graph G is (η, d, K)-**dense** if $0 < |V(G)| \leq Kd$ and $\delta(G) \geq (1 - \eta)d$.*

Our second definition records properties of path forests, where we require each forest to not contain too many paths and that the endpoints of the paths are relatively well spread out around a graph.

Definition 2.2. *Say a collection of path forests $\mathcal{P}_1, \ldots, \mathcal{P}_\ell$ is (m, Δ_0, Δ_1)-**bounded** if each $\mathcal{P}_i$, $i \in [\ell]$, has at most m paths, each vertex appears as the endvertex of at most Δ_0 paths in total in all of the path forests, and, for each $i \in [\ell]$, each vertex has at most Δ_1 neighbours among the endvertices of $\mathcal{P}_i$.*

We can now give our lemma which we will use to find a good collection of path forests for step ii) in Section 2.2. The lemma applies more generally than to a d-regular graph as we apply it to a regular graph with a maximal set of dense spots (and a 'sample set') removed.

Lemma 2.3. *Let $1/d \ll \gamma \ll \varepsilon \leq 1$. Let G be an n-vertex graph with maximum degree at most d in which all but at most γn vertices have degree at least $(1 - \gamma)d$. Then, G contains a $((1 + \varepsilon)n/d, d^{1/4}, d^{1/4})$-bounded edge-disjoint collection of $d/2$ path forests which cover all but at most εnd of the edges of G.*

Note that (as seen by the disjoint union of $(d + 1)$-vertex cliques), the bound $(1 + \varepsilon)n/d$ in Lemma 2.3 is close to optimal. Given $d/2$ edge-disjoint path forests which each contain $(1 + \varepsilon)n/d \leq 2n/d$ paths in an n-vertex d-regular graph, note that there are in total at most $2n$ endvertices, so that on average each vertex appears in total as the endvertex of at most 2 paths. Thus, the corresponding upper bound of $d^{1/4}$ in Lemma 2.3 is relatively unambitious; while it could be pushed much further with our methods this will be sufficient for our purpose. Similarly, for just one of these path forests, on average a vertex can expect to have at most 2 neighbours among the endpoints of the at most $(1 + \varepsilon)n/d \leq 2n/d$ paths. Again, the corresponding upper bound of $d^{1/4}$ in Lemma 2.3 is chosen loosely, only so that it is comfortably enough for our later proofs.

Next, we give the lemma that represents the heart of our proof. It shows the existence of our 'sample set' X such that, given a collection of path forests as produced by Lemma 2.3, we can join paths together using vertices from X and find further paths to get a collection of path forests in which the paths are connected into the dense spots (in a well spread manner similar to the conditions in Definition 2.2) or can be collectively mostly decomposed into paths with length $(1 - \varepsilon)d$.

Lemma 2.4. *Let $1/d \ll p, \eta \ll 1/K \ll \varepsilon$. Let G be a d-regular n-vertex graph. Let $\mathcal{F} = \{G_1, \ldots, G_t\}$ be a maximal family of vertex-disjoint (η, d, K)-dense subgraphs. Then, there is a set $X \subset V(G)$ with*

A1 *$|X| \leq 2pn$ and, for each $v \in V(G)$, $d_G(v, X) = (1 \pm \eta)pd$, and*

A2 *for each $j \in [t]$, $|X \cap V(G_j)| \leq 2p|V(G_j)|$ and, for each $v \in V(G_j)$, $d_G(v, X \cap V(G_i)) = (1 \pm 3\eta)2pd$,*

such that the following holds.

For any $(2n/d, d^{1/4}, d^{1/4})$-bounded edge-disjoint collection of path forests $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_{d/2}$ in $G - V(\mathcal{F}) - X$ we can find in $G - (V(\mathcal{F}) \setminus X)$ an edge-disjoint collection of path forests $\mathcal{P}'_1, \mathcal{P}'_2, \ldots, \mathcal{P}'_{d/2}$, such that the following hold.

A3 *For each $i \in [d/2]$, every path in $\mathcal{P}'_i$ has both of its endvertices in $X \cap V(\mathcal{F})$.*

A4 *Each vertex in G is in total an endvertex of at most $\sqrt{d}$ paths in $\mathcal{P}'_i$, $i \in [d/2]$.*

A5 *For each $j \in [t]$ and $i \in [d/2]$ at most $\sqrt{d}$ of the paths in $\mathcal{P}'_i$ end in G_j.*

A6 *All but at most $\varepsilon nd/4$ of the edges of $E(\mathcal{P}_1 \cup \ldots \cup \mathcal{P}_{d/2}) \setminus E(\mathcal{P}'_1 \cup \ldots \cup \mathcal{P}'_{d/2})$ can be decomposed into copies of $P_{(1-\varepsilon)d}$.*

Finally, we give our lemma which can decompose the dense spots along with some paths which all have their endpoints among the dense spots in a relatively well spread manner, as follows.

Lemma 2.5. *Let $1/d \ll \eta \ll p, 1/K \ll \varepsilon$. Let G be a graph consisting of a vertex-disjoint family $\mathcal{F} = \{G_1, \ldots, G_t\}$ of (η, d, K)-dense graphs and path forests $\mathcal{P}_1, \ldots, \mathcal{P}_{d/2}$ where each path in each path forest has both its endpoints in $V(\mathcal{F})$ and all its internal vertices not in $V(\mathcal{F})$. Suppose also that the following properties hold.*

B1 *For each $i \in [t]$, there is some $X_i \subseteq V(G_i)$ so that X_i contains all of the vertices of $V(\mathcal{P}_1) \cup \ldots \cup V(\mathcal{P}_{d/2})$ in $V(G_i)$ (which are necessarily endpoints), and, for each $v \in V(G_i)$, $d_{G_i}(v, X_i) = (1 \pm \eta)pd$.*

B2 *Each vertex $v \in V(\mathcal{F})$ is an endpoint of in total at most $\sqrt{d}$ paths from $\mathcal{P}_1, \ldots, \mathcal{P}_{d/2}$.*

B3 *For each $j \in [t]$ and $i \in [d/2]$ at most $\sqrt{d}$ of the paths in $\mathcal{P}_i$ have at least one endpoint in G_j.*

Then, all but at most $\varepsilon d \cdot |V(\mathcal{F})|$ of the edges of G can be decomposed into copies of $P_{(1-\varepsilon)d}$.

Subject to the proof of these three key lemmas, we can now prove Theorem 1.2, following the steps i)–v) in Section 2.2.

Proof of Theorem 1.2. Let $K \in \mathbb{N}$ and $\eta, p > 0$ satisfy

$$1/d \ll \eta \ll p \ll 1/K \ll \varepsilon \leq 1.$$

Let G be an n-vertex d-regular graph. Let $\mathcal{F} = \{G_1, \ldots, G_t\}$ be a maximal collection of vertex-disjoint (η, d, K)-dense graphs in G. Using Lemma 2.4, let $X \subset V(G)$ be such that the following hold.

C1 $|X| \leq 2pn$, and, for each $v \in V(G)$, $d_G(v, X) \leq 2pd$.

C2 For each $i \in [t]$, $|X \cap V(G_i)| \leq 2p|V(G_i)|$ and, for each $v \in V(G_i)$, $d_G(v, X \cap V(G_i)) = (1 \pm \eta)pd$.

C3 For any $(2n/d, d^{1/4}, d^{1/4})$-bounded edge-disjoint collection of path forests $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_{d/2}$ in $G - V(\mathcal{F}) - X$, we can find in $G - V(\mathcal{F} \setminus X)$ an edge-disjoint collection of path forests $\mathcal{P}'_1, \mathcal{P}'_2, \ldots, \mathcal{P}'_{d/2}$ such that the following hold.

    i) For each $i \in [d/2]$, every path in $\mathcal{P}'_i$ has both of its endvertices in $X \cap V(\mathcal{F})$.

    ii) Each vertex in $V(G)$ is an endvertex of at most $\sqrt{d}$ paths in $\mathcal{P}'_i$, $i \in [d/2]$.

    iii) For each $j \in [t]$, and each $i \in [d/2]$, at most $\sqrt{d}$ of the paths in $\mathcal{P}'_i$ end in G_j.

    iv) All but at most $\varepsilon nd/4$ edges of $E(\mathcal{P}_1 \cup \ldots \cup \mathcal{P}_{d/2}) \setminus E(\mathcal{P}'_1 \cup \ldots \cup \mathcal{P}'_{d/2})$ can be decomposed into copies of $P_{(1-\varepsilon)d}$.

Let $G' = G - V(\mathcal{F}) - X$. If $|V(G')| \leq \varepsilon n/2$, then let $\mathcal{P}_1 = \ldots = \mathcal{P}_{d/2} = \emptyset$, noting that these cover all but at most $\varepsilon nd/4$ edges of G' and they therefore trivially form a $(2n/d, d^{1/4}, d^{1/4})$-bounded collection of paths which cover all but at most $\varepsilon nd/4$ of the edges of G'. Suppose, otherwise, that $|V(G')| > \varepsilon n/2$. As each G_i, $i \in [t]$, is (η, d, K)-dense, the number of edges incident on $V(G')$ with, for some $i \in [t]$, one vertex in some G_i is at most $\sum_{i \in [t]} |V(G_i)| \cdot \eta d = \eta d|V(\mathcal{F})|$. Therefore, by **C1**,

$$|E(G')| \geq \frac{1}{2}d|V(G')| - \eta d|V(\mathcal{F})| - 2pd|X| \geq \frac{1}{2}d(1 - 5p)|V(G')|.$$

7

Let $\gamma = \sqrt{5p}$. As $\Delta(G') \leq d$, if N is the number of vertices in G' with degree at most $(1-\gamma)d$ in G', then $N \cdot \gamma d \leq 5dp|V(G')|$, so that $N \leq \gamma|V(G')|$. Thus, as $1/d \ll p \ll \varepsilon$, by Lemma 2.3, G' contains a $(2n/d, d^{1/4}, d^{1/4})$-bounded edge-disjoint collection of path-forests $\mathcal{P}_1, \ldots, \mathcal{P}_{d/2}$ which together cover all but at most $\varepsilon nd/4$ of the edges of $G' = G - V(\mathcal{F}) - X$.

Then, by **C3**, we can find an edge-disjoint collection of path-forests $\mathcal{P}'_1, \ldots, \mathcal{P}'_{d/2}$ in $G - V(\mathcal{F})$ such that **i)**–**iv)** hold. Using **iv)**, let $\mathcal{Q}_1$ be a set of edge-disjoint copies of $P_{(1-\varepsilon)d}$ which decompose all but at most $\varepsilon nd/4$ edges of $E(\mathcal{P}_1 \cup \ldots \cup \mathcal{P}_{d/2}) \setminus E(\mathcal{P}'_1 \cup \ldots \cup \mathcal{P}'_{d/2})$.

Let $G'' = \left(\bigcup_{i\in[t]} G_i\right) \cup \left(\bigcup_{i\in[d/2]} \bigcup_{P\in\mathcal{P}'_i} P\right)$. By Lemma 2.5 (with $X_i = X \cap V(G_i)$ for each $i \in [t]$, $\eta' = 3\eta$, and using **C2** and **i)**–**iii)**), let $\mathcal{Q}_2$ be a set of edge-disjoint copies of $P_{(1-\varepsilon)d}$ which decomposes all but at most $\varepsilon nd/4$ of the edges of G''.

Then, $\mathcal{Q}_1 \cup \mathcal{Q}_2$ is an edge-disjoint collection of copies of $P_{(1-\varepsilon)d}$ which decomposes all but at most $\varepsilon nd/2$ edges of $\mathcal{P}_1, \ldots, \mathcal{P}_{d/2}, \mathcal{F}$, and therefore all but at most $3\varepsilon nd/4$ of the edges in $E(G') \cup E(\mathcal{F})$. As G_i is (η, d, K)-dense for each $i \in [t]$, each vertex in $V(\mathcal{F})$ has at most ηd neighbours in $V(G')$, so that there are at most ηnd edges in G between $V(\mathcal{F})$ and $V(G')$. Furthermore, as $|X| \leq 2pn$ by **C1**, there are at most $2pnd$ edges in G between X and $V(G) \setminus X$. Therefore, as the only edges of G which are not in $E(G') \cup E(\mathcal{F})$ are those with a vertex in X or which are between G' and $V(\mathcal{F})$, $\mathcal{Q}_1 \cup \mathcal{Q}_2$ decomposes all but at most $3\varepsilon nd/4 + \eta nd + 2pnd \leq \varepsilon nd$ of the edges of G into copies of $P_{(1-\varepsilon)d}$, as required. $\qquad\square$

2.4 Tools

We collect here some standard results we will use in our proofs.

2.4.1 Concentration inequalities

We will use the following standard version of Chernoff's bound (see, for example, [24, Corollary 2.2 and Theorem 2.10]).

Lemma 2.6 (Chernoff's bound)**.** *Let X be a random variable with mean μ which is binomially distributed or hypergeometrically distributed. Then, for any $0 < \gamma < 1$, we have that $\mathbb{P}(|X - \mu| \geq \gamma\mu) \leq 2e^{-\mu\gamma^2/3}$.*

Given a product probability space $\Omega = \prod_{i\in[n]} \Omega_i$, a random variable $X\colon \Omega \to \mathbb{R}$ is called C-Lipschitz if $|X(\omega) - X(\omega')| \leq C$ whenever ω and ω' differ in at most 1 co-ordinate. We will use the following standard version of Azuma's inequality (see, for example, [24, Corollary 2.27]).

Lemma 2.7 (Azuma's inequality)**.** *Let X be C-Lipschitz random variable on a product probability space with n co-ordinates. Then, for any $t > 0$,*
$$\mathbb{P}(|X - \mathbb{E}(X)| > t) \leq 2e^{\frac{-t^2}{2nC^2}}.$$

2.4.2 The local lemma

We will use the local lemma of Lovász (see [2]) many times throughout our proofs, so after stating the form that we use we will make some remarks on how we use it.

Lemma 2.8 (The local lemma)**.** *Let $B_1, B_2, \ldots, B_n$ be events in an arbitrary probability space. Suppose that, for each $i \in [n]$, the event B_i is mutually independent of a set of at most Δ of the other events and $\mathbb{P}(B_i) \leq p$. Then, if $ep(\Delta + 1) \leq 1$, the probability that none of the events B_i, $i \in [n]$, occurs is strictly positive.*

Each time we apply the local lemma we will do so within a d-regular (or approximately d-regular) n-vertex graph G. The 'bad events' B_i, $i \in I$, that we use, where I is some appropriate index set, will all be of the form that

8

'a binomial or hypergeometric variable is not concentrated around its mean', where these variables will depend on some partition of the vertices (or in one case of the edges) for which each vertex (or edge) is placed in a set of the partition independently at random. Thus, using Lemma 2.6, we can bound the probability of each 'bad event' occurring, which in each instance will be at most $p := e^{-d^{0.1}/100}$.

Each bad event will, for some central vertex, depend only on the placement of vertices within a distance k of the central vertex, for some k with $1/d \ll 1/k$. Changing the placement of a single vertex (or edge) in the partition will affect at most d^k of the 'bad events'. Therefore, for each 'bad event', B say, after identifying the central vertex, we take the set of events which depend on any vertex within a distance k of the central vertex, and note that B is mutually independent of the set of all the other 'bad events'. As we have a graph with maximum degree d, the number of events we omitted from the set is at most $\Delta := 2d^k \cdot d^k$. Therefore, as $1/d \ll 1/k$, we comfortably have that $ep(\Delta + 1) = e \cdot e^{-d^{0.1}/100}(2d^k + 1) \cdot d^k \le 1$, and we can apply the local lemma, Lemma 2.8, to show that there is some instance in which none of the 'bad events' occur.

Let us also note that, when for example choosing a subset $X \subset V(G)$ by including vertices independently at random with probability p, the 'bad events' that we wish to consider may include that $|X| \ne (1 \pm \gamma)pn$ (where $1/d \ll p, \gamma$). This event will not be independent of any other bad event we define as it is influenced by every vertex in G. We could deal with this using the asymmetric local lemma, but, instead, to keep using Lemma 2.8, we will take an arbitrary partition of $V(G)$ into sets $V_1, \ldots, V_t$ which each have size between $d/2$ and d. If, for each $i \in [t]$, B_i is the event that $V_i \cap X \ne (1 \pm \gamma)p|V_i|$ then, when no event B_i, $i \in [t]$, holds then $|X| = (1 \pm \gamma)pn$ and, furthermore, each of these events only depends on the location of at most d vertices.

As an example of how we use the local lemma, we will give here the details of our first application, from Lemma 3.1. Let $s = 2\lceil d^{0.15}\rceil$ and $\eta = 4d^{-0.4}$ with $1/d \ll \eta \ll 1$. Suppose that G has vertices degrees which are $(1 \pm \eta/2)d$ and let $V(G) = V_1 \cup \ldots \cup V_t$ be a partition of $V(G)$ into sets of size between $d/2$ and d. Let $V(G) = A_1 \cup \ldots \cup A_s$ be a partition of $V(G)$ formed by taking each $v \in V(G)$ and, independently at random, placing it in each A_i, $i \in [s]$, with probability $1/s$. For each $v \in V(G)$ and $i \in [s]$, let $B_{v,i}$ be the event that $d_G(v, A_i) \ne (1 \pm \eta)d/s$. For each $i \in [s]$ and $j \in [t]$, let $B_{i,j}$ be the event that $|A_i \cap V_j| \ne (1 \pm \eta)|V_j|/s$. Note that, by Chernoff's bound, each of these events occur with probability at most $\exp(-\eta^2 \cdot (d/2s)/3) \le \exp(-d^{0.1})$. For each $v \in V(G)$ and $i \in [s]$, the event $B_{v,i}$ only depends on whether each vertex in $N_G(v)$ is in A_i or not, so therefore on the placement of at most d vertices in the partition $A_1 \cup \ldots \cup A_s$. Similarly, for each $i \in [s]$ and $j \in [t]$, $B_{i,j}$ only depends on the placement of at most d vertices in the partition $A_1 \cup \ldots \cup A_s$. For any vertex $w \in V(G)$, the placement of w only affects an event $B_{v,i}$ with $v \in V(G)$ and $i \in [s]$ if v is a neighbour of w, so at most ds pairs (v, i). Furthermore, the placement of w only affects an event $B_{i,j}$ with $i \in [s]$ and $j \in [t]$ if $w \in V_j$, and thus affects only at most s such events. Therefore, the placement of each vertex affects at most $ds + s \le d^2$ events. Therefore, each event in $\{B_{v,i} : v \in V(G), i \in [s]\} \cup \{B_{i,j} : i \in [s], j \in [t]\}$ is mutually independent of a set of all but at most $\Delta := d \cdot d^2 = d^3$ other events. Thus, as $ep\Delta = e \cdot \exp(-d^{0.1}) \cdot d^3 \le 1$, by Lemma 2.8, we have that, with positive probability, none of the events in $\{B_{v,i} : v \in V(G), i \in [s]\} \cup \{B_{i,j} : i \in [s], j \in [t]\}$ hold. Therefore, we can take a partition of $V(G)$ into $A_1 \cup \ldots \cup A_s$ for which none of these events hold, and hence, for each $v \in V(G)$ and $i \in [s]$ we have $d_G(v, A_i) = (1 \pm \eta)d/s$ and, for each $i \in [s]$, as no event $B_{i,j}$, $j \in [t]$, holds, we have $|A_i| = (1 \pm \eta)n/s$.

Finally, here let us highlight the instance in which we use the most 'bad events', in the proof of Lemma 4.3. There, working with a d-regular graph G, we will consider the collection $\mathcal{S}$ of all subsets of $V(G)$ of size at most k such that these vertices are all pairwise at most k apart in G, where $1/d \ll 1/k$. Each vertex in G will be allocated to some set X or not independently at random with probability p. There will be some bad events B_i, $i \in I$, where, for each $U \in \mathcal{S}$, at most $d|U| + 1 \le dk + 1$ of the indices $i \in I$ will contain U. Any event B_i whose index includes $U \in \mathcal{S}$ will depend only whether or not the vertices in $\bigcup_{v \in U} N_G(v)$ are in X, where this set has size at most $|U|d \le dk$. As each vertex w will appear as a neighbour of at most $(2d^{2k+1})^{k-1}$ sets in $\mathcal{S}$, we will have that at most $d^{2k^2} \cdot (dk + 1)$ events B_i, $i \in I$, depend on whether or not w is in X. As $1/d \ll 1/k$, then, we can think of this as every event depending on at most a polynomial of d (whose power is a function of k) many other bad events, while the probability of each bad event will be $\exp(-d^{0.1}/100)$, allowing us to apply Lemma 2.8 once again.

2.4.3 Matchings in bipartite graphs

In order to find the matchings $M_{ij,k}$ described in Section 2.2, we will use the following result.

Lemma 2.9. *Let $d, n > 0$ and let $0 < \gamma < 1$ satisfy $\gamma d \geq 1$. Let G be a bipartite graph with vertex class sizes $(1 \pm \gamma)n$ and $d(v) = (1 \pm \gamma)d$ for each $v \in V(G)$. Then, G contains $(1 - 10\sqrt{\gamma})d$ edge-disjoint matchings which have size $(1 - 10\sqrt{\gamma})n$.*

Proof. Note that $e(G) \geq (1 - \gamma)d \cdot (1 - \gamma)n > (1 - 2\gamma)nd$. By Vizing's theorem, the edge set of G can be partitioned into at most $(1 + \gamma)d + 1 \leq (1 + 2\gamma)d$ matchings, where we used that $\gamma d \geq 1$. If the statement of the lemma fails, then

$$e(G) \leq (1 - 10\sqrt{\gamma})d \cdot (1 + \gamma)n + (10\sqrt{\gamma} + 2\gamma)d \cdot (1 - 10\sqrt{\gamma})n = (1 - 97\gamma + 30\gamma^{3/2})nd \leq (1 - 2\gamma)nd,$$

a contradiction. □

2.4.4 Decomposing complete graphs into paths

In decomposing our dense spots, we will use the following result from [11], which is the special case for paths of a result for bounded degree trees.

Theorem 2.10 (Böttcher-Hladký-Piguet-Taraz). *Let $1/n \ll \varepsilon$. Let $\mathcal{P}$ be a collection of paths of length at most n whose total length is at most $\binom{n}{2}$. Then, $\mathcal{P}$ packs into $K_{(1+\varepsilon)n}$.*

3 Decomposing outside the dense bits

In this section, we prove Lemma 2.3. We begin, in Section 3.1, by finding an initial collection of path forests which are, together, weakly bounded (see Definition 2.2), before using this to find a collection of path forests in Section 3.2 which are more strongly bounded. In Section 3.3, we give a short proof of Theorem 1.4 from this. In Section 3.4, we take near-regularisation results from [12, 13] and develop them to find very nearly regular subgraphs in a roughly regular graph. Finally, we put this all together in Section 3.5 to prove Lemma 2.3.

3.1 Initial path forests

We prove the following lemma by partitioning the vertex set of a regular graph using the local lemma and putting together matchings between different vertex classes, as sketched in Section 2.2.

Lemma 3.1. *Let $1/d \ll \varepsilon \leq 1$ and $\gamma = 2d^{-0.4}$. Let G be an n-vertex graph with vertex degrees in $(1 \pm \gamma)d$. Then, G contains an $(n/d^{1/8}, d^{7/8}, d^{7/8})$-bounded edge-disjoint collection $\mathcal{P}_1, \ldots, \mathcal{P}_{d/2}$ of path forests which cover all but at most εnd edges of G.*

Proof. Let $s = 2\lceil d^{0.15} \rceil$ and $\eta = 4d^{-0.4}$. Let $V(G) = A_1 \cup \ldots \cup A_s$ be a partition of $V(G)$ formed by taking each $v \in V(G)$ and, independently at random, placing it in each A_i, $i \in [s]$, with probability $1/s$.

By the local lemma (see Section 2.4.2 for details), with positive probability, we can have the following properties.

D1 For each $v \in V(G)$ and $i \in [s]$, $d_G(v, A_i) = (1 \pm \eta)d/s$.
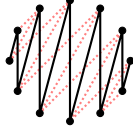
D2 For each $i \in [s]$, $|A_i| = (1 \pm \eta)n/s$.

Figure 2: For $s = 12$, an embedding of a path of length $s - 1 = 11$ which can be rotated $s/2 = 6$ times to form together a complete graph with $s = 12$ vertices, with one clockwise rotation depicted in red, giving a decomposition of K_s into paths of length $s - 1$ in which each vertex appears as an endpoint exactly once.

Let $\eta' = 20\sqrt{\eta} = 40d^{-0.2}$ and $d' = (1 - \eta')d/s$. For each edge $e = jk$ in the complete s-vertex graph K_s, using Lemma 2.9, **D1** and **D2**, find d' edge-disjoint matchings in $G[A_j, A_k]$ which each have at least $(1 - \eta')n/s$ edges. Call these matchings $M_{e,i}$, $i \in [d']$. Now, take paths $Q_1, \ldots, Q_{s/2}$ of length $s - 1$ in the complete graph K_s in which each vertex is the endpoint of exactly one path. (Such a decomposition is well-known to exist when s is even. See, for example, Figure 2.) For each $j \in [s/2]$ and $k \in [d']$, let $\mathcal{P}_{j,k}$ be the union of all the matchings $M_{e,k}$ with $e \in E(Q_j)$ and let $\mathcal{P}'_{j,k}$ be the subcollection of paths in $\mathcal{P}_{j,k}$ of length $s - 1$.

Claim 3.2. *For each $j \in [s/2]$ and $k \in [d']$, there are at most $\varepsilon n/3$ edges of $\mathcal{P}_{j,k}$ which are not in $\mathcal{P}'_{j,k}$.*

Proof. Set $j \in [s/2]$ and $k \in [d']$. For each $i \in [s]$ which is not an endpoint of Q_k, note that the number of vertices in A_i which can be an endpoint of a path in $\mathcal{P}_{j,k}$ is at most $2(|A_i| - (1 - \eta')n/s) \leq 3\eta'n/s$, using **D2**. Therefore, at most $3\eta'n$ paths in $\mathcal{P}_{j,k}$ can have an endpoint in some A_i where i is a middle vertex of Q_k. As each path in $\mathcal{P}_{j,k} \setminus \mathcal{P}'_{j,k}$ has such an endpoint, there are at most $(s - 1) \cdot 3\eta'n \leq \varepsilon n/3$ edges of $\mathcal{P}_{j,k}$ which are not in $\mathcal{P}'_{j,k}$, where we have used that $s = 2\lfloor d^{0.15} \rfloor$, $\eta' = 40d^{-0.2}$ and $1/d \ll \varepsilon$. ⊡

Relabel the path forests $\mathcal{P}'_{j,k}$, $j \in [s/2]$ and $k \in [d']$, as $\mathcal{P}_i$, $i \in [sd'/2]$. Noting that $sd'/2 = (1 - \eta')d/2 \leq d/2$, let $\mathcal{P}_i = \emptyset$ for each $sd'/2 < i \leq d/2$. The number of edges in $\mathcal{P}_i$, $i \in [d/2]$, is, by Claim 3.2 and as $1/s, \eta' \ll \varepsilon$, at least

$$\frac{sd'}{2} \cdot (s - 1) \cdot \frac{(1 - \eta')n}{s} - \frac{sd'}{2} \cdot \frac{\varepsilon n}{3} \geq \frac{sd'}{2}\left(\left(1 - \frac{\varepsilon}{3}\right)n - \frac{\varepsilon n}{3}\right) = \frac{(1 - \eta')d}{2} \cdot \left(1 - \frac{2\varepsilon}{3}\right)n \geq \frac{n(1 + \eta)d}{2} - \varepsilon nd$$
$$\geq |E(G)| - \varepsilon nd.$$

Thus, it is only left to show that the path forests $\mathcal{P}_i$, $i \in [d/2]$, form an $(n/d^{1/8}, d^{7/8}, d^{7/8})$-bounded collection.

For each $i \in [d/2]$, if $\mathcal{P}_i \neq \emptyset$, then there are some distinct $j, k \in V(K_s)$ such that all the endvertices of $\mathcal{P}_i$ are in $A_j \cup A_k$. Thus, by **D1**, each vertex in G has at most $4d/s \leq d^{7/8}$ neighbours among the endvertices of $\mathcal{P}_i$ and, by **D2**, $\mathcal{P}_i$ contains at most $2n/s \leq n/d^{1/8}$ paths. Furthermore, for each $v \in V(G)$, if j, k are such that $v \in A_j$ and j is an endvertex of Q_k, then the only paths in $\mathcal{P}_i$, $i \in [d/2]$, that can end in v are those from $\mathcal{P}_{k,k'}$, $k' \in [d']$, so that v is the endvertex of at most $d' \leq d/s \leq d^{7/8}$ paths in $\mathcal{P}_i$, $i \in [d/2]$. Therefore, the path forests $\mathcal{P}_i$, $i \in [d/2]$, form an $(n/d^{1/8}, d^{7/8}, d^{7/8})$-bounded collection, as required. □

3.2 Improving a bounded collection of path forests

Having set aside a set Y (which fulfils a similar 'sample set' function as X in the sketch in Section 2.2) in a d-regular graph G, we can then improve a collection of path forests in $G - Y$ by taking each path forest in turn and iteratively connecting paths in any path forest using a path of length two whose middle vertex comes from Y (for practical reasons, we actually write this in the proof through taking some maximal set I). In order to do this efficiently, we use the local lemma to partition Y into ten parts and find as many such paths using a middle vertex from these sets, exhausting each set in turn. Furthermore, for all but only a few of the remaining endpoints in the resulting path forest, we attach an extra edge randomly from this endpoint to the last set in the partition of Y, using the local lemma with regard to all these choices to make sure that no vertex has too many

neighbours among these new endpoints. This last step is critical in ensuring that (once a few paths are removed from the resulting path forests) the endpoints of the path forests will be well spread.

Lemma 3.3. *Let $1/d \ll \gamma \ll p \ll \varepsilon$. Let G be a n-vertex graph with $\Delta(G) \leq d$. Let $Y \subset V(G)$ be such that $|Y| \leq (1 + \gamma)pn$ and, for each $v \in V(G)$, $d_G(v, Y) = (1 \pm \gamma)pd$. Let $\mathcal{P}_1, \ldots, \mathcal{P}_{d/2}$ be edge-disjoint path forests in $G - Y$ which are $(n/d^{1/9}, d^{8/9}, d^{8/9})$-bounded.*

Then, G contains an edge-disjoint collection of path forests $\mathcal{P}'_1, \ldots, \mathcal{P}'_{d/2}$ such that at most $\varepsilon n/d$ paths can be removed from each $\mathcal{P}'_i$, $i \in [d/2]$, to get altogether a $((1 + \varepsilon)n/d, d^{1/4}, d^{1/4})$-bounded edge-disjoint collection of path forests, and, for each $i \in [d/2]$, $E(\mathcal{P}_i) \subset E(\mathcal{P}'_i)$.

Proof. Let $\Delta_0 = d^{1/4}$ and $\Delta_1 = d^{1/4}$. Take a maximal set $I \subset [d/2]$ for which there are edge-disjoint collections $\mathcal{Q}_i$, $\mathcal{R}_i$, $\mathcal{P}'_i$, $i \in I$, of path forests in G, and, for each $i \in I$, a path forest $\mathcal{P}''_i$ consisting of some of the paths in $\mathcal{P}'_i$, such that, for each $i \in I$,

E1 each path in $\mathcal{Q}_i$ has length 2 with endvertices among the endvertices of $\mathcal{P}_i$ and its middle vertex in Y,

E2 each path in $\mathcal{R}_i$ has length 1 with an endvertex among the endvertices of $\mathcal{P}_i$ and an endvertex in Y,

E3 the paths in $\mathcal{Q}_i$ are vertex disjoint from the paths in $\mathcal{R}_i$,

E4 $\mathcal{P}_i$, $\mathcal{Q}_i$, $\mathcal{R}_i$ combine to form $\mathcal{P}'_i = \mathcal{P}_i + \mathcal{Q}_i + \mathcal{R}_i$,

E5 $\mathcal{P}''_i$ contains all but at most of the $\varepsilon n/d$ paths in $\mathcal{P}'_i$,

and the collection $\mathcal{P}''_i$, $i \in I$, is $((1 + \varepsilon)n/d, \Delta_0, \Delta_1)$-bounded.

Take such path forests $\mathcal{Q}_i, \mathcal{R}_i, \mathcal{P}'_i, \mathcal{P}''_i$, $i \in I$. Note that if $I = [d/2]$, then taking $\mathcal{P}'_i$, $i \in [d/2]$, gives us a collection of edge-disjoint path forests in G which satisfies the conditions in the lemma (as seen by considering $\mathcal{P}''_i$, $i \in [d/2]$). Therefore, assume for contradiction that $I \neq [d/2]$.

Let G' be the graph of all of the edges of G which are not in any of the paths in $\mathcal{Q}_i$ or $\mathcal{R}_i$ for any $i \in I$. As the collection $\mathcal{P}_i$, $i \in [d/2]$, is $(n/d^{1/9}, d^{8/9}, d^{8/9})$-bounded, each vertex $v \in V(G) \setminus Y$ appears as the endpoint of at most $d^{8/9}$ paths in total among the paths of $\mathcal{P}_i$, $i \in [d/2]$, and therefore v has at most $d^{8/9}$ neighbouring edges in total in the path forests $\mathcal{Q}_i$ and $\mathcal{R}_i$, $i \in I$. Thus, as $d_G(v, Y) \geq (1 - \gamma)pd$ and $1/d \ll \gamma, p$, we have $d_{G'}(v, Y) \geq (1 - 2\gamma)pd$. Take a random partition $Y = Y_1 \cup \ldots \cup Y_{10}$ by choosing the location of each $v \in Y$ independently and uniformly at random. Using the local lemma, we can assume that, for each $v \in V(G) \setminus Y$ and $j \in [10]$, $d_{G'}(v, Y_j) \geq (1 - 3\gamma)pd/10$ and $|Y_j| \leq (1 + 2\gamma)pn/10$.

Using that $I \neq [d/2]$, pick $i \in [d/2] \setminus I$. For each $1 \leq j \leq 10$ in turn, let $\mathcal{Q}_{i,j}$ be a maximal set of vertex-disjoint paths of length 2 in G' such that each path in $\mathcal{Q}_{i,j}$ has its middle vertex in Y_j and its endvertices among the endvertices of $\mathcal{P}_i + \mathcal{Q}_{i,1} + \ldots + \mathcal{Q}_{i,j-1}$, and $\mathcal{P}_i + \mathcal{Q}_{i,1} + \ldots + \mathcal{Q}_{i,j}$ is a path forest. We will show the following.

Claim 3.4. *For each $0 \leq j \leq 10$, $\mathcal{P}_i + \mathcal{Q}_{i,1} + \ldots + \mathcal{Q}_{i,j}$ contains at most $(1 + \varepsilon/2)n/d + n/d^{(j+1)/10}$ paths.*

Proof. We will prove this by induction on j, noting that it follows for $j = 0$ as $\mathcal{P}_i$ contains at most $n/d^{1/9} \leq n/d^{1/10}$ paths by the $(n/d^{1/9}, d^{8/9}, d^{8/9})$-boundedness condition. Then, suppose $j \in [10]$, and that $\mathcal{P}_i + \mathcal{Q}_{i,1} + \ldots + \mathcal{Q}_{i,j-1}$ contains at most $(1 + \varepsilon/2)n/d + n/d^{j/10}$ paths. Therefore, $\mathcal{Q}_{i,j}$ contains at most $3n/d^{j/10}$ paths, and so, if $Y'_j \subset Y_j$ is the set of vertices in Y_j not appearing in any path in $\mathcal{Q}_{i,j}$, then $|Y_j \setminus Y'_j| \leq 3n/d^{j/10}$.

Let $X_j \subset V(G) \setminus Y$ be a maximal set of endvertices of $\mathcal{P}_i + \mathcal{Q}_{i,1} + \ldots + \mathcal{Q}_{i,j}$ which are in $V(G) \setminus Y$ and which are such that no two vertices in X_j are endpoints of the same path in $\mathcal{P}_i + \mathcal{Q}_{i,1} + \ldots + \mathcal{Q}_{i,j}$. By the maximality of $\mathcal{Q}_{i,j}$, no two vertices in X_j have a common neighbour in Y'_j in G'. Thus, the number of edges in G' between X_j and Y'_j is at most $|Y'_j| \leq |Y_j| \leq (1 + 2\gamma)pn/10$.

12

On the other hand, each vertex $v \in X_j \subset V(G) \setminus Y$ satisfies $d_{G'}(v, Y_j) \geq (1 - 3\gamma)pd/10$, so the number of edges in G' between X_j and Y_j' is (as each vertex in $Y_j \setminus Y_j'$ has at most $d^{8/9}$ neighbours among the endvertices of $\mathcal{P}_i$ by the $(n/d^{1/9}, d^{8/9}, d^{8/9})$-boundedness of the $\mathcal{P}_k$, $k \in [d/2]$) at least

$$|X_j| \cdot (1 - 3\gamma)pd/10 - |Y_j \setminus Y_j'| \cdot d^{8/9} \geq |X_j| \cdot (1 - 3\gamma)pd/10 - d^{8/9} \cdot 3n/d^{j/10},$$

so that, as $1/d \ll p$ and $\gamma \ll \varepsilon$,

$$|X_j| \leq \frac{(1 + 2\gamma)pn/10 + d^{8/9} \cdot 3n/d^{j/10}}{(1 - 3\gamma)pd/10} \leq (1 + 6\gamma)n/d + n/d^{(j+1)/10} \leq (1 + \varepsilon/2)n/d + n/d^{(j+1)/10}.$$

Then, by the definition of X_j, we have that $\mathcal{P}_i + \mathcal{Q}_{i,1} + \ldots + \mathcal{Q}_{i,j}$ contains at most $(1 + \varepsilon/2)n/d + n/d^{(j+1)/10}$ paths, as required. □

Let $\mathcal{Q}_i = \mathcal{Q}_{i,1} + \ldots + \mathcal{Q}_{i,10}$ so that, by the claim with $j = 10$, $\mathcal{P}_i + \mathcal{Q}_i$ contains at most $(1 + \varepsilon)n/d$ paths as $1/d \ll \varepsilon$. Let Z_i be the set of vertices in Y_{10} which appear in $\mathcal{Q}_{i,10}$, so that, as $\mathcal{P}_i + \mathcal{Q}_{i,1} + \ldots + \mathcal{Q}_{i,9}$ contains at most $(1 + \varepsilon/2)n/d + n/d$ paths by Claim 3.4 with $j = 9$, we have $|Z_i| \leq 3n/d$. Let $Z \subset Y$ be the set of vertices in Y which are the endvertices of more than $\Delta_0/2 = d^{1/4}/2$ of the paths in $\mathcal{P}_j''$, $j \in I$. Then, as $\mathcal{P}_j''$, $j \in I$, is $((1 + \varepsilon)n/d, \Delta_0, \Delta_1)$-bounded and hence contains at most $|I| \cdot (1 + \varepsilon)n/d$ paths in total, we have, as $|I| \leq d/2$,

$$|Z| \leq \frac{2 \cdot |I| \cdot (1 + \varepsilon)n/d}{\Delta_0/2} \leq \frac{4n}{\Delta_0}. \tag{1}$$

Let A_i be the set of endvertices of the paths in $\mathcal{P}_i + \mathcal{Q}_i$ for which both endvertices have at least $pd/30$ neighbours in $Y_{10} \setminus (Z \cup Z_i)$ in G'. Note that, by the maximality of $\mathcal{Q}_{i,10}$, the sets $N_{G'}(u, Y_{10}) \setminus Z_i$ and $N_{G'}(u', Y_{10}) \setminus Z_i$ are disjoint for $u, u' \in A_i$ unless u and u' are endpoints of the same path in $\mathcal{P}_i + \mathcal{Q}_i$. Note then that A_i partitions naturally into a set of unordered pairs $\mathcal{F}_i$ where if $\{u, u'\} \in \mathcal{F}_i$ then u and u' are endpoints of the same path in $\mathcal{P}_i + \mathcal{Q}_i$. For each $\{u, u'\} \in \mathcal{F}_i$, pick a vertex v_u from $N_{G'}(u, Y_{10} \setminus (Z \cup Z_i))$ and a vertex $v_{u'} \in N_{G'}(u', Y_{10} \setminus (Z \cup Z_i))$ uniformly and independently at random subject to $v_u \neq v_{u'}$ and add uv_u and $u'v_{u'}$ to $\mathcal{R}_i$.

For each $v \in V(G)$, let B_v be the event that v has more than $\Delta_1 = d^{1/4}$ neighbours among the set $\{v_u, v_{u'} : \{u, u'\} \in \mathcal{F}_i\}$. For each $w \in N_G(v, Y_{10})$, there is at most one pair $\{u, u'\} \in \mathcal{F}_i$ for which $w \in N_{G'}(u)$(by the maximality again of $\mathcal{Q}_{i,10}$). Therefore, as $u, u' \in A_i$ here, $\mathbb{P}(w \in \{v_u, v_{u'}\}) \leq 2/(pd/30) = 60/pd$. Therefore, by Chernoff's bound, $\mathbb{P}(B_v) \leq e^{-\sqrt{d}}$ as $1/d \ll p$. Note that, for each $\{u, u'\} \in \mathcal{F}_i$, the choice of v_u and $v_{u'}$ can affect B_v only if v is within a distance 2 of u or u' in G. Therefore, by the local lemma, we can assume that each vertex in $V(G)$ has at most $\Delta_1 = d^{1/4}$ neighbours among the set $\{v_u, v_{u'} : \{u, u'\} \in \mathcal{F}_i\}$. Let $\mathcal{P}_i' = \mathcal{P}_i + \mathcal{Q}_i + \mathcal{R}_i$, and let $\mathcal{P}_i''$ be the paths in $\mathcal{P}_i'$ with both endvertices in Y_{10}. Then, by the choice from the local lemma, we have that every vertex in $V(G)$ has at most Δ_1 neighbours among the endvertices of $\mathcal{P}_i''$. As, furthermore, none of the endvertices of the paths in $\mathcal{P}_i''$ are in Z, and there are at most $(1 + \varepsilon)n/d$ paths in $\mathcal{P}_i'$ and therefore $\mathcal{P}_i''$, we have that $\mathcal{P}_j''$, $j \in I \cup \{i\}$, is $((1 + \varepsilon)n/d, \Delta_0, \Delta_1)$-bounded, while **E1**–**E4** hold for each $j \in I \cup \{i\}$ by construction. Therefore, if we can show **E5** for i we will get a contradiction to the choice of I, and then we can conclude that $I = [d/2]$, which, as noted above, finishes the proof.

To prove **E5**, we simply need to show that at most $\varepsilon n/d$ of the paths in $\mathcal{P}_i'$ have an endpoint which is not in A_i. Take a maximal set X of endvertices of the paths in $\mathcal{P}_i$ which come from different paths, such that each vertex in X has at most $pd/30$ neighbours in $Y_{10} \setminus (Z \cup Z_i)$. Recall that each vertex in $V(G) \setminus Y$ has at least $(1 - 3\gamma)pd/10$ neighbours in Y_{10} in G'. Thus, each vertex in X has at least $pd/20$ neighbours in $Z_i \cup Z$ in G'.

Now, as the vertices in X are endpoints of different paths in $\mathcal{P}_i + \mathcal{Q}_i$, we have concluded that they share no neighbours in $Y_{10} \setminus Z_i$, and therefore each vertex in $Z \setminus Z_i$ has at most 1 neighbour in G', so there are at most $|Z| \leq 4n/\Delta_0$ edges between X and $Z \setminus Z_0$ in G' by (1). Furthermore, we deduced that $|Z_i| \leq 3n/d$ and so, as $\mathcal{P}_j$, $j \in [d/2]$, is $(n/d^{1/9}, d^{8/9}, d^{8/9})$-bounded, and every vertex in X is an endvertex of some path in $\mathcal{P}_i$, there are at most $|Z_i| \cdot d^{8/9} \leq 3n/d^{1/9}$ edges from X to Z_i. Therefore, as $1/d \ll p, \varepsilon$,

$$|X| \leq \frac{(4n/\Delta_0) + (3n/d^{1/9})}{pd/30} \leq \frac{\varepsilon n}{d},$$

as required. Thus, by the definition of X, there are at most $\varepsilon n/d$ paths in $\mathcal{P}_i'$ which are not in $\mathcal{P}_i''$. This completes the proof of **E5**, and hence the lemma. $\qquad\square$

3.3   Proof of Theorem 1.4

Lemma 3.1 and Lemma 3.3 already allow us to prove Theorem 1.4.

Proof of Theorem 1.4. Let d_0, γ and p be such that $1/d_0 \ll \gamma \ll p \ll \varepsilon$ and let $d \geq d_0$. Let G be a d-regular n-vertex graph. Let $Y \subset V(G)$ be a random set chosen by including each vertex independently at random with probability p. Using the local lemma, and that G is d-regular and $1/d \ll \gamma, p$, we can assume that $d_G(v, Y) = (1 \pm \gamma)pd$ for each $v \in V(G)$ and $|Y| = (1 \pm \gamma)pn$. Note that $G - Y$ has vertex degrees in $(1 \pm \gamma)(1 - p)d$. Therefore, by Lemma 3.1, and as $1/d \ll p$, $G - Y$ contains an $(n/d^{1/9}, d^{8/9}, d^{8/9})$-bounded edge-disjoint collection $\mathcal{P}_1, \ldots, \mathcal{P}_{d/2}$ of path forests which cover all but at most $\varepsilon nd/8$ of the edges of $G - Y$. As there are at most $n \cdot (1 + \gamma)pd \leq \varepsilon nd/8$ edges with a vertex in Y, these path forests contain all but at most $\varepsilon nd/4$ of the edges of G. Then, applying Lemma 3.3 we get an edge-disjoint collection $\mathcal{P}_1', \ldots, \mathcal{P}_{d/2}'$ of path forests which cover all but at most $\varepsilon nd/4$ edges of G such that (simplifying the conclusion) for each $i \in [d/2]$, $\mathcal{P}_i'$ has at most $(1 + \varepsilon/4)n/d$ paths.

Taking $j \in [d/2]$ to maximise $|E(\mathcal{P}_j')|$, we have that $|E(\mathcal{P}_j')| \geq (nd/2 - \varepsilon nd/4)/(d/2) = (1 - \varepsilon/2)n$. Letting k be the number of paths in $\mathcal{P}_j'$, we have that $k \leq (1 + \varepsilon/4)n/d$, and note that we are done if $k \leq n/(d+1)$, so we can assume otherwise. Then, observe that the total number of edges in the $\lfloor n/(d+1) \rfloor$ longest paths in $\mathcal{P}_j'$ is at least

$$(1 - \varepsilon/2)n \cdot \frac{\lfloor n/(d+1) \rfloor}{k} \geq (1 - \varepsilon/2)n \cdot \frac{n/(d+1) - 1}{(1 + \varepsilon/4)n/d} \geq (1 - \varepsilon/2)^2 n \geq (1 - \varepsilon)n.$$

Thus, $\mathcal{P}_j'$ contains a set of at most $n/(d+1)$ paths which cover all but at most εn vertices of G, as required. $\quad\square$

3.4   Almost-spanning nearly-regular subgraphs

The purpose of this section is to prove the following lemma, Lemma 3.5. As discussed in Section 2.2, the removal of dense spots can significantly weaken the regularity of the graph we are decomposing. Given a graph with such weaker conditions, the following lemma partitions most of the edges into a few subgraphs which are all very close to being regular, so that we can later apply Lemma 3.1 to each of them.

Lemma 3.5. *Let $1/d \ll \mu \ll \varepsilon \leq 1$ and $\eta = d^{-0.9}$. Let G be an n-vertex graph with $\Delta(G) \leq d$. Then, there is some $k \in \mathbb{N}$ such that there are edge-disjoint subgraphs $G_1, \ldots, G_k$ in G and some even $d_1, \ldots, d_k \geq \mu d$ such that all but at most εnd edges of G are in $\bigcup_{i \in [k]} G_i$, $\sum_{i \in [k]} d_i \leq (1 + \varepsilon)d$, and, for each $i \in [k]$, G_i is $(1 \pm \eta)d_i$-regular.*

To prove Lemma 3.5, we will use the following two lemmas, which follow using very recent methods of Chakraborti, Janzer, Methuku and Montgomery [12, 13]. For completion, we prove the lemmas in an appendix. Each lemma takes a graph which is roughly regular (with degrees varying by at most a constant multiple C, and a multiple $(1 + \gamma)$, respectively) and finds within it a subgraph whose vertex degrees only differ by at most the addition of a constant multiple of the logarithm of the average degree and which has average degree comparable to the original graph. In the second lemma, the average degree of the subgraph is particularly close to the original average degree and the subgraph contains almost all of the original vertices.

Lemma 3.6. *Let $1/C' \ll 1/C \leq 1$. For any d, if a graph has degrees between d and Cd, then it contains a subgraph with degrees between d' and $d' + C' \log d'$ for some $d' \geq d/C'$.*

Lemma 3.7. *There exists some $C > 0$ such that for each $1/d \ll \gamma \leq 1/100$ the following holds. Let G be a graph with $d \leq \delta(G), \Delta(G) \leq (1 + \gamma)d$. Then, for some $d' \geq (1 - 40\gamma)d$, G contains a subgraph G' with $|V(G')| \geq (1 - 40\gamma)|V(G)|$ and $d' \leq \delta(G'), \Delta(G') \leq d' + C \log d'$.*

To prove Lemma 3.5, we start by taking a maximal collection $H_1, \ldots, H_t$ of nearly-regular subgraphs of G whose average degree is not too far below that of G. It will follow easily from Lemma 3.6 that they will cover most of the edges of G, however, these subgraphs may contain significantly fewer vertices than G (and so, in particular, the sum of the average degree of the subgraphs may exceed d). However, each vertex of G will be in very few of the subgraphs H_i, so we can randomly choose k (for some particular k) almost-spanning subgraphs of G by, for the jth subgraph, allocating each vertex to a different subgraph H_i which contains that vertex and including in the jth subgraph any edge of any H_i which has both vertices allocated to H_i for the jth subgraph. These k subgraphs will not be edge-disjoint. In fact, using the local lemma, we will choose such subgraphs so that each edge in each H_i will appear in many of the k subgraphs; for each of these edges we will then randomly allocate it to one such subgraph and delete it from all the others. By carefully choosing all the probabilities involved, and using again the local lemma, we will be able to get subgraphs G'_i, $i \in [k]$, which contain almost all of the vertices of G and which are nearly regular. Finally, as these subgraphs will not be quite regular enough (as the application of the local lemma cannot preserve as much of the near-regularity as we would like), we will boost the near-regularity of each G'_i using Lemma 3.7 to get a more regular subgraph $G_i \subset G'_i$ with only the loss of relatively few edges or vertices.

Proof of Lemma 3.5. Let γ and β be such that $1/d \ll \gamma \ll \mu \ll \beta \ll \varepsilon$ and note that we can assume $\varepsilon \ll 1$. Let $H_1, \ldots, H_t$ be a maximal edge-disjoint collection of subgraphs of G such that, for each $i \in [t]$, there is some $d'_i \geq \beta d$ such that H_i is $(1 \pm \gamma)d'_i$-regular. Let $H = \bigcup_{i \in [t]} H_i$.

Claim 3.8. *We have $|E(G) \setminus E(H)| \leq \varepsilon nd/3$.*

Proof. Suppose, for contradiction, that $|E(G) \setminus E(H)| > \varepsilon nd/3$. Then, by the well-known folklore result, $G \setminus H$ has a subgraph G' with minimum degree at least $\varepsilon d/3$. Using that $\Delta(G) \leq d$ and $\beta \ll \varepsilon$, apply Lemma 3.6 with $C = 3/\varepsilon$ to find a subgraph H_{t+1} of G' that is $(1 \pm \gamma)d_{t+1}$-regular for some $d_{t+1} \geq \beta d$, contradicting the maximality of the subgraphs H_i, $i \in [t]$. □

For each $i \in [t]$, let $\beta_i = (1 - \gamma)d'_i/d \geq \beta/2$. For each $v \in V(G)$, let I_v be the set of $i \in [t]$ with $v \in V(H_i)$, noting that
$$d \geq d_G(v) \geq \sum_{i \in I_v} d_{H_i}(v) \geq \sum_{i \in I_v} \beta_i d \geq |I_v| \cdot \beta d/2,$$
so that $|I_v| \leq 2/\beta$. Take $k = 1/2\mu$. For every $j \in [k]$, take disjoint subsets $R_1^{(j)}, \ldots, R_t^{(j)}$ of $V(G)$, by determining the location of each $v \in V(G)$ independently and uniformly at random subject to the following rule.

- For each $i \in I_v$, there are exactly $\lfloor \beta_i k \rfloor$ values of $j \in [k]$ for which $v \in R_i^{(j)}$.

To see that this is possible, note that, for each $v \in V(G)$, $\sum_{i \in I_v} \beta_i d \leq \sum_{i \in I_v} d_{H_i}(v) \leq d$, so that $\sum_{i \in I_v} \beta_i \leq 1$.

For each $i \in [t]$, let $H'_i \subset H_i$ be the subgraph of edges $e \in E(H_i)$ such that the number of $j \in [k]$ with $V(e) \subset R_i^{(j)}$ is $(1 \pm \beta)\beta_i^2 k$. Let $H' = \bigcup_{i \in [t]} H'_i$.

Claim 3.9. *With positive probability, $|E(H) \setminus E(H')| \leq \varepsilon nd/6$ and, for each $i \in [t]$ and $j \in [k]$, if $R_i^{(j)} \neq \emptyset$, then $H'_i[R_i^{(j)}]$ is $(1 \pm \beta)\beta_i^2 d$-regular.*

Proof. For each $v \in V(G)$ and $i \in I_v$, let $B_{v,i}$ be the event that at least $\varepsilon d'_i/3$ edges incident to v in H_i are not in H'_i. Let $v \in V(G)$ and $i \in I_v$, and suppose we know which sets $R_i^{(j)}$, $j \in [k]$, contain v. Then, for each $u \in N_{H_i}(v)$ the probability that $uv \in E(H')$, or, equivalently, that u is not in $(1 \pm \beta)\beta_i^2 k$ of the $\lfloor \beta_i k \rfloor$ sets $R_i^{(j)}$, $j \in [k]$ which contain v, is, by Chernoff's bound applied to the appropriate hypergeometric variable, at most $\exp(-\beta^2 \beta_i^2 k/12) \leq \varepsilon/6$, where we have used that $1/k \ll \beta, \varepsilon$ and $\beta_i \geq \beta/2$. Furthermore, these events are

15

independent over each $u \in N_{G_i}(v)$. Thus, by Chernoff's bound, $\mathbb{P}(B_v) \leq 2e^{-\varepsilon^2(1-\gamma)d_i'/10^3} \leq e^{-\sqrt{d}}$, as $d_i' \geq \beta d$ and $1/d \ll \beta, \varepsilon$.

For each $v \in V(G)$, $i \in I_v$ and $j \in [k]$, let $B_v^{i,j}$ be the event that $v \in R_i^{(j)}$ but $d_{H_i'}(v, R_i^{(j)}) \neq (1 \pm \beta)\beta_i^2 d$. Let $v \in V(G)$, $i \in I_v$ and $j \in [k]$ and suppose we know which sets v lies in, and that one of these sets is $R_i^{(j)}$. Then, for each $u \in N_{H_i}(v)$ and $j \in [k]$ the probability that $u \in R_i^{(j)}$ is $\lfloor \beta_i k \rfloor / k = (1 \pm \gamma)\beta_i$, and, when this happens, the probability that, of the $\lfloor \beta_i k \rfloor - 1$ other sets $R_i^{(j)}$ which contain v, u is not in $(1 \pm \gamma)(\beta_i^2 k - 1)$ of them is, by Chernoff's bound applied to the appropriate hypergeometric variable, at most $\exp(-\gamma^2 \beta_i^2 k / 12) \leq \gamma$. Thus, for each $u \in N_{H_i}(v)$ and $j \in [k]$ with $v \in R_i^{(j)}$, the probability that $u \in R_i^{(j)}$ and $u \in N_{H_i'}(v)$ is $(1 \pm 2\gamma)\beta_i$. Note furthermore, that these events are independent over each $u \in N_{G_i}(v)$ and $d_{H_i}(v) = (1 \pm \gamma)d_i' = (1 \pm 3\gamma)\beta_i d$. Thus, $\mathbb{P}(B_v^{i,j}) \leq e^{-\sqrt{d}}$ by Chernoff's bound.

Note that any bad event we have defined depends on the placement of a vertex w only if the vertex v in the subscript is within a distance 1 of each other in G, where $\Delta(G) \leq d$, and the placement of each vertex w therefore affects at most $(d+1) \cdot (2/\beta) \cdot k \leq d^2$ events (where we have used that, for each $v \in V(G)$, $|I_v| \leq 2/\beta$, and $1/d \ll \mu, \beta$). Thus, by the local lemma, with positive probability none of the events $B_{v,i}$, $v \in V(G)$ and $i \in I_v$, or $B_v^{i,j}$, $v \in V(G)$, $i \in I_v$ and $j \in [k]$ hold. If none of the events $B_v^{i,j}$, $v \in V(G)$, $i \in I_v$ and $j \in [k]$ hold, then, for each $i \in [t]$ and $j \in [k]$, if $R_i^{(j)} \neq \emptyset$, then $H_i'[R_i^{(j)}]$ is $(1 \pm \beta)\beta_i^2 d$-regular. We will show that if none of the events $B_{v,i}$, $v \in V(G)$ and $i \in I_v$, hold, then $|E(H) \setminus E(H')| \leq \varepsilon nd/6$, which completes the proof of the claim.

For this, note that, if none of the events $B_{v,i}$, $v \in V(G)$ and $i \in I_v$, hold, then

$$|E(H) \setminus E(H')| = \frac{1}{2} \sum_{v \in V(G)} (d_H(v) - d_{H'}(v)) = \frac{1}{2} \sum_{v \in V(G)} \sum_{i \in I_v} (d_{H_i}(v) - d_{H_i'}(v)) \leq \frac{1}{2} \sum_{v \in V(G)} \sum_{i \in I_v} \frac{\varepsilon d_i'}{3}$$

$$\leq \frac{1}{2} \sum_{v \in V(G)} \sum_{i \in I_v} \frac{\varepsilon d_{H_i}(v)}{3} \leq \frac{1}{2} \sum_{v \in V(G)} \frac{\varepsilon d}{3} = \frac{\varepsilon nd}{6}.$$

Thus, with positive probability, the properties in the claim hold. $\boxdot$

Fix a choice of $R_i^{(j)}$, $i \in [t]$ and $j \in [k]$, satisfying the properties in the claim. For each $i \in [t]$ and $e \in E(H_i')$, independently and uniformly at random assign e to some $j \in [t]$ with $V(e) \subset R_i^{(j)}$ and let $f(e) = j$. Then, for each $j \in [k]$, let G_j' be the graph with vertex set $\bigcup_{i \in [t]} R_i^{(j)}$ and edge set $\bigcup_{i \in [t]} \{e \in E(H_i') : f(e) = j\}$. Let $d' = d/k$ and note that the graphs G_j', $j \in [k]$, are, by design, edge disjoint.

Claim 3.10. *With positive probability, for each $j \in [k]$, G_j' is $(1 \pm 4\beta)d'$-regular.*

Proof. Let $v \in V(G)$, $i \in I_v$ and $j \in [k]$ with $v \in R_i^{(j)}$. Recall that the degree of v in $H_i'[R_i^{(j)}]$ is $(1 \pm \beta)\beta_i^2 d$, and each edge e in H_i' has $(1 \pm \beta)\beta_i^2 k$ indices $j' \in [k]$ with $V(e) \subset R_i^{(j')}$. Therefore, the expected degree of v in G_j' is at least $(1 - \beta)\beta_i^2 d / (1 + \beta)\beta_i^2 k$ and at most $(1 + \beta)\beta_i^2 d / (1 - \beta)\beta_i^2 k$, so that

$$\mathbb{E}(d_{G_j'}(v)) = (1 \pm 3\beta)d/k = (1 \pm 3\beta)d'.$$

Thus, as $1/d \ll \mu, \beta$ and $k = 1/2\mu$, using Chernoff's bound, for each $v \in V(G)$, $j \in [k]$, we have with probability at most $e^{-\sqrt{d}}$ that if $v \in V(G_j')$ then $d_{G_j'}(v) \neq (1 \pm 4\beta)d'$. For each edge e, which graph G_j', $j \in [k]$, is in affects the degrees of at most $2d - 1$ vertices in each of the k graphs, and each degree is affected by the location of at most d edges, and therefore an application of the local lemma implies the claim. $\boxdot$

Using this claim, we can thus partition $H' \subset G$ into edge-disjoint graphs G_i, $i \in [k]$, in G such that, for each $i \in [k]$, G_i' is $(1 \pm 4\beta)d'$-regular. For each $i \in [k]$, by Lemma 3.7 (with $\gamma = 10\beta$ and $d = 1 - 4\beta$) we can find

some $G_i \subset G_i'$ and even d_i with $(1 - 400\beta)(1 - 4\beta)d' \leq d_i \leq (1 + 5\beta)d'$ such that $|V(G_i)| \geq (1 - 400\beta)|V(G_i')|$ is $(1 \pm \eta)d_i$-regular (where we have used that $\eta = d^{-0.9}$ and $1/d \ll \beta, 1/k$). Note that,

$$\sum_{i \in [k]} |E(G_i') \setminus E(G_i)| \leq \sum_{i \in [k]} ((1 + 4\beta)d' \cdot |V(G_i) \setminus V(G_i')| + n \cdot ((1 + 4\beta) - (1 - 400\beta)(1 - 4\beta))d'$$

$$\leq k \cdot (1 + 4\beta)d' \cdot 400\beta n + k \cdot n \cdot 410\beta d' \leq 1000\beta dn \leq \varepsilon n/3,$$

as $\beta \ll \varepsilon$ and $d' = d/k$. Thus, combined with the properties from Claim 3.8 and Claim 3.9, we have that the graphs $G_1, \ldots, G_k$ cover all but at most $2\varepsilon nd/3 + \varepsilon nd/6 \leq \varepsilon nd$ edges of G, so that, as $\mu \ll \beta$, they satisfy the conditions of the lemma as

$$\sum_{i \in [k]} d_i \leq (1 + 5\beta) \sum_{i \in [k]} d' = (1 + 5\beta)d \leq (1 + \varepsilon)d. \qquad \square$$

3.5 Proof of Lemma 2.3

Given Lemmas 3.1, 3.3 and 3.5 it is now short work to prove Lemma 2.3.

Proof of Lemma 2.3. Recall that we have $1/d \ll \gamma \ll \varepsilon \leq 1$ and G is an n-vertex graph with maximum degree at most d in which all but at most γn vertices have degree at least $(1 - \gamma)d$. Let p satisfy $\gamma \ll p \ll \varepsilon$. Let $Y \subset V(G)$ be a random set chosen by including each vertex independently at random with probability p. Using the local lemma, we can assume that, for every $v \in V(G)$ with $d_G(v) \geq (1 - \gamma)d$, we have $d_G(v, Y) = (1 \pm 2\gamma)pd$, and that $|Y| \leq (1 + \gamma)pn$.

Let X be the set of vertices $v \in V(G) \setminus Y$ with $d_G(v, Y) = (1 \pm 2\gamma)pd$, so that $|X| \geq n - (1 + \gamma)pn - \gamma n$, and let $G' = G[X]$. Note that

$$|E(G) \setminus E(G')| \leq ((1 + \gamma)pn + \gamma n) \cdot d \leq \varepsilon nd/4.$$

Let $\eta = d^{-0.9}$. By Lemma 3.5, there is some $k \in \mathbb{N}$ such that there are edge-disjoint subgraphs $G_1, \ldots, G_k$ in G' and some even $d_1, \ldots, d_k \geq \gamma d$ such that all but at most $\varepsilon nd/4$ edges of G' are in $\bigcup_{i \in [k]} G_i$, $\sum_{i \in [k]} d_i \leq (1 + \varepsilon/10)d$ and, for each $i \in [k]$, G_i is $(1 \pm \eta)d_i$-regular. Note that $k \leq 2/\gamma$. For each $i \in [k]$, using Lemma 3.1, find in G_i an $(n/d_i^{1/8}, d_i^{7/8}, d_i^{7/8})$-bounded edge-disjoint collection $\mathcal{P}_1^i, \ldots, \mathcal{P}_{d_i/2}^i$ of path forests which cover all but at most $\varepsilon nd_i/3$ edges of G. If $(\sum_{i \in [k]} d_i)/2 < d/2$, then relabel these path forests as $\mathcal{P}_j$, $1 \leq j \leq (\sum_{i \in [k]} d_i)/2$ and, for each $(\sum_{i \in [k]} d_i)/2 \leq i \leq d/2$, let $\mathcal{P}_i$ be an empty path forest. If $(\sum_{i \in [k]} d_i)/2 \geq d/2$, then let $\mathcal{P}_j$, $j \in [d/2]$, be path forests from $\mathcal{P}_j^i$, $j \in [d_i/2]$ and $i \in [k]$, which maximise the total number of edges in $\mathcal{P}_j$, $j \in [d/2]$. Note that, as $\sum_{i \in [k]} d_i \leq (1 + \varepsilon/10)d$, these path forests contain all but at most $\varepsilon nd/4$ of the edges in $\mathcal{P}_j^i$, $j \in [d_i/2]$ and $i \in [k]$.

Note that, as $k \leq 2/\gamma$ and $1/d \ll \gamma$, the collection of path forests $\mathcal{P}_i$, $i \in [d/2]$, is $(n/d^{1/9}, d^{8/9}, d^{8/9})$-bounded, and contains all but at most $\varepsilon n \cdot \sum_{i \in [t]} d_i/4 \leq \varepsilon nd/4$ of the edges of $\mathcal{P}_j^i$, $j \in [d_i/2]$ and $i \in [k]$. Then, by Lemma 3.3, there is a $((1 + \varepsilon)n/d, d^{1/4}, d^{1/4})$-bounded collection of edge-disjoint path forests in G' which covers all of the edges of $\mathcal{P}_i$, $i \in [d/2]$. Thus, this collection of paths forests covers all but at most $\varepsilon nd/4$ of the edges of $\mathcal{P}_j^i$, $j \in [d_i/2]$ and $i \in [k]$, and hence all but at most $\varepsilon nd/2$ of the edges of $\bigcup_{i \in [t]} G_i$, and hence all but at most $3\varepsilon nd/4$ of the edges of G', and hence all but at most εnd of the edges of G, as required. $\qquad \square$

4 Connecting to the dense bits

In this section we prove Lemma 2.4. In Section 4.1, we show that there will exist a 'good sample' X which samples certain important dense spots appropriately, before using this to connect paths to a maximal set of dense spots through X in Section 4.2.

4.1 Sampling of dense spots

In order to facilitate our use of the local lemma, we need some control over the number of dense spots that we consider for our 'sampling'. We do this using the following definition.

Definition 4.1. *Say a (μ, d, K)-dense graph $H \subset G[X]$ is (k, η, d)-**approximable in G by neighbourhoods in X** if G contains a set V of at most k vertices, which are pairwise a distance at most k apart in G, and such that $|V(H) \triangle \bigcup_{v \in V} N_G(v, X)| \leq \eta d$.*

We now formalise the notion of a good sample. The following definition will be used with $\mathcal{V}$ as the set of vertex sets of the dense spots in the maximal collection $\mathcal{F}$ (see Section 2.2). The set X should be thought of as a random set (with properties chosen to exist together with strictly positive probability by the local lemma) with vertex density p in $V(G)$, which then has approximately the expected size, as well as approximately the expected intersection with each of the sets in $V \in \mathcal{V}$, and, moreover, similar properties with respect to the neighbourhoods of each vertex in all these subsets (see **F1** and **F2**). Finally, and most importantly, if there is an (appropriately approximable) dense spot in $G[X]$, then there must be a larger dense spot H in G such that if H intersects on quite a few vertices in $V \in \mathcal{V}$ then the original smaller dense spot must also intersect that V (this will be used to join the smaller dense spot to one of the dense spots in $\mathcal{F}$ in our application).

Definition 4.2. *A set $X \subset V(G)$ is a $(d, \gamma, k, \eta, K, p, \mathcal{V})$-**good-sample in** G, where $\mathcal{V}$ is a set of subsets of $V(G)$, if the following hold.*

F1 $|X| = (1 \pm \gamma)p|V(G)|$ *and, for any $v \in V(G)$ we have that $d_G(v, X) = (1 \pm \gamma)pd$.*

F2 *For each $V \in \mathcal{V}$, $|V \cap X| = (1 \pm \gamma)p|V|$, and, for any $v \in V(G)$ we have that $d_G(v, V \cap X) = p \cdot d_G(v, V) \pm \gamma pd$.*

F3 *For any $H \subset G[X]$ which is (η, pd, K)-dense and (k, η, pd)-approximable in G by neighbourhoods in X, there exists some $H' \subset G$ which is $(6\eta, d, 2K)$-dense such that, for each $V \in \mathcal{V}$, if $|V \cap V(H')| \geq d/2$, then $V \cap V(H) \neq \emptyset$.*

Given these definitions, using the local lemma (as discussed in Section 2.4.2), it is relatively straightforward to show the existence of the good samples that we will need, as follows.

Lemma 4.3. *Let $1/d \ll \gamma, 1/k, \eta, 1/K, p$. Let G be a d-regular graph and let $\mathcal{V}$ be a set of vertex-disjoint subsets of $V(G)$ which each have size at least $d/2$ and at most $2Kd$. Then, there is a $(d, \gamma, k, \eta, K, p, \mathcal{V})$-good-sample X in G.*

Proof. Let $t = \lceil |V(G)|/d \rceil$, and let A_i, $i \in [t]$, be a partition of $V(G)$ into sets which each have size between $d/2$ and d. Let $X \subset V(G)$ be chosen by including each vertex independently at random with probability p. For each $v \in V(G)$, let B_v be the event that $d_G(v, X) \neq (1 \pm \gamma)pd$. For each $V \in \mathcal{V}$, let B_V be the event that $|V \cap X| \neq (1 \pm \gamma)p|V|$. For each $i \in [t]$, let B_i be the event that $|A_i \cap X| \neq (1 \pm \gamma)p|A_i|$. For each $V \in \mathcal{V}$ and $v \in V(G)$, let $B_{V,v}$ be the event that $d_G(v, V \cap X) \neq p \cdot d_G(v, V) \pm \gamma pd$.

Let $\mathcal{S}$ consist of all of the subsets of $V(G)$ of size at most k such that these vertices are all pairwise at most k apart in G. For each $U \in \mathcal{S}$, let $Y_U = \cup_{u \in U} N_G(u)$ and let B_U be the event that the following do not all hold.

G1 For each $v \in V(G)$, we have $d_G(v, Y_U \cap X) \leq p \cdot d_G(v, Y_U) + \eta pd$.

G2 If Z_U is the set of vertices in Y_U with at least $(1 - 3\eta)d$ neighbours in $G[Y_U]$, then $|Z_U \cap X| \leq p|Z_U| + \eta pd$.

G3 $|Y_U \cap X| = p|Y_U| \pm \eta pd$.

For each $U \in \mathcal{S}$ and $V \in \mathcal{V}$ with $Y_U \cap V \neq \emptyset$, let $B_{U,V}$ be the event that the following does not hold.

18

G4 $|(Y_U \cap V) \cap X| \geq p|Y_U \cap V| - \eta pd$.

Using the local lemma, we can assume that none of the events we have defined holds (see Section 2.4.2).

As no event B_i, $i \in [t]$, or B_v, $v \in V(G)$, holds, we have that **F1** holds. As no event B_V, $V \in \mathcal{V}$, or $B_{V,v}$, $V \in \mathcal{V}$ and $v \in V(G)$, holds, we have that **F2** holds. It is left then only to show that **F3** holds.

For this, suppose we have $H \subset G[X]$ which is (η, pd, K)-dense and (k, η, pd)-approximable in G by neighbourhoods in X. By the approximability of H, there exists some $U \in \mathcal{S}$ such that $|V(H) \triangle (Y_U \cap X)| \leq \eta pd$. Note that every vertex in $V(H) \cap Y_U$ has at least $(1-2\eta)pd$ neighbours in $Y_U \cap X$ as $H \subset G[X]$ is (η, pd, K)-dense, and, by **G1**, if a vertex $v \in V(G)$ has at least $(1-2\eta)pd$ neighbours in $Y_U \cap X$, then it has at least $\frac{1}{p}(d_G(v, Y_U \cap X) - \eta pd) \geq (1-3\eta)d$ neighbours in Y_U in G, and hence belongs to Z_U (as defined in **G2**). Therefore, $Z_U \cap X \subset Y_U \cap X$ has size at least $|V(H) \cap Y_U| \geq |Y_U \cap X| - \eta pd$ and hence

$$p|Z_U| + \eta pd \overset{\mathbf{G2}}{\geq} |Z_U \cap X| \geq |Y_U \cap X| - \eta pd \overset{\mathbf{G3}}{\geq} p|Y_U| - 2\eta pd,$$

so that $|Z_U| \geq |Y_U| - 3\eta d$. Thus, setting $H' := G[Z_U]$, we have $\delta(H') \geq (1-6\eta)d$ by the definition of Z_U in **G2**. Furthermore,

$$|Z_U| \leq |Y_U| \overset{\mathbf{G3}}{\leq} \frac{1}{p}(|Y_U \cap X| + \eta pd) \leq \frac{1}{p}(|V(H)| + 2\eta pd) \leq \frac{1}{p}(Kpd + 2\eta pd) \leq 2Kd,$$

and therefore H' is $(6\eta, d, 2K)$-dense.

To show the last part of **F3** holds for $H' = G[Z_U]$, suppose that $V(H')$ intersects with at least $d/2$ vertices in V, for some $V \in \mathcal{V}$. Then, as Y_U contains Z_U as so also interests with at least $d/2$ vertices in V, and as the event $B_{U,V}$ does not hold, we have $|(Y_U \cap V) \cap X| \geq pd/4$. As $|(Y_U \cap X) \setminus V(H)| \leq \eta pd$, we have by **G4** that $V \cap V(H) \neq \emptyset$. Thus, **F3** holds, and therefore X is a $(d, \gamma, k, \eta, K, p, \mathcal{V})$-good-sample in G, as required. $\qquad \square$

4.2 Proof of Lemma 2.4

We come now to perhaps the most important part of our proof. Using a good sample, say X, provided by Lemma 4.3, we show that X has the property required in Lemma 2.4. That is, given a bounded edge-disjoint collection of path-forests $\mathcal{P}_1, \ldots, \mathcal{P}_{d/2}$ in $G - V(\mathcal{F}) - X$ (as in the set up of Lemma 2.4), we join lots of these paths together in an edge-disjoint fashion before joining all but at most a few of the resulting paths into the dense spots, all in some relatively well spread manner (that is, **A4** and **A5** will hold). To do this, we join as many of the paths together in each path forest using vertices in X (using paths with length at most $2k$ where $1/d \ll 1/k \ll \eta, p$, which together form the path forests denoted in the proof by $\mathcal{Q}_i$, $i \in [d/2]$), and then try to join as many of the paths into the dense spots in a well spread fashion using vertices in X (using paths with length at most k, which together form the path forests denoted by $\mathcal{R}_i$, $i \in [d/2]$). We then show (for Claim 4.4) that all of the $d/2$ resulting path forests will not have that many paths which do not end in some dense spot.

We will show this by contradiction: If we can pick many endvertices of some $\mathcal{P}_i$ that have not had a path joined to them in $\mathcal{Q}_i$ or $\mathcal{R}_i$, then, looking at the iterative neighbourhoods in X from these vertices, if they intersect within distance k we will have a path we could perhaps add to $\mathcal{Q}_i$ (if the path uses no vertices in $\mathcal{Q}_i$). This will imply that not many of these iterating neighbourhoods can grow very large, and thus allow us to connect one of the path ends to an approximable dense spot in $G[X]$, and, hence, to one of the dense spots in the maximal collection $\mathcal{F}$, using **F3**. In order to get a contradiction, we will need these iterating neighbourhoods to avoid vertices in $\mathcal{Q}_i$ and $\mathcal{R}_i$, the edges of all the connecting paths we have found, and some further vertices in order to ensure the final paths will be well spread (for **A4** and **A5**) – all of these vertices we collect into the set Z_i in the proof of Claim 4.4 (see **I1**–**I3**) and all of these edges we delete from G to get G'. We cannot find a subgraph of $G'[X \setminus Z_i]$ with all its degrees close to pd (as we would need for a simple iterating neighbourhood argument to work), but what we can do is find a large subset X_k of almost all of $X \setminus Z_i$ such that we can iteratively find many

neighbours in G' in X from vertices in X_k as long as we only iterate up to k times (by finding neighbours in subsequent sets $X_{k-1} \subset \ldots \subset X_1 \subset X \setminus Z_i$ that we found in reverse order, showing they are large by induction, as at (4)). Then, if there are too many unconnected endvertices of some $\mathcal{P}_i$, from at least one such endpoint we can do this iteration without growing in total very much (as intersecting iterated neighbourhoods would provide a short path connect two more original endpoints together).

Not only will this allow us to connect one more of the path ends in $\mathcal{P}_i$ to a dense spot in $G'[X \setminus Z_i]$, but, in this iteration we can easily find a sequence of vertices which are at distance at most $2k$ apart pairwise in G whose neighbourhoods in X together approximate the dense spot we find in $G'[X \setminus Z_i]$. This allows us to use **F3** to connect the endpoint of $\mathcal{P}_i$ into a dense spot in $\mathcal{F}$, gaining the contradiction that will conclude the proof of the key claim, Claim 4.4, from which we easily deduce Lemma 2.4.

Proof of Lemma 2.4. Let $\mathcal{V} = \{V(G_i) : i \in [t]\}$ and take γ and k such that $1/d \ll \gamma \ll 1/k \ll p, \eta$. Using Lemma 4.3, let $X \subset V(G)$ be a $(d, \gamma/4, k, \eta/6, K/2, p, \mathcal{V})$-good-sample in G. That is, the following hold.

H1 $|X| = (1 \pm \gamma/4)pn$ and, for any $v \in V(G)$ we have that $d_G(v, X) = (1 \pm \gamma/4)pd$.

H2 For each $i \in [t]$, $|V(G_i) \cap X| = (1 \pm \gamma/4)p|V(G_i)|$, and, for any $v \in V(G)$ we have that $d_G(v, V(G_i) \cap X) = p \cdot d_G(v, V(G_i)) \pm \gamma pd/4$.

H3 For any $H \subset G[X]$ which is $(\eta/6, pd, K/2)$-dense and $(k, \eta/6, pd)$-approximable in G by neighbourhoods in X, there exists some $H' \subset G$ which is (η, d, K)-dense such that if H' intersects with some G_i, $i \in [t]$, in at least $d/2$ vertices then $V(G_i) \cap V(H) \neq \emptyset$.

Due to the maximality of $\mathcal{F} = \{G_1, \ldots, G_t\}$, and as any two (η, d, K)-dense spots in G that intersect on some vertex must intersect in at least $(1 - 2\eta)d \geq d/2$ vertices (as $\Delta(G) \leq d$), **H3** immediately implies the following.

H4 For any $H \subset G[X]$ which is $(\eta/6, pd, K/2)$-dense and $(k, \eta/6, pd)$-approximable in G by neighbourhoods in X, there exists some $i \in [t]$ with $V(G_i) \cap V(H) \neq \emptyset$.

We will show that X satisfies the required property. Firstly, **A1** follows from **H1**, and **A2** follows from **H2** and each G_i being (η, d, K)-dense. Now, take any $(2n/d, d^{1/4}, d^{1/4})$-bounded edge-disjoint collection of path-forests $\mathcal{P}_1, \ldots, \mathcal{P}_{d/2}$ in $G - V(\mathcal{F}) - X$. We will show that there exists in $G - (V(\mathcal{F}) \setminus X)$ an edge-disjoint collection of path forests $\mathcal{P}'_1, \mathcal{P}'_2, \ldots, \mathcal{P}'_{d/2}$ such that **A3**–**A6** hold, completing the proof of the lemma.

Let $Y = V(G) \setminus (V(\mathcal{F}) \cup X)$. Let $\mathcal{Q}_1, \ldots, \mathcal{Q}_{d/2}$ be edge-disjoint path forests of paths in $G[X \cup Y]$ with length at most $2k$ and no internal vertices in Y, such that, for each $i \in [d/2]$, all of the endvertices of $\mathcal{Q}_i$ are among the endvertices of $\mathcal{P}_i$, and $\mathcal{Q}_i + \mathcal{P}_i$ creates a path forest. Subject to this, maximise the total number of all of the paths in $\mathcal{Q}_1, \ldots, \mathcal{Q}_{d/2}$. Then, take edge-disjoint path forests $\mathcal{R}_1, \ldots, \mathcal{R}_{d/2}$ of paths in $G[Y \cup X]$ with length at most k, such that, for each $i \in [d/2]$, the paths in $\mathcal{R}_i$ each have one endvertex among the endvertices of $\mathcal{P}_i + \mathcal{Q}_i$ and one endvertex in $X \cap V(\mathcal{F})$ and all of their internal vertices in $X \setminus V(\mathcal{F} \cup \mathcal{P}_i \cup \mathcal{Q}_i)$, and such that each vertex in $V(G)$ is an endvertex in total of at most $\sqrt{d}$ paths in $\mathcal{R}_i$, $i \in [d/2]$, and, for each $j \in [t]$ and $i \in [d/2]$, at most $\sqrt{d}$ of the paths in $\mathcal{R}_i$ end in G_j. Subject to this, maximise the total number of all of the paths in $\mathcal{R}_1, \ldots, \mathcal{R}_{d/2}$. For each $i \in [d/2]$, let $\mathcal{P}'_i$ be the set of paths in $\mathcal{P}_i + \mathcal{Q}_i + \mathcal{R}_i$ which have both endvertices in $V(\mathcal{F})$. By construction, we have that **A3**, **A4** and **A5** hold, so it is left only to prove **A6**.

For each $i \in [d/2]$, let $\mathcal{P}''_i$ be the path forest of paths in $\mathcal{P}_i + \mathcal{Q}_i + \mathcal{R}_i$ which are not in $\mathcal{P}'_i$. We will show the following.

Claim 4.4. *For each $i \in [d/2]$, $\mathcal{P}''_i$ contains at most $16n/Kd$ paths.*

Proof. Suppose, for a contradiction, that there is some $i \in [d/2]$ for which $\mathcal{P}''_i$ contains at least $16n/Kd$ paths. Let Z_i be the set of vertices $v \in X$ such that one of the following holds.

20

I1 v is in a path in $\mathcal{Q}_i + \mathcal{R}_i$.

I2 v is in G_j for some $j \in [t]$ such that at least $\sqrt{d}/2$ of the paths in $\mathcal{P}_i'$ end in G_j.

I3 v is an endpoint of at least $\sqrt{d}/2$ paths in the collection $\mathcal{P}_{i'}'$, $i' \in [d/2]$.

Then,

$$|Z_i| \leq \frac{2n}{d} \cdot 2k + Kd \cdot \frac{2 \cdot n/d}{\sqrt{d}/2} + \frac{d/2 \cdot 2n/d \cdot 2}{\sqrt{d}/2} \leq \frac{10Kn}{\sqrt{d}}. \tag{2}$$

Let G' be the graph G with every edge in $\bigcup_{i \in [d/2]} \mathcal{Q}_i$ and $\bigcup_{i \in [d/2]} \mathcal{R}_i$ removed, so that

$$|E(G) \setminus E(G')| \leq \frac{d}{2} \cdot 2 \cdot \frac{2n}{d} \cdot 2k = 4nk. \tag{3}$$

Let $X_0 = X \setminus Z_i$. Iteratively, for each $1 \leq j \leq k$, let X_j be the set of vertices in X_{j-1} with at least $(1 - \gamma/2)pd$ neighbours in X_{j-1} in G'. We will show, by induction, that, for each $0 \leq j \leq k$,

$$|X \setminus X_j| \leq (16/\gamma)^j \cdot \frac{10Kn}{\sqrt{d}}. \tag{4}$$

Note first that this holds for $j = 0$ by (2). Let then $j \in [k]$ and assume that $|X \setminus X_{j-1}| \leq (16/\gamma)^{j-1} \cdot 10Kn/\sqrt{d}$. Then, as every vertex in X_{j-1} has at least $(1 - \gamma/4)pd$ neighbours in X in G by **H1**, every vertex in $X_{j-1} \setminus X_j$ has at least $\gamma pd/4$ neighbouring edges which are either in $G[X] \setminus G'$ or which lead to a vertex of $X \setminus X_{j-1}$. Thus, using **H1** for each vertex in $X \setminus X_{j-1}$,

$$|X_{j-1} \setminus X_j| \cdot \frac{\gamma pd}{4} \leq |X \setminus X_{j-1}| \cdot 2pd + E(G[X] \setminus G') \stackrel{(3)}{\leq} (16/\gamma)^{j-1} \cdot \frac{10Kn}{\sqrt{d}} \cdot 2pd + 4nk \leq 3 \cdot (16/\gamma)^{j-1} \cdot \frac{10Kn}{\sqrt{d}} \cdot pd,$$

so that $|X_{j-1} \setminus X_j| \leq (12/\gamma) \cdot (16/\gamma)^{j-1} \cdot 10Kn/\sqrt{d}$, and hence

$$|X \setminus X_j| \leq (12/\gamma) \cdot (16/\gamma)^{j-1} \cdot \frac{10Kn}{\sqrt{d}} + (16/\gamma)^{j-1} \cdot \frac{10Kn}{\sqrt{d}} \leq (16/\gamma)^j \cdot \frac{10Kn}{\sqrt{d}},$$

as required. This completes the proof of the induction hypothesis, and therefore, (4) holds for every $0 \leq j \leq k$. In particular, then, $|X \setminus X_k| \leq (16/\gamma)^k \cdot 10Kn/\sqrt{d} \leq n/d^{1/3}$, as $1/d \ll \gamma, 1/k, 1/K$.

Now, recalling that $\mathcal{P}_i''$ contains at least $16n/Kd$ paths, we will show that at least $8n/Kd$ of these paths has an endpoint which is not in $V(\mathcal{F})$ which has at least $pd/4$ neighbours in G' in X_k. Suppose to the contrary that there is a set A of at least $8n/Kd$ of the endpoints of $\mathcal{P}_i''$ which are each not in $V(\mathcal{F})$ and have at most $pd/4$ neighbours in G' in X_k. For each $v \in A$, as $v \notin V(\mathcal{F})$, and as $\mathcal{P}_1, \ldots, \mathcal{P}_{d/2}$ is $(2n/d, d^{1/4}, d^{1/4})$-bounded, v is the endvertex of at most $d^{1/4}$ paths in $\mathcal{P}_1, \ldots, \mathcal{P}_{d/2}$, and, hence, has degree at most $d^{1/4}$ in $G - G'$. Thus, by the choice of A and by **H1**, v has at least $pd/2$ neighbours in G in $X \setminus X_k$. On the other hand, again as $\mathcal{P}_1, \ldots, \mathcal{P}_{d/2}$ is $(2n/d, d^{1/4}, d^{1/4})$-bounded, each vertex in $X \setminus X_k$ is the neighbour of at most $d^{1/4}$ of the vertices in A (a subset of the endpoints of $\mathcal{P}_i$). Therefore,

$$\frac{4pn}{K} = \frac{8n}{Kd} \cdot \frac{pd}{2} \leq |A| \cdot \frac{pd}{2} \leq |X \setminus X_k| \cdot d^{1/4} \leq \frac{n}{d^{1/12}},$$

a contradiction as $1/d \ll 1/K, p$.

Therefore, $\mathcal{P}_i''$ contains at least $8n/Kd$ paths with an endpoint which is not in $V(\mathcal{F})$ and has at least $pd/4$ neighbours in G' in X_k. Thus, taking $r = 8n/Kd$, we can let $v_1, \ldots, v_r \notin V(\mathcal{F})$, be endvertices of different paths in $\mathcal{P}_i''$ which each have at least $pd/4$ neighbours in X_k in G'.

Now, by the maximality of the path forests $\mathcal{Q}_1, \ldots, \mathcal{Q}_{d/2}$ and **I1**, there is no path of length at most $2k$ in $G' - Z_i$ with internal vertices in X between any two of the vertices v_j, $j \in [r]$. Therefore, we can pick some $j \in [r]$

for which there are at most $|X|/r \leq 2pd/r \leq Kpd/2$ vertices in $X \setminus Z_i$ which can be reached by a path of length at most k from v_j in G' with vertices otherwise in $X \setminus Z_i$. Let $w_1 = v_j$ and $A_1 = N_{G'}(v_j, X_k)$, so that $|A_1| \geq pd/4$. Then, for each $2 \leq i' \leq k$, if possible, pick $w_{i'} \in A_{i'}$ with $|N_{G'}(w_{i'-1}, X_{k-i'+1} \setminus A_{i'-1})| \geq \eta pd/12$ and let $A_{i'} = A_{i'-1} \cup N_{G'}(w_{i'}, X_{k-i'+1})$, and, otherwise, stop. As each $|A_{i'}| \leq Kpd/2$ and $|A_{i'} \setminus A_{i'-1}| \geq \eta pd/12$, this must stop for some $i' < k$ as $1/k \ll \eta, 1/K$. Let $H = G'[A_{i'}]$. We will show that H is $(\eta/6, pd, K/2)$-dense and $(k, \eta/6, pd)$-approximable in G by neighbourhoods in X.

As $i' < k$, for each $v \in A_{i'} \subset X_{k-i'+1}$, v has at least $(1 - \gamma/2)pd$ neighbours in G' in $X_{k-i'}$ and at most $\eta pd/12$ neighbours in $X_{k'-i'+1} \setminus A_{i'}$, so at least $(1 - \eta/6)pd$ neighbours in $A_{i'}$. Thus, as $|V(H)| = |A_{i'}| \leq Kpd/2$, H is $(\eta/6, pd, K/2)$-dense. Furthermore, each vertex $w_1, \ldots, w_{i'} \in X_1$ has at most $2\gamma pd$ neighbours in $G[X]$ outside of H by **H1** and the definition of X_1, and therefore, as $\gamma \ll 1/k, \eta, p$, the vertices altogether have at most $k \cdot (2\gamma pd) \leq \eta pd/6$ vertices outside of $V(H)$ in X. Therefore, as $V(H) \subset \bigcup_{i'' \in [i']} N_G(w_{i''}, X)$, $H \subset G[X]$ is $(k, \eta/6, pd)$-approximable in G by neighbourhoods in X.

Therefore, by **H4**, there exists some $m \in [t]$ with $V(G_m) \cap V(H) \neq \emptyset$. Then, we can find a path from v_j to $V(G_m) \cap V(H)$ with length at most k, which contradicts the maximality of the path forests $\mathcal{R}_{i''}$, $i'' \in [d/2]$, where we recall that $V(H)$ contains no vertex in Z_i and use **I1**, **I2** and **I3**. $\boxdot$

As all but at most $(1 - \varepsilon)d$ edges of each path in $\mathcal{P}_i''$, $i \in [d/2]$, can be decomposed into paths of length $(1 - \varepsilon)d$, and $1/K \ll \varepsilon$, by Claim 4.4 all but at most $(d/2) \cdot (1 - \varepsilon)d \cdot (16n/Kd) \leq \varepsilon nd/4$ edges of $E(\mathcal{P}_1'' \cup \ldots \cup \mathcal{P}_{d/2}'')$ can be decomposed into copies of $P_{(1-\varepsilon)d}$, so that **A6** holds, as required. $\square$

5 Decomposing inside the dense bits

In this section, we will (mostly) decompose the dense spots with attached paths (see Section 2.2). Firstly we note that so far we have no conditions on the dense spots that imply they are connected, which we will need to connect various paths together. Therefore, in Section 5.1 we give the notation of connectivity we will use (Definition 5.1) and show that we can partition the vertices of our dense spots into (essentially) connected dense spots (Lemma 5.3) and in each such connected dense spot find a small vertex subset which can connect each pair of vertices in the connected dense spot using many possible short edge-disjoint paths using internal vertices only in that vertex subset. In Section 5.2, we show we can decompose a dense spot into many path forests of some specified numbers of vertices and few overall paths (which we later connect up, including to an attached path where appropriate, using our connectivity property). Finally, in Section 5.3, we put this all together and prove Lemma 2.5, thus completing the proof of all our key lemmas, and hence Theorem 1.2.

5.1 Connectivity

First, we define our connectivity.

Definition 5.1. *A graph G is (ζ, λ, d)-**connected** if, for all $A, B \subseteq V(G)$ with $|A|, |B| \geq \zeta d$, and for all $F \subseteq E(G)$ such that $|F| \leq \lambda d^2$, there exists a path between A and B in $G - F$.*

Observe that any graph with fewer than $2\zeta d$ vertices is trivially (ζ, λ, d)-connected. We also have the following observation.

Observation 5.2. *If G is (ζ, λ, d)-connected and $F \subseteq E(G)$ satisfies $|F| \leq \tau d^2$, then $G - F$ is $(\zeta, \lambda - \tau, d)$-connected.* $\square$

We now show how to partition a dense graph into highly connected subgraphs.

Lemma 5.3. *Let $1/d \ll \lambda \ll 1/K, \beta$ with $\beta \leq 1/4$. Let G be a (β, d, K)-dense graph. Then, for some $t \leq 2K$, there exists a set $J \subseteq V(G)$ and vertex-disjoint subgraphs $G_1, \ldots, G_t \subset G$ with $V(G) = V(G_1) \cup \ldots \cup V(G_t)$ such that*

J1 $|J| \leq \sqrt{\lambda} d,$

J2 $G_i - J$ *is $(2\beta, d, K)$-dense for each $i \in [t]$, and*

J3 G_i *is $(\lambda^{1/4}, \lambda, d)$-connected for each $i \in [t]$.*

Proof. Set $\zeta := \lambda^{1/4}$. Initialise $\mathcal{S} = \{G\}$, and, while there is some $G' \in \mathcal{S}$ which is not (ζ, λ, d)-connected, take such a G' and use this to find a set $F \subseteq E(G')$ and partition $V(G') = V_1 \cup V_2$ such that $|F| \leq \lambda d^2$, $|V_1|, |V_2| \geq \zeta n$ and $e_{G-F}(V_1, V_2) = 0$, before deleting the edges of F and replacing G' in $\mathcal{S}$ by $G'[V_1]$ and $G'[V_2]$.

Once this process terminates, for $t := |\mathcal{S}|$, enumerate $\mathcal{S}$ as $\{G_1, \ldots, G_t\}$. Note that, from the process, if $i \neq j$, then we have deleted all of the edges in G between $V(G_i)$ and $V(G_j)$, while each G_i is (ζ, λ, d)-connected (and so, in particular, **J3** holds) with size at least ζd, and we deleted at most $t \cdot \lambda d^2$ edges from G to get $\bigcup_{i \in [t]} G_i$. Noting that $V(G) = V(G_1) \cup \ldots \cup V(G_t)$ is a partition, we will define a J satisfying **J1** and **J2**, and show that $t \leq 2K$, completing the proof of the lemma.

Note that, $t \leq |V(G)|/(\zeta d) \leq K/\zeta = K\lambda^{-1/4}$, and the number of deleted edges is at most $t \cdot \lambda d^2$. Denote by J the vertices of G which are incident to more than $\lambda^{1/5} d$ deleted edges. Note that

$$|J| \leq \frac{2t \cdot \lambda d^2}{\lambda^{1/5} d} \leq \frac{2K\lambda^{-1/4} \cdot \lambda d^2}{\lambda^{1/5} d} \leq \sqrt{\lambda} d, \tag{5}$$

where we have used that $\lambda \ll 1/K$. Thus, **J1** holds.

Now, for each $i \in [t]$ and $v \in V(G_i) \setminus J$, v is adjacent to at most $\lambda^{1/5} d$ edges in G that were deleted (as $v \notin J$) and at most $|J| \leq \sqrt{\lambda} d$ edges to J. Thus, as G is (β, d, K)-dense and $\lambda \ll \beta$, v has at least $(1 - \beta)d - \sqrt{\lambda} d - \lambda^{1/5} d \geq (1 - 2\beta)d$ neighbours in $G_i - J$. As $|V(G_i)| \leq |V(G)| \leq Kd$, $G_i - J$ is therefore $(2\beta, d, K)$-dense, so **J2** holds. As, for each $i \in [t]$, we now have $|V(G_i)| \geq (1 - 2\beta)d \geq d/2$, we get the improved bound $t \leq 2|V(G)|/d \leq 2K$, completing the proof of the required properties. $\square$

Next, we show that our connectivity implies we can find a short connecting path between any two vertex sets which are not too small.

Lemma 5.4. *Let $1/d, \zeta, \lambda \leq 1$ and $K \geq 1$. Let G be (ζ, λ, d)-connected and have at most Kd vertices, and suppose $U, V \subset V(G)$ are sets with size at least ζn.*

Then, G contains a path from U to V with length at most $8K/\lambda$.

Proof. Let $\ell = 4/(K\lambda)$. Let $U_0 = U$. For each $1 \leq i \leq 2\ell$ in turn, let $U_i = U_{i-1} \cup N_G(U_{i-1})$. Then, $U_0 \subset U_1 \subset U_2 \subset \ldots \subset U_{2\ell}$ is a nested sequence of sets in $V(G)$, and $|V(G)| \leq Kd$, so there must be $> \ell$ values of $i \in [2\ell]$ for which $|U_i \setminus U_{i-1}| \leq 2Kd/\ell$. Therefore, we can find some $j \in [2\ell - 1]$ such that $|U_j \setminus U_{j-1}| \leq 2Kd/\ell$ and $|U_{j+1} \setminus U_j| \leq 2Kd/\ell$. Then,

$$e_G(U_j, V(G) \setminus U_j) = e_G(U_j \setminus U_{j-1}, U_{j+1} \setminus U_j) \leq \left(\frac{2Kd}{\ell}\right)^2 \leq \lambda d^2.$$

As G is (ζ, λ, d)-connected, we therefore must have that either $|U_j| < \zeta d$ or $|V(G) \setminus U_j| < \zeta d$. As $|U_j| \geq |U_0| \geq \zeta d$, we thus must have $|V(G) \setminus U_j| < \zeta d$, and hence $V \cap U_j \neq \emptyset$. That is, there is a path from U to V in G with length at most $j \leq 2\ell - 1 \leq 8K/\lambda$, as required. $\square$

Combined with Observation 5.2, this tells us that many such (almost as short) paths will exist, as follows.

Corollary 5.5. *Let $1/d, \zeta, \lambda \leq 1$ and $K \geq 1$. Let G be (ζ, λ, d)-connected and have at most Kd vertices, and suppose $U, V \subset V(G)$ are sets with size at least ζn.*

Then, G contains at least $\lambda^2 d^2 / 32K$ edge-disjoint paths from U to V with length at most $16K/\lambda$.

Proof. Take a maximal collection $\mathcal{R}$ of edge-disjoint paths from U to V in G which each have length at most $16K/\lambda$ and let the set of all of their edges be E. Then, by Lemma 5.4, the maximality of $\mathcal{R}$ implies that $G - E$ is not $(\zeta, \lambda/2, d)$-connected. Therefore, by Observation 5.2, we must have $|E| \geq \lambda d^2 / 2$. Thus, $|\mathcal{R}| \geq |E|/(16K/\lambda) \geq \lambda^2 d^2 / 32K$, as required. $\square$

From this corollary, we can now show that we can even find many of these paths which only use internal vertices in a pre-selected, small vertex set (chosen randomly within the proof).

Lemma 5.6. $1/d \ll q \ll \lambda, \zeta, 1/K, \varepsilon \leq 1$. *Let G be a (ζ, λ, d)-connected graph such that $\Delta(G) \leq d$, $|V(G)| \leq Kd$, and $\delta(G) \geq \zeta d$.*

Then, there is some $W \subseteq V(G)$ with $|W| \leq 2\varepsilon|V(G)|$ and the following properties.

K1 *For each distinct $v, w \in V(G)$, $G[W]$ contains at least qd^2 edge-disjoint paths of length at most $1/q$ between $N_G(v)$ and $N_G(w)$.*

K2 *For all $v \in V(G)$, $d_G(v, W) \leq 2\varepsilon d$.*

Proof. Let W be a random subset of $V(G)$, sampled by including each vertex independently at random with probability ε. For each $u, v \in V(G)$, we will show that **K1** does not hold for that pair of vertices with probability at most $e^{-\sqrt{d}}$. Taking a union bound over all pairs of vertices will then show that **K1** holds with probability at least $3/4$.

Fix then $u, v \in V(G)$. Using Corollary 5.5, take a collection $\mathcal{R}_{u,v}$ of at least $\lambda^2 d^2 / 32K$ edge-disjoint paths from $N_G(u)$ to $N_G(v)$ in G which each have length at most $16K/\lambda$. Let $X_{u,v}$ be the number of paths in $\mathcal{R}_{u,v}$ whose vertices all lie in W. Note that, as the paths are edge-disjoint and $\Delta(G) \leq d$, changing whether a vertex is in W or not can change $X_{u,v}$ by at most d, so therefore $X_{u,v}$ is d-Lipschitz. Furthermore, each path in $\mathcal{R}_{u,v}$ has all of its vertices in W with probability at least $\varepsilon^{1+16K/\lambda}$, so $\mathbb{E}X_{u,v} \geq (\lambda^2 d^2 / 32K) \cdot \varepsilon^{1+16K/\lambda} \geq 2qd^2$ as $q \ll \lambda, 1/K, \varepsilon$. Therefore, by Lemma 2.7, we have

$$\mathbb{P}(X_{u,v} < qd^2) \leq 2\exp\Big(-\frac{1}{2|V(G)|} \cdot \Big(\frac{qd^2}{d}\Big)^2\Big) \leq 2\exp\Big(-\frac{(qd)^2}{2Kd}\Big) \leq \exp(-\sqrt{d}),$$

as claimed. Thus, as described above, by a suitable union bound, we have that **K1** holds with probability at least $3/4$.

Furthermore, by a simple application of Chernoff's bound and a union bound, as $\Delta(G) \leq d$, with probability at least $3/4$ we have that **K2** holds. Finally, again by a simple application of Chernoff's bound, with probability at least $3/4$ we have that $|W| \leq 2\varepsilon|V(G)|$. Together, then, W satisfies all of the properties in the lemma with positive probability, and hence some such W as desired must exist. $\square$

5.2 Decomposing dense spots into path forests with specified sizes

Using the work in Section 5.1 we can decompose each dense spot in Lemma 2.5 into a few connected dense spots and join up any paths in the same path forest coming into the same connected dense spot. This will leave at most $d/2$ paths coming into each dense spot, as there are at most $d/2$ distinct path forests to begin with. For example, suppose we have paths $P_1, \ldots, P_{d/2}$ which are attached to a well connected dense spot G, where every path P_i has length $d/2$. Having set aside a vertex subset of G to make connections (as in Section 5.1), we will

then wish to find in G plenty (almost $d/2$) edge-disjoint paths of length $(1-\varepsilon)d - d/2$ to join to the paths P_i to get (roughly) a path of length $(1-\varepsilon)d$, while decomposing the rest of the edges mostly into paths of length $(1-\varepsilon)d$. To make this easier, we find instead an edge-disjoint path *forest* of few paths which have the right number of vertices, before using the connectivity property to join them up (including possibly to some path P_i).

To find edge-disjointly path forests with specified numbers of vertices in our dense spots we will use the following lemma. Its proof is similar to that of Lemma 3.1: we partition the vertices into sets A_i, $i \in [s]$, with $s = 2d^{0.15}$ and find many large matchings between each of them, putting them together to find large path forests without too many paths. Instead of decomposing the auxiliary graph K_s into paths of length $(s-1)$, however, we will use Theorem 2.10 to mostly decompose K_s into paths of different lengths, chosen so that the path forests we produce will have the desired size. For this, we start the proof by batching together the desired path forest sizes into groups with similar sizes, as each path we find in K_s will produce many path forests with a similar size.

Lemma 5.7. *Let $1/d \ll \beta, 1/K \ll \varepsilon \le 1$. Let G be a (β, d, K)-dense graph. Let $r \in \mathbb{N}$, and suppose $n_1, \ldots, n_r \in [\varepsilon d, (1-\varepsilon)d]$ are such that $\sum_{i \in [r]} n_i \le (1 - 10^3 \varepsilon)|V(G)|d/2$.*

Then, G contains edge-disjoint path forests $F_1, \ldots, F_r$ such that, for each $i \in [r]$, $|V(F_i)| = n_i$ and F_i contains at most $d^{9/10}$ paths.

Proof. Note that we can assume that $\varepsilon \le 1/10^3$. Let $n = |V(G)| \le Kd$ and note that $r \le |E(G)|/(\varepsilon d) \le Kd/2\varepsilon$. Let $s = 2d^{0.15}$ and $\bar{n} = \lfloor n/s \rfloor \ge (1-\varepsilon)n/s$. Take μ such that $1/d \ll \mu \ll \beta, 1/K$ and $\mu n \in \mathbb{N}$ with $\mu n = 0 \bmod \bar{n}$. We will batch the lengths n_i together so that in each batch they have length varying by only up to μn. For this, for each $i \in [r]$, let $n'_i = \lceil n_i/\mu n \rceil \cdot \mu n \le n_i + \mu n$. Note that there are at most $d/(\mu d) + 1 \le 2/\mu$ different values taken by n'_i, $i \in [r]$, and

$$\sum_{i \in [r]} n'_i \le \sum_{i \in [r]} (n_i + \mu n) \le (1 - 10^3 \varepsilon)n \cdot \frac{d}{2} + \frac{Kd}{2\varepsilon} \cdot \mu n \le (1 - 5 \cdot 10^2 \varepsilon)n \cdot \frac{d}{2}, \tag{6}$$

as $1/d \ll \mu \ll \beta, 1/K, \varepsilon$.

Let $i_{\min} = \min\{n'_i/\bar{n} : i \in [r]\}$ and $i_{\max} = \max\{n'_i/\bar{n} : i \in [r]\}$, so that $i_{\min} \ge \varepsilon n/\bar{n} \ge \varepsilon s$ and $i_{\max} \le (1 - \varepsilon + \mu)n/\bar{n} \le (1 - \varepsilon/2)s$. Let $I = \{i : \exists j \in [r] \text{ s.t. } n'_j = i \cdot \bar{n}\}$, so that $i_{\min} \le i \le i_{\max}$ for each $i \in I$, and $|I| \le 2/\mu$. For each $i \in I$, let m_i be the number of n'_j, $j \in [r]$, with $n'_j = i \cdot \bar{n}$. Let $\eta' = 40d^{-0.2}$ and $d' = (1 - \eta')(1 - 100\beta)d/s$. For each $i \in I$, let $s_i = \lceil m_i/d' \rceil$. Then,

$$\sum_{i \in I} i \cdot s_i \le \sum_{i \in I} i + \sum_{i \in I} i \cdot \frac{m_i}{d'} \le \frac{2s}{\beta} + \sum_{i \in I} i \cdot \frac{|\{j \in [r] : n'_j = i \cdot \bar{n}\}|}{(1 - \eta')(1 - 100\beta)d/s}$$

$$\le \frac{\varepsilon s^2}{10} + \sum_{j \in [r]} \frac{n'_j}{\bar{n}} \cdot \frac{s}{(1 - \eta')(1 - 100\beta)d}$$

$$\le \frac{\varepsilon s^2}{10} + \sum_{j \in [r]} \frac{n'_j}{(1-\varepsilon)n/s} \cdot \frac{s}{(1 - \eta')(1 - 100\beta)d}$$

$$\overset{(6)}{\le} \frac{\varepsilon s^2}{10} + (1 - 5 \cdot 10^2 \varepsilon)n \cdot \frac{d}{2} \cdot \frac{1}{(1-\varepsilon)n/s} \cdot \frac{s}{(1 - \eta')(1 - 100\beta)d}$$

$$\le (1 - \varepsilon/4)\binom{s}{2} \le \binom{(1 - \sqrt{\varepsilon}/3)s}{2}.$$

Thus, by Theorem 2.10, we can take a collection $\mathcal{Q}$ of edge-disjoint paths in K_s such that, for each $i \in I$, there are s_i paths with i vertices in $\mathcal{Q}$. For each $i \in I$, let $\mathcal{Q}_i$ be the set of paths in $\mathcal{Q}$ with i vertices, so that $|\mathcal{Q}_i| = s_i$.

Let $\gamma = 2d^{-0.4}$. Apply Lemma 3.7 to G with $(1-\varepsilon)d$ in place of d and 2ε in place of γ to find a subgraph $G' \subset G$ with $|V(G')| \ge (1 - 100\varepsilon)|V(G)|$ and some $d_0 \ge (1 - 100\varepsilon)d$ such that, for each $v \in V(G')$, $d_{G'}(v) = (1 \pm \gamma)d_0$.

Let $\eta = 4d^{-0.4}$. As in the proof of Lemma 3.1, using the local lemma, take a partition of $V(G')$ as $A_1, \ldots, A_s$ for which the following properties hold.

L1 For each $v \in V(G)$ and $i \in [s]$, $d_G(v, A_i) = (1 \pm \eta)d/s$.

L2 For each $i \in [s]$, $|A_i| = (1 \pm \eta)n/s$.

Note that $\eta' = 40d^{-0.2} = 20\sqrt{\eta}$ and $d' = (1 - \eta')d_0/s$. For each edge $e = jk$ in the complete s-vertex graph K_s, using Lemma 2.9, **L1**, and **L2**, find d' edge-disjoint matchings in $G[A_j, A_k]$ which each have at least $(1 - \eta')n/s$ edges. Call these matchings $M_{e,i}$, $i \in [d']$.

For each $P \in \mathcal{Q}$ and $i \in [d']$, let $F_{P,i}$ be the path forest with vertex set $\cup_{j \in V(P)} A_j$ and edge set $\cup_{e \in E(P)} M_{e,i}$. Observe that, as the paths in $\mathcal{Q}$ are edge-disjoint, the subgraphs $F_{P,i}$, $P \in \mathcal{Q}$ and $i \in [d']$ are edge disjoint. Moreover, observe that, for each $P \in \mathcal{Q}$ and $i \in [d']$, if j is an interior vertex of P, then there are at most $2(|A_j| - (1 - \eta')n/s)$ vertices in A_j which are an endvertex of a path in $F_{P,i}$. Thus, the number of paths in $F_{P,i}$ is at most, by **L2**,

$$2(1 + \eta)\frac{n}{s} + (s - 2) \cdot 2((1 + \eta) - (1 - \eta'))\frac{n}{s} \leq \frac{3n}{s} + 4\eta' n \leq \frac{4n}{s} \leq d^{9/10},$$

as $n \leq Kd$, $s = 2d^{0.15}$ and $1/d \ll 1/K$.

Thus, the path forests $F_{P,i}$, $P \in \mathcal{Q}$ and $i \in [d']$, are edge-disjoint, each have at most $d^{9/10}$ paths, and, for each $i \in I$, there are at least $s_i \cdot d' \geq m_i$ path forests with $\bar{n} \cdot i$ vertices. Thus, by the choice of the m_i, $i \in I$, we can find r edge-disjoint path forests F_i, $i \in [r]$, in G such that, for each $i \in [r]$, F_i has at most $d^{9/10}$ paths and at least $n'_i \geq n_i$ vertices. Iteratively removing $n'_i - n_i$ leaves from F_i, for each $i \in [r]$, then gives the desired paths. $\square$

For our application, we now prove a simple variant of Lemma 5.7 in which the endvertices of the paths additionally satisfy some weak spreadness condition.

Corollary 5.8. *Let $1/d \ll \beta, 1/K \ll \varepsilon \leq 1$. Let G be a (β, d, K)-dense graph. Let $r \in \mathbb{N}$, and suppose $n_1, \ldots, n_r \in [\varepsilon d, (1 - \varepsilon)d]$ are such that $\sum_{i \in [r]} n_i \leq (1 - 2 \cdot 10^3 \varepsilon)|V(G)|d/2$.*

Then, G contains edge-disjoint path forests $F_1, \ldots, F_r$ such that, for each $i \in [r]$, $n_i \leq |V(F_i)| \leq (1 + \varepsilon)n_i$ and F_i contains at most $d^{9/10}$ paths, and each vertex in G appears as the endpoint of at most $d^{19/20}$ paths in $F_1, \ldots, F_r$.

Proof. For each $i \in [r]$, let $n'_i = (1 + \varepsilon)n_i$, noting that

$$\sum_{i \in [r]} n'_i = (1 + \varepsilon) \sum_{i \in [r]} n_i \leq (1 - \varepsilon)(1 - 2 \cdot 10^3 \varepsilon) \cdot \frac{|V(G)|}{2} \leq (1 - 10^3 \varepsilon) \cdot \frac{|V(G)|}{2}.$$

Therefore, by Lemma 5.7, we can find edge-disjoint path forests $F'_1, \ldots, F'_r$ in G such that, for each $i \in [r]$, $|F'_i| = n'_r$ and F'_i contains at most $d^{9/10}$ paths.

For each $1 \leq i \leq r$ in turn do the following to find $F_i \subset F'_i$ with at most $d^{9/10}$ paths. Let Z_i be the vertices of G which appear as the endpoints of at least $d^{19/20}$ different paths in F_j, $j < i$, so that $|Z_i| \leq r \cdot 2d^{9/10}/d^{19/20} \leq (Kd^2/\varepsilon d) \cdot 2d^{-1/20} \leq \varepsilon^2 d \leq \varepsilon n_i$. Thus, we can iteratively delete any endpoints in Z_i from the paths in F'_i until they are all outside of Z_i, to get F_i, while deleting at most $|Z_i| \leq \varepsilon n_i$ vertices, so that, hence, $|F_i| \geq n_i$. By the choice of Z_i, $i \in [r]$, the path forests F_i, $i \in [r]$, have our desired properties. $\square$

We now use Corollary 5.8 to deduce an almost decomposition of a connected dense spot with a few attached paths, as follows.

Lemma 5.9. *Let $1/d \ll \lambda \ll p, 1/K \ll \varepsilon \le 1$. Let G be a $(\lambda^{1/4}, \lambda/2, d)$-connected graph which contains a set $J \subset V(G)$ with $|J| \le \varepsilon d/4$ such that $G - J$ is (p, d, K)-dense, and suppose that $\delta(G) \ge d/8K$ and $\Delta(G) \le d$. Let $P_1, \ldots, P_{d/2}$ be edge-disjoint paths which each have exactly one vertex in G, which is, moreover, an endvertex and in J. Suppose that no vertex of G is the endvertex of more than $\sqrt{d}$ paths P_i, $i \in [d/2]$.*

Then, all but at most $\varepsilon |V(G)| d$ edges of $G \cup P_1 \cup \ldots \cup P_{d/2}$ can be decomposed into copies of $P_{(1-\varepsilon)d}$.

Proof. Let q be such that $1/d \ll q \ll \lambda$. By Lemma 5.6, there is a set $W \subset V(G)$ with $|W| \le \lambda |V(G)|$ and the following property.

M For each distinct $v, w \in V(G)$, there are at least qn^2 edge-disjoint paths with length at most $1/q$ between $N_G(v)$ and $N_G(w)$ in $G[W]$ with all their vertices in W.

Now, as $G - J$ is (p, d, K)-dense, and $|W| \le \lambda d$, $G' := G - J - W$ is $(2p, d, K)$-dense, and

$$|V(G')| \ge |V(G) \setminus J| - |W| \ge \delta(G - J) - \lambda d \ge (1 - p)d - pd = (1 - 2p)d. \tag{7}$$

For each $i \in [d/2]$, let ℓ_i be the length of P_i, and let $n_i \in [(1 - \varepsilon)d]$ be such that $\ell_i + n_i = 0 \mod (1 - \varepsilon)d$. Note that

$$\sum_{i \in [d/2]} n_i \le (1 - \varepsilon)d \cdot d/2 \overset{(7)}{\le} (1 - \varepsilon/4)|V(G')|d/2.$$

Using this, let s be the largest integer such that

$$(s - d/2) \cdot (1 - \varepsilon)d + \sum_{i \in [d/2]} n_i \le (1 - \varepsilon/4)|V(G')|d/2, \tag{8}$$

and, for each $d/2 < i \le s$, let $\ell_i = 0$ and $n_i = (1 - \varepsilon)d$. Note that, by the choice of s,

$$(1 - \varepsilon/4)|V(G')|d/2 \overset{(8)}{\ge} \sum_{i \in [s]} n_i = (s - d/2) \cdot (1 - \varepsilon)d + \sum_{i \in [d/2]} n_i \ge (1 - \varepsilon/4)|V(G')|d/2 - (1 - \varepsilon)d$$

$$\ge (1 - \varepsilon)|V(G)|d/2 \ge |E(G)| - \varepsilon|V(G)|d/2, \tag{9}$$

where we have used that $\Delta(G) \le d$.

By Corollary 5.8, there are edge-disjoint path forests $F_1, \ldots, F_s \subset G'$, each of at most $d^{9/10}$ paths, such that $n_i \le |V(F_i)| \le (1 + \varepsilon)n_i$ for each $i \in [s]$ and every vertex in $V(G_j)$ appears as the endvertex of at most $d^{19/20}$ of the paths in $F_1, \ldots, F_s$. Now, for each $i \in [s]$, let E_i be a set of pairs of vertices from the endvertices of F_i and P_i so that $P_i \cup F_i + E_i$ is a path, and note that $|E_i| \le d^{9/10}$ as F_i contains at most $d^{9/10}$ paths. Let $F \subset \{(i, xy) : i \in [d/2], xy \in E_i\}$ be a maximal set for which there are edge-disjoint paths $R_{i,xy}$, $(i, xy) \in F$, for which the following hold.

N1 For each $(i, xy) \in F$, $R_{i,xy}$ is an x, y-path in G with length at most $2/q$ and internal vertices in W.

N2 For each $i \in [d/2]$, the internal vertices of $R_{i,e}$, $(i, e) \in F$, are all distinct.

N3 Every vertex in $V(G)$ appears in at most $2d^{199/200}$ edges in $\bigcup_{(i,xy) \in F : v \notin \{x,y\}} R_{i,xy}$.

Suppose for contradiction that there is some (i, xy) with $i \in [d/2]$ and $xy \in E_i$ such that $(i, xy) \notin F$. By **N3** and as each vertex in G appears as the endvertex of at most $\sqrt{d}$ paths P_i, $i \in [d/2]$, and at most $d^{19/20}$ of the paths in $F_1, \ldots, F_s$, each vertex is in at most $2d^{199/200} + 2\sqrt{d} + 2d^{19/20} \le qd/4$ edges in paths in $\bigcup_{(i,xy) \in F} R_{i,xy}$. Using **M**, find a set $\mathcal{R}$ of qd^2 edge-disjoint paths of length at most $1/q$ between $N_G(x)$ and $N_G(y)$ and internal vertices in W. Note that, due to the maximality of F, every path in $\mathcal{R}$ must either contain an edge of $R_{i',e}$,

27

$(i', e) \in F$, or contain an internal vertex in $R_{i,e}$, $(i, e) \in F$, or contain a vertex appearing in at most $2d^{199/200}$ edges in $\bigcup_{(i', e) \in F} R_{i', e}$, or there is an edge from its endpoints to $\{x, y\}$ which is in $\bigcup_{(i, xy) \in F} R_{i, xy}$. However, in order, this implies that

$$|\mathcal{R}| \leq s \cdot d^{9/10} \cdot \frac{2}{q} + d^{9/10} \cdot \frac{2}{q} \cdot d + \frac{s \cdot 2 \cdot d^{9/10} \cdot (2/q)}{2d^{199/200}} \cdot d + 2 \cdot d^{9/10} \cdot d \leq \frac{qd^2}{2}, \tag{10}$$

where we have used that $s \leq |V(G)| \cdot (d/2)/((1 - \varepsilon)d) \leq Kd$. As (10) contradicts that $|\mathcal{R}| \geq qd^2$, we must then have that $F = \{(i, xy) : xy \in E_i\}$.

Now, note that, for each $i \in [s]$, $P_i \cup F_i \cup (\bigcup_{e \in E_i} R_{i,e})$ is a path with length at least $\ell_i + n_i$. As, for each $i \in [s]$, $\ell_i + n_i = 0 \mod (1 - \varepsilon)d$, $P_i \cup F_i \cup (\bigcup_{e \in E_i} R_{i,e})$ contains edge-disjoint copies of $P_{(1-\varepsilon)d}$ which cover at least $\ell_i + n_i$ edges. Therefore, in total, we can find edge-disjoint copies of $P_{(1-\varepsilon)d}$ which cover at least $\sum_{i \in [s]} (\ell_i + n_i)$ edges of $G \cup (P_1 \cup \ldots \cup P_{d/2})$. As the number of edges in $P_1 \cup \ldots \cup P_{d/2}$ is $\sum_{i \in [s]} \ell_i$, and, from (9), these paths cover all but at most $\varepsilon |V(G)| d$ edges of $G \cup P_1 \cup \ldots \cup P_{d/2}$, as required. $\qquad \square$

5.3   Proof of Lemma 2.5

Finally, then, we can prove Lemma 2.5.

Proof of Lemma 2.5. Take q and λ with $1/d \ll q \ll \lambda \ll \eta$. We will prove the lemma by induction on $t = |\mathcal{F}|$. If $|\mathcal{F}| = 0$, then we must have that all of the path forests are empty, as there is nowhere for their endvertices to go, and thus the result is trivial in this case. Suppose then that $t \geq 1$ and that the lemma is true if $|\mathcal{F}| = t - 1$, and let $|\mathcal{F}| = t$.

Let $H = G_t$, and, using **B1**, let $X_H \subseteq V(H)$ be such that X_H contains all the vertices of $\cup_{i \in [d/2]} P_i$ in $V(G_i)$ (which are necessarily endpoints), and, for each $v \in V(H)$, $d_H(v, X_H) = (1 \pm \eta)pd$. From this, and as H is (η, d, K)-dense, we have that $H - X_H$ is $(2p, d, K)$-dense. Therefore, by Lemma 5.3, we can find some $s \in \mathbb{N}$ and a set $J \subset V(H) \setminus X_H$ and vertex-disjoint subgraphs $H_1, \ldots, H_s$ whose vertex sets partition $V(H) \setminus X_H$ such that each H_i, $i \in [s]$, is $(\lambda^{1/4}, \lambda, d)$-connected, each $H_i - J$ is $(4p, d, K)$-dense, $s \leq 2K$, and $|J| \leq \sqrt{\lambda} d$. Furthermore, using that H is (η, d, K)-dense, and so each vertex in X_H has at least $(1 - \eta)d - (1 + \eta)pd \geq d/2$ neighbours in $V(H) \setminus X_H$, partition X_H as $\bigcup_{i \in [s]} X_{H,i}$ so that, for each $i \in [s]$ and $v \in X_{H,i}$, v has at least $d/4K$ neighbours in G in $V(H_i)$.

Let $\mathcal{Q}^0 = \emptyset$ and, for each $i \in [d/2]$, let $\mathcal{P}_i^0 = \mathcal{P}_i$. We will show by induction the following claim, where, essentially, for each j in turn, we look to connect paths in $\mathcal{P}_i^{j-1}$ (for each $i \in [d/2]$) with endvertices in H_j together while decomposing at most 1 path from each $\mathcal{P}_i^{j-1}$, $i \in [d/2]$, along with the unused edges in H_j into copies of $P_{(1-\varepsilon)d}$ using Lemma 5.9 (below, these copies of $P_{(1-\varepsilon)d}$ appear as those in the set $\mathcal{Q}^j$, while $\mathcal{P}_i^j$ will be some of the paths in $\mathcal{P}_i^{j-1}$ which have been joined up where necessary so that they have no endpoints in H_j).

Claim 5.10. *For each $0 \leq j \leq s$, there are edge-disjoint path forests $\mathcal{P}_i^j$, $i \in [d/2]$, and $\mathcal{Q}^j$ such that the following hold (where **O1**–**O3** allow us to maintain the properties for the paths $\mathcal{P}_i^j$ that will later allow us to apply the induction hypothesis to these paths and $G_1, \ldots, G_{t-1}$).*

O1 *For each $i \in [t-1]$, there is some $X_i \subseteq V(G_i)$ so that X_i contains all of the vertices of $V(\mathcal{P}_1^j) \cup \ldots \cup V(\mathcal{P}_{d/2}^j)$ in $V(G_i)$, and, for each $v \in V(G_i)$, $d_{G_i}(v, X_i) = (1 \pm \eta)pd$.*

O2 *Each vertex $v \in V(G_1) \cup \ldots \cup V(G_{t-1})$ is an endpoint of in total at most $\sqrt{d}$ paths from $\mathcal{P}_1^j, \ldots, \mathcal{P}_{d/2}^j$.*

O3 *For each $j' \in [t-1]$ and $i \in [d/2]$ at most $\sqrt{d}$ of the paths in $\mathcal{P}_i^j$ have at least one endpoint in $G_{j'}$.*

O4 *For each $i \in [d/2]$, the path forest $\mathcal{P}_i^j$ is contained in $\mathcal{P}_i \cup (\bigcup_{j' \leq j} H_{j'})$ and its paths have both endpoints in $\cup_{j' > j} X_{j'}$.*

28

O5 $\mathcal{Q}^j$ is a collection of edge-disjoint copies of $P_{(1-\varepsilon)d}$ with edges in $E(\bigcup_{j' \leq j} H_{j'}) \cup (\bigcup_{i \in [d/2]} E(\mathcal{P}_i))$ which is edge-disjoint from each $\mathcal{P}_i^j$, $i \in [d/2]$.

O6 The number of edges in $\bigcup_{j' \leq j} H_{j'}$ and $\bigcup_{i \in [d/2]} \mathcal{P}_i$ which are not in $\mathcal{P}_i^j$, $i \in [d/2]$, or $\mathcal{Q}^j$ is at most $\sum_{j' \leq j} \varepsilon |V(H_{j'})| d/2$.

Proof. Note that this is easily true for $j = 0$ using that we set $\mathcal{P}_i^0 = \mathcal{P}_i$ for each $i \in [d/2]$ and $\mathcal{Q}^0 = \emptyset$, where **O1**–**O3** follow from **B1**–**B3**. Assume then that $j \in [s]$ and that we have edge-disjoint path forests $\mathcal{P}_i^{j-1}$, $i \in [d/2]$, and $\mathcal{Q}^{j-1}$ satisfying **O1**–**O6** with $j - 1$ in place of j. For each $i \in [d/2]$, let $E_i \subset X_{H,j}^{(2)}$ be a maximal set of pairs of vertices from the endpoints of $\mathcal{P}_i^{j-1}$ in $X_{H,j}$ such that $\mathcal{P}_i^{j-1} + E_i$ is a path forest (where it may be that none of the pairs in E_i are edges in the graph). Note that in $\mathcal{P}_i^{j-1} + E_i$ at most 1 path will have an endpoint in $V(H_j)$, by the maximality of E_i.

We will now greedily find edge-disjoint paths $R_{i,e}$, $i \in [d/2]$ and $e \in E_i$, such that, the following hold.

P1 For each $i \in [d/2]$ and $xy \in E_i$, $R_{i,xy}$ is an x, y-path in H_j with length at most $2/q$.

P2 For each $i \in [d/2]$, the internal vertices of $R_{i,e}$, $e \in E_i$, are all pairwise disjoint.

P3 Every vertex in $V(H_j)$ appears in at most $2d^{3/4}$ edges in $\bigcup_{i \in [d/2], e \in E_i} R_{i,e}$.

To see this is possible, suppose we have found paths $R_{i,e}$, $(i, e) \in F$, for some $F \subset \{(i, e) : i \in [d/2], e \in E_j^i\}$ and are looking to find the path $R_{i',xy}$, for some $i' \in [d/2]$ and $xy \in E_i$ with $(i', xy) \notin F$. Let $Z_{i',xy}$ be the set of vertices v in H_j which either

- are in at least $d^{3/4}$ edges in $\bigcup_{(i,e) \in F} R_{i,e}$, or

- are an internal vertex of one of the paths $R_{i',e}$, $(i', e) \in F$, or

- are such that xv or yv are in some $R_{i,e}$, $(i, e) \in F$.

By **O3** we have $|F| \leq d^{3/2}$, so that, using **O3** and **O2**, we have

$$|Z_{i',xy}| \leq \frac{d^{3/2} \cdot (2/q)}{d^{3/4}} + (2/q) \cdot \sqrt{d} + 2\sqrt{d} \leq \frac{qd}{4}. \tag{11}$$

Let E_F be the set of all the edges of all the paths $R_{i,e}$, $i \in [d/2]$ and $e \in E_i$, so that $|E_F| \leq (2/q) \cdot d^{3/2} \leq qd^2/2$ by **O3**. By the choice of $X_{H,j}$, and as $x, y \in X_{H,j}$ implies $d_G(x, V(H_j)), d_G(y, V(H_j))$, and as $\lambda \ll 1/K$ and H_j is $(\lambda^{1/4}, \lambda, d)$-connected, by Corollary 5.5 there is a set $\mathcal{R}$ of at least qd^2 edge-disjoint paths of length at most $1/q$ between $N_G(x, V(H_j) \setminus Z_{i',xy})$ and $N_G(y, V(H_j) \setminus Z_{i',xy})$ in G. As $\Delta(G) \leq d$, for each $v \in V(G)$ at most d of the paths in $\mathcal{R}$ go through v. Therefore, as $|E_F| \leq qd^2/2$ and as $|Z_{i',xy}| \leq qd/4$ by (11), we can find a path $R'_{i',xy} \in \mathcal{R}$ which has no vertex in $Z_{i',xy}$ or edge in E_F. Say $R'_{i',xy}$ has endpoints $x' \in N_G(x, V(H_j) \setminus Z_{i',xy})$ and $y' \in N_G(y, V(H_j) \setminus Z_{i',xy})$. Add edges xx' and yy' to $R'_{i',xy}$ and call this path $R_{i',xy}$. Then, by the choice of $Z_{i',xy}$ (so that $xx', yy' \notin E_F$) and E_F, this path avoids all of the edges in E_F, and its internal vertices are disjoint from those of $R_{i',e}$, $(i', e) \in F$, and no vertex in $V(H_j)$ appears in more than $2d^{3/4}$ edges in $\bigcup_{(i,e) \in F \cup \{(i',xy)\}} R_{i,e}$. Therefore, we can pick $R_{i',xy}$ as required, and thus can pick paths $R_{i,e}$, $i \in [d/2]$ and $e \in E_i$ satisfying **P1**–**P3**.

Now, for each $i \in [d/2]$, by the choice of E_i and by **P1** and **P2**, we have that $\mathcal{P}_i^{j-1} \cup (\bigcup_{e \in E_i} R_{i,e})$ is a path forest. Let $\mathcal{P}_i^j$ be the path forest of paths in $\mathcal{P}_i^{j-1} \cup (\bigcup_{e \in E_i} R_{i,e})$ with no endvertices in X_j, so that, therefore, **O4** holds. For each $i \in [d/2]$, by the maximality of E_i we have that there is at most one path in $\mathcal{P}_i^{j-1} \setminus \mathcal{P}_i^j$ – call this path P_i where it exists, and otherwise set $P_i = \emptyset$. Let H'_j be H_j with the edges of the paths in $\mathcal{P}_i^j$ removed. Then, as $H_j - J$ is $(4p, d, K)$-dense, by **P3**, we have that $H'_j - J$ is $(5p, d, K)$-dense. Furthermore, as we removed at most

29

$(2/q) \cdot d^{3/2} \leq \lambda d^2/2$ edges from H_j to get H_j', we have that H_j' is $(\lambda^{1/4}, \lambda/2, d)$-connected. Finally, each path P_i which is non-empty has an endvertex with at least $d/4K - \sqrt{d} \geq d/8K$ neighbours in $V(H_j)$ in G'. Thus, by Lemma 5.9, there is a set, $\mathcal{Q}$ say, of copies of $P_{(1-\varepsilon)d}$ which decomposes all but at most $\varepsilon|V(H_j)|d/2$ of the edges of $G_j' \cup (\cup_{i \in [d/2]} P_i)$. Let $\mathcal{Q}^j = \mathcal{Q}^{j-1} \cup \mathcal{Q}$, and note that both **O5** and **O6** are satisfied. Finally, then, note that **O1**–**O3** hold as, for each $i \in [t]$, all the endvertices of $\mathcal{P}_1^j, \ldots, \mathcal{P}_{d/2}^j$ in $V(G_i)$ are endvertices of $\mathcal{P}_1^{j-1}, \ldots, \mathcal{P}_{d/2}^{j-1}$. Thus, **O1**–**O6** hold, as required. $\quad\boxdot$

Setting $j = s$ in Claim 5.10, take edge-disjoint path forests $\mathcal{P}_i^s$, $i \in [d/2]$, and $\mathcal{Q}^s$, such that **O1**–**O6** hold. Now, **O4** implies that the collections of paths all have no endpoints in H, and therefore $\mathcal{Q}^s$ is a collection of edge disjoint copies of $P_{(1-\varepsilon)d}$ in G by **O5** which, by **O6**, decomposes all but at most $\varepsilon \sum_{j \in [s]} |V(H_j)|d/2$ of the edges of $\bigcup_{j \in [s]} G_j$ and $\mathcal{P}_i$, $i \in [d/2]$, which are not in $\mathcal{P}_i^s$, $i \in [d/2]$. As there are at most $|J|d + |V(G_t)| \cdot pd \leq \varepsilon|V(G_t)|d/2$ edges of G_j which are not in $\bigcup_{j \in [s]} H_j$, all but at most $\varepsilon|V(G_t)|d$ edges of G_t and $\mathcal{P}_i$, $i \in [d/2]$, are contained in $\mathcal{Q}^s$ or $\mathcal{P}_i^s$, $i \in [2]$. Then, using **O1**–**O3**, by the induction hypothesis on $\mathcal{F}' = \{G_1, \ldots, G_{t-1}\}$ and the paths $\mathcal{P}_i^s$, $i \in [d/2]$, there is a set $\mathcal{Q}$ of copies of $\mathcal{P}_{(1-\varepsilon)d}$ which decomposes all but at most $\varepsilon \sum_{i=1}^{t-1} |V(G_i)|d$ edges of $\mathcal{F}' = \{G_1, \ldots, G_{t-1}\}$ and the paths $\mathcal{P}_i^s$, $i \in [d/2]$. Combined with $\mathcal{Q}^s$, this gives a set of copies of $P_{(1-\varepsilon)d}$ which decomposes all but at most $\varepsilon \sum_{i=1}^{t} |V(G_i)|d = \varepsilon|V(G)|d$ of the edges of $\mathcal{F}$ and $\mathcal{P}_i$, $i \in [d/2]$. This completes the proof of the induction step, and hence the lemma. $\quad\square$

6 Concluding remarks

In this paper, we showed that it is possible to approximately decompose d-regular graphs into paths with length approximately d. Improving this to give a full answer to Kotzig's orginal problem (Problem 1.1), even for large d, appears very hard and certainly requires further new ideas and methods. This is true even for strengthening Theorem 1.2 to find paths of length d instead of $\lceil(1-\varepsilon)d\rceil$, where it should be noted that our 'dense spots' may have (slightly) fewer than d vertices and thus may not approximately decompose into paths of length d. Though this paper is motivated by the paucity of results on the decomposition of sparse graphs, we note that Kotzig's problem is unsolved even in the dense regime, and showing that, when $d = \Omega(n)$, any d-regular n-vertex graph can be decomposed into copies of P_d is an interesting open problem.

Decompositions of regular sparse graphs into other subgraphs have also been studied, where the existence of sparse regular graphs with high girth mean that we can only study comparable questions to Problem 1.1 when the graph will be decomposed into trees. Here, Graham and Häggkvist [23] conjectured in 1989 that any $2d$-regular graph decomposes into any d-edge tree, giving a far reaching generalisation of Ringel's conjecture. This problem is wide open. Theorem 1.2 implies that every $2d$-regular graph decomposes approximately into copies of P_d, and it would be very interesting to generalise this to obtain an approximate decomposition of any $2d$-regular graph into any d-edge tree, which appears to be beyond the capabilities of the methods used here.

References

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

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

[3] N. Alon, V. J. Teague, and N. C. Wormald. Linear arboricity and linear k-arboricity of regular graphs. *Graphs and Combinatorics*, 17:11–16, 2001.

[4] J. Bensmail, A. Harutyunyan, T.-N. Le, M. Merker, and S. Thomassé. A proof of the Barát–Thomassen conjecture. *Journal of Combinatorial Theory, Series B*, 124:39–55, 2017.

[5] J. A. Bondy. Perfect path double covers of graphs. *Journal of Graph Theory*, 14(2):259–272, 1990.

[6] F. Botler and L. Hoffmann. Decomposition of $(2k+1)$-regular graphs containing special spanning $2k$-regular Cayley graphs into paths of length $2k + 1$. *Discrete Mathematics*, 345(8):112906, 2022.

[7] F. Botler, G. O. Mota, M. T. I. Oshiro, and Y. Wakabayashi. Decomposing highly edge-connected graphs into paths of any given length. *Journal of Combinatorial Theory, Series B*, 122:508–542, 2017.

[8] F. Botler, G. O. Mota, M. T. I. Oshiro, and Y. Wakabayashi. Decomposing regular graphs with prescribed girth into paths of given length. *European Journal of Combinatorics*, 66:28–36, 2017.

[9] F. Botler, G. O. Mota, and Y. Wakabayashi. Decompositions of triangle-free 5-regular graphs into paths of length five. *Discrete Mathematics*, 338(11):1845–1855, 2015.

[10] J. Böttcher. Graph and hypergraph packing. In *Proc. Int. Cong. Math*, volume 6, pages 4542–4566, 2022.

[11] J. Böttcher, J. Hladkỳ, D. Piguet, and A. Taraz. An approximate version of the tree packing conjecture. *Israel Journal of Mathematics*, 211(1):391–446, 2016.

[12] D. Chakraborti, O. Janzer, A. Methuku, and R. Montgomery. Regular subgraphs at every density. *in preparation*.

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

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

[15] P. Erdős and H. Hanani. On a limit theorem in combinatorical analysis. *Publ. Math. Debrecen*, 10(10-13):2–2, 1963.

[16] O. Favaron, F. Genest, and M. Kouider. Regular path decompositions of odd regular graphs. *Journal of Graph Theory*, 63(2):114–128, 2010.

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

[18] U. Feige, R. Ravi, and M. Singh. Short tours through large linear forests. In *Integer Programming and Combinatorial Optimization: 17th International Conference, IPCO 2014, Bonn, Germany, June 23-25, 2014. Proceedings 17*, pages 273–284. Springer, 2014.

[19] A. Ferber, J. Fox, and V. Jain. Towards the linear arboricity conjecture. *Journal of Combinatorial Theory, Series B*, 142:56–79, 2020.

[20] S. Glock, D. Kühn, A. Lo, and D. Osthus. The existence of designs via iterative absorption: hypergraph F-designs for arbitrary F. *Memoirs of the American Mathematical Society*, 284(1406), 2023.

[21] S. Glock, D. Kühn, and D. Osthus. Extremal aspects of graph and hypergraph decomposition problems. In *Surveys in Combinatorics*, pages 235–266, 2021.

[22] V. Gruslys and S. Letzter. Cycle partitions of regular graphs. *Combinatorics, Probability and Computing*, 30(4):526–549, 2021.

[23] R. Häggkvist. Decompositions of complete bipartite graphs. *Surveys in Combinatorics*, 141:115–147, 1989.

[24] S. Janson, T. Łuczak, and A. Rucinski. *Random graphs.* John Wiley & Sons, 2011.

[25] P. Keevash. The existence of designs. *arXiv preprint arXiv:1401.3665*, 2014.

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

[27] A. Kotzig. Aus der Theorie der endlichen regulären Graphen dritten und vierten Grades. *Casopis Pest. Mat*, 82:76–92, 1957.

[28] R. Lang and L. Postle. An improved bound for the linear arboricity conjecture. *Combinatorica*, 43(3):547–569, 2023.

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

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

[31] R. Montgomery, A. Pokrovskiy, and B. Sudakov. A proof of Ringel's conjecture. *Geometric and Functional Analysis*, 31(3):663–720, 2021.

[32] V. Rödl. On a packing and covering problem. *European Journal of Combinatorics*, 6(1):69–78, 1985.

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

A   Near regularisation: Proof of Lemmas 3.6 and 3.7

In this appendix, we prove the two lemmas we need to efficiently find near-regular subgraphs, which we restate here for convenience.

Lemma 3.6. *Let $1/C' \ll 1/C \leq 1$. For any d, if a graph has degrees between d and Cd, then it contains a subgraph with degrees between d' and $d' + C' \log d'$ for some $d' \geq d/C'$.*

Lemma 3.7. *There exists some $C > 0$ such that for each $1/d \ll \gamma \leq 1/100$ the following holds. Let G be a graph with $d \leq \delta(G), \Delta(G) \leq (1 + \gamma)d$. Then, for some $d' \geq (1 - 40\gamma)d$, G contains a subgraph G' with $|V(G')| \geq (1 - 40\gamma)|V(G)|$ and $d' \leq \delta(G'), \Delta(G') \leq d' + C \log d'$.*

As noted in [13, Section 8], these lemmas can be proved using a recent technique of Chakraborti, Janzer, Methuku and Montgomery [12, 13]. In this, we take any graph G which is approximately regular and carefully take a random subgraph G' which we can show is slightly closer to being regular with positive probability (see [13, Section 2.4] for a more detailed sketch). This will give us the following lemma, which we can then apply iteratively to ultimately find a very nearly regular subgraph without losing very much in the average degree.

Lemma A.1. *Let $1/d \ll 1$, $\varepsilon \leq 1/100$ and $\gamma \geq 10\varepsilon$ such that $\varepsilon d \geq 10^3 \log d$. Let G be a graph in which $d \leq d(v) \leq (1 + \gamma)d$ for each $v \in V(G)$. Then, for some $d' \geq (1 - 2\varepsilon)d$, G contains a subgraph G' with $d' \leq d_{G'}(v) \leq (1 + \gamma)(1 - \varepsilon/2)d'$ and $|V(G')| \geq (1 - 2\varepsilon)|V(G)|$.*

Proof. Let $n = |V(G)|$ and assume that G has no edges between any vertex with degree at least $d + 1$ (for otherwise we could delete such an edge and maintain that $\delta(G) \geq d$). Let $U_L = \{v \in V(G) : d(v) \leq (1 + \gamma/2)d\}$ and $U_H = \{v \in V(G) : d(v) > (1 + \gamma/2)d\}$ be the set of low and high degree vertices in G respectively, and note that there are no edges in $G[U_H]$. Let G' be a random subgraph of G given by

- deleting each edge from U_L to U_H independently at random with probability ε, and
- deleting each vertex in U_L independently at random with probability ε.

For each $v \in V(G)$, let B_v be the event that $v \in V(G')$ but $d_{G'}(v) \notin [(1 - 5\varepsilon/4)d, (1 - 7\varepsilon/4)(1 + \gamma)d]$. Let $t = \lceil |V(G)|/d \rceil$ and let $V(G) = A_1 \cup \ldots \cup A_t$ be an arbitrary partition of $V(G)$ into sets with size between d and $d/2$. For each $i \in [t]$, let B_i be the event that $|V(G') \cap A_i| < (1 - 2\varepsilon)|A_i|$. Let $d' = (1 - 5\varepsilon/4)d$, so that, as $(1 - 7\varepsilon/4) \leq (1 - 5\varepsilon/4)(1 - \varepsilon/2)$, if no event B_v, $v \in V(G)$, occurs, then the degrees of G' are in $[d', (1 - \varepsilon/2)(1 + \gamma)d']$. Furthermore, if no event B_i, $i \in [t]$, occurs, then $|V(G')| \geq (1 - 2\varepsilon)|V(G)|$. Therefore, it is sufficient to show that, with positive probability, no event B_v, $v \in V(G)$, or B_i, $i \in [t]$, holds.

For each $i \in [t]$, as $|A_i| \geq d/2$, we have, by Chernoff's bound, that

$$\mathbb{P}(B_i) = \mathbb{P}(|A_i \setminus V(G')| > 2\varepsilon|A_i|) \leq 2\exp(-\varepsilon d/24) \leq d^{-3}. \tag{12}$$

If $v \in U_H$, then, for each $u \in V(G)$ with $uv \in E(G)$, we have $u \in U_L$, and so $uv \in E(G')$ exactly when u and uv are not deleted, so the probability that $uv \in E(G')$ is $(1 - \varepsilon)^2$. Thus, if $v \in U_H$, then $\mathbb{E}(d_{G-G'}(v)) = (2\varepsilon - \varepsilon^2) \cdot d_G(v)$. As $\gamma \geq 10\varepsilon$ and $\varepsilon \leq 1/100$, we have here that $(1 - 9\varepsilon/4) \cdot d_G(v) \geq (1 - 9\varepsilon/4)(1 + \gamma/2)d \geq (1 - 5\varepsilon/4)d$. Thus,

$$\mathbb{P}(B_v) \leq \mathbb{P}(d_{G'}(v) \notin [(1 - 9\varepsilon/4) \cdot d_G(v), (1 - 7\varepsilon/4) \cdot d_G(v)]) = \mathbb{P}(d_{G-G'}(v) \notin [(7\varepsilon/4) \cdot d_G(v), (9\varepsilon/4) \cdot d_G(v)]).$$

Then, by Chernoff's bound (in particular, Lemma 2.6 applied with $\gamma = 1/10$)

$$\mathbb{P}(B_v) \leq 2\exp(-(2\varepsilon - \varepsilon^2)d/300) \leq d^{-3}. \tag{13}$$

Now, suppose $v \in U_L$. If v survives into $V(G')$, then, note that, for each $uv \in E(G)$, if $v \in U_L$, then the probability that $uv \notin E(G')$ is ε (the probability that v is deleted), while if $v \in U_H$ then the probabilty that $uv \in E(G')$ is also ε (the probability that uv is deleted). Thus, if $v \in U_L$, then $\mathbb{E}(d_{G-G'}(v)|v \in V(G')) = \varepsilon \cdot d_G(v)$. As $\gamma \geq 10\varepsilon$ and $\varepsilon \leq 1/100$, we have that $d_G(v) \leq (1 + \gamma/2)d \leq (1 - 7\varepsilon/4)(1 + \gamma)d$, so that, by Chernoff's bound

$$\mathbb{P}(B_v) \leq \mathbb{P}(d_{G-G'}(v) \geq 5\varepsilon d_G(v)/4|v \in V(G')) \leq \exp(-\varepsilon d_G(v)/48) \leq \exp(-\varepsilon d/48) \leq d^{-3}. \tag{14}$$

By (12), (13), and (14), all the 'bad events' we have defined occur with probability at most d^{-3}. Each 'bad event' is affected by the possible deletion of at most $(1 + \gamma)d$ vertices and $(1 + \gamma)d$ edges, and the possible deletion of each vertex/edge affects at most $(1 + \gamma)d + 1$ 'bad events'. As $1/d \ll 1$, we have $e \cdot d^{-3} \cdot (2(1 + \gamma)d((1 + \gamma)d + 1) + 1) \leq 1$, so by the local lemma there is some choice of G' for which none of the 'bad events' hold, as required. □

Note that if $\gamma \leq 1/10$, then we can iteratively apply Lemma A.1 with $\varepsilon = \gamma/20$ at most 50 times to immediately get the following corollary.

Corollary A.2. *Let $1/d \ll 1$ and $\gamma \leq 1/10$ such that $\gamma d \geq 10^5 \log d$. Let G be a graph in which $d \leq d(v) \leq (1 + \gamma)d$ for each $v \in V(G)$. Then, for some $d' \geq (1 - 10\gamma)d$, G contains a subgraph G' with $d' \leq d_{G'}(v) \leq (1 + \gamma/2)d'$ and $|V(G')| \geq (1 - 10\gamma)|V(G)|$.*

We can now apply Corollary A.2 iteratively to prove Lemma 3.7.

Proof of Lemma 3.7. Let $G_0 = G$, and let r be the smallest integer with $\gamma d/2^r \leq 10^5 \log d$. For each $0 \leq i < r$ in turn, apply Corollary A.2 with $\gamma' = \gamma/2^i$ to G_i to get G_{i+1} with $|V(G_{i+1})| \geq (1 - 10\gamma/2^i)|V(G_{i+1})|$ and vertex degrees in $[d_{i+1}, (1 + \gamma/2^{i+1})d_{i+1}]$ for some $d_{i+1} \geq (1 - 10\gamma/2^i)d_i$. Then, we have

$$|V(G_r)| \geq \prod_{i=0}^{r-1}(1 - 10\gamma/2^i)|V(G)| \geq \left(1 - 10\gamma \cdot \sum_{i=0}^{\infty} 2^{-i}\right)|V(G)| = (1 - 20\gamma)|V(G)|$$

33

and the degrees in G_r are between $[d_r, (1 + \gamma/2^{r+1})d_r] \subset [d_r, d_r + 10^5 \log d]$, where

$$d_r \geq \prod_{i=0}^{r-1}(1 - 10\gamma/2^i)d \geq (1 - 20\gamma)d,$$

so that $G' = G_r$ satisfies the required conditions with $C = 2 \cdot 10^5$ as $\log d_r \geq (\log d)/2$. $\qquad \square$

Finally, we can prove Lemma 3.6.

Proof of Lemma 3.6. Note that we can assume that $C \geq 10$ and $1/d \ll 1$. Let $\varepsilon = 1/1000$ and $k = 10^5 \log C$. Let G be a graph with vertex degrees between d and Cd. Apply Lemma A.1 iteratively k times to G to get a subgraph G'' in which the vertex degrees differ by a factor of at most $\max\{1 + 10\varepsilon, (1 - \varepsilon/4)^k \cdot C\} \leq 1 + \frac{1}{100}$ with $d(G') \geq (1 - 2\varepsilon)^k d(G) \geq 2d/C'$. Then, applying Lemma 3.7 gives the required subgraph. $\qquad \square$