Spanning and induced subgraphs in graphs and digraphs

by

Amarja Kathapurkar

A thesis submitted to the University of Birmingham for the degree Doctor of Philosophy in Pure Mathematics

Supervisor: Dr. Richard Mycroft

Co-supervisor: Professor Daniela Kühn

School of Mathematics

College of Engineering and Physical Sciences

University of Birmingham

June 2023

Abstract

In this thesis, we make progress on three problems in extremal combinatorics, particularly in relation to finding large spanning subgraphs, and removing induced subgraphs.

First, we prove a generalisation of a result of Komlós, Sárközy and Szemerédi and show that for n sufficiently large, any n-vertex digraph with minimum semidegree at least $n/2 + o(n)$ contains a copy of every n-vertex oriented tree with underlying maximum degree at most $O(n/\log n)$.

For our second result, we prove that when k is an even integer and n is sufficiently large, if G is a k-partite graph with vertex classes $V_1, \dots, V_k$ each of size n and $\delta(G[V_i, V_{i+1}]) \geq (1 + 1/k)n/2$, then G contains a transversal C_k-factor, that is, a C_k-factor in which each copy of C_k contains exactly one vertex from each vertex class. In the case when k is odd, we reduce the problem to proving that when G is close to a specific extremal structure, it contains a transversal C_k-factor. This resolves a conjecture of Fischer for even k.

Our third result falls into the theory of edit distances. Let C_h^t be the t-th power of a cycle of length h, that is, a cycle of length h with additional edges between vertices at distance at most t on the cycle. Let $\mathrm{Forb}(C_h^t)$ be the class of graphs with no induced copy of C_h^t. For $p \in [0, 1]$, what is the minimum proportion of edges which must be added to or removed from a graph of density p to eliminate all induced copies of C_h^t? The maximum of this quantity over all graphs of density p is measured by the edit distance function, $\mathrm{ed}_{\mathrm{Forb}(C_h^t)}(p)$, a function which provides a natural metric between graphs and hereditary properties. For our third result, we determine $\mathrm{ed}_{\mathrm{Forb}(C_h^t)}(p)$ for all $p \geq p_0$ in the case when $(t+1) \mid h$, where $p_0 = \Theta(1/h^2)$, thus improving on earlier work of Berikkyzy, Martin and Peck.

DECLARATION

Chapter 2 is joint work with Richard Montgomery and appears in the Journal of Combinatorial Theory Series B [38]. Chapter 3 and Chapter 4 are joint work with Richard Mycroft and the manuscripts are in preparation for submission.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Richard Mycroft. Thank you for your patience, guidance and unwavering support throughout my PhD. I have enjoyed all of our discussions over the past four years, be they mathematical or about trains, and I hope we will have many more.

Thanks also to Richard Montgomery for taking over as my supervisor for a while, and suggesting such a lovely problem to work on and of course for all the cups of tea!

Of course, the past four years would not have been possible without my family, friends and colleagues. It has been an immense pleasure to work alongside such a diverse and intelligent group of people, whether it be playing board games or quizzing over Zoom through the pandemic. Thanks in particular to my parents and Vincent. I am incredibly grateful for all the years of your unquestioned belief in me, often when I needed it the most.

Special thanks go to the School of Mathematics at the University of Birmingham, which I have had the immense pleasure of being a part of for the past eight years, and which introduced me to the world of combinatorics. Special thanks also go to the mathematics department at Chelmsford County High School, which helped to nurture my love of mathematics from an early stage of my education.

I would also like to acknowledge the generous funding provided by the Engineering and Physical Sciences Research Council.

CONTENTS

CHAPTER 1

INTRODUCTION

In 1736, the problem of the seven bridges of Königsberg was resolved by Euler [25], and this is now widely regarded as the origin of *graph theory*. This theory is founded on separating the notion of physical positions of objects from their relative positions or relationships. In order to do this, we represent the objects we are dealing with as *vertices* in our graph, and the connections between them as *edges*.

While there are many other puzzles which can be solved using graph theory, there are also numerous real-world applications of this theory. Perhaps most present-day applications lie in computer science where graphs can be used to represent the flow of information through complex networks, such as social networks or even the internet itself. For example, the PageRank algorithm [17] used by Google uses graph theory to evaluate the importance of webpages. Another extremely topical application is within the area of epidemiology, where dynamics on graphs can be used to model interactions between people, and the spread of disease. In general any field which involves the study of relationships between substructures will naturally lead to an application for graph theory.

While graphs are a useful modelling tool, it is equally important and interesting to study these structures in their own right. Since its conception, graph theory has greatly developed, and now the term encompasses a much wider variety of structures, such as *directed graphs*, where edges can have a direction, or *hypergraphs*, where edges can contain more than two vertices. There has also been a steady evolution of the tools and techniques being used in this area. A notable example of this change is the Four Colour Theorem in planar graph theory which was proved by Appel and Haken through a series of papers (see for example [4] and [5]) with the aid of computers. Later, a significantly simplified version of this proof was found by Robertson, Sanders, Seymour and Thomas [65] who still used similar ideas as Appel and Haken, together with a computer-based approach.

In this thesis we will be focusing on problems in extremal graph theory. This area is concerned with determining how large or small some parameter of a graph needs to be in order to guarantee the existence of a given substructure. In particular, the problems we study will fall under two broad categories, namely finding spanning structures in graphs and digraphs (which we discuss further in Section 1.1), and removing induced subgraphs from graphs (discussed in Section 1.2).

1.1 Spanning structures in graphs and digraphs

Given two graphs H and G, when may we expect to find a copy of H in G? In general, this decision problem is NP-complete (for instance see [40] for H-factors), and therefore we seek simple conditions on G which imply it contains a copy of H. An important early result is Dirac's theorem [23] from 1952 which says that, when $n \geq 3$, any n-vertex graph with minimum degree at least $n/2$ contains a cycle through every vertex, that is, a Hamilton cycle. This is a particular instance of the following meta-question, which has seen much subsequent study. Given an n-vertex graph H, what is the lowest minimum degree condition on an n-vertex graph G which guarantees it contains a copy of H? As such a copy of H would contain every vertex in G, we say it is a *spanning* copy of H.

This question has been studied for many different graphs H, for example when H is the k-th power of a Hamilton cycle for any $k \geq 2$ [42] and when H has bounded chromatic number and maximum degree, and sublinear bandwith [16]. For more details on these results, and those for other graphs, see the survey by Kühn and Osthus [47].

In this thesis, we concentrate on the cases when H is a spanning tree and when H is a K-factor, for a small fixed graph K. We now give a more detailed background of the first of these in Section 1.1.1 and of the second of these in Section 1.1.2. In Section 1.1.3, we discuss some further directions in which this work can be extended.

1.1.1 Spanning trees

Komlós, Sárközy and Szemerédi [41] proved in 1995 that, for each $\alpha, \Delta > 0$, there is some n_0 such that, if $n \geq n_0$, then every n-vertex graph with minimum degree at least $(1/2+\alpha)n$ contains a copy of every n-vertex tree with maximum degree at most Δ, thus confirming a conjecture of Bollobás [12]. This result is furthermore notable as one of the earliest applications of the blow-up lemma. In 2001, Komlós, Sárközy and Szemerédi [43] relaxed the maximum degree condition, showing that, for each $\alpha > 0$, there is some $c > 0$ and n_0 such that, if $n \geq n_0$, then every n-vertex graph with minimum degree at least $(1/2 + \alpha)n$ contains a copy of every n-vertex tree with maximum degree at most $cn/\log n$. This is tight up to the constant c. In 2010, Csaba, Levitt, Nagy-György and Szemerédi [20] showed that, in the other direction, the degree bound in the graph can be reduced for trees with constant maximum degree. That is, they showed that, for each $\Delta > 0$, there is some $C = C(\Delta)$ such that every n-vertex graph with minimum degree at least $n/2 + C\log n$ contains a copy of every n-vertex tree with maximum degree at most Δ. This is tight up the constant C, and, moreover, unlike the previous results, did not use Szemerédi's regularity lemma.

In this thesis, we will prove the corresponding version of the result of Komlós, Sárközy and Szemerédi [43] from 2001 for *directed graphs (digraphs)* instead of graphs. The *minimum semidegree* of a digraph D, denoted by $\delta^0(D)$, is the smallest in- or out-degree over the vertices in D, that is, $\delta^0(D) = \min_{v\in V(D),\diamond\in\{+,-\}} d^\diamond(v)$. Ghouila-Houri [28] solved the minimum semidegree problem for the directed Hamilton cycle, showing that, if an n-vertex digraph D has $\delta^0(D) \geq n/2$, then it contains a directed Hamilton cycle. That is, an n-vertex cycle with the edges oriented in the same direction. DeBiasio, Kühn, Molla, Osthus and Taylor [21] showed that, when n is sufficiently large, this holds in fact for any n-vertex cycle with any orientations on its edges, except for when the edges change direction at every vertex around the cycle. This latter cycle, known as the *anti-directed Hamilton cycle*, is only guaranteed to appear if $\delta^0(D) \geq n/2 + 1$, as shown by DeBiasio and Molla [22].

Recently, Mycroft and Naia [61, 62] gave the first bound on the minimum semidegree required for the appearance of different spanning trees. Here, H is an oriented n-vertex tree, with some bound on the degree of its underlying (undirected) tree. Mycroft and Naia [61, 62] proved that, for each $\alpha, \Delta > 0$, there is some n_0 such that, if $n \geq n_0$, then every n-vertex digraph with minimum semidegree at least $(1/2 + \alpha)n$ contains a copy of every oriented n-vertex tree T with $\Delta^\pm(T) \leq \Delta$. Moreover, their result holds for a slightly wider class of trees, allowing them to show that, for each $\alpha > 0$, almost every oriented n-vertex tree appears in every n-vertex digraph with minimum semidegree at least $(1/2 + \alpha)n$.

In this thesis, we introduce new methods to embed oriented trees in digraphs, relaxing the maximum degree condition to give a full directed version of Komlós, Sárközy and Szemerédi's result, as follows.

Theorem 1.1.1. *For each $\alpha > 0$, there exists $c > 0$ and $n_0 \in \mathbb{N}$ such that the following holds for every $n \geq n_0$. Any n-vertex digraph D with $\delta^0(D) \geq (1/2 + \alpha)n$ contains a copy of every oriented n-vertex tree T with $\Delta^\pm(T) \leq cn/\log n$.*

This result appears in [38]. The proof of this result will be given in Chapter 2.

1.1.2 Tilings

The problem of finding a *perfect H-tiling*, also known as *an H-factor* in a graph G has been widely studied. Clearly, a requirement is that $|H|$ divides $|G|$, so we assume this. In 1970, Hajnal and Szemerédi [30] proved what is now a fundamental theorem in extremal graph theory, showing that whenever G is an n-vertex graph with $\delta(G) \geq (1 - 1/r)n$ then G contains a K_r-factor. This was an improvement on an earlier work of Corrádi and Hajnal [19] who showed this for triangle-factors. Another generalisation of Corrádi and Hajnal's result was to cycles of length ℓ, and a special case of a result of Abbasi [1] confirmed a conjecture of El-Zahar, showing that in an n-vertex graph G, a minimum degree of at least $n/2$ guarantees a C_ℓ-factor when ℓ is even, and $\delta(G) \geq (\ell + 1)n/2\ell$ guarantees a C_ℓ-factor when ℓ is odd. More generally, Kühn and Osthus [50] determined, up to an additive constant, the minimum degree which ensures that G contains an H-factor for general H, showing that this depends on either the *chromatic number of H*, or the *critical chromatic number of H*, depending on the structure of H. As this is a technical result, we do not state this formally here.

In 1999, Fischer [27] conjectured a multipartite version of Hajnal and Szemerédi's result. We give some basic definitions in order to state this. Let G be a k-partite graph with vertex classes $V_1, \ldots, V_k$. For a graph F with vertex set $\{1, \ldots, k\}$, define $\delta^*_F(G) = \min_{ij \in E(F)} \delta(G[V_i, V_j])$ (where we omit the graph in the subscript when it is clear from context). We say a copy of F in G is *transverse* if it contains one vertex from each vertex class; a *transversal F-tiling* in G is a collection of vertex-disjoint transverse copies of F in G, and is a transversal F-factor in G if it covers every vertex of G (note that a necessary condition for this is that all vertex classes have the same size).

Fischer [27] conjectured that if G is an r-partite graph, with vertex classes $V_1, \ldots, V_r$ each of size n and $\delta^*_{K_r}(G) \geq (1 - 1/r)n$, then G contains a transversal

K_r-factor. In the same paper, Fischer showed that when $r = 3, 4$, such a graph G contains a transversal K_r-tiling of size $n - M_r$ for a large constant M_r. The conjecture was proved asymptotically by Johannson [36] in the case when $r = 3$, and later, Lo and Markström [51], and independently, Keevash and Mycroft [39] proved the conjecture asymptotically for all $r \geq 4$.

Fischer [27] also conjectured that $\delta^*_{C_k}(G) \geq (1+1/k)n/2$ is enough to guarantee a transversal C_k-factor. A similar conjecture by Häggkvist with an additional $+1$ term in $\delta^*_{C_k}(G)$ also appeared in [36]. Our main result is the following theorem, which resolves this conjecture completely in the case when k is even.

Theorem 1.1.2. *For every even integer $k \geq 4$ there exists n_0 such that if G is a k-partite graph whose vertex classes each have size $n \geq n_0$ with $\delta^*_{C_k}(G) \geq (1 + \frac{1}{k})\frac{n}{2}$, then G contains a transversal C_k-factor.*

For even k, this improves on an earlier asymptotic version by Ergemlidze and Molla [24], and uses significantly different methods to this earlier work. We remark that in the case when k is odd, the class of extremal graphs is wider. So to get the analogous result for odd k, it only remains to show that any graph which is 'close to' being extremal does in fact contain a transversal C_k-factor. We aim to return to this problem. The proof of Theorem 1.1.2 will be given in Chapter 3.

1.1.3 Further directions

Extending results from graphs into hypergraphs is widely studied. For example, Rödl, Ruciński and Szemerédi [66] generalised Dirac's theorem into the hypergraph setting, showing that any k-uniform hypergraph on n vertices with minimum codegree at least $n/2 + o(n)$ contains a Hamilton cycle. Here, the minimum codegree is the minimum number of k-edges that any set of $k-1$ vertices must be contained in, and is only one of many notions of degree in hypergraphs. Another such notion is the minimum (or maximum) 1-degree, which is the minimum (or maximum) number of

k-edges that any vertex is contained in. The result of Rödl, Ruciński and Szemerédi is also notable for introducing the method of absorption, which is now a key tool in proofs within extremal graph theory, and indeed is a tool we use in the proof of Theorem 1.1.1.

Therefore, it is just as natural to ask whether we can find an analogue of Theorem 1.1.1 for hypergraphs. However, not only are there varying notions of degree in hypergraphs, but there are also multiple definitions of a tree, and therefore the first difficulty is deciding which to use. Pavez–Signé, Sanhueza–Matamala and Stein [63] showed that in a k-uniform hypergraph on n vertices, a minimum codegree of $n/2 + o(n)$ guarantees any spanning tight k-tree of bounded maximum 1-degree. Here, a tight k-tree is a tree which can be created by starting with a k-edge and then at each stage adding a new vertex and a new edge that contains this vertex and $k-1$ vertices from another edge. This generalises the original result of Komlós, Sárközy and Szemerédi [41], and the result of Mycroft and Naia [62] into the hypergraph setting. With these notions of minimum degree and of trees, a full generalisation to hypergraphs of the latter result of Komlós, Sárközy and Szemerédi [43] and of Theorem 1.1.1 would require us to show that any k-graph with minimum codegree $n/2 + o(n)$ contains a copy of any tight k-tree with maximum 1-degree $O(n/\log n)$. If this is true, the work of [63] confirms that this would be best possible. Indeed, the authors adapt the extremal example given by Komlós, Sárközy and Szemerédi and show that it is possible to find a k-tree which has maximum 1-degree $\Theta(n/\log n)$ but asymptotically almost surely (a.a.s), that is, with probability tending to 1, is not contained in the binomial random k-graph with edge probability $p = 0.9$. Therefore, it would be interesting to explore this further.

Another interesting direction to take this work is into the randomly perturbed setting. Randomly perturbed graphs were introduced by Bohman, Frieze and Martin [11] in order to formalise the problem of how many random edges must be added to a graph with given minimum degree in order for it to satisfy some property a.a.s.

For an n-vertex graph G and $p \in [0,1]$, we generally consider $G \cup G(n,p)$ to be the randomly perturbed graph model. In this model, there is a trade-off between δ and p, and therefore we are particularly interested in results where we can start with a lower minimum degree and also a value of p significantly below the threshold at which the structure would appear in a purely random model. In [11], Bohman, Frieze and Martin translate Dirac's theorem to the randomly perturbed setting. These structures have been studied for a large variety of spanning subgraphs (see for example [9], [15], [14], [32]) and the theory has also been extended into digraphs and hypergraphs (see for example [18], [33], [58]).

In the spanning tree setting, Krivelevich, Kwan and Sudakov [45] showed that for any α, Δ, there exists some constant c such that if an n-vertex graph G has minimum degree at least αn, then the graph $G \cup G(n, c/n)$ contains a copy of any n-vertex tree with maximum degree at most Δ, and Joos and Kim [37] made further progress by extending this to trees with maximum degree at most $O(n/\log n)$. However, the result of Joos and Kim requires the trees to have maximum degree at least $n^{o(1)}$, which leaves a gap to the bounded degree case. Krivelevich, Kwan and Sudakov [46] also introduced a notion of randomly perturbing digraphs and hypergraphs. Therefore, finding a full analogue of Theorem 1.1.1 into the randomly perturbed digraphs setting is still open and an interesting question.

For the tilings in multipartite graphs, there are plenty of interesting problems which arise. The first is a conjecture of Ergemlidze and Molla [24] who suggest that in Theorem 1.1.2, it might be possible to allow minimum degrees between classes to vary, provided that their average still satisfies the bound in Theorem 1.1.2. More precisely, they conjecture that if G is a k-partite graph with parts $V_1, \dots, V_k$ each of size n, and there exist $\delta_1, \dots, \delta_k \geq n/2$ such that $\delta(G[V_i, V_{i+1}]) \geq \delta_i$ for each $i \in [k]$ (indices taken modulo k), and furthermore if $\sum_{i\in[k]} \delta_i/k \geq (1 + 1/k + \varepsilon)n/2$, then G contains a transversal C_k-factor. In the same work, Ergemlidze and Molla prove this in the case when $k = 3$. Determining whether this conjecture holds, or whether an exact version of this conjecture holds would be interesting. More generally, it is of interest to develop the theory of tilings in multipartite graphs further, by determining the minimum degree threshold $\delta^*_H(G)$ for general small fixed graphs H, and obtaining a generalisation of the result of Kühn and Osthus [50] into the multipartite setting.

1.2 The edit distance function

The notion of the edit distance was first conceived by Alon and Stav [2] to study problems on property testing, and independently by Axenovich, Kezdy and Martin [6] for its applications in evolutionary biology. At its heart is the following question. Given a graph G and a class of graphs $\mathcal{H}$, how 'far' is G from belonging to $\mathcal{H}$? The *edit distance* is a way to quantify this distance, and measures how many edges we must add to or remove from G in order for it to belong to the class $\mathcal{H}$. While this question is valid for all classes of graphs, it is most natural to consider this in the case of *hereditary graph properties*, that is, a class of graphs $\mathcal{H}$ which is closed under isomorphism and taking induced subgraphs. Indeed, Alon and Stav suggest it is most interesting to study these since many of the classes of graphs which appear within both graph theoretical research and within wider applications in the sciences fall into this category.

Formally, we define the edit distance between two graphs G and G' on the same vertex set to be the size of the symmetric difference between their edge sets normalised by the total number of possible edges, that is, if $|G| = |G'| = n$, then

$$\mathrm{dist}(G, G') = \frac{|E(G)\Delta E(G')|}{\binom{n}{2}}.$$

$\mathrm{Forb}(H)$ represents the class of all graphs G which do not have H as an induced subgraph. A reason why it is intuitive to study hereditary properties is that these can be classified in terms of their forbidden subgraphs, that is, for any hereditary property $\mathcal{H}$, there is a family $\mathcal{F}(\mathcal{H})$ of forbidden graphs, that is,

$$\mathcal{H} = \bigcap_{H \in \mathcal{F}(\mathcal{H})} \mathrm{Forb}(H).$$

We say a hereditary property $\mathcal{H}$ is *trivial* if there is an $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, there is no n-vertex graph contained in $\mathcal{H}$. In other words, a class is trivial if and only if it is finite. Otherwise, we say $\mathcal{H}$ is *non-trivial*. For instance, an example of a non-trivial hereditary property is $\mathrm{Forb}(C_h)$, the class of graphs with no C_h as an induced subgraph. We can extend the notion of distance between graphs to define the distance between a graph G and a hereditary property $\mathcal{H}$, which we define to be the minimum distance from G to some graph G' in $\mathcal{H}$ on the same vertex set, that is,

$$\mathrm{dist}(G, \mathcal{H}) = \min \left\{ \mathrm{dist}(G, G') \colon G' \in \mathcal{H}, V(G) = V(G') \right\}.$$

Study in this area was initiated by the problem of determining, for any hereditary property $\mathcal{H}$, which graph is furthest from belonging to $\mathcal{H}$. This question was motivated by problems in theoretical computer science, and indeed, Alon and Stav [2] showed that for any property $\mathcal{H}$, there is some $p^* = p^*_{\mathcal{H}} \in [0, 1]$ such that asymptotically, $G(n, p^*)$ is the furthest graph from belonging to $\mathcal{H}$. It was this which led to the conception of the *edit distance function* by Balogh and Martin [8]. For any $p \in [0, 1]$, we define

$$\mathrm{ed}_{\mathcal{H}}(p) = \lim_{n \to \infty} \max \left\{ \mathrm{dist}(G, \mathcal{H}) \colon \ |G| = n, |E(G)| = \left\lfloor p \binom{n}{2} \right\rfloor \right\}, \tag{1.2.1}$$

if this limit exists. Note that the normalisation is implicit in the definition of $\mathrm{dist}(G, \mathcal{H})$ and so if this limit exists, it takes values in $[0, 1]$. So in other words, for any p, the edit distance function for a hereditary property $\mathcal{H}$ tells us the furthest distance a graph of density p can be from belonging to $\mathcal{H}$. Balogh and Martin [8] later generalised the result of Alon and Stav to show that the limit in (1.2.1) does exist for all non-trivial hereditary properties $\mathcal{H}$. In addition to this, they showed the following result.

Theorem 1.2.1 (Balogh-Martin [8])**.**

$$\mathrm{ed}_{\mathcal{H}}(p) = \lim_{n\to\infty} \mathbb{E}\left[\mathrm{dist}(G(n,p),\mathcal{H})\right].$$

That is, asymptotically, for any p and hereditary property $\mathcal{H}$, we can use the random graph $G(n,p)$ to estimate the edit distance function. Balogh and Martin [8] also showed that the edit distance function is continuous and concave down. Methods to determine the edit distance function $\mathrm{ed}_{\mathcal{H}}(p)$ make implicit use of these properties, as well as Theorem 1.2.1, as we will see in Chapter 4. It is worth remarking that the proof of Balogh and Martin is not constructive, and does not indicate how to determine the $\mathrm{dist}(G(n,p),\mathcal{H})$ for a given $\mathcal{H}$, for example.

The edit distance function has been studied for a range of hereditary properties of the form $\mathcal{H} = \mathrm{Forb}(H)$ for a fixed graph H. In the case when $H = K_r$, this was completely determined by Martin [53]. The function was also calculated for some small fixed graphs by Marchant and Thomason [52] and Martin [54]. More recently, Martin and Riasanovsky [57] also considered the case when $H = G(n_0, p_0)$, where n_0 and p_0 are fixed in terms of n and p.

In this thesis, we study the edit distance function when $\mathcal{H} = \mathrm{Forb}(C_h^t)$ for some $h, t \in \mathbb{N}$. Here, C_h is a cycle on h vertices and C_h^t is defined to be the graph on the same vertex set as C_h such that there is an edge between two vertices of C_h^t if and only if these vertices were at distance at most t in C_h. In particular, when $t = 1$, this is just the cycle on h vertices. Thus, we aim to determine $\mathrm{ed}_{\mathrm{Forb}(C_h^t)}(p)$, where $\mathrm{Forb}(C_h^t)$ is the class of graphs which contain no C_h^t as an induced subgraph. This is a very natural property to consider, and the question has received a lot of interest in the past. Marchant and Thomason [52] determined $\mathrm{ed}_{\mathrm{Forb}(C_h)}(p)$ for all $p \in [0,1]$ when $h = 4$. Martin [53] extended this and explicitly determined $\mathrm{ed}_{\mathrm{Forb}(C_h)}(p)$ for

$h \in \{3, \ldots, 9\}$. Peck [64] gave the first general version of this result, determining the edit distance function from $\mathrm{Forb}(C_h)$ for all p when h is odd, and for sufficiently large p for even h. Berikkyzy, Martin and Peck [10] generalised this result by determining $\mathrm{ed}_{\mathrm{Forb}(C_h^t)}(p)$ for all p when $(t+1) \nmid h$ and for sufficiently large p when $(t+1) \mid h$. More precisely, they proved the following.

Theorem 1.2.2 (Berikkyzy, Martin and Peck [10]). *Let $t \geq 1$ and $h \geq 2t(t+1)+1$ be integers, and let $\mathcal{H} = \mathrm{Forb}(C_h^t)$.*

(i) *If $(t+1) \not| \, h$, then for all $p \in [0,1]$, we have*

$$\mathrm{ed}_{\mathrm{Forb}(C_h^t)}(p) = \min\left\{\frac{p}{t+1}, \min_{r\in\{0,1,\ldots,t\}}\left\{\frac{p(1-p)}{r+\left(\left\lceil\frac{h}{2t+1}\right\rceil - r - 1\right)p}\right\}\right\}.$$

(ii) *If $(t+1) \mid h$, then for all $p \in [1/\lceil h/(2t+1)\rceil, 1]$, we have*

$$\mathrm{ed}_{\mathrm{Forb}(C_h^t)}(p) = \min_{r\in\{0,1,\ldots,t\}}\left\{\frac{p(1-p)}{r+\left(\left\lceil\frac{h}{2t+1}\right\rceil - r - 1\right)p}\right\}.$$

However, this result leaves a gap for small p in the case when $(t+1) \mid h$. While this is explained in more detail in Chapter 4, we briefly discuss the case when $t = 1$, that is for cycles. In this case, intuitively, the gap in Theorem 1.2.2 comes from certain constructions which can be used to eliminate all cycles of length h from $G(n,p)$. In the case when h is odd, there are three such constructions which are candidates. For the first construction, we partition $G(n,p)$ into two equal sized parts, and remove all edges that lie within any of the parts. Then in expectation, we remove $2p \cdot \binom{n/2}{2}$ edges in total, and this corresponds to the first term in the minimum of Theorem 1.2.2 when h is odd. Furthermore, the resulting graph is bipartite and therefore contains no cycles of odd length. Similarly, there are two more constructions, each of which correspond to the other terms in the minimum. However, in the case when h is even, the first construction described no longer works, since the resulting bipartite graph could indeed contain even length cycles. We extend on Theorem 1.2.2 for small p in the case when $(t+1) \mid h$ to show the following.

Theorem 1.2.1. *Let $t \geq 1$ and $h \geq 4t(2t+1)$ be integers, with $(t+1) \mid h$. Let $c_0 = \lfloor (\lfloor h/t \rfloor + 1)/3 \rfloor$, $\ell_0 = \lceil h/(2t+1) \rceil$, and let $p_0 = t/(c_0\ell_0 - c_0 - \ell_0 + t + 1)$. Then for all $p \in [p_0, 1/\lceil h/(2t+1) \rceil]$, we have that*

$$\mathrm{ed}_{\mathrm{Forb}(C_h^t)}(p) = \frac{p(1-p)}{t + \left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1 \right) p}.$$

We also prove that for all t and h satisfying Theorem 1.1.2, the range of values $[p_0, 1/\lceil h/(2t+1) \rceil]$ is non-empty (observe, in particular, that $p_0 = \Theta(1/h^2)$), and therefore Theorem 1.1.2 strictly extends on the work of Berikkyzy, Martin and Peck [10].

1.2.1 Further directions

Another natural hereditary property of interest is $\mathrm{Forb}(K_{s,t})$, the class of graphs with no $K_{s,t}$ as an induced subgraph. In this case, Marchant and Thomason [52] determined this partially for $K_{3,3}$. Martin and McKay [56] also gave some results for $K_{2,t}$, giving complete results when $t = 3, 4$ and partial results for larger t. It would be interesting to know more about the edit distance function for this class of graphs. In [55], Martin suggests that a starting point could be to determine the edit distance function when $\mathcal{H} = \mathrm{Forb}(K_{s,s})$.

More generally, there are plenty of developments which could be made to the theory. For instance, Axenovich and Martin [7] also extended this theory into edge-coloured graphs and directed graphs, and an interesting open question raised by Martin [55] is whether this notion of the edit distance function can be extended into the setting of hypergraphs. A starting point for this may be to find a generalisation of the result of Alon and Stav [2] into this setting.

CHAPTER 2

SPANNING TREES IN DENSE DIGRAPHS

The aim of this chapter is to prove Theorem 1.1.1 which appears in [38], and which we recall here.

Theorem 1.1.1. *For each $\alpha > 0$, there exists $c > 0$ and $n_0 \in \mathbb{N}$ such that the following holds for every $n \geq n_0$. Any n-vertex digraph D with $\delta^0(D) \geq (1/2 + \alpha)n$ contains a copy of every oriented n-vertex tree T with $\Delta^{\pm}(T) \leq cn/\log n$.*

We begin by noting that the undirected version follows immediately from Theorem 1.1.1. Indeed, given any n-vertex tree T and an n-vertex graph G, we can apply Theorem 1.1.1 to a copy of T with each edge oriented arbitrarily and a digraph formed from G by replacing each edge uv with an edge from u to v and an edge from v to u. This demonstrates that, as with Komlós, Sárközy and Szemerédi's result, Theorem 1.1.1 is tight up to the constant c. Furthermore, through Theorem 1.1.1 we give a new proof of the undirected result without using Szemerédi's regularity lemma, in contrast to the work of both Komlós, Sárközy and Szemerédi [41], and Mycroft and Naia [61, 62], adding to the non-regularity proof for trees with constant maximum degree by Csaba, Levitt, Nagy-György and Szemerédi [20] described above. Key to our result is to use a random embedding of part of the tree using 'guide sets' and embedding many leaves (and small subtrees) of the tree using 'guide graphs'. This replaces the regularity methods of [41, 61, 62], and is sketched in Section 2.1, where we also outline the rest of this proof.

2.1 Preliminaries

2.1.1 Notation

Let D be a digraph. We denote by $V(D)$ and $E(D)$ the vertex set and edge set of D, respectively, where every element of the edge set of D is an ordered pair of vertices. We let $|D| = |V(D)|$, which we call the *size* of D, and let $e(D) = |E(D)|$. Letting $u, v \in V(D)$, if $uv \in E(D)$, then we say that u is an *in-neighbour* of v and v is an *out-neighbour* of u. Denote by $N_D^-(v)$ and $N_D^+(v)$, respectively, the set of all in- and out-neighbours of v. We let $d_D^-(v) = \left|N_D^-(v)\right|$ and $d_D^+(v) = \left|N_D^+(v)\right|$, and we refer to these as the *in-* and *out-degree of* v, respectively. For each $\diamond \in \{+, -\}$, we let $\delta^\diamond(D)$ and $\Delta^\diamond(D)$ be, respectively, the *minimum* and *maximum* $\diamond$*-degree* of D. For any $A, B \subseteq V(D)$, and each $\diamond \in \{+, -\}$, let $N_D^\diamond(A, B) = \bigcup_{a \in A}(N_D^\diamond(a) \cap B)$, and let $d_D^\diamond(A, B) = |N_D^\diamond(A, B)|$. We omit the subscript when the graph is clear from context. Note that, for simplicity of notation, we use '$-$' and 'in' interchangeably, and, similarly, we use '$+$' and 'out' interchangeably. We use '$\pm$' to represent that a property holds for both '$-$' and '$+$'.

Suppose that A and B are disjoint subsets of $V(D)$. We write $D[A]$ to mean D induced on the set A, that is, the graph obtained from D by deleting all vertices which are not in A. For each $\diamond \in \{+, -\}$, a $\diamond$*-matching* from A into B is a set of vertex-disjoint edges such that every edge in the set has one endpoint in A and one endpoint in B, and the endpoint in B is a $\diamond$-neighbour of the endpoint in A, that is, every edge is a $\diamond$-edge from A into B. We say this matching *covers* A if every vertex of A belongs to some edge in the matching, and we call this a *perfect* $\diamond$*-matching* if it covers both A and B. A *bare path* of length m in a tree is a path with m edges such that each of the internal vertices have degree 2 in the tree. When P is a path in D, we let $D - P$ denote the subgraph of D obtained by removing the internal vertices of P.

For any $n \in \mathbb{N}$, we let $[n] := \{1, \dots, n\}$. In order to simplify notation, we use hierarchies to state our results. That is, for $a, b \in (0, 1]$, whenever we write that a statement holds for $a \ll b$ (or $b \gg a$), we mean that there exists a non-decreasing function $f\colon (0, 1] \to (0, 1]$ such that the statement holds whenever $a \leq f(b)$. We define similar expressions with multiple variables analogously. We say a random event occurs *with high probability* if the probability of the event occurring tends to 1 as n tends to infinity. In our proofs, when we have shown that a property holds with high probability, we will implicitly assume that this property holds from that point onwards. For simplicity, we ignore floors and ceilings wherever this does not affect the argument.

2.1.2 Proof sketch

When $1/n \ll c \ll \alpha$, we will embed any oriented n-vertex tree T with $\Delta^{\pm}(T) \leq cn/\log n$ into any n-vertex digraph D with $\delta^0(D) \geq (1/2 + \alpha)n$. We embed T using the absorption method, an approach first introduced in general by Rödl, Ruciński and Szemerédi [66] which has been effective on a range of embedding problems for spanning graphs and digraphs (see, for example, the survey [13]). We first partially embed a subtree T'' of T into a set A such that, given any subset $B \subset V(D)$ with $A \subset B$ and $|B| = |T''|$, we can complete this embedding of T'' into $D[B]$ (see Theorem 2.1.1).

We then use an almost-spanning embedding to embed the vertices in $V(T) \setminus V(T'')$ to extend the partial embedding of T'' (see Theorem 2.1.2). We will have chosen T'' so that in this stage a tree, called T', is attached to an embedded vertex of T''. Using the property of the partial embedding of T'', we then complete the embedding of T'' with the unused vertices in D. The decomposition of T that we need follows from a simple proposition (Proposition 2.1.3).

In Section 2.1.2.1, we state these three results, Theorem 2.1.1, Theorem 2.1.2 and Proposition 2.1.3, before deducing Theorem 1.1.1 from them. In Section 2.1.2.2, we then discuss in detail the proof of Theorem 2.1.2, which is the major challenge overcome by this proof.

In the rest of Section 2.1, we restate the probabilistic tools we will use, and give a basic structural decomposition of trees and some simple results on matchings. In Section 2.2, we prove Theorem 2.1.2. In Section 2.3, we prove Theorem 2.1.1.

2.1.2.1 Main tools and deduction of Theorem 1.1.1

For Theorem 1.1.1, we will first find a suitable subtree $T'' \subset T$ and a set $A \subset V(D)$ with slightly fewer than $|T''|$ vertices, so that, given any set B of $|T''|$ vertices containing A, we can embed T'' in $D[B]$. Furthermore, we will ensure that some pre-specified vertex $t \in V(T'')$ is always embedded to some fixed vertex $v \in A$, as follows.

Theorem 2.1.1. *Let $1/n \ll c \ll \varepsilon \ll \mu \ll \alpha$. Let D be an n-vertex digraph with minimum semidegree at least $(1/2 + \alpha)n$. Let T be an oriented tree with μn vertices and $\Delta^{\pm}(T) \leq cn/\log n$, and let $t \in V(T)$.*

Then, $V(D)$ contains a vertex set A with size $(\mu - \varepsilon)n$ containing a vertex $v \in A$ such that the following holds. For any set $B \subset V(D)$ with $A \subset B$ and $|B| = \mu n$, $D[B]$ contains a copy of T in which t is copied to v.

Theorem 2.1.1 is proved in Section 2.3 by randomly embedding most of T and taking A to be the image of this embedding. We then show that the partial embedding of T can be extended using any new vertex in $y \in V(D) \setminus A$ by switching y into the partial embedding in place of some vertex in A that can instead be used to embed a new vertex of T. Repeatedly doing this will allow the embedding of T to be completed using any set of $|T| - |A|$ new vertices in $V(D) \setminus A$. This is sketched in more detail at the start of Section 2.3, before Theorem 2.1.1 is proved.

We will embed the majority of the tree for Theorem 1.1.1, using the following almost-spanning embedding.

Theorem 2.1.2. *Let $1/n \ll c \ll \varepsilon, \alpha$. Let D be an n-vertex digraph with minimum semidegree at least $(1/2+\alpha)n$ and let $v \in V(D)$. Let T be an oriented tree with at most $(1-\varepsilon)n$ vertices and $\Delta^{\pm}(T) \leq cn/\log n$, and let $t \in V(T)$.*

Then, D contains a copy of T in which t is copied to v.

Using in addition the following simple proposition (see, for example, [59, Proposition 3.22]), we can now deduce Theorem 1.1.1.

Proposition 2.1.3. *Let $n, m \in \mathbb{N}$ satisfy $1 \leq m \leq n/3$. Given any n-vertex tree T and a vertex $t \in V(T)$, there are two edge-disjoint trees $T_1, T_2 \subset T$ such that $E(T_1) \cup E(T_2) = E(T)$, $t \in V(T_1) \cap V(T_2)$ and $m \leq |T_2| \leq 3m$.*

Proof of Theorem 1.1.1 from Theorems 2.1.1 and 2.1.2. Let $\varepsilon, \mu > 0$ be such that $c \ll \varepsilon \ll \mu \ll \alpha$. Let D be an n-vertex digraph with $\delta^0(D) \geq (1/2+\alpha)n$. Let T be an oriented n-vertex tree with $\Delta^{\pm}(T) \leq cn/\log n$.

Using Proposition 2.1.3 with $m = \mu n$, find edge-disjoint trees $T', T'' \subset T$ such that $E(T') \cup E(T'') = E(T)$ and $\mu n \leq |T''| \leq 3\mu n$. Let t be the vertex which is in both T' and T''. By Theorem 2.1.1 applied with $\mu' = |T''|/n$, there is a set $A \subset V(D)$ such that $|A| = |T''| - \varepsilon n$, and a vertex $v \in A$ such that, for any set $B \subset V(D)$ with $A \subset B$ and $|B| = |T''|$, $D[B]$ contains a copy of T'' in which t is copied to v.

Let $D' = D - (A \setminus \{v\})$. Let $n' = |D'|$, so that $(1-3\mu)n \leq n' \leq n$. Let α' be such that D' has minimum semidegree $(1/2+\alpha')n'$. Note that $(1/2+\alpha')n \geq (1/2+\alpha')n' \geq (1/2+\alpha-3\mu)n$, so that $\alpha' \geq \alpha/2$. Furthermore, $n' = n - |T''| + \varepsilon n + 1 = |T'| + \varepsilon n$, and therefore

$$\frac{|T'|}{n'} = \frac{|T'|}{|T'|+\varepsilon n} \leq \frac{|T'|}{|T'|(1+\varepsilon)} \leq 1 - \varepsilon/2.$$

Thus, by Theorem 2.1.2 with $\varepsilon \ll \mu$, we can find a copy, S' say, of T' in D' in which t is copied to v. By applying the property of A from Theorem 2.1.1, we can then find a copy of T'' in $D - (V(S') \setminus \{v\})$ in which t is copied to v. Together, these give us a copy of T. $\square$

2.1.2.2 Proof Sketch of Theorem 2.1.2

We will embed a $(1-\varepsilon)n$-vertex tree T for Theorem 2.1.2 by dividing most of T into a small core forest $T_0 \subset T$ and a collection of constant-sized subtrees, which are either attached to T_0 by a single edge or by two short paths. It is the trees attached to T_0 by a single edge that will be the most challenging to embed, and so we dedicate most of our attention in the proof sketch to this.

More precisely, we will find a tree $T' \subset T$, containing a *core forest* $T_0 \subset T'$ and vertex-disjoint trees $S_1, \ldots, S_\ell \subset T' - V(T_0)$, for some $\ell \in \mathbb{N}$, such that T' is formed from T_0 by, for each $i \in [\ell]$,

(1) either attaching S_i to T_0 using two bare paths with length 2,

(2) or attaching S_i to T_0 with a single edge.

Furthermore, for some $\mu > 0$ and $K \in \mathbb{N}$, with $1/n \ll 1/K, \mu \ll \alpha, \varepsilon$, we will have that

- $|T_0| \leq \mu n$ (i.e., T_0 is small),
- $|T'| \geq |T| - \mu n$ (i.e., T' is most of T),
- there are at most μn trees S_i which are in Case (1), and
- each tree S_i has at most K vertices.

In Case (1), we say S_i is added to T_0 as a path, and in Case (2) we say S_i is added to T_0 as a leaf. The crux of our method is to embed T_0 along with the trees S_i in Case (2) connected to the embedding of T_0 by the appropriate edge. This is encapsulated in the following lemma, which is proved in Section 2.2.1.

Lemma 2.1.4. *Let $1/n \ll c \ll \mu \ll \alpha, \varepsilon$, let $c \ll 1/K$ and let $\ell \in \mathbb{N}$. Suppose D is an n-vertex digraph with $\delta^0(D) \geq (1/2+\alpha)n$ and $v \in V(D)$.*

Suppose that T is an oriented tree with $|T| \leq (1-\varepsilon)n$ and $\Delta^\pm(T) \leq cn/\log n$. Suppose that $T', S_1, \ldots S_\ell \subset T$ are vertex-disjoint subtrees with $|T'| \leq \mu n$, and

$|S_i| \leq K$ for each $i \in [\ell]$. Suppose further that T is formed from T' by attaching each S_i, $i \in [\ell]$, to T' by an edge, and let $t \in V(T')$.

Then, D contains a copy of T in which t is copied to v.

We will now briefly sketch how Theorem 2.1.2 can be proved from Lemma 2.1.4. Let m be the total number of vertices that appear in the trees S_i in Case (1) above. To embed these trees, we use the fact that two random sets in D of the same (linear) size are likely to have a perfect matching from one to the other (see Proposition 2.1.12). Taking $p \gg 1/n$ and $Kp \leq 1$, we can, with high probability, find pn copies of an oriented tree with K vertices in a random set of Kpn vertices in D by taking randomly K disjoint subsets within this set of size pn and finding appropriate matchings between them (see Section 2.1.5). Collecting isomorphic trees S_i together, and applying this to each of the constantly many (depending on K) isomorphism classes, allows us to embed the trees S_i in Case (1) with high probability in a random set with size $m + \varepsilon n/4$. Here, the extra $\varepsilon n/4$ vertices allow us to find a linear number of trees in each isomorphism class by finding some additional trees if required.

Thus, in a partition of $V(D)$ into sets $V_1 \cup V_2 \cup V_3 \cup V_4$ chosen uniformly at random so that $|V_1| = n - m - 3\varepsilon n/4$, $|V_2| = m + \varepsilon n/4$, $|V_3| = |V_4| = \varepsilon n/4$, with high probability, the following occur.

- $\delta^{\pm}(D[V_1]) \geq (1/2 + \alpha/2)|V_1|$, so that, applying Lemma 2.1.4, we can embed T_0 along with the trees S_i in Case (2) connected to the embedding of T_0 by the appropriate edge.

- We can embed the trees S_i in Case (1) in $D[V_2]$.

- Then, using that there are at most μn trees in Case (1), we can greedily attach them to the embedding of T_0 using two paths with length 2 whose interior vertex is an unused vertex in V_3 (see Section 2.2.2).

- Finally, as $|T| - |T'| \le \mu n$, we can greedily extend the resulting embedding of T' to one of T, by adding a sequence of leaves using vertices in V_4 (see Section 2.2.3).

Here, the last two steps are (with high probability) possible using the semi-degree condition of D. Note that, as $\mu \ll \varepsilon$, we only embed a small proportion of vertices into V_3 and V_4.

We will now give a detailed proof sketch of Lemma 2.1.4.

Proof sketch of Lemma 2.1.4

To simplify our discussion, let us assume that each tree S_i in Lemma 2.1.4 consists of only a single vertex, which is an out-neighbour in the tree T of a vertex of T_0, and that every vertex in T_0 is attached to exactly one such tree. That is, T consists of T_0 with an out-matching attached. Our embedding of T_0 is randomised, which will allow the methods described to be used to find matchings attached from different subsets of the image of the embedding of T_0 to different random sets. This will allow the embedding below for T_0 to be used for the general case.

Let us detail the example situation precisely. Suppose we have a μn-vertex tree T_0 and choose two disjoint random sets $V_0, V_1 \subset V(D)$ with size $p_0 n$ and $p_1 n$ respectively, where $p_0 \gg \mu$ and $p_1 = (1 + o(1))\mu$. We will randomly embed T_0 into V_0, so that there is an out-matching from the vertex set of the embedding of T_0 into V_1. More generally, we may have to attach matchings into several different sets from T_0, so we can only use a small proportion of spare vertices in D (and so p_1 is only a little larger than μ). On the other hand, as these matchings will be all attached to the same small tree, T_0, we can use many spare vertices when embedding T_0 (and so we take $p_0 \gg \mu$).

We will embed T_0 vertex-by-vertex, say in order $t_1, \ldots, t_\ell$, so that each new vertex is embedded as an in- or out-leaf of the previously embedded subtree. Having chosen the random sets V_0, V_1, and before beginning the embedding, we will find *guide* *sets* $A_{v,\diamond} \subset N^{\diamond}_D(v, V_0)$, $v \in V_0$ and $\diamond \in \{+,-\}$, which we use to guide the random embedding. We then start the random embedding, under the rule that if, for some $v \in V_0$ and $\diamond \in \{+,-\}$, we are attaching a $\diamond$-edge as a leaf to v, then we choose this leaf uniformly at random from the unused vertices in $A_{v,\diamond}$.

The guide sets ensure that, with high probability, there will be a matching from the embedding of T_0 into V_1. These guide sets are found using Lemma 2.2.5, and they exist (with high probability for the choice of V_0, V_1) due to the semi-degree condition in D. Essentially, for some constants β, γ, we find, for each $v \in V(D)$ and $\diamond \in \{+,-\}$, a set $A_{v,\diamond} \subset N^{\diamond}_D(v, V_0)$ with size βn and bipartite digraphs $H^{\circ}_{v,\diamond} \subset D^{\circ}[A_{v,\diamond}, V_1]$, $\circ \in \{+,-\}$, so that in $H^{\circ}_{v,\diamond}$ each vertex in $A_{v,\diamond}$ has around $\gamma p_1 n$ $\circ$-neighbours in V_1, and each vertex in V_1 has around $\gamma\beta n$ $\circ$-edges leading into it. That is, $H^{\circ}_{v,\diamond}$ is approximately regular on each side with edge density approximately γ.

We use the guide graphs $H^{+}_{v,\diamond}$ to find the matching from the embedding of T_0 by constructing an auxiliary bipartite digraph Q with vertex classes $\{s_1, \ldots, s_\ell\}$ and V_ℓ. In this example situation, Q is a subgraph of D, and a matching in Q corresponds exactly to a matching from the image of $V(T_0)$ to V_1. In the more general case we attach multiple different matchings simultaneously and Q has vertices from the image of $V(T_0)$ copied different numbers of times (see Section 2.2). The digraph Q does not contain all the edges in D from $\{s_1, \ldots, s_\ell\}$ to V_ℓ. Instead, we add edges using the guide graphs so that when we have constructed Q it will be, with high probability, approximately regular, so that we can find our required matching via Hall's matching criterion.

When a vertex t_i is embedded using a guide set $A_{v_i,\diamond_i}$, to some vertex s_i say, we add only the edges in $H^{+}_{v_i,\diamond_i}$ adjacent to s_i to Q – note that approximately $\gamma p_1 n$ edges are added next to s_i. Note further that, as most of the vertices in $A_{v_i,\diamond_i}$ will be unused, each $w \in V_1$ will have an edge added from s_i to w with probability approximately

$$\frac{d^{-}_{H^{+}_{v_i,\diamond_i}}(w)}{|A_{v_i,\diamond_i}|} \approx \frac{\gamma\beta n}{\beta n} = \gamma. \tag{2.1.1}$$

When this is complete, Q is a bipartite digraph with vertex classes $\{s_1, \ldots, s_\ell\}$ and V_1. Each vertex s_i will have out-degree approximately $\gamma p_1 n$, and, due to the randomness of the embedding and (2.1.1), each vertex in V_1 will have in-degree which is approximately $\gamma\ell = \gamma|T_0| \approx \gamma p_1 n$.

Thus, Q will be a bipartite graph with the in-degrees in one vertex class approximately equal to the out-degrees in the other. Via Hall's matching criterion, an out-matching will exist from $\{s_1, \ldots, s_\ell\}$ to V_1 which covers most of the vertices in $\{s_1, \ldots, s_\ell\}$. By ensuring that V_1 is likely to be a little larger than ℓ, we in fact will get with high probability that such an out-matching can cover $\{s_1, \ldots, s_\ell\}$.

Note that, in the sketch above, we do not use the graph $H^-_{v,\diamond}$. However, in practice, we find such guide sets and guide graphs with $V_1 = V(D) \setminus V_0$ (see Lemma 2.2.3), before taking random subsets of V_1. We will find out-matchings into some of these random sets, and in-matchings into some others. Therefore, it is important to have both guide graphs $H^-_{v,\diamond}$ and $H^+_{v,\diamond}$, and, furthermore, that the same set $A_{v,\diamond}$ is used for both graphs.

Finally, let us note where the condition $\Delta^\pm(T) \leq cn/\log n$ is used in our proof of Lemma 2.1.4. In the sketch above the set V_1 will always have size which is linear in n, but we may need to attach the trees in Lemma 2.1.4 to few vertices in T. The maximum in- or out-degree condition on T ensures, that, if the trees S_i in Lemma 2.1.4 together comprise linearly (in n) many vertices in T, then they are attached to at least $C\log n$ different vertices, for some large constant C, which gives us sufficient probability concentration when these vertices are randomly embedded for the corresponding versions of Hall's criterion to hold (see the proof of Claim 2.2.7).

2.1.3 Probabilistic tools

Let $n, m, k \in \mathbb{N}$ be such that $\max\{m, k\} \leq n$. Let A be a set of size n, and $B \subseteq A$ be such that $|B| = m$. Let A' be a uniformly random subset of A of size k. Then the random variable $X = |A' \cap B|$ is said to have hypergeometric distribution with parameters n, m and k, which we denote by $X \sim \mathrm{Hyp}(n, m, k)$. We will use the following Chernoff-type bound.

Lemma 2.1.5 (see, for example, [35])**.** *Suppose* $X \sim \mathrm{Hyp}(n, m, k)$*. Then for any* $0 < \alpha < 3/2$*, we have*

$$\mathbb{P}\left[|X - \mathbb{E}[X]| \geq \alpha \mathbb{E}[X]\right] \leq 2 \exp\left(-\alpha^2 \mathbb{E}[X]/3\right).$$

A sequence of random variables $(X_i)_{i \geq 0}$ is a martingale if $\mathbb{E}[X_i] < \infty$ and $\mathbb{E}[X_{i+1} \mid X_0, \ldots, X_i] = X_i$ for each $i \geq 0$. We will use the following Azuma-type bound for martingales.

Lemma 2.1.6 (see, for example, [29])**.** *Let* $(X_i)_{i \geq 0}$ *be a martingale and let* $c_i > 0$ *for each* $i \geq 1$*. If* $|X_i - X_{i-1}| < c_i$ *for each* $i \geq 1$*, then, for each* $n \geq 1$*,*

$$\mathbb{P}[|X_n - X_0| \geq t] \leq 2 \exp\left(-\frac{t^2}{2 \cdot \sum_{i=1}^n c_i^2}\right).$$

We will use this bound for supermartingales and submartingales. A sequence of random variables $(X_i)_{i \geq 0}$ is a supermartingale if $\mathbb{E}[X_{i+1} \mid X_0, \ldots, X_i] \leq X_i$ for each $i \geq 0$, and a submartingale if $\mathbb{E}[X_{i+1} \mid X_0, \ldots, X_i] \geq X_i$ for each $i \geq 0$. The bound on the upper tail in Lemma 2.1.6 holds for supermartingales, while the bound on the lower tail holds for submartingales. We will always use this to bound the sum of random variables using the following simple corollary.

Corollary 2.1.7. *Let* $(Z_i)_{i=1}^n$ *be a sequence of random variables. For each* $i \in [n]$*, let* $a_i, c_i \in \mathbb{R}$ *be constants such that* $|Z_i - a_i| \leq c_i$*.*

(i) *If* $\mathbb{E}[Z_i \mid Z_1, \ldots, Z_{i-1}] \leq a_i$*, then for each* $t > 0$*,*

$$\mathbb{P}\left[\sum_{i=1}^n Z_i \geq \sum_{i=1}^n a_i + t\right] \leq 2 \exp\left(-\frac{t^2}{2 \cdot \sum_{i=1}^n c_i^2}\right).$$

(ii) *If* $\mathbb{E}[Z_i \mid Z_1, \dots, Z_{i-1}] \geq a_i$, *then for each* $t > 0$,

$$\mathbb{P}\left[\sum_{i=1}^{n} Z_i \leq \sum_{i=1}^{n} a_i - t\right] \leq 2\exp\left(-\frac{t^2}{2 \cdot \sum_{i=1}^{n} c_i^2}\right).$$

Proof. We prove (i), and note that (ii) follows by applying (i) to the sequence $(-Z_i)_{i=1}^n$. Let $Y_0 = 0$ and, for each $i \in [n]$, let $Y_i = \sum_{i' \in [i]}(Z_{i'} - a_{i'})$. Then, $\mathbb{E}[Y_{i+1} \mid Y_1, \dots, Y_i] = \mathbb{E}[Z_{i+1} - a_{i+1} + Y_i \mid Y_1, \dots, Y_i] \leq Y_i$. Furthermore, for each $i \in [n]$, $|Y_i - Y_{i-1}| = |Z_i - a_i| \leq c_i$. Therefore, we can apply Lemma 2.1.6 for supermartingales to show that

$$\mathbb{P}\left[\sum_{i=1}^{n} Z_i \geq \sum_{i=1}^{n} a_i + t\right] = \mathbb{P}\left[Y_i - Y_0 \geq t\right] \leq 2\exp\left(-\frac{t^2}{2 \cdot \sum_{i=1}^{n} c_i^2}\right).$$

□

2.1.4 Structural lemmas

In this section we decompose undirected trees. Note that we will later apply this to directed trees as the edge directions do not affect the decompositions. We will use the following simple but useful lemma (see [60, Lemma 4.1]) which tells us that either a tree has many leaves, or it has many bare paths.

Lemma 2.1.8. *Let* $t, s \geq 2$, *and suppose that* T *is a tree with at most* t *leaves. Then there is some* m *and some vertex-disjoint bare paths* P_i, $i \in [m]$, *in* T *with length* s *so that* $|T - P_1 - \cdots - P_m| \leq 6st + 2|T|/(s+1)$.

We can now prove the following key lemma, in which we decompose a tree for our embedding.

Lemma 2.1.9. *Let* $0 \ll 1/n \ll 1/K \ll 1/k \ll \eta$. *Let* T *be a tree on* n *vertices with* $t \in V(T)$. *Then,* T *contains induced subgraphs* $T_0 \subset T_1 \subset T_2 \subset T_3 = T$, *such that* T_2 *is a tree, and the following hold.*

P1 $|T_0| \leq \eta n$ *and* $t \in V(T_0)$*.*

P2 T_1 *is formed from* T_0 *by the vertex-disjoint addition of trees,* S_v*,* $v \in V(T_0)$*, so that, for each* $v \in V(T_0)$*,* $S_v - v$ *is a forest consisting of trees of size at most* K*.*

P3 T_2 *is formed from* T_1 *by the addition of trees with size at least* k *and at most* K *attached to* T_1 *with exactly two bare paths of length 2.*

P4 $|T_3| - |T_2| \leq \eta n$*.*

Proof. Take ε and k' such that $1/K \ll \varepsilon \ll 1/k' \ll 1/k$. We start by finding a subtree T' of T which includes t and has few leaves, and is such that $T - V(T')$ is a forest of components with size at most K. We do this by including in T' every vertex which appears on the path in T from t to many other vertices. That is, for each $v \in V(T)$, let $w(v)$ be the number of vertices $u \in V(T)$ whose path from t to u includes v (in particular, v is such a vertex). Let T' be the subgraph of T induced on all the vertices $v \in V(T)$ with $w(v) \geq K + 1$.

For each $v \in V(T')$, let S_v be the tree containing v in $T - (V(T') \setminus \{v\})$. Note that $S_v - v$ is a forest which consists of trees with at most K vertices. Indeed, suppose T'' is a tree in $S_v - v$, and let v' be the neighbour of v in T''. Since every path from a vertex $u \in V(T'')$ to t in T goes through v' (and then v), we have that $K \geq w(v') \geq |T''|$ (and, in fact, the final inequality is an equality). Here, the first inequality holds because $v' \notin V(T')$. Observe further that, for any leaf v of T', $|S_v - v| = w(v) - 1 \geq K$, and, therefore, T' can have at most $n/K \leq \varepsilon n$ leaves.

Thus, by Lemma 2.1.8, for some $m \leq n/(k'+1)$, T' contains vertex disjoint bare paths $P_1, \ldots, P_m$ with length k' such that $t \notin V(P_i)$ for each $i \in [m]$ and

$$|T' - P_1 - \cdots - P_m| \leq 6k' \cdot \varepsilon n + 2n/(k'+1) + k' + 1 \leq \eta n/4. \tag{2.1.2}$$

For each $i \in [m]$, if possible, find within P_i a path P_i' with length at least $k' - 2\eta^3 k'$, such that, labelling its endvertices x_i and y_i, the following hold:

(i) for each x_i and y_i, $|S_{x_i} - x_i| \leq \eta k'/4$ and $|S_{y_i} - y_i| \leq \eta k'/4$, and

(ii) letting Q_i be the component of $T - \{x_i, y_i\}$ containing $P'_i - \{x_i, y_i\}$, we have $|Q_i| \leq K$.

Say, with relabelling, these paths are $P'_1, \ldots, P'_{m'}$. We will show that $m' \geq m - \eta n/2k'$. Note first that the number of $i \in [m]$ with no vertices x_i and $y_i \in V(P_i)$ respectively within $\eta^3 k'$ of the two endvertices of P_i, such that each of x_i and y_i had a forest with at most $\eta k'/4$ vertices deleted from them, is at most $n/(\eta^3 k' \cdot \eta k'/4) \leq \eta n/4k'$. Note further that the number of $i \in [m]$ with at least K vertices in $V(Q_i)$ is at most $n/K \leq \eta n/4k'$. Therefore, we can find such a path P'_i for all but at most $\eta n/2k'$ values of $i \in [m]$, so that $m' \geq m - \eta n/2k'$.

Let $T_0 = T[V(T') \setminus (\bigcup_{i \in [m']} V(P'_i))]$. We will show that $|T_0| \leq \eta n$. Note that, for each $i \in [m']$, $|V(P_i) \setminus V(P'_i)| \leq 2\eta^3 k'$. Therefore, as $m \leq n/k'$,

$$|T_0| \leq |T' - P'_1 - \cdots - P'_{m'}| \leq |T' - P_1 - \cdots - P_m| + k' \cdot \eta n/2k' + m \cdot 2\eta^3 k' \overset{(2.1.2)}{\leq} \eta n.$$

Furthermore, clearly $t \in V(T_0)$, and thus **P1** holds.

Let $T_1 = T[V(T_0) \cup (\bigcup_{v \in V(T_0)} V(S_v))]$. Recall that for each $v \in V(T')$, $S_v - v$ is a forest which consists of trees with at most K vertices. Therefore, **P2** holds.

Let $T_2 = T[V(T_1) \cup (\bigcup_{i \in [m']}(\{x_i, y_i\} \cup V(Q_i)))]$, and let $T_3 = T$. Here, we obtain T_2 by attaching the trees Q_i to vertices of T_1 by two bare paths of length 2, which have middle vertices given by the vertices x_i and y_i. Since Q_i contains the path P'_i for every $i \in [m']$, each of these trees contain at least $k' - 2\eta^3 k' - 2 \geq k$ vertices. On the other hand, by (ii), $|Q_i| \leq K$ for each $i \in [m']$ and so each such tree has size at most K. Therefore, **P3** holds. Furthermore, the only missing vertices from T are those in $S_v - v$ for each $v \in \{x_i, y_i : i \in [m']\}$, and thus T_2 is a tree. For each such v, $|S_v| \leq \eta k'/4$ by (i). Therefore, $|T_3| - |T_2| \leq (n/k') \cdot (2\eta k'/4) \leq \eta n$, and hence **P4** holds. □

2.1.5 Matchings between random sets

With high probability, any random subset of vertices in the digraph in Theorem 1.1.1 satisfies a similar minimum semidegree condition, as follows.

Lemma 2.1.10. *Let $1/n \ll c, \alpha$, and suppose D is an n-vertex digraph with $\delta^0(D) \geq (1/2+\alpha)n$. Let $A \subseteq V(D)$ be chosen uniformly at random subject to $|A| = cn$. Then, with high probability, for every vertex $v \in V(D)$, we have $\left|N_D^{\pm}(v, A)\right| \geq (1/2+\alpha/2)\,|A|$.*

Proof. Let v be an arbitrary vertex of D and let $A \subseteq V(D)$ be a uniformly random subset with $|A| = cn$. For $\diamond \in \{+, -\}$, we let $Z_v^{\diamond}$ be the random variable which measures $|N^{\diamond}(v) \cap A|$. Then $Z_v^{\diamond}$ has hypergeometric distribution with expectation

$$\mathbb{E}[Z_v^{\diamond}] = \frac{|N^{\diamond}(v)|\,|A|}{n} \geq \left(\frac{1}{2} + \alpha\right) cn.$$

Therefore, by Lemma 2.1.5, we have

$$\begin{aligned}\mathbb{P}\left[|Z_v^{\diamond} - \mathbb{E}[Z_x^{\diamond}]| > \frac{\alpha/2}{1/2+\alpha}(1/2+\alpha)cn\right] &\leq 2\exp\left(-\left(\frac{\alpha/2}{1/2+\alpha}\right)^2 \frac{(1/2+\alpha)cn}{3}\right)\\ &= 2\exp\left(\frac{-\alpha^2 cn}{6+12\alpha}\right).\end{aligned}$$

Then, applying a union bound, with probability at least

$$1 - 2n\exp\left(-\alpha^2 cn/(6+12\alpha)\right) = 1 - o(1),$$

we have that $Z_v^{\diamond} \geq (1/2+\alpha/2)\,|A|$ for each $\diamond \in \{+, -\}$ and $v \in V(D)$. □

The following digraph version of Hall's matching criterion implies a matching exists, as follows directly from the same result for undirected graphs.

Lemma 2.1.11. *Let D be a bipartite digraph with vertex classes A and B, and let $\diamond \in \{+, -\}$. Suppose that for every $S \subset A$, $|N_D^{\diamond}(S, B)| \geq |S|$. Then there is a $\diamond$-matching from A into B which covers A.*

We will refer to the condition in Lemma 2.1.11 as *Hall's criterion*. In combination with Lemma 2.1.10, Lemma 2.1.11 shows that with high probability there is a perfect matching between a large random pair of disjoint equal-sized vertex subsets in the digraph, as follows.

Proposition 2.1.12. *Let $1/n \ll p, \alpha$, and $p \leq 1/2$, and suppose D is an n-vertex digraph with $\delta^0(D) \geq (1/2 + \alpha)n$. Let A, B be chosen uniformly at random from all disjoint pairs of subsets of $V(D)$, each with size pn, and let $\diamond \in \{+, -\}$. Then, with high probability, there is a perfect $\diamond$-matching from A into B.*

Proof. By Lemma 2.1.10, with high probability we can assume the following. For all $v \in A$, we have $|N^{\pm}(v, B)| \geq (1/2 + \alpha/2)\,|B|$, and, for all $v \in B$, we have $|N^{\pm}(v, A)| \geq (1/2 + \alpha/2)\,|A|$. We will now show that Hall's criterion holds.

Let $S \subseteq A$, such that $S \neq \emptyset$ and $|S| \leq (1/2 + \alpha/2)pn$, and let $x \in S$. Then, $|N^{\diamond}(S, B)| \geq |N^{\diamond}(x, B)| \geq (1/2 + \alpha/2)pn \geq |S|$, so Hall's condition is trivially satisfied. Now take $S \subseteq A$, $|S| > (1/2 + \alpha/2)pn$, and assume for a contradiction that $|N^{\diamond}(S, B)| < |S|$. Then in particular, $B \setminus N^{\diamond}(S, B) \neq \emptyset$. Take $b \in B \setminus N^{\diamond}(S, B)$, and let $\circ \in \{+, -\}$ be such that $\circ \neq \diamond$. We have $|N^{\circ}(b, A)| \geq (1/2 + \alpha/2)pn$. However, since $b \notin N^{\diamond}(S, B)$, we have $N^{\circ}(b, A) \cap S = \emptyset$. So,

$$pn = |A| \geq |N^{\circ}(b, A)| + |S| \geq (1/2 + \alpha/2)pn + (1/2 + \alpha/2)pn = (1 + \alpha)pn > pn,$$

giving a contradiction. Thus, Hall's criterion is satisfied for all $S \subseteq A$ and so, since $|A| = |B|$, by Lemma 2.1.11, there is a perfect $\diamond$-matching from A into B. □

We use Proposition 2.1.12 to embed many vertex disjoint small trees, via the following two lemmas. In Lemma 2.1.13, we embed linearly many copies of a given constant-sized tree into specified subsets of our digraph. In Lemma 2.1.14, we embed a forest of constant-sized trees covering almost all the vertices in our digraph.

Lemma 2.1.13. *Let $1/n \ll 1/K, p, \alpha$ with $pK \leq 1$. Suppose T is an oriented K-vertex tree containing $t \in V(T)$. Let D be an n-vertex digraph with $\delta^0(D) \geq$*

$(1/2+\alpha)n$. Let V_1, V_2 be vertex disjoint subsets of $V(D)$ chosen uniformly at random subject to $|V_1| = pn$ and $|V_2| = (K-1)pn$.

Then, with high probability, $D[V_1 \cup V_2]$ contains pn vertex disjoint copies of T, in which t is copied into V_1 in each copy of T.

Proof. Let $V_1 = U_1$, and let $U_2 \cup \cdots \cup U_K$ be a partition of V_2 chosen uniformly at random so that $|U_i| = pn$ for each $i \in \{2, \ldots, K\}$. Note that the distribution of any pair of sets U_i, U_j with $1 \leq i < j \leq K$ is that of two disjoint vertex sets with size pn in $V(D)$, drawn uniformly at random from all such pairs.

Label the vertices of T by $t_1, \ldots, t_K$ so that $t_1 = t$ and $T[\{t_1, \ldots, t_i\}]$ is a tree for each $i \in \{1, \ldots, K\}$. For each $i \in \{2, \ldots, K\}$, let $j_i \in \{1, \ldots, i-1\}$ be such that t_{j_i} is the in- or out-neighbour in $T[\{t_1, \ldots, t_{i-1}\}]$ of the vertex t_i, and let $\diamond_i \in \{+, -\}$ be such that $t_i \in N_T^{\diamond_i}(t_{j_i})$.

Now by Proposition 2.1.12, for each $i \in \{2, \ldots, K\}$, with high probability, we can find a $\diamond_i$-matching from U_{j_i} into U_i. By applying a union bound, we see that, with high probability, for every $i \in \{2, \ldots, K\}$, there is a $\diamond_i$-matching, M_i say, from U_{j_i} into U_i.

Note that the union of these matchings, $\bigcup_{2 \leq i \leq K} M_i \subset D[V_1 \cup V_2]$ is the disjoint union of pn copies of T, in which, for each $i \in [K]$, the copy of t_i is in V_i. Thus, in each of these pn copies of T, $t = t_1$ is copied into $V_1 = U_1$, as required. $\square$

Lemma 2.1.14. *Let $1/n \ll 1/K, \varepsilon, \alpha$ and suppose F is a digraph with at most $(1-\varepsilon)n$ vertices which is the disjoint union of trees with size at most K. Let D be an n-vertex digraph with $\delta^0(D) \geq (1/2+\alpha)n$. Then, with high probability, D contains a copy of F.*

Proof. Arrange the components of F into isomorphism classes of trees $\mathbb{R}_1, \ldots, \mathbb{R}_\ell$, noting that we may take $\ell \leq (2K)^{K-1}$. For each $i \in [\ell]$, let $t_i = |\mathbb{R}_i|$ and let s_i be the size of each component in $\mathbb{R}_i$. Uniformly at random, take, in $V(D)$, disjoint subsets $V_{i,1}$ and $V_{i,2}$, $i \in [\ell]$, with $|V_{i,1}| = p_i n$ and $|V_{i,2}| = (s_i - 1)p_i n$, where $p_i = t_i/n + \varepsilon/\ell s_i$, for each $i \in [\ell]$. Note that this is possible, since

$$\sum_{i=1}^{\ell} s_i p_i n = \sum_{i=1}^{\ell} \left(s_i t_i + \frac{\varepsilon n}{\ell} \right) \leq n.$$

For each $i \in [\ell]$, we can apply Lemma 2.1.13 to show that, with high probability, there are $p_i n$ copies of the underlying tree of $\mathbb{R}_i$ in $D_i = D[V_{i,1} \cup V_{i,2}]$. Since $p_i n \geq t_i$, this implies that with high probability, we can find a copy of $\mathbb{R}_i$ in D_i for each $i \in [\ell]$. By applying a union bound and using that $1/n \ll 1/\ell$, we have, with high probability, that there is a copy of F in D. □

2.2 Almost-spanning trees

The key aim of this section is to prove Theorem 2.1.2, that is, to prove we can embed an almost-spanning tree T in our digraph. By Lemma 2.1.9, we can find $T_0 \subset T_1 \subset T_2 \subset T_3 = T$, satisfying **P1** to **P4**. In Section 2.2.1, we show that we can embed T_1. In Section 2.2.2, we show that we can embed $T_2 \setminus T_1$, and $T_3 \setminus T_2$. We conclude in Section 2.2.3 by combining this to obtain an embedding of T.

2.2.1 Embedding constant-sized trees as stars

As sketched in Section 2.1.2, we will embed T_0 randomly, leaf by leaf, using a guide set to embed each new vertex. Each guide set has an accompanying guide graph, which we later use to find a matching. The property of the guide graph that we use to find the matching is that it is *skew-bounded*, as follows.

Definition 2.2.1. *A digraph D with vertex sets $A, B \subset V(D)$ is $(a, b, \diamond)$*-skew-bounded *on (A, B) if $d^{\diamond}_D(v, B) \geq a$ for each $v \in A$ and $d^{\circ}_D(v, A) \leq b$ for each $v \in B$, where $\circ \in \{+, -\}$ and $\circ \neq \diamond$.*

This property can imply a matching exists via Hall's criterion, as follows.

Proposition 2.2.2. *Let $a \geq b$ and $\diamond \in \{+,-\}$. Suppose D is a digraph containing disjoint vertex sets $A, B \subset V(D)$, such that D is $(a,b,\diamond)$-skew-bounded on (A,B). Then, there is a $\diamond$-matching from A into B in D which covers A.*

Proof. Let $U \subset A$. As D is $(a,b,\diamond)$-skew-bounded on (A,B), there are at least $a\,|U|$ and at most $b\,|N^{\diamond}_D(U,B)|$ $\diamond$-edges from U to $N^{\diamond}_D(U,B)$. Thus, $|N^{\diamond}_D(U,B)| \geq a\,|U|\,/b \geq |U|$. Therefore, by Lemma 2.1.11, there is a $\diamond$-matching from A into B which covers A. □

In the following lemmas, we find our guide sets and guide graphs. We start by finding in D, for each $v \in V(D)$ and $\diamond \in \{+,-\}$, a guide set A and guide graphs which are skew-bounded on $(A, V(D))$.

Lemma 2.2.3. *Let $1/n \ll \varepsilon \ll \alpha, \eta \leq 1$ and $1/n \ll \mu \leq \alpha^2/2$. Let D be an n-vertex digraph with $\delta^0(D) \geq (1/2+\alpha)n$, let $v \in V(D)$ and let $\diamond \in \{+,-\}$.*

Then, there is a set $A \subset N^{\diamond}_D(v)$ with size μn and digraphs $H^+, H^- \subset D$ such that, for each $\circ \in \{+,-\}$, H° is $(\varepsilon n, (1+\eta)\mu\varepsilon n, \circ)$-skew-bounded on $(A, V(D))$.

Proof. We start by showing that we can label the vertices of $V(D)$ as $V(D) = \{x_1,\dots,x_n\} = \{y_1,\dots,y_n\}$ so that, for each $i \in [n]$,

$$|N^-_D(x_i) \cap N^{\diamond}_D(v) \cap N^+_D(y_i)| \geq \alpha^2 n. \tag{2.2.1}$$

To do this, create an auxiliary graph, as follows. For each $w \in V(D)$, create distinct new vertices w^- and w^+, and let $V^+ = \{w^+ : w \in V(D)\}$ and $V^- = \{w^- : w \in V(D)\}$. Consider the auxiliary bipartite graph H with vertex set $V^+ \cup V^-$, where for each $x, y \in V(D)$, there is an edge between x^+ and y^- if and only if $\left|N^-_D(x) \cap N^{\diamond}_D(v) \cap N^+_D(y)\right| \geq \alpha^2 n$.

Claim 2.2.4. $\delta(H) \geq (1/2+\alpha/2)n$.

Proof of Claim 2.2.4. Let $x \in V(D)$. We have $|N^-_D(x) \cap N^{\diamond}_D(v)| \geq n-(n-d^-_D(x))-(n-d^{\diamond}_D(v)) \geq 2\alpha n$. Let $B = N^-_D(x) \cap N^{\diamond}_D(v)$ and $Y = \{y \in V(D) : |N^+_D(y) \cap B| \geq$

$\alpha^2 n\}$, and note that $d_H(x^+) = |Y|$.

For each $u \in B$, we have $|N_D^-(u)| \geq (1/2+\alpha)n$, and thus $e_D(V(D), B) \geq (1/2+\alpha)|B|n$. By the choice of Y, we have $e_D(V(D), B) \leq |Y||B| + \alpha^2 n^2$. Therefore, as, in addition, $2\alpha n \leq |B|$, we have

$$(1/2+\alpha)|B|n \leq |Y||B| + \alpha^2 n^2 \leq |Y||B| + \alpha|B|n/2.$$

Thus, $(1/2+\alpha/2)|B|n \leq |Y||B|$, so that $|Y| \geq (1/2+\alpha/2)n$. Therefore, $d_H(x^+) = |Y| \geq (1/2+\alpha/2)n$.

A similar argument, with the signs reversed, shows that $d_H(y^-) \geq (1/2+\alpha/2)n$ for each $y \in V(D)$, completing the proof of the claim. $\square$

As in the proof of Proposition 2.1.12, Claim 2.2.4 easily implies that Hall's criterion is satisfied, so that there is a matching from V^+ to V^- in H. That is, we can label the vertices of $V(D)$ as $V(D) = \{x_1, \dots, x_n\} = \{y_1, \dots, y_n\}$ so that, for each $i \in [n]$, (2.2.1) holds.

We will now show by induction that, for each $0 \leq i \leq \mu n$, there is a set $A_i \subset N_D^\diamond(v)$ with size i and graphs $H_i^+, H_i^- \subset D$ such that, for each $\circ \in \{+,-\}$, H_i° is $(\varepsilon n, (1+\eta)\mu\varepsilon n, \circ)$-skew-bounded on $(A_i, V(D))$, $e(H_i^\circ) = i\varepsilon n$, and, for each $j \in [n]$, $d^-_{H_i^+}(x_j) = d^+_{H_i^-}(y_j)$.

Note that if $A_0 = \emptyset$ and if H_0^+, H_0^- have no edges and vertex set $V(D)$, then the conditions hold, so assume that $0 \leq i < \mu n$ and we have $A_i \subset N_D^\diamond(v)$ and $H_i^+, H_i^- \subset D$ as described.

Let $J_i \subset [n]$ be the set of $j \in [n]$ for which $d^-_{H_i^+}(x_j) = d^+_{H_i^-}(y_j) \leq (1+\eta/2)\mu\varepsilon n$. Note that, as $e(H_i^+) = e(H_i^-) = i\varepsilon n < \mu\varepsilon n^2$, we have

$$(n - |J_i|)(1+\eta/2)\mu\varepsilon n \leq \mu\varepsilon n^2.$$

Thus, as $\eta \leq 1$, $(n - |J_i|) \leq n/(1+\eta/2) \leq n(1-\eta/4)$, so that $|J_i| \geq \eta n/4$.

For each $j \in J_i$, let $W_{i,j} = (N^-_D(x_j) \cap N^\diamond_D(v) \cap N^+_D(y_j)) \setminus A_i$, noting that, by (2.2.1), $|W_{i,j}| \geq \alpha^2 n - i > \alpha^2 n - \mu n \geq \alpha^2 n/2$. By averaging, choose some $w_i \in V(D)$ such that

$$|\{j \in J_i : w_i \in W_{i,j}\}| \geq \frac{\sum_{j \in J_i} |W_{i,j}|}{n} \geq \frac{\eta n/4 \cdot \alpha^2 n/2}{n} \geq \varepsilon n,$$

using that $\alpha, \eta \gg \varepsilon$. Choose a set $J'_i \subset \{j \in J_i : w_i \in W_{i,j}\}$ with size εn. Let $A_{i+1} = A_i \cup \{w_i\}$. Let H^+_{i+1} be the digraph H^+_i with edges $w_i x_j$, $j \in J'_i$, added. Note that, as $d^-_{H^+_i}(x_j) \leq (1 + \eta/2)\mu\varepsilon n$ for each $j \in J'_i$, H^+_{i+1} is $(\varepsilon n, (1+\eta)\mu\varepsilon n, +)$-skew-bounded on $(A_{i+1}, V(D))$. Furthermore, by the definition of $W_{i,j}$, the edges added to H^+_i are in D, and therefore $H^+_{i+1} \subset D$.

Let H^-_{i+1} be the digraph H^-_i with the edges $y_j w_i$, $j \in J'_i$, added. Note that, similarly, H^-_{i+1} is $(\varepsilon n, (1+\eta)\mu\varepsilon n, -)$-skew-bounded on $(A_{i+1}, V(D))$. Finally, noting that A_{i+1} has size $i+1$, that $e(H^+_{i+1}) = e(H^-_{i+1}) = (i+1)\varepsilon n$ and that, for each $j \in [n]$, $d^-_{H^+_{i+1}}(x_j) = d^+_{H^-_{i+1}}(y_j)$, completes the inductive step, and hence the proof. $\square$

We now show that the guide sets and guide graphs found by Lemma 2.2.3 have a similar skew-bounded property when restricted to random vertex subsets, as follows.

Lemma 2.2.5. *Let $1/n \ll \varepsilon \ll \alpha, \eta \leq 1$ and $1/n \ll 1/k, p_0, p_1, \ldots, p_k \leq 1$. Let $\mu = \alpha^2 p_0/4$. Let D be an n-vertex digraph with $\delta^0(D) \geq (1/2 + \alpha)n$. Let $V_0, V_1, \ldots, V_k \subset V(D)$ be disjoint random sets chosen uniformly at random subject to $|V_i| = p_i n$ for each $i \in \{0, \ldots, k\}$.*

Then, with high probability, for each $v \in V(D)$ and $\diamond \in \{+, -\}$, there is a set $A_{v,\diamond} \subset N^\diamond_D(v) \cap V_0$ with size μn and digraphs $H^\circ_{v,\diamond} \subset D$, $\circ \in \{+, -\}$, such that, for each $\circ \in \{+, -\}$ and $i \in [k]$, $H^\circ_{v,\diamond}$ is $(\varepsilon p_i n, (1+\eta)\varepsilon\mu n, \circ)$-skew-bounded on $(A_{v,\diamond}, V_i)$.

Proof. By Lemma 2.2.3, applied with $\varepsilon' = (1 + \eta/4)\varepsilon$, $\eta' = \eta/4$ and $\mu' = (1 + \eta/4)\alpha^2/4$, for each $v \in V(D)$ and $\diamond \in \{+, -\}$, there is a set $\bar{A}_{v,\diamond} \subset N^\diamond_D(v)$ with size $(1 + \eta/4)\alpha^2 n/4$ and digraphs $H^+_{v,\diamond}, H^-_{v,\diamond} \subset D$ such that, for each $\circ \in \{+, -\}$, $H^\circ_{v,\diamond}$ is $((1 + \eta/4)\varepsilon n, (1 + \eta/4)^3 \varepsilon\alpha^2 n/4, \circ)$-skew-bounded on $(\bar{A}_{v,\diamond}, V(D))$.

Select $V_0, V_1, \ldots, V_k \subset V(D)$ according to the distribution in the statement of the lemma. Using Lemma 2.1.5, and a union bound, we have that, with high probability, the following hold.

Q1 For each $v \in V(D)$ and $\diamond \in \{+, -\}$, $|\bar{A}_{v,\diamond} \cap V_0| \geq \alpha^2 p_0 n/4 = \mu n$.

Q2 For each $v \in V(D)$, $\diamond, \circ \in \{+, -\}$, and $w \in \bar{A}_{v,\diamond}$, $|N^{\circ}_{H^{\circ}_{v,\diamond}}(w, V_i)| \geq \varepsilon p_i n$.

Q3 For each $v \in V(D)$, $\diamond, \circ \in \{+, -\}$, and $w \in V(D)$, $|N^{\bar{\circ}}_{H^{\circ}_{v,\diamond}}(w, \bar{A}_{v,\diamond}) \cap V_0)| \leq (1+\eta)\varepsilon\alpha^2 p_0 n/4 = (1+\eta)\varepsilon\mu n$, where $\bar{\circ} \in \{+, -\}$ is such that $\bar{\circ} \neq \circ$.

Indeed, by Lemma 2.1.5, as $\varepsilon, \eta, \alpha, p_0, p_1, \ldots, p_k \gg 1/n$, for any instance of $v \in V(D)$, $\diamond, \circ \in \{+, -\}$, and $w \in V(D)$, the property **Q1** above holds with probability $1 - \exp(-\Omega(n))$, and the same is true for **Q2** and **Q3**. Therefore, by a union bound, with high probability, the properties **Q1**, **Q2** and **Q3** hold.

Now, for each $v \in V(D)$ and $\diamond \in \{+, -\}$, using **Q1**, choose $A_{v,\diamond} \subset \bar{A}_{v,\diamond} \cap V_0$ with $|A_{v,\diamond}| = \mu n$. By **Q2** and **Q3**, we have, for each $\circ \in \{+, -\}$ and $i \in [k]$, that $H^{\circ}_{v,\diamond}$ is $(\varepsilon p_i n, (1+\eta)\varepsilon\mu n, \circ)$-skew-bounded on $(A_{v,\diamond}, V_i)$, as required. □

We will now use the guide sets produced by Lemma 2.2.5 to randomly embed T_0, the small core of the original tree, and then use the guide graphs to find matchings from certain subsets of the image of the embedding to other random sets, as follows.

Lemma 2.2.6. *Let $1/n \ll c \ll \beta \ll \eta, q, \alpha \leq 1$ and $1/n \ll c \ll p \ll 1/m$. Let D be an n-vertex digraph with $\delta^0(D) \geq (1/2 + \alpha)n$.*

Let T be an oriented tree with $\Delta^{\pm}(T) \leq cn/\log n$ consisting of a subtree $T_0 \subset T$ with $|T_0| \leq \beta n$, such that every vertex in $V(T) \setminus V(T_0)$ is attached as a leaf to T_0. Let $t \in V(T_0)$. Let $U_0 = V(T_0)$ and let $U_1 \cup \ldots \cup U_m$ be a partition of $V(T) \setminus V(T_0)$ such that, for each $i \in [m]$, either $e_T(V(T_0), U_i) = 0$ or $e_T(U_i, V(T_0)) = 0$. Let $V_0, V_1, \ldots, V_m \subset V(D)$ be disjoint random sets chosen uniformly at random subject to $|V_0| = qn$, and, for each $i \in [m]$, $|V_i| = \lfloor (1+\eta)|U_i| \rfloor + pn$.[1]

[1]Note that this gives an implicit bound on $|T|$.

Then, with high probability, for each $s \in V_0$, there is an embedding of T into D such that t is embedded to s, and, for each $i \in \{0, 1, \ldots, m\}$, U_i is embedded into V_i.

Proof. Choose ε such that $\beta \ll \varepsilon \ll \eta, q, \alpha$. For each $j \in [m]$, let $p_j = (\lfloor(1+\eta)|U_j|\rfloor/n) + p$. Choose $V_0, V_1, \ldots, V_m$ according to the distribution in the lemma. By Lemma 2.2.5 applied with $\eta' = \eta/2$ and $p_0 = q$, with high probability, for each $v \in V(D)$ and $\diamond \in \{+, -\}$, there is

R1 a set $A_{v,\diamond} \subset N_D^\diamond(v) \cap V_0$ with size $q\alpha^2 n/4$, and

R2 digraphs $H_{v,\diamond}^\circ \subset D$, $\circ \in \{+, -\}$, such that, for each $j \in [m]$, $H_{v,\diamond}^\circ$ is $(\varepsilon p_j n, (1+\eta/2)\varepsilon q\alpha^2 n/4, \circ)$-skew-bounded on $(A_{v,\diamond}, V_j)$.

We will now show that, given only **R1** and **R2**, we can embed T as required in the lemma for each $s \in V_0$. Let then $s \in V_0$. We will randomly embed T_0 into $D[V_0]$, as follows, before showing that, with high probability, it can be extended into the required copy of T. Let $\ell = |T_0|$ and label $V(T_0) = \{t_1, \ldots, t_\ell\}$, so that $t_1 = t$ and $T_0[\{t_1, \ldots, t_i\}]$ is a tree for each $i \in [\ell]$. Let $s_1 = s$ and embed t_1 to s_1. For each $i \in \{2, \ldots, \ell\}$ in turn, let $j_i \in \{1, \ldots, i-1\}$ be such that t_{j_i} is the in- or out-neighbour of t_i in $T_0[\{t_1, \ldots, t_i\}]$ and let $\diamond_i \in \{+, -\}$ be such $t_i \in N_{T_0}^{\diamond_i}(t_{j_i})$, and embed t_i to $s_i \in A_{s_{j_i},\diamond_i} \setminus \{s_1, \ldots, s_{i-1}\}$ uniformly at random. Such an embedding is possible since, for every $v \in V(D)$ and $\diamond \in \{+, -\}$, $|A_{v,\diamond}|$ is much larger than $|T_0|$ as $\beta \ll q, \alpha$.

Claim 2.2.7. *For each $j \in [m]$, with high probability, the embedding of T_0 can be extended to an embedding of $T[V(T_0) \cup U_j]$ by embedding U_j into V_j.*

As $p \gg 1/n$, and $m \leq 1/p$, we can take a union bound over all $j \in [m]$, to show that, with high probability, for each $j \in [m]$, the embedding of T_0 can be extended to $T[V(T_0) \cup U_j]$ by embedding U_j into V_j, and hence T can be embedded as required in the lemma. Therefore, there is some choice of the embedding of T_0 for which this can be done. It is left then to prove Claim 2.2.7.

Proof of Claim 2.2.7. Let $j \in [m]$ and let $\circ_j \in \{+,-\}$ be such that all the edges from $V(T_0)$ to U_j in T are $\circ_j$-edges. For each $i \in [\ell]$, let $d_{j,i} = |N_T^{\circ_j}(t_i, U_j)|$. For each $i \in [\ell]$, take $d_{j,i}$ new vertices from V_j and call them $w_{j,i,i'}$, $i' \in [d_{j,i}]$. Let $W_j = \{w_{j,i,i'} : i \in [\ell], i' \in [d_{j,i}]\}$. Let K_j be the directed graph with vertex set $W_j \cup V_j$, containing only $\circ_j$-edges from W_j to V_j, and where, for each $i \in [\ell]$, $i' \in [d_{j,i}]$ and $v \in V_j$, there is a $\circ_j$-edge from $w_{j,i,i'}$ to v in K_j if, and only if, $s_i v \in E(H^{\circ_j}_{s_{j_i},\diamond_i})$.

We will show that, with high probability, K_j is $(\varepsilon p_j n, \varepsilon p_j n, \circ_j)$-skew-bounded on (W_j, V_j). This is enough to prove the claim, as, by Proposition 2.2.2, there is a $\circ_j$-matching from W_j into V_j in K_j which covers W_j. Thus, we can label distinct vertices $v'_{j,i,i'}$, $i \in [\ell]$, $i' \in [d_{j,i}]$ in V_j so that $w_{j,i,i'}v'_{j,i,i'}$, $i \in [\ell]$ and $i' \in [d_{j,i}]$, is a matching in K_j. For each $i \in [\ell]$, use the vertices $v'_{j,i,i'}$, $i' \in [d_{j,i}]$, to embed the $d_{j,i}$ $\circ_j$-neighbours of t_i in U_j into V_j. This is possible as, by the definition of K_j and $H^{\circ_j}_{s_{j_i},\diamond_i}$, $s_i v'_{j,i,i'}$ is a $\circ_j$-edge in D. Therefore, this extends the embedding of T_0 to an embedding of $T_0 \cup T[U_j]$ with U_j embedded into V_j, as required.

Thus, it is sufficient to prove that, with high probability, K_j is $(\varepsilon p_j n, \varepsilon p_j n, \circ_j)$-skew-bounded on (W_j, V_j). Now, for each $i \in [\ell]$, $s_i \in A_{j_i,\diamond_i}$, and therefore s_i has at least $\varepsilon p_j n$ $\circ_j$-neighbours in V_j in $H^{\circ_j}_{s_{j_i},\diamond_i}$ by **R2**. Therefore, for each $i \in [\ell]$ and $i' \in [d_{j,i}]$, $w_{j,i,i'}$ has at least $\varepsilon p_j n$ $\circ_j$-neighbours in K_j. That is, each $v \in W_j$ has at least $\varepsilon p_j n$ $\circ_j$-neighbours in K_j. Thus, letting $\bar{\circ}_j \in \{+,-\}$ with $\bar{\circ}_j \neq \circ_j$, it is sufficient to prove that, for each $v \in V_j$, with probability $1 - o(n^{-1})$, $d^{\bar{\circ}_j}_{K_j}(v, W_j) \leq \varepsilon p_j n$.

Let then $v \in V_j$. For each $i \in [\ell]$, let

$$X_i^{j,v} = \begin{cases} d_{j,i} & \text{if } s_i v \in E(H^{\circ_j}_{s_{j_i},\diamond_i}) \\ 0 & \text{otherwise,} \end{cases}$$

so that $d^{\bar{\circ}_j}_{K_j}(v, W_j) = \sum_{i \in [\ell]} X_i^{j,v}$. Note that, when $s_i \in A_{s_{j_i},\diamond_i} \setminus \{s_1, \dots, s_{i-1}\}$ is chosen uniformly at random, by **R1** and **R2**, and as $\beta \ll \eta, \alpha, q$ and $i \leq \ell \leq \beta n$, if $d_{i,j} > 0$, then $X_i^{j,v} = d_{j,i}$ with probability at most

$$\frac{d_{H^{\circ_j}_{s_{j_i},\diamond_i}}(v)}{|A_{s_{j_i},\diamond_i} \setminus \{s_1, \ldots, s_{i-1}\}|} \leq \frac{(1+\eta/2)\varepsilon q\alpha^2 n/4}{q\alpha^2 n/4 - (i-1)} \leq (1+\eta)\varepsilon.$$

Let $\gamma = (1+\eta)\varepsilon$. Then, for each $i \in [\ell]$, $\mathbb{E}[X_i^{j,v}|X_1^{j,v}, \ldots, X_{i-1}^{j,v}] \leq \gamma \cdot d_{j,i}$. Note that the inequality

$$a^2 + b^2 \leq (a-1)^2 + (b+1)^2 \tag{2.2.2}$$

holds whenever $1 \leq a \leq b$. Repeated application of this inequality shows that $\sum_{i\in[\ell]} d_{j,i}^2$ is maximised when as many of the $d_{j,i}$ are maximised as possible. Therefore, as $d_{j,i} \leq cn/\log n$ for each $i \in [\ell]$, and $\sum_{i\in[\ell]} d_{j,i} \leq |U_j| \leq n$, we have $\sum_{i\in[\ell]} d_{j,i}^2 \leq (n/(cn/\log n))(cn/\log n)^2 = cn^2/\log n$. Using this and the fact that $|X_i^{j,v} - \gamma \cdot d_{j,i}| \leq d_{j,i}$ for each $i \in [\ell]$ and $c \ll p$, we can apply Corollary 2.1.7 (i) with $a_i = \gamma \cdot d_{j,i}$, $c_i = d_{j,i}$ and $t = pn/3$ to get

$$\begin{aligned}\mathbb{P}\left[\sum_{i\in[\ell]} X_i^{j,v} \geq \gamma \cdot \left(\sum_{i\in[\ell]} d_{j,i}\right) + pn/3\right] &\leq 2\exp\left(\frac{-(pn/3)^2}{2\cdot\sum_{i\in[\ell]} d_{j,i}^2}\right) \\ &\leq 2\exp\left(\frac{-p^2\log n}{18c}\right) \leq o(n^{-1}).\end{aligned}$$

Thus, with probability $1 - o(n^{-1})$, we have

$$\begin{aligned}d_{K_j}^{\bar{\circ}_j}(v, W_j) &= \sum_{i\in[\ell]} X_i^{j,v} < \gamma \cdot \left(\sum_{i\in[\ell]} d_{j,i}\right) + pn/3 < \gamma|U_j| + pn/3 \leq \gamma p_j n/(1+\eta) \\ &= (1+\eta)\varepsilon p_j n/(1+\eta) = \varepsilon p_j n,\end{aligned}$$

completing the proof of the claim. □

Thus, this concludes the proof of the lemma. □

Finally, by combining Lemma 2.2.6 and Lemma 2.1.13, we can prove Lemma 2.1.4.

Proof of Lemma 2.1.4. Let p satisfy $1/n \ll c \ll p \ll \varepsilon, 1/K$. For each $j \in [\ell]$, let s_j

be the vertex of S_j with an in- or out-neighbour in $V(T')$ in T. Let $\mathcal{R}$ be a maximal set of pairs (R, r) for which R is a directed tree with at most K edges and $r \in V(R)$, such that the pairs (R, r) are unique up to isomorphism. Let $m = |\mathcal{R}|$ and enumerate $\mathcal{R}$ as $(R_1, r_1), \dots, (R_m, r_m)$. Note that $p \ll 1/m$ since m is a function of K.

Let $T'' = T[V(T') \cup N_T^+(V(T')) \cup N_T^-(V(T'))]$. For each $i \in [m]$ and $\diamond \in \{+, -\}$, let $U_{i,\diamond} \subset V(T'')$ be the set of vertices s_j, $j \in [\ell]$, for which (S_j, s_j) is isomorphic to (R_i, r_i) and the edge from $V(T')$ to s_j in T is a $\diamond$-edge.

In $V(D)$, take disjoint random sets V_0 and $V_{i,\diamond,j}$, $i \in [m]$, $\diamond \in \{+, -\}$ and $j \in \{1, 2\}$, uniformly at random subject to the following.

- $|V_0| = \varepsilon n/2$.
- For each $i \in [m]$ and $\diamond \in \{+, -\}$, we have that $|V_{i,\diamond,1}| = \lfloor (1 + \varepsilon/6)|U_{i,\diamond}| \rfloor + pn$ and $|V_{i,\diamond,2}| = (\lfloor (1 + \varepsilon/6)|U_{i,\diamond}| \rfloor + pn)(|R_i| - 1)$.

Note that this is possible, as

$$\begin{aligned}
|V_0| + \sum_{i \in [m], \diamond \in \{+,-\}} (|V_{i,\diamond,1}| + |V_{i,\diamond,2}|) &= |V_0| + \sum_{i \in [m], \diamond \in \{+,-\}} (\lfloor (1 + \varepsilon/6)|U_{i,\diamond}| \rfloor + pn)|R_i| \\
&\leq \varepsilon n/2 + (1 + \varepsilon/6) \sum_{j \in [\ell]} |S_j| + \sum_{i \in [m]} 2pn \cdot |R_i| \\
&\leq \varepsilon n/2 + (1 + \varepsilon/6)|T| + (2pn) \cdot m \cdot (K + 1) \\
&\leq n.
\end{aligned}$$

Now, with probability $\varepsilon/2$, $v \in V_0$. By Lemma 2.2.6, with high probability, if $v \in V_0$, then there is an embedding of T'' into D such that t is embedded to v, $V(T') \subset V_0$, and, for each $i \in [m]$ and $\diamond \in \{+, -\}$, $U_{i,\diamond}$ is embedded into $V_{i,\diamond,1}$. By Lemma 2.1.13, for each $i \in [m]$ and $\diamond \in \{+, -\}$, $D[V_{i,\diamond,1} \cup V_{i,\diamond,2}]$ contains $|V_{i,\diamond,1}|$ vertex disjoint copies of R_i, in which r_i is copied into $V_{i,\diamond,1}$. For each $i \in [m]$ and $\diamond \in \{+, -\}$, add each copy of R_i containing an embedded vertex of $U_{i,\diamond}$ to the embedding of T''. Note that this results in a copy of T. □

2.2.2 Embedding constant-sized trees as paths

Given our decomposition $T_0 \subset T_1 \subset T_2 \subset T_3 = T$, we have now embedded T_1. We now embed the vertices from $V(T_2) \setminus V(T_1)$, recalling that we obtain T_2 from T_1 by adding constant-sized trees, where each tree is attached to T_1 by exactly two bare paths of length 2. In the following lemma, we embed $T_2 \setminus T_1$ so that the vertices in $V(T_2) \cap V(T_1)$ are embedded to preselected vertices (labelled a_i, b_i, $i \in [\ell]$). This allows us to extend our embedding of T_1 to one of T_2.

Lemma 2.2.8. *Let $1/n \ll 1/K \leq 1/k \ll \alpha, \varepsilon$. Suppose T is a forest formed of vertex-disjoint oriented trees T_i, $i \in [\ell]$, with at most $(1-\varepsilon)n$ vertices in total, and so that $k \leq |T_i| \leq K$, for each $i \in [\ell]$, and each tree T_i contains distinct vertices r_i and s_i which are leaves in T_i whose neighbour has total in- and out-degree 2.*

Suppose D is an n-vertex digraph with $\delta^0(D) \geq (1/2+\alpha)n$, containing the distinct vertices a_i, b_i, $i \in [\ell]$. Then, D contains a copy of T in which, for each $i \in [\ell]$, r_i is embedded to a_i and s_i is embedded to b_i.

Proof. Let β be such that $1/k \ll \beta \ll \alpha, \varepsilon$. For each $i \in [\ell]$, let r_i' and s_i' be the neighbours in T_i of r_i and s_i, respectively, and let $T_i' = T_i - \{r_i, r_i', s_i, s_i'\}$. Let T' be the forest composed of connected components T_i', $i \in [\ell]$, so that $|T_i'| \leq (1-\varepsilon)n$. Let $A = \{a_i, b_i : i \in [\ell]\}$. Then $|A| = 2\ell \leq 2n/k$. Let $B \subset V(D) \setminus A$ be a random subset of vertices with $|B| = \beta n$.

Let $D' = D - A - B$. As $1/k, \beta \ll \alpha, \varepsilon$, we have $|D'| \geq (1-\varepsilon/4)n$ and $\delta^0(D') \geq (1/2 + \alpha/2)\,|D'|$. Since

$$|T'| \leq (1-\varepsilon)n \leq \frac{(1-\varepsilon)}{(1-\varepsilon/4)}|D'| \leq (1-\varepsilon/2)\,|D'|\,,$$

by Lemma 2.1.14, we can find a copy, S' say, of T' inside D' with high probability.

Let r_i'' and s_i'' be the neighbours in T' of r_i' and s_i', respectively, for each $i \in [\ell]$, and let a_i'' and b_i'' be the copy of r_i'' and s_i'' in S', respectively.

Claim 2.2.9. *The following holds with high probability. For any pair of vertices $u, v \in V(D)$ and $\diamond, \circ \in \{+,-\}$, we have that $|N^\diamond(u) \cap N^\circ(v) \cap B| \geq \alpha\beta n$.*

Proof of Claim 2.2.9. Let $u, v \in V(D)$ and $\diamond, \circ \in \{+,-\}$. Note that, by the semi-degree condition on D, $|N^\diamond(u) \cap N^\circ(v)| \geq 2\alpha n$, and hence $|N^\diamond(u) \cap N^\circ(v) \cap B|$ has a hypergeometric distribution with $\mathbb{E}\,|N^\diamond(u) \cap N^\circ(v) \cap B| \geq 2\alpha\beta n$. By Lemma 2.1.5, and a union bound over all pairs $u, v \in D$ and $\diamond, \circ \in \{+,-\}$, the statement in the claim thus holds with probability $1 - o(1)$. □

Thus, with high probability, we can assume the property in the claim holds. Now, for each $i \in [\ell]$, embed r_i and s_i to a_i and b_i, respectively. Let $\diamond_i, \circ_i, \diamond_i', \circ_i' \in \{+,-\}$ be such that $r_i' \in N^{\diamond_i}(r_i) \cap N^{\circ_i}(r_i'')$, and $s_i' \in N^{\diamond_i'}(s_i) \cap N^{\circ_i'}(s_i'')$. Greedily and disjointly, for each $i \in [\ell]$, embed r_i' to a vertex in $N^{\diamond_i}(a_i) \cap N^{\circ_i}(a_i'') \cap B$ and embed s_i' to a vertex in $N^{\diamond_i'}(b_i) \cap N^{\circ_i'}(b_i'') \cap B$. Note that this is possible, since, from the property in the claim we have, for each $i \in [\ell]$

$$|N^{\diamond_i}(a_i) \cap N^{\circ_i}(a_i'') \cap B|\,, \left|N^{\diamond_i'}(b_i) \cap N^{\circ_i'}(b_i'') \cap B\right| \geq \alpha\beta n \geq \frac{2n}{k} \geq 2\ell.$$

This completes the embedding of T with the property required in the lemma. □

2.2.3 Proof of Theorem 2.1.2

We now combine Lemma 2.1.4 and Lemma 2.2.8 to find a copy of any almost-spanning tree.

Proof of Theorem 2.1.2. Take K, k and η so that $c \ll 1/K \ll 1/k \ll \eta \ll \varepsilon, \alpha$. Let D be an n-vertex graph with $\delta^0(D) \geq (1/2 + \alpha)n$. Let T be an oriented tree on at most $(1-\varepsilon)n$ vertices with $\Delta^\pm(T) \leq cn/\log n$. By Lemma 2.1.9, we can find forests $T_0 \subset T_1 \subset T_2 \subset T_3 = T$ satisfying **P1** to **P4**. Randomly partition $V(D)$ into three parts, $V(D) = V_1 \cup V_2 \cup V_3$ so that $|V_1| = |T_1| + \varepsilon n/3$, $|V_2| = |T_2| - |T_1| + \varepsilon n/3$, and $|V_3| = |T| - |T_2| + \varepsilon n/3$. Note that, with probability at least $\varepsilon/3$, we have $v \in V_1$.

By applying Lemma 2.1.10 with $A = V_1$, we have $\delta^0(D[V_1]) \geq (1/2 + \alpha/2)\,|V_1|$ with high probability. Thus, by **P1** and **P2**, we can apply Lemma 2.1.4 to $D = D[V_1]$ and $T = T_1$ and find a copy of T_1 in V_1 in which t is copied to v. By **P3**, for some $\ell \in \mathbb{N}$, T_2 is formed from T_1 by the addition of trees F_i, $i \in [\ell]$, where $k \leq |F_i| \leq K$, which are each attached to T_1 by exactly two bare paths of length 2, P_i and Q_i say. For each $i \in [\ell]$, let p_i and q_i be the endpoint of P_i and Q_i, respectively, which belongs to T_1. Let a_i and b_i be the embedding in V_1 of p_i and q_i, respectively, and let $A = \{a_i, b_i : i \in [\ell]\}$.

By Lemma 2.1.10 again, we have, with high probability, $\delta^0(D[A \cup V_2]) \geq (1/2 + \alpha/2)\,|A \cup V_2|$. Applying Lemma 2.2.8 to $D[A \cup V_2]$ with $T_i = F_i \cup P_i \cup Q_i$, $r_i = p_i$, and $s_i = q_i$, for each $i \in [\ell]$, we can find a copy of T_2 in $D[V_1 \cup V_2]$. Now since T_2 is a tree, any vertex in $T_3 \setminus T_2$ can have at most one neighbour in T_2. Note that, by Lemma 2.1.10, we know that with high probability every vertex in D has at least $(1/2 + \alpha/2)\,|V_3| \geq \eta n$ in-neighbours in V_3 and at least $(1/2 + \alpha/2)\,|V_3| \geq \eta n$ out-neighbours in V_3. Let $j = |T_3| - |T_2| \leq \eta n$ (where the inequality holds by **P4** and order the vertices of $T_3 \setminus T_2$ as $u_1, \ldots, u_j$, so that $T[V(T_2) \cup \{u_1, \ldots, u_i\}]$ is a tree for each $i \in [j]$. Embed the vertices $u_1, \ldots, u_j$ greedily into V_3, to complete the copy of T in D. Noting that this embedding was successful with probability at least $\varepsilon/3 - o(1) > 0$, there must always be such a copy of T. $\square$

2.3 Absorption from switching

The aim of this section is to prove Theorem 2.1.1. The main idea is as follows. Given a small tree T, we split it into two trees T' and T'' and randomly embed T' vertex by vertex. With positive probability, the resulting tree is such that, given the right number of other vertices in the graph, we can embed T'' to extend this into a copy of T while making some small modifications to the copy of T'. Essentially, we show that, for each vertex y, there are many vertices in the embedding of T' which we can switch with y and still get a copy of T. We then embed T'' vertex-by-vertex, at each step switching an unused vertex into the copy of T' in place of a vertex which we can instead use to extend the (partial) embedding of T''.

Proof of Theorem 2.1.1. Take λ such that $\varepsilon \ll \lambda \ll \mu$. Using Proposition 2.1.3, let $T = T' \cup T''$, where $t \in V(T')$ and $\varepsilon n < |T''| \leq 3\varepsilon n$. Let $\ell = |T'|$, and label $V(T')$ as $t_1, \ldots, t_\ell$ so that $t_1 = t$, $T'[t_1, \ldots, t_i]$ is a tree for each $i \in [\ell]$, and the leaves of T' appear last in this order (except possibly for t) and in any bare path of length 6 the middle 3 vertices appear consecutively. For each $i \in [\ell]$, let $T_i = T'[\{t_1, \ldots, t_i\}]$, so t_i is a leaf of T_i.

Pick an arbitrary vertex $v \in V(D)$, and let R_1 be the graph with only the vertex v. For each $i = 2, \ldots, \ell$, do the following. Let $\diamond_i \in \{+,-\}$ be such that $N^{\diamond_i}_{T_i}(t_i)$ is non-empty (and thus contains exactly one vertex). Let $\circ_i \in \{+,-\}$ with $\circ_i \neq \diamond_i$. Take R_{i-1}, which is a copy of T_{i-1}, and let w_i be the copy of the sole vertex in $N^{\diamond_i}_{T_i}(t_i)$ in R_{i-1}. Pick a vertex v_i independently at random from $N^{\circ_i}_D(w_i) \setminus V(R_{i-1})$. Embed t_i to v_i to get R_i, a copy of T_i.

Note that this process always ends with a copy of T', as $N^{\circ_i}_D(w_i) \setminus V(R_{i-1})$ always has size at least $d^{\circ_i}_D(w_i) - |T'| \geq d^{\circ_i}_D(w_i) - |T| \geq (1/2 + \alpha)n - \mu n$ and $\mu \ll \alpha$. Let $R = R_\ell$, so that R is a copy of T'. We will show that, with positive probability the following property holds.

S For each distinct $x, y \in V(D)$ and $\diamond \in \{+,-\}$,

$$|\{i \in [\ell] : v_i \in N^{\diamond}_D(x) \text{ and } N^{\pm}_R(v_i) \subset N^{\pm}_D(y)\}| \geq \lambda n.$$

Noting $|R| = |T'| \leq |T| - |T''| + 1 \leq (\mu - \varepsilon)n$, let $A \subset V(D)$ contain $V(R)$ so that $|A| = (\mu - \varepsilon)n$, and let v be the copy of t. We will show in two claims that, with positive probability **S** holds, and that, if **S** holds, then A and v satisfy the property in the theorem. Thus, the theorem follows from these two claims.

Claim 2.3.1. *With positive probability,* ***S*** *holds.*

Proof of Claim 2.3.1. Fix $x, y \in V(D)$ and $\diamond \in \{+, -\}$ with $x \neq y$. We will show that $\mathbf{S}$ holds for x, y and $\diamond$ with probability at least $1 - 1/4n^2$, so that the result follows by a union bound.

For convenience, let us take two cases. Either T' has $2\mu^2 n$ leaves (Case I) or $\mu^2 n$ vertex-disjoint bare paths with length 6 (Case II). One of these cases must hold, as, suppose that Case I does not hold and thus T' has fewer than $2\mu^2 n$ leaves. Then, by Lemma 2.1.8, we know that there is some s and some vertex-disjoint bare paths P_i, $i \in [s]$, in T' of length 6 so that $|T' - P_1 - \cdots - P_s| \leq 72\mu^2 n + 2\ell/7$. Removing the internal vertices of each path P_i, $i \in [s]$, from T' removes 5 vertices, and $|T'| = \ell$, so that $\ell - 5s \leq 72\mu^2 n + 2\ell/7$, and therefore

$$s \geq (\ell - 2\ell/7)/5 - 72\mu^2 n/5 \geq \ell/7 - 15\mu^2 n \geq (\mu - 3\varepsilon)n/7 - 15\mu^2 n \geq \mu^2 n,$$

where the final inequality holds since $\varepsilon \ll \mu$.

Case I. Assume that at least $\mu^2 n$ leaves of T' are out-leaves, where the proof whenever T' has at least $\mu^2 n$ in-leaves follows similarly. Let ℓ' be the smallest integer such that, for each $i > \ell'$, t_i is a leaf of T' which is at distance at least 2 from t in T'. We will analyse the embedding of T' in two stages. First, for the embedding of $t_1, \ldots, t_{\ell'}$, we show that with high probability there will be plenty of these vertices which are adjacent to out-leaves in $t_{\ell'+1}, \ldots, t_\ell$ that are embedded to in-neighbours of y. Then, we will analyse the embedding of $t_{\ell'+1}, \ldots, t_\ell$, and show that plenty of these vertices whose in-neighbour in $t_1, \ldots, t_{\ell'}$ was embedded to an in-neighbour of y are themselves embedded to a $\diamond$-neighbour of x. Here, we require that the leaves we consider are at distance at least 2 from t in T'. Let $\mu' = \mu/2$. By the degree condition, at least $(\mu')^2 n$ of these out-leaves are at distance at least 2 from t.

For each $i \in [\ell']$, let c_i be the number of out-leaves of t_i in T'. For each $i \in [\ell']$, let X_i be the random variable which takes value c_i if $v_i \in N^-_D(y)$, and 0 otherwise. Note that, for each $i \in [\ell]$, if $c_i > 0$, then, when the process selects v_i, having chosen

$v_1, \dots, v_{i-1}$, $X_i = c_i$ with probability at least

$$\frac{|(N_D^{\circ_i}(w_i) \setminus V(R_i)) \cap N_D^-(y)|}{n} \geq \frac{|(N_D^{\circ_i}(w_i)) \cap N_D^-(y)| - |R_i|}{n} \geq \frac{2\alpha n - \mu n}{n} \geq \alpha, \tag{2.3.1}$$

as $\alpha \gg \mu$ and since R_i is a copy of the tree T_i, which has at most μn vertices. Thus, for each $i \in [\ell]$, $\mathbb{E}[X_i \mid X_1, \dots\ X_{i-1}] \geq \alpha c_i$.

Note that $\sum_{i\in[\ell']} c_i$ is the number of out-leaves of T', so that $\sum_{i\in[\ell']} c_i \geq (\mu')^2 n$. On the other hand, clearly $\sum_{i\in[\ell']} c_i \leq n$, and for every $i \in [\ell']$, $c_i \leq \Delta(T) \leq cn/\log n$. Thus, as before by repeated application of (2.2.2), we have $\sum_{i\in[\ell']} c_i^2 \leq cn^2/\log n$. Note that $|X_i - \alpha c_i| \leq c_i$ for each $i \in [\ell']$. Therefore, we can apply Corollary 2.1.7 (ii) with $q = \alpha(\mu')^2 n/2$ to get

$$\mathbb{P}\left[\sum_{i\in[\ell']} X_i \leq \sum_{i\in[\ell']} \alpha c_i - q\right] \leq 2\exp\left(\frac{-q^2}{2\cdot\sum_{i\in[\ell']} c_i^2}\right) \leq 2\exp\left(\frac{-q^2\log n}{2cn^2}\right) \leq \frac{1}{8n^2}. \tag{2.3.2}$$

Here, the final inequality holds because $c \ll \mu, \alpha$. Therefore, with probability at least $1 - 1/8n^2$, we have $\sum_{i\in[\ell']} X_i \geq \sum_{i\in[\ell']} \alpha c_i - \alpha(\mu')^2 n/2 \geq \alpha(\mu')^2 n/2$.

Let $m = \sum_{i\in[\ell']} X_i \geq \alpha(\mu')^2 n/2$. Consider now the embedding of $t_{\ell'+1}, \dots, t_\ell$. Let $j_1, \dots, j_m \in \{\ell'+1, \dots, \ell\}$ be such that t_{j_i} is an out-leaf of T' and the image of $N_{T'}^-(t_{j_i})$ is an in-neighbour of y for each $i \in [m]$. For each $i \in [m]$, let Y_i be the random variable which takes value 1 if v_{j_i} is in $N_D^{\diamond}(x)$, and 0 otherwise. Note that, similarly to the calculation in (2.3.1), $\mathbb{E}[Y_i \mid Y_1, \dots Y_{i-1}] \geq \alpha$ for each $i \in [m]$. As before, since $|Y_i - \alpha| \leq 1 - \alpha$ for each $i \in [m]$ as $\alpha \leq 1$, we can apply Corollary 2.1.7 (ii) with $t = \alpha m/2$ to get

$$\mathbb{P}\left[\sum_{i\in[m]} Y_i < \alpha m - t\right] \leq 2\exp\left(\frac{-t^2}{2(1-\alpha)^2 m}\right) \leq \frac{1}{8n^2}, \tag{2.3.3}$$

where the final inequality holds because $1/n \ll \mu, \alpha$. Hence, with probability at least

$1 - 1/8n^2$, we have $\sum_{i\in[m]} Y_i \geq \alpha m/2$. Note that

$$\Big|\{i \in [\ell] : v_i \in N_D^{\diamond}(x) \text{ and } N_R^{\pm}(v_i) \subset N_D^{\pm}(y)\}\Big| \geq \sum_i Y_i.$$

Thus, by taking a simple union bound over the events in (2.3.2) and (2.3.3) and using $\lambda \ll \alpha, \mu$, we see that in total, with probability at least $1 - 1/4n^2$,

$$\Big|\{i \in [\ell]\colon v_i \in N_D^{\diamond}(x) \text{ and } N_R^{\pm}(v_i) \subset N_D^{\pm}(y)\}\Big| \geq \alpha m/2 \geq \lambda n.$$

Taking a union bound over all possible $x, y \in V(D)$ and $\diamond \in \{+,-\}$, we see that in this case $\mathbf{S}$ holds with probability at least $1/2$.

Case II. Let $m = \mu^2 n$. Let $P_1, \dots, P_m$ be vertex disjoint paths of length 6 in T, such that t_{j_i} is the middle vertex of P_i for each $i \in [m]$, and the vertices t_{j_i} appear in order of increasing i in the sequence $t_1, \dots, t_\ell$.

For each $i \in [m]$, let X_i be the random variable taking value 1 if

$$v_{j_i} \in N_D^{\diamond}(x) \text{ and } N_R^{\pm}(v_{j_i}) \subset N_D^{\pm}(y) \tag{2.3.4}$$

and 0 otherwise. Note that, by virtue of the labelling of $t_1, \dots, t_\ell$, the vertices that appear in $N_R^{\pm}(v_{j_i})$ are exactly the vertices v_{j_i-1} and v_{j_i+1}. When we choose each of $v_{j_i-1}, v_{j_i}, v_{j_i+1}$, the probability that all three satisfy the condition in (2.3.4) (however the previous vertices $v_{i'}$ are chosen) is at least α, in a calculation similar to (2.3.1). Therefore, we have, for each $i \in [m]$, that $\mathbb{E}[X_i \mid X_1, \dots\ X_{i-1}] \geq \alpha^3$. Since $|X_i - \alpha^3| \leq 1$ for each $i \in [m]$ as $\alpha \leq 1$, we can apply Corollary 2.1.7 (ii) with $t = \alpha^3 m/2$ to get

$$\mathbb{P}\left[\sum_{i\in[m]} X_i \leq \alpha^3 m - t\right] \leq 2\exp\left(\frac{-t^2}{2m}\right) = 2\exp\left(\frac{-\alpha^6 m}{8}\right) \leq \frac{1}{4n^2},$$

as $1/n \ll \alpha, \mu$. Therefore, with probability at least $1 - 1/4n^2$, as $\lambda \ll \mu, \alpha$,

$$\begin{aligned}&\left|\{i \in [\ell] \colon v_i \in N^{\diamond}_D(x) \text{ and } N^{\pm}_R(v_i) \subset N^{\pm}_D(y)\}\right| \\ \geq &\left|\{i \in [m] : v_{j_i} \in N^{\diamond}_D(x) \text{ and } N^{\pm}_R(v_{j_i}) \subset N^{\pm}_D(y)\}\right| \\ = &\sum_{i \in [m]} X_i \geq \alpha^3 m/2 \geq \lambda n.\end{aligned}$$

Taking a union bound over all possible $x, y \in V(D)$ and $\diamond \in \{+, -\}$, we see that in this case $\mathbf{S}$ holds with probability at least $1/2$. □

Claim 2.3.2. *If $\boldsymbol{S}$ holds then A and v satisfy the property in the theorem. That is, for any set $B \subseteq V(D)$ with $A \subset B$ and $|B| = \mu n$, then $D[B]$ contains a copy of T in which t is copied to v.*

Proof of Claim 2.3.2. Let $B \subset V(D)$ with $A \subset B$ and $|B| = \mu n$. Let $k = |T''| - 1 \leq 3\varepsilon n$ and label the vertices of $V(T'') \setminus V(T')$ as $s_1, \ldots, s_k$, so that, for each $i \in [k]$, $T'_i := T' \cup T''[\{s_1, \ldots, s_i\}]$ is a tree. Note that $|B \setminus V(R)| = k$ and label the vertices of $B \setminus V(R)$ as $y_1, \ldots, y_k$.

Let $S_0 = R$. Now, for each $i = 1, \ldots, k$ in turn, do the following. Let $x_i \in V(S_{i-1})$ and $\diamond_i \in \{+, -\}$ be such that we need to add a $\diamond_i$-neighbour to x_i as a leaf to get a copy of T'_i. If possible, choose some $j'_i \in [\ell] \setminus \{1, j'_1, \ldots, j'_{i-1}\}$ such that

$$v_{j'_i} \in N^{\diamond_i}_D(x_i) \;\text{ and }\; N^{\pm}_{S_{i-1}}(v_{j'_i}) \subset N^{\pm}_D(y_i) \;\text{ and }\; d^{+}_{S_{i-1}}(v_{j'_i}) + d^{-}_{S_{i-1}}(v_{j'_i}) \leq 4/\lambda.$$

Replace $v_{j'_i}$ with y_i in S_{i-1} and add $v_{j'_i}$ as a $\diamond_i$-neighbour of x_i to get S_i, a copy of T'_i with vertex sets $V(S_{i-1}) \cup \{y_i\}$.

We need only show that there is such a vertex $v_{j'_i}$ in each case, since if this process finds S_k, then we have a copy of $T'_k = T$. Fix then $i \in [k]$. By $\mathbf{S}$, we know there are at least λn choices of $i' \in [\ell]$ such that $v_{i'} \in N^{\diamond_i}_D(x_i)$ and $N^{\pm}_R(v_{i'}) \subset N^{\pm}_D(y_i)$. By the construction of S_{i-1}, there are at most $(4/\lambda) \cdot 3\varepsilon n \leq \lambda n/4$ vertices adjacent to the vertices $v_{j'_1}, \ldots, v_{j'_{i-1}}$ in S_{i-1}, and so at most $\lambda n/4$ vertices of R

can be adjacent to the vertices $v_{j'_1}, \ldots, v_{j'_{i-1}}$ in S_{i-1}. Therefore for all but at most $\lambda n/4$ values of $i' \in [\ell]$, we have $N^+_R(v_{i'}) = N^+_{S_{i-1}}(v_{i'})$ and $N^-_R(v_{i'}) = N^-_{S_{i-1}}(v_{i'})$. Furthermore, as $\sum_{i' \in [\ell]}(d^+_T(t_{i'}) + d^-_T(t_{i'})) \leq 2n$, at most $\lambda n/2$ values of $i \in [k]$ can have $d^+_{S_{i-1}}(v_{j'_i}) + d^-_{S_{i-1}}(v_{j'_i}) > 4/\lambda$. Indeed, suppose that I' is the set of all such $i \in [k]$. Then $\sum_{i' \in I'} d^+_{S_{i-1}}(v_{j'_i}) + d^-_{S_{i-1}}(v_{j'_i}) \geq (\lambda n/2)(4/\lambda) > 2n$, but since S_i is a copy of T'_i, this is a contradiction. Thus, we know that there will be at least $\lambda n - n\lambda n/4 - \lambda n/2 \geq \lambda n/4$ choices for j'_i, and so such a j'_i will always exist by **S**. $\square$

Thus we have proved the claim and therefore the lemma. $\square$

CHAPTER 3

TRANSVERSAL CYCLE FACTORS

The aim of this section is to prove Theorem 1.1.2, which we recall below.

Theorem 1.1.2. *For every even integer $k \geq 4$ there exists n_0 such that if G is a k-partite graph whose vertex classes each have size $n \geq n_0$ with $\delta^*_{C_k}(G) \geq (1 + \frac{1}{k})\frac{n}{2}$, then G contains a transversal C_k-factor.*

3.1 Introduction

The degree condition of Theorem 1.1.2 is best possible, as shown by the following extremal example (see Figure 3.1), which works for any k, both even and odd. For each $i \in [k]$ let X_i and Y_i be sets each of size $\lceil(\frac{1}{2} + \frac{1}{2k})n\rceil - 1$ such that $V_i := |X_i \cup Y_i| = n$ and so that the sets V_i are pairwise vertex-disjoint. Form a graph G with vertex classes $V_1, \ldots, V_k$ in which $G[X_i, X_{i+1}]$ and $G[Y_i, Y_{i+1}]$ are complete for each $i \in [k-1]$, and $G[X_1, Y_k]$ and $G[Y_1, X_k]$ are also complete, but G has no edges other than those in this complete subgraphs. We then have $\delta^*_G(G) \geq \lceil(\frac{1}{2} + \frac{1}{2k})n\rceil - 1$ since every vertex in V_i is adjacent to all vertices in either X_{i+1} or Y_{i+1}, and to all vertices in either X_{i-1} or Y_{i-1} (with addition on indices taken modulo k). Note, however, that every copy of C_k in G must contain a vertex from $X_i \cap Y_i$ for some

$i \in [k]$. Since $|X_i| = |Y_i| < (\frac{1}{2} + \frac{1}{2k})n$, we have

$$\sum_{i\in[k]} |X_i \cap Y_i| = \sum_{i\in[k]} |X_i| + |Y_i| - |V_i| < k\left(2\left(\frac{1}{2} + \frac{1}{2k}\right)n - n\right) = n,$$

so every transversal C_k-tiling in G contains fewer than n copies of C_k and therefore is not a transversal C_k-factor.

It is worth noting that when k is odd, there is a wider class of extremal examples. For example suppose n is divisible by $2k$ and, for each $i \in [k] \setminus \{1\}$, let X_i and Y_i be sets of size $(k-1)n/2k$, and Z_i a set of size n/k. Let X_1 be a set of size $(k-1)n/2k - 1$, let Y_1 be a set of size $(k-1)n/2k+1$, and let Z_1 be a set of size n/k. Now suppose that for each $i \in [k]$, $V_i = X_i \cup Y_i \cup Z_i$, and let G be a k-partite graph with vertex classes $V_1, \ldots, V_k$. Note that each class has exactly n vertices. Add edges of G as follows. For each $i \in [k]$ (where $[k] = \{1, \ldots, k\}$ and indices are taken modulo k), suppose that each vertex in Z_i sees every vertex in $X_{i-1} \cup Y_{i-1}$ and every vertex in $X_{i+1} \cup Y_{i+1}$. For each $i \in [k-1]$, suppose $G[X_i \cup X_{i+1}]$, $G[Y_i \cup Y_{i+1}]$, $G[X_1 \cup Y_k]$ and $G[Y_1 \cup X_k]$ are each complete bipartite graphs. In this case, $\delta^*(G) \geq (1+1/k)n/2 - 1$.

Any C_k from this graph must contain a vertex from Z_i for some $i \in [k]$. We say a path is *partition-respecting* if it contains at most one vertex from each vertex class, and if each edge of the path is between V_i and V_{i+1} for some $i \in [k]$. Since there are exactly n copies of C_k in any C_k-factor and there are exactly n vertices in $\bigcup_{i\in[k]} Z_i$, there must in fact be exactly one vertex from Z_i in each copy of C_k. This implies that $G[\bigcup_{i\in[k]} X_i \cup Y_i]$ must have a partition-respecting P_{k-2}-factor, where for any j, P_j is a path with $j+1$ vertices and j edges.

However, when k is odd, each partition-respecting path of length $k-2$ contains $k-1$ vertices. Now let $A = \bigcup_{i\in[(k-1)/2]}(X_{2i-1} \cup X_k \cup Y_{2i})$ and $B = \bigcup_{i\in[(k-1)/2]}(Y_{2i-1} \cup Y_k \cup X_{2i})$. Any partition-respecting path of length $k-2$ has exactly $k-1$ vertices, and since k is odd, $k-1$ must be even. Also, any partition-respecting path of length $k-2$ must alternate between vertices of A and B. Therefore, each such path must

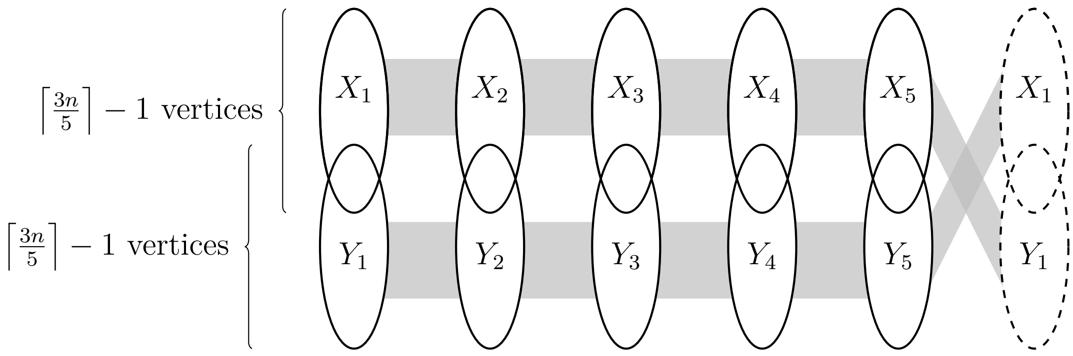


Figure 3.1: An extremal example when $k = 5$

hit A as many times as it hits B. However, A contains two fewer vertices than B since $|X_1| = |Y_1| - 2$. Therefore, there cannot be a partition-respecting P_{k-2}-factor in $G[\bigcup_{i\in[k]} X_i \cup Y_i]$, and so G contains no transversal C_k-factor. Intuitively the same example does not work in the case when k is even, since $k-1$ would be odd, and therefore a P_{k-2} would either contain one more vertex from A or from B, and so in any case, the same argument cannot be made. In fact we will see at the end of Section 3.5 that when k is even, this structure does indeed contain a perfect fractional C_k-tiling.

At the top level, our proof of Theorem 1.1.2 splits into two cases, according to whether the graph G is 'close' to the extremal example described above. The next definition specifies precisely what we mean by this; to compare this to the extremal example, note that when k is even, the sets A_i and B_i take the roles of the sets $X_i \setminus Y_i$ and $Y_i \setminus X_i$ respectively, and when k is odd, we have a slightly broader definition to reflect the family of examples described in the previous paragraph.

Definition 3.1.1. *Let G be a k-partite graph with vertex classes $V_1, \dots, V_k$ each of size n. We say that G is ε-*extremal *if there exist sets A_i, B_i and Z_i for each $i \in [k]$ such that*

- *A_i, B_i, Z_i partition V_i for each $i \in [k]$,*
- *$|A_i| = |B_i| = (k-1)n/2k$ and $|Z_i| = n/k$.*

- *$d(A_i, A_{i+1}), d(B_i, B_{i+1}) \geq 1 - \varepsilon$ for every $i \in [k-1]$,*

- *$d(A_k, B_1), d(B_k, A_1) \geq 1 - \varepsilon$,*

- *If k is even, then $d(Z_i, A_{i+1}), d(Z_{i+1}, A_i), d(Z_i, B_{i+1}), d(Z_{i+1}, B_i) \geq 1 - \varepsilon$ for every $i \in [k-1]$, and also $d(Z_1, A_k), d(Z_k, A_1), d(Z_1, B_k), d(Z_k, B_1) \geq 1 - \varepsilon$.*

- *If k is odd, then for each $i \in [k-1]$, there exist $X, Y \in \{A, B\}$ with $X \neq Y$ such that $d(X_i, Z_{i+1}), d(Y_{i+1}, Z_i) \geq 1 - \varepsilon$, and also there exists $X \in \{A, B\}$ such that $d(X_k, Z_1), d(X_1, Z_k) \geq 1 - \varepsilon$.*

In the case where G is close to the extremal example, we use *ad hoc* methods to obtain the following lemma, stating that the exact degree condition of Theorem 1.1.2 suffices to ensure the existence of a transversal C_k-factor in the case when k is even; the proof of this lemma is presented in Section 3.3.

Lemma 3.1.2. *Suppose that $1/n \ll \psi \ll 1/k$ and that k is even, and let G be a balanced k-partite graph whose vertex classes each have size n. If $\delta^*(G) \geq (\frac{1}{2} + \frac{1}{2k})n$ and G is ψ-extremal, then G contains a transversal C_k-factor.*

To prove the analoge of Theorem 1.1.2 in the case when k is odd, it only remains to prove an analogous version of Lemma 3.1.2 and this will give a full proof of Fischer's conjecture. Most of the work in proving Theorem 1.1.2 is concerned with the other case, where G is not close to the extremal example. For this case our goal is the following lemma, which says that in this case a slightly weaker degree condition is sufficient to ensure a transversal C_k-factor.

Lemma 3.1.3. *Suppose that $1/n \ll \gamma \ll \psi, 1/k$, and let G be a balanced k-partite graph whose vertex classes each have size n. If $\delta^*(G) \geq (\frac{1}{2} + \frac{1}{2k})n - \gamma n$, then either G contains a transversal C_k-factor or G is ψ-extremal.*

Observe that Lemma 3.1.3 and Lemma 3.1.2, taken together, immediately prove Theorem 1.1.2. In fact, the principal novel contribution involved in the proof of

Lemma 3.1.3 is to prove our next result, a fractional version of the lemma. A *perfect fractional C_k-tiling* in a graph G is an assignment of a weight $w_C \in [0,1]$ to each copy C of C_k in G so that for every $v \in V(G)$ we have $\sum_{C:v\in V(C)} w_C = 1$ (so a perfect C_k-tiling can be viewed as the special case in which $w_C \in \{0,1\}$ for every copy C of C_k in G).

Theorem 3.1.4. *Suppose that $1/n \ll \gamma \ll \psi, 1/k$, and let G be a balanced k-partite graph whose vertex classes each have size n. If $\delta^*(G) \geq (\frac{1}{2} + \frac{1}{2k})n - \gamma n$, then either G contains a perfect fractional C_k-tiling or G is ψ-extremal.*

The proof of Theorem 3.1.4 is presented in Section 3.5. In Section 3.4 we derive Lemma 3.1.3 from Theorem 3.1.4 by a standard approach using the Szemerédi regularity lemma [67] along with the following absorbing lemma due to Ergemlidze and Molla [24].

Lemma 3.1.5 (The Absorbing Lemma [24])**.** *Let $1/n \ll \sigma \ll \gamma, 1/k$. Let G be a k-partite graph whose vertex classes each have size n. If $\delta^*(G) \geq n/2 + \gamma n$, then for some $z \leq \sigma n$ there is a set $A \subseteq V(G)$, which we call the* absorbing set*, with $|A \cap V_i| = z$ for each $i \in [k]$, such that if $G - A$ has a transversal C_k-tiling of size at least $n - z - \sigma^2 n$, then G contains a transversal C_k-factor.*

3.2 Preliminaries

3.2.1 Notation

For a graph G, we say $V(G)$ is the set of vertices of G and $E(G)$ is the set of edges of G. We define $|G| = |V(G)|$ and $e(G) = |E(G)|$. Let $A, B \subset V(G)$. Then the set $E(A, B)$ is the set of all edges with one endpoint in A and the other in B, and $e(A, B) = |E(A, B)|$. $E(A)$ is defined as the set of all edges with both endpoints in A. The graph $G[A]$ is the induced subgraph of G with vertex set A and edge set $E(A)$, and the graph $G - A$ is defined as $G[(V(G) \setminus A)]$. Similarly, for any subgraph

$H \subseteq G$, $G - H := G[V(G) \setminus V(H)]$. We say that the length of a path or cycle is the number of edges it contains. We may refer to a path or cycle of length k as P_k or C_k, respectively.

In order to simplify notation, we use hierarchies to state our results. That is, for $a, b \in (0, 1]$, whenever we write that a statement holds for $a \ll b$ (or $b \gg a$), we mean that there exists a non-decreasing function $f : (0, 1] \to (0, 1]$ such that the statement holds whenever $a \leq f(b)$. We define similar expressions with multiple variables analogously. Whenever $1/n$ appears in this hierarchy, we assume n is an integer. For simplicity, we ignore floors and ceilings wherever this does not affect the argument.

3.2.1.1 Multipartite graphs

Throughout the paper, we will be considering k-partite graphs G with vertex classes $V_1, \dots, V_k$. Here we give some specific notation that we will be using within these structures. We describe vertex classes V_i and V_{i+1} as *consecutive vertex classes* for each $i \in [k]$ (indices taken modulo k). Unless otherwise specified, any cycles we consider throughout this paper will be *transversal cycles*, that is, cycles containing exactly one vertex from each vertex class V_i with each edge of the cycle lying between consecutive vertex classes. Any path we consider will be a *partition-respecting path*, that is, containing at most one vertex from each vertex class V_i and such that each edge of the path lies between consecutive vertex classes. Furthermore, for a path P and vertices $u \in V_i$ and $v \in V_j$ for some $i, j \in [k]$, we say that u is the initial vertex of P if u is an endvertex of P, and the edge containing u in P lies in $G[V_i, V_{i+1}]$. Similarly, we say that v is the final vertex of P if v is an endvertex of P, and if the edge containing v in P lies in $G[V_j, V_{j-1}]$.

For any integer m, let $[m] = \{1, \dots, m\}$. We will often wish to iterate over the vertex classes in a certain order, and in order to do this, we also define the following. For integers $a, b \in [k]$ with $a \neq b$, we let $\{a \uparrow b\} = \{a, a + 1, \dots, b - 1, b\}$,

$\{a \downarrow b\} = \{a, a-1, \dots, b+1, b\}$, and $\{a \uparrow a\} = \{a \downarrow a\} = \{a\}$. We remark that $\{a \uparrow b\} = \{b \uparrow a\}$. Let $[a \uparrow b]$ be the list $(a, a+1, \dots, b-1, b)$ and $[a \downarrow b]$ be the list $(a, a-1, \dots, b+1, b)$ (with entries taken modulo k in each case) and we say that $[a \uparrow a] = [a \downarrow a] = (a)$. When we say we do something *for each* $i \in [a \uparrow b]$, we mean we iterate over the values in $[a \uparrow b]$ in the order in which they appear in the list (and analogously for $[a \downarrow b]$).

3.2.2 Robust expanders

Robust expanders are a key tool we use in the paper. These were first introduced in directed graphs in the form of robust outexpanders by Kühn, Osthus and Treglown, and were used notably in the proof of Kelly's conjecture for large tournaments by Kühn and Osthus [48]. Later, the same authors introduced an undirected version of robust expanders in [49], and used this to prove an approximate version of a conjecture of Nash-Williams. Many of the results in this subsection will be standard and appear through the literature, though we provide proofs for completeness.

We will be considering the following bipartite notion of robust expansion. Let G be a bipartite graph with parts A and B, such that $|A| = |B| = n$, and let ν, τ be constants. For a set $S \subseteq V(G)$, the *ν-robust neighbourhood of S*, denoted by $RN_{\nu,G}(S)$, is the set of vertices of G with at least νn neighbours in S. We say that G is a *robust (ν, τ)-expander* if $|RN_{\nu,G}(S)| \geq |S| + \nu n$ whenever $\tau n \leq |S| \leq (1-\tau)n$ with either $S \subseteq A$ or $S \subseteq B$. For a pair of sets $U, V \subseteq G$, we say that (U, V) is a *(ν, τ)-robust pair* if $G[U, V]$ is a robust (ν, τ)-expander. Otherwise, we say that (U, V) is a *non-(ν, τ)-robust pair.*

As discussed, many of the graphs in this paper will be multipartite graphs. Suppose G is a k-partite graph with vertex classes $V_1, \dots, V_k$ each of size n. Let

- $\mathcal{N}_{\nu,\tau}(G) = \{i \in [k] \colon G[V_i, V_{i+1}]$ is a robust (ν, τ)-expander$\}$,
- $\mathcal{L}_{\nu,\tau}(G) = \{i \in [k] \colon G[V_i, V_{i+1}]$ is not a robust (ν, τ)-expander$\}$,

- $\mathcal{R}_{\nu,\tau}(G) = \{i \in [k] : G[V_{i-1}, V_i]$ is not a robust (ν, τ)-expander$\}$, and

- $\mathcal{B}_{\nu,\tau}(G) = \mathcal{L}_{\nu,\tau}(G) \cap \mathcal{R}_{\nu,\tau}(G)$.

Note that in each of the definitions above, we may omit the ν and τ when they are clear from context. It may be helpful to observe that in the definitions above, $\mathcal{L}$ coresponds to the part being the 'left' part of a robust pair, $\mathcal{R}$ corresponds to it being the 'right' part, $\mathcal{B}$ corresponds to it being both and $\mathcal{N}$ corresponds to it being neither. We now prove some useful properties of robust expanders.

Lemma 3.2.1. *Let $1/n \ll \nu' \leq \nu \ll \tau \leq \tau'$. Let $G = (A, B)$ with $|A| = |B| = n$. If G is a robust (ν, τ)-expander, then it is also a robust (ν', τ')-expander.*

Proof. Suppose G is a robust (ν, τ)-expander. Let $S \subseteq A$ with $\tau' n \leq |S| \leq (1 - \tau')n$. We will prove that $|RN_{\nu'}(S)| \geq |S| + \nu' n$. The equivalent statement when $S \subseteq B$ holds by symmetry. Since $\tau' \geq \tau$, we have $\tau n \leq |S| \leq (1 - \tau)n$. Therefore, $|RN_\nu(S)| \geq |S| + \nu n$. Since $\nu' \leq \nu$, we know that $RN_\nu(S) \subseteq RN_{\nu'}(S)$. So,

$$|RN_{\nu'}(S)| \geq |RN_\nu(S)| \geq |S| + \nu n \geq |S| + \nu' n,$$

as required. □

We would like to show that a robust expander contains a perfect matching. In order to prove this, we first state Hall's condition [31].

Lemma 3.2.2 (Hall's condition)**.** *Let G be a bipartite graph with vertex classes A and B. Suppose that for every $S \subseteq A$, $|N_G(S, B)| \geq |S|$. Then G contains a matching which covers A.*

Lemma 3.2.3. *Let $1/n \ll \varepsilon \ll \nu \ll \tau \ll 1/2$. Let G be a bipartite graph with vertex classes A and B, such that $|A| = |B| = n$. Suppose that $\delta(G) \geq (1/2 - \varepsilon)n$ and that G is a robust (ν, τ)-expander. Then G contains a perfect matching.*

Proof. Let $S \subseteq A$. We would like to show that Hall's criterion is satisfied. Suppose first that $|S| \leq (1/2 - \varepsilon)n$. Then by the minimum degree condition, $|N(S)| \geq (1/2 - \varepsilon)n \geq \tau n \geq |S|$. On the other hand, suppose $|S| > (1/2 + \varepsilon)n$. In this case, $N(S) = B$. Indeed, suppose for a contradiction there is some vertex $v \in B$ which is not contained in $N(S)$. Then $|N(v) \cap S| = 0$, and so $n = |A| \geq |N(v)| + |S| > (1/2 - \varepsilon)n + (1/2 + \varepsilon)n = n$, a contradiction. Therefore, $n = |N(S)| \geq |S|$, and once again, Hall's criterion is satisfied. It only remains to consider the case when $(1/2 - \varepsilon)n \leq |S| \leq (1/2 + \varepsilon)n$. In this case, since $\varepsilon \ll \tau \ll 1/2$, we have that $\tau n \leq |S| \leq (1 - \tau)n$. Therefore, since $RN_\nu(S) \subseteq N(S)$ we have $|N(S)| \geq |RN_\nu(S)| \geq |S| + \nu n$, and so in each case, Hall's criterion is satisfied, and G contains a matching covering A. Since $|A| = |B|$, this is a perfect matching. □

The following lemma says that if a balanced bipartite graph with parts of size n has minimum degree close to $n/2$, then either G is a robust expander or it is 'close to' the union of two almost complete bipartite graphs. In order to state this, we first need the following definition.

Definition 3.2.4. *Let $\alpha, \beta > 0$ and for $n \in \mathbb{N}$, let $G = (A, B)$ be a bipartite graph with vertex classes of size n. Then an (α, β)-*bipartition *of G is a partition of A into sets $A_1 \cup A_2$ and B into sets $B_1 \cup B_2$ so that the following hold for each $i \in \{1, 2\}$.*

(i) $(1/2 - \beta)n \leq |A_i|, |B_i| \leq (1/2 + \beta)n$.

(ii) *All but at most βn vertices in A_i have at least $(1 - \alpha)|B_i|$ neighbours in B_i.*

(iii) *All but at most βn vertices in B_i have at least $(1 - \alpha)|A_i|$ neighbours in A_i.*

We say this is a high-degree (α, β)-bipartition *if every vertex in A_i has at least $(1-\alpha)|B_i|$ neighbours in B_i and every vertex in B_i has at least $(1-\alpha)|A_i|$ neighbours in A_i.*

Lemma 3.2.5. *Let $1/n \ll \varepsilon \ll \nu \ll \tau$, and let $\nu \ll \beta \ll \alpha \leq 1$. Let $G = (A, B)$ be a bipartite graph with $|A| = |B| = n$, and suppose that $\delta(G) \geq n/2 - \varepsilon n$. Suppose that G is not a robust (ν, τ)-expander. Then G has an (α, β)-bipartition.*

Proof. Let μ satisfy $\nu \ll \mu \ll \beta$.

Since G is not a robust (ν, τ)-expander, there is some set $X_1 \subset A$ such that $\tau n \leq |X_1| \leq (1-\tau)n$ and $|RN(X_1)| < |X_1| + \nu n$. Let $Y_1 = RN(X_1)$, $Y_2 = B \setminus Y_1$ and $X_2 = A \setminus X_1$. We claim that $((X_1, X_2), (Y_1, Y_2))$ is an (α, β)-bipartition of (A, B). First we would like to determine the sizes of each of X_i and Y_i for $i \in [2]$. We begin by counting the edges between X_1 and Y_1 in two ways to obtain lower bounds on $|X_1|$ and $|Y_1|$. Since every vertex in Y_2 can have at most νn neighbours in X_1, we know that $e(X_1, Y_2) \leq \nu n |Y_2|$. Therefore,

$$\begin{aligned} e(X_1, Y_1) &= e(X_1, B) - e(X_1, Y_2) \geq (1/2 - \varepsilon)n |X_1| - \nu n |Y_2| \\ &\geq (1/2 - \varepsilon)n |X_1| - \nu n^2 \geq (1/2 - \varepsilon)n |X_1| - \nu n |X_1| / \tau \\ &= (1/2 - \varepsilon - \nu/\tau)n |X_1| . \end{aligned} \tag{3.2.1}$$

On the other hand, we also have that $e(X_1, Y_1) \leq |X_1| |Y_1|$. Therefore, combining these, we get that $|Y_1| \geq (1/2 - \varepsilon - \nu/\tau)n$. Recall that $|Y_1| \leq |X_1| + \nu n$, therefore we get that $|X_1| \geq (1/2 - \varepsilon - \nu/\tau + \nu)n$.

Now we will count the edges between X_2 and Y_2 in two ways to obtain upper bounds on $|X_1|$ and $|Y_1|$. First,

$$e(X_2, Y_2) = e(A, Y_2) - e(X_1, Y_2) \geq (1/2 - \varepsilon)n |Y_2| - \nu n |Y_2| = (1/2 - \varepsilon - \nu)n |Y_2| . \tag{3.2.2}$$

On the other hand, $e(X_2, Y_2) \leq |X_2| |Y_2|$. Therefore, we get that $|X_2| \geq (1/2 - \varepsilon - \nu)n$. Therefore, $|X_1| \leq (1/2 + \varepsilon + \nu)n$. This also implies that $|Y_1| \leq |X_1| + \nu n \leq (1/2 + \varepsilon + 2\nu)n$.

Since $\nu \ll \tau, \mu$ we have $\nu/\tau < \sqrt{\nu} < \mu/3$. Using this combined with the fact that $|A| = |B| = n$, we get that $(1/2 - \beta)n \leq (1/2 - \mu)n \leq |X_i| , |Y_i| \leq (1/2 + \mu)n \leq (1/2 + \beta)n$ for each $i \in [2]$. Now we want to show that 'almost all' vertices in X_i see 'almost everything' in Y_i and vice versa, for $i \in [2]$. First, by (3.2.1) and by the fact that $\nu \ll \tau, \mu$, we have

$$\frac{e(X_1, Y_1)}{|X_1|\,|Y_1|} \geq \frac{(1/2 - \varepsilon - \nu/\tau)n\,|X_1|}{|X_1|\,|Y_1|} \geq \frac{(1/2-\mu)n}{(1/2+\mu)n} \geq 1 - 4\mu.$$

Now let $S = \{v \in X_1 \colon d_{Y_1}(v) \leq (1-\alpha)\,|Y_1|\}$. Then $e(X_1, Y_1) \leq (1-\alpha)\,|Y_1|\,|S| + (|X_1| - |S|)\,|Y_1|$. Therefore, combining these inequalities gives

$$\begin{aligned}(1-4\mu)\,|X_1|\,|Y_1| &\leq (1-\alpha)\,|Y_1|\,|S| + (|X_1| - |S|)\,|Y_1| \\ \iff \alpha\,|S| &\leq 4\mu\,|X_1| \\ \iff |S| &\leq \frac{4\mu}{\alpha}\,|X_1|\,.\end{aligned}$$

Therefore, since $\mu \ll \alpha$, we get that $|S| \leq \sqrt{\mu}\,|X_1|$. Therefore, we know that at least $(1-\sqrt{\mu})\,|X_1| \geq (1-\beta)\,|X_1|$ vertices in X_1 have at least $(1-\alpha)\,|Y_1|$ neighbours in Y_1. By symmetry, we know there must be at least $(1-\sqrt{\mu})\,|Y_1| \geq (1-\beta)\,|Y_1|$ vertices in Y_1 which have at least $(1-\alpha)\,|X_1|$ neighbours in X_1.

Observe also that by (3.2.2),

$$\frac{e(X_2, Y_2)}{|X_2|\,|Y_2|} \geq \frac{(1/2 - \varepsilon - \nu)n\,|Y_2|}{|X_1|\,|Y_1|} \geq \frac{(1/2-\mu)n}{(1/2+\mu)n} \geq 1 - 4\mu.$$

Therefore, we can use the same argument as before to show that at least $(1-\sqrt{\mu})\,|X_2| \geq (1-\beta)\,|X_2|$ vertices in X_2 have at least $(1-\alpha)\,|Y_2|$ neighbours in Y_2. By symmetry, we know there must be at least $(1-\sqrt{\mu})\,|Y_2| \geq (1-\beta)\,|Y_2|$ vertices in Y_2 which have at least $(1-\alpha)\,|X_2|$ neighbours in X_2, and this concludes the proof. □

Lemma 3.2.6. *Let $1/n \ll \varepsilon \ll \nu \ll \tau \ll 1$. Let $G = (A, B)$ be a robust (ν, τ)-expander, with $|A| = |B| = n$. Let $A' \subset A$ and $B' \subseteq B$ with $|A'| = |B'| \geq (1-\varepsilon)n$. Then $G' = G[A', B']$ is a robust $(\nu - \varepsilon, \tau)$-expander.*

Proof. Let G' be as above, let $n' = |A'| = |B'| \geq (1-\varepsilon)n$ and take $S \subset A'$ satisfying $\tau n' \leq |S| \leq (1-\tau)n'$. There are two cases to consider. Suppose first that $\tau n \leq |S| \leq (1-\tau)n' \leq (1-\tau)n$. Then as G is a robust (ν, τ)-expander, $|RN_{\nu,G}(S)| \geq |S| + \nu n$.

In particular, each vertex in $RN_{\nu,G}(S)$ has at least $\nu n \geq (\nu-\varepsilon)n'$ neighbours in S. Then $RN_{\nu-\varepsilon,G'}(S) \geq |S| + \nu n - \varepsilon n \geq |S| + (\nu-\varepsilon)n'$.

Now suppose that $\tau n' \leq |S| < \tau n$. Then let $S' = S \cup (A \setminus A')$. Then $\tau n \leq \tau n' + n - n' \leq |S'| < (\tau+\varepsilon)n \leq (1-\tau)n$. Therefore, we have $|RN_{\nu,G}(S')| \geq |S'| + \nu n$. Again, this is the set of vertices in B with at least νn neighbours in S'. Each of these vertices must have at least $(\nu-\varepsilon)n \geq (\nu-\varepsilon)n'$ neighbours in S. Therefore, $|RN_{\nu-\varepsilon,G'}(S)| \geq |S'| + \nu n - \varepsilon n \geq |S| + (\nu-\varepsilon)n'$. □

3.2.3 Fractional tilings

A key tool we will be using is fractional tilings. Recall that a *fractional H-tiling* in a graph G is a function f which assigns a weight in $[0,1]$ to each copy of H in G such that the sum of weights of copies of H adjacent to any vertex is at most 1. We say this is *perfect* if $\sum_{T \text{ a copy of } H, v \in T} f(T) = 1$ for every vertex $v \in G$. Note that we can describe an H-tiling to be a function g which assigns a weight in $\{0,1\}$ to each copy of H in G so that the weight at any vertex is at most 1, and we say this is a perfect H-tiling if the weight at each vertex is exactly 1. Therefore, the problem of finding a perfect fractional H-tiling can be thought of as an LP-relaxation of the problem of finding an H-tiling.

Let G be a graph on n vertices. Then for any $S \subseteq V(G)$, we say that the *characteristic vector* $\chi(S)$ is the vector in $\{0,1\}^n$ which has 1 in the coordinates corresponding to vertices in S, and 0 elsewhere. For vectors $\mathbf{x}_1, \ldots, \mathbf{x}_r \in \mathbb{R}^n$, we define the *positive cone* of these vectors to be $PC(\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}) = \{\sum_{j\in[r]} \lambda_j \mathbf{x}_j \colon \lambda_1, \ldots, \lambda_r \geq 0\}$. Then a perfect fractional H-tiling is just an assignment of weights w_T to each copy T of H in G such that $\sum_{T \text{ a copy of } H} w_T \cdot \chi(V(T)) = \mathbf{1}$, where $\mathbf{1}$ is the all 1s vector. Therefore, we can state the following fact.

Fact 3.2.7. *G has a perfect fractional H-tiling if and only if*

$$\mathbf{1} \in PC(\chi(V(T)) \colon T \textit{ is a copy of } H \textit{ in } G).$$

This is crucial to our proof, since the following lemma gives us a concrete way to use this fact.

Lemma 3.2.8 (Farkas' Lemma [26]). *Suppose* $\mathbf{x} \in \mathbb{R}^n \setminus PC(Y)$ *for some finite* $Y \subseteq \mathbb{R}^n$. *Then there is some vector* $\boldsymbol{a} \in \mathbb{R}^n$ *such that* $\boldsymbol{a} \cdot \mathbf{y} \geq 0$ *for each* $\mathbf{y} \in Y$*, and* $\boldsymbol{a} \cdot \mathbf{x} < 0$.

In our context, this lemma implies that G has no perfect fractional H-tiling if and only if there is some vector $a \in \mathbb{R}^n$ such that $\boldsymbol{a} \cdot \chi(T) \geq 0$ for each copy T of H in G, while $\boldsymbol{a} \cdot \mathbf{1} < 0$. In the proof, we will use this to find a contradiction, therefore implying the existence of a perfect H-tiling.

3.3 The extremal case

Our goal in this section is to prove that if G is close to extremal and meets our exact minimum degree condition, then G contains a transversal C_k-factor. We assume throughout that $k \geq 4$ is even. Our aim in this section is then to prove Lemma 3.1.2.

The key structural feature of near-extremal graphs is that the sets $A_1, \dots, A_k, B_1, \dots, B_k$, in that order, induce a subgraph of G which is close to a blow-up of C_{2k}, in the sense that there is a high density of edges between each consecutive pair of sets. Since for much of the time we will work with this feature, we frequently relabel the sets A_i and B_i as $X_1, \dots, X_{2k}$, where $X_i = A_i$ for $i \in [k]$ and $X_i = B_{i-k}$ for $i \in \{k+1, \dots 2k\}$, with arithmetic on the indices of the sets X_i always taken modulo $2k$. For $i \in \mathbb{Z}$ we also define $c(i)$ to be the unique integer in $[k]$ which is congruent to i modulo k, so in particular for each $i \in [2k]$ we have $X_i \subseteq V_{c(i)}$. Note that the density conditions in the definition of ψ-extremal can then be restated as saying that $d(X_i, X_{i+1})$, $d(X_i, Z_{c(i+1)})$, $d(X_i, Z_{c(i-1)})$ are all at least $1 - \psi$ for each $i \in [2k]$. The above is encapsulated in the following formal setup, to which we frequently refer.

Setup 3.3.1. *Let G be a k-partite graph with vertex classes $V_1, V_2, \dots, V_k$ each of size n. For each $i \in [k]$ let A_i, B_i and Z_i be disjoint sets with $A_i \cup B_i \cup Z_i = V_i$. Define sets $X_1, \dots, X_{2k}$ by setting $X_i = A_i$ for each $i \in [k]$ and $X_i = B_{i-k}$ for each $i \in \{k+1, \dots, 2k\}$. Arithmetic on the indices of the sets X_i should always be taken modulo $2k$.*

Our definition of extremal graphs only establishes density conditions between the sets X_i and Z_i rather than describing the neighbourhoods of individual vertices. The following simple proposition allows us to deduce that across high-density pairs most vertices have large neighbourhoods.

Proposition 3.3.2. *Let G be a bipartite graph with vertex classes A and B. If $d(A, B) \geq 1 - \psi$, then there are at most $\sqrt{\psi}|A|$ vertices $x \in A$ with fewer than $(1 - \sqrt{\psi})|B|$ neighbours in B.*

Proof. If, on the contrary, more than $\sqrt{\psi}|A|$ vertices $x \in A$ have fewer than $(1 - \sqrt{\psi})|B|$ neighbours in B, then we find that

$$\begin{aligned}(1-\psi)|A||B| \leq d(A,B)|A||B| &= e(A,B) \\ &< (1-\sqrt{\psi})|A| \cdot |B| + \sqrt{\psi}|A| \cdot (1-\sqrt{\psi}) \cdot |B| \\ &= (1-\psi)|A||B|,\end{aligned}$$

a contradiction. □

Proposition 3.3.2 ensures that if G is ψ-extremal for small ψ, then almost all vertices in X_i have few non-neighbours in X_{i+1}. However there remains the possibility that some vertices have atypical neighbourhoods, and indeed there may be vertices in X_i with no neighbours at all in X_{i+1}. The first step in the proof of Lemma 3.1.2 is to edit the partition witnessing that G is ψ-extremal to ensure that this is not the case. For this we make the following definitions. Let a graph G, vertex classes V_i and sets A_i, B_i, Z_i and X_i be as in Setup 3.3.1. Then we describe vertices as follows.

1. For $i \in [2k]$ we say that $x \in X_i$ is α-suitable if $|N(x) \cap X_{i+1}| \geq \alpha |X_{i+1}|$.

2. For $i \in [k]$ we say that $x \in Z_i$ is α-good if $|N(x) \cap X_j| \geq (1-\alpha)|X_j|$ for each $j \in \{i-1, i+1, i-k+1, i+k-1\}$.

3. For $i \in [2k]$ we say that $x \in X_i$ is α-good if $|N(x) \cap X_j| \geq (1-\alpha)|X_j|$ for each $j \in \{i-1, i+1\}$ and $|N(x) \cap Z_j| \geq (1-\alpha)|Z_j|$ for each $j \in \{c(i-1), c(i+1)\}$.

Note that whether a vertex is suitable or good depends on which set it is contained in, but this will always be clear from the context. The next proposition encapsulates how we will edit the partition to ensure that all vertices in the sets X_i are suitable.

Proposition 3.3.3. *Suppose that $k \geq 4$ is even, that $1/n \ll \psi \ll \alpha \ll 1/k$, and that n is divisible by $2k$. Let G be a balanced k-partite graph whose vertex classes $V_1, \ldots, V_k$ each have size n with $\delta^*(G) \geq (k+1)n/2k$, and let $m := \lceil 4\sqrt{\psi} n \rceil$. If G is ψ-extremal, then there exist sets A_i, B_i, Z_i for $i \in [k]$ and X_i for $i \in [2k]$ meeting the conditions of Setup 3.3.1, and integers d_i for $i \in [k]$, with the properties that*

- $|A_i| = (k-1)n/2k + d_i + m$,
- $|B_i| = (k-1)n/2k - d_i + m$,
- $|Z_i| = n/k - 2m$,
- *for each $i \in [2k]$, every vertex in X_i is $1/3$-suitable, and*
- *for each $i \in [k]$ at most $3m$ vertices in V_i are not α-good.*

Proof. For each $i \in [k]$ let $A'_i, B'_i, Z'_i \subseteq V_i$ be pairwise disjoint sets witnessing that G is ψ-extremal, in the sense that the conditions of Definition 3.1.1 are satisfied with A'_i, B'_i and Z'_i in place of A_i, B_i and Z_i respectively. Write $X'_i := A'_i$ for $i \in [k]$ and $X'_i := B'_{i-k}$ for $i \in \{k+1, \ldots, 2k\}$. In particular we then have $|X'_i| = (k-1)n/2k$ for each $i \in [2k]$ and $|Z'_i| = n/k$ for each $i \in [k]$. Moreover, the density conditions together with Proposition 3.3.2 imply the following.

1. For each $i \in [k]$, all but at most $4\sqrt{\psi}|Z'_i|$ vertices of Z'_i have at least $(1 - \sqrt{\psi})(k-1)n/2k$ neighbours in each of $X'_{i-1}, X'_{i+1}, X'_{i+k-1}$ and X'_{i+k+1}; we call these vertices *provisionally good*, and those which do not satisfy this condition *bad*.

2. For each $i \in [2k]$ all but at most $4\sqrt{\psi}|X'_i|$ vertices of X'_i have at least $(1 - \sqrt{\psi})(k-1)n/2k$ neighbours in each of X'_{i-1} and X'_{i+1} and at least $(1 - \sqrt{\psi})n/k$ neighbours in each of $Z'_{c(i-1)}$ and $Z'_{c(i-1)}$. Again we call these vertices *provisionally good*, and those which do not satisfy this condition *bad*.

Since $\delta^*(G) \geq (k+1)n/2k$ we also know that for each $i \in [2k]$, each vertex $x \in V_i$ either has at least $|X'_{i+1}|/2$ neighbours in $|X'_{i+1}|$ or at least $|X'_{i+k+1}|/2$ neighbours in X'_{i+k+1}; we say that vertices satisfying the former condition are A-suitable, whilst those satisfying the latter *but not the former* are B-suitable (so that every vertex is either A-suitable or B-suitable, but not both). Observe moreover that for each $i \in [k]$, all provisionally good vertices in $A'_i = X'_i$ are A-suitable, and all provisionally good vertices in $B'_i = X'_{i+k}$ are B-suitable.

Observe also that for each $i \in [k]$ the number of bad vertices in V_i is at most $4\sqrt{\psi}(|Z'_i| + |X'_i| + |X'_{i+k}|) = 4\sqrt{\psi}n \leq m$. So we may choose for each $i \in [k]$ a set ZA_i of m vertices of Z'_i which includes all bad vertices of Z'_i which are A-suitable, and a set ZB_i of m vertices of Z'_i which includes all bad vertices of Z'_i which are B-suitable, in such a way that ZA_i and ZB_i are disjoint (note that $ZA_i \cup ZB_i$ must then include all bad vertices in Z'_i). Let AB_i consist of all vertices of A'_i which are not A-suitable (and so are B-suitable), and let BA_i consist of all vertices of B'_i which are not B-suitable (and so are A-suitable). Now for each $i \in [k]$ define

$$A_i := (A'_i \setminus AB_i) \cup BA_i \cup ZA_i,$$

$$B_i := (B'_i \setminus BA_i) \cup AB_i \cup ZB_i,$$

$$Z_i := Z'_i \setminus (ZA_i \cup ZB_i).$$

For each $i \in [k]$ let $X_i = A_i$, and for $i \in \{k+1, \ldots, 2k\}$ let $X_i = B_{i-k}$. Also, for each $i \in [k]$ let $d_i := |BA_i| - |AB_i|$. Observe that the graph G, the vertex classes V_i and the sets A_i, B_i, Z_i and X_i then meet the requirements of Setup 3.3.1, and furthermore that the size conditions follow from the sizes of the sets $A'_i, B'_i, Z'_i, AB_i, BA_i, ZA_i$ and ZB_i. So it remains only to check the final two conditions.

Observe that for each $i \in [2k]$ we have $|X'_i \triangle X_i| \leq 3m$ and $|Z'_{c(i)} \triangle Z_{c(i)}| \leq 2m$. For each $i \in [k]$, by choice of the sets AB_i, BA_i, ZA_i and ZB_i, every vertex of $A_i = X_i$ is A-suitable (in the sense defined above) so has at least $|X'_{i+1}|/2 - 3m \geq (|X_{i+1}| - 3m)/2 - 3m \geq |X_{i+1}|/3$ neighbours in X_{i+1}, and likewise every vertex of $B_i = X_{i+k}$ is B-suitable so has at least $|X'_{i+k+1}|/2 - 3m \geq |X_{i+k+1}|/3$ neighbours in X_{i+k+1}. We conclude that for each $i \in [2k]$ every vertex in X_i is $1/3$-suitable (with respect to the graph G and the sets X_i). Similar calculations show that for each $i \in [k]$ every vertex of $V_i \setminus (ZA_i \cup ZB_i)$ which is provisionally good (as defined above) is in fact α-good (with respect to the graph G and the sets X_i). Since $|ZA_i| = |ZB_i| = m$ and at most m vertices of V_i are not provisionally good, this completes the proof. □

Proposition 3.3.4. *Fix $\ell \geq 3$, let $Y_1, \ldots, Y_\ell$ be pairwise-disjoint sets of vertices, and for each $i \in [\ell]$ let $d_i \geq 0$ be an integer. Also let G be a graph with vertex set $\bigcup_{i\in[\ell]} Y_i$, and let $Q \subseteq V(G)$ have $|Z \cap Y_i| < |Y_i| - (d_{i-1} + d_i + d_{i+1})$ for each $i \in [\ell]$ (indices taken modulo ℓ). If every vertex in Y_i has at least d_i neighbours in Y_{i+1}, then there exists a family $(M_1, \ldots, M_\ell)$ of matchings in G such that for each $i \in [\ell]$ the matching M_i is a matching in $G[Y_i \setminus Q, Y_{i+1}]$ with $|M_i| = d_i$, and the matchings M_i are pairwise vertex-disjoint.*

We note that for each $i \in [\ell]$ a matching of size d_i in $G[Y_i \setminus Q, Y_{i+1}]$ can be chosen by arbitrarily choosing d_i vertices of $Y_i \setminus Q$ and then greedily selecting a distinct neighbour for each; the difficulty is to ensure that the chosen matchings have no vertices in common.

Proof. Let $\mathcal{M}$ be the set of all families $(M_1, \dots, M_\ell)$ of matchings in G such that for each $i \in [\ell]$ the matching M_i is a matching in $G[Y_i \setminus Q, Y_{i+1}]$ with $|M_i| \leq d_i$, and the matchings M_i are pairwise vertex-disjoint. Choose a family $M = (M_1, \dots, M_\ell) \in \mathcal{M}$ for which $\sum_{i \in [\ell]} |M_i|$ is maximal among all elements of $\mathcal{M}$; we will write $V(M)$ for $\bigcup_{i \in [\ell]} V(M_i)$.

Suppose for a contradiction that $\sum_{i \in [\ell]} |M_i| < \sum_{i \in [\ell]} d_i$. This implies that we may fix $i \in [\ell]$ with $|M_i| < d_i$, whereupon every vertex $x \in Y_i \setminus Q$ has at least one neighbour in $Y_{i+1} \setminus V(M_i)$. If there exists a vertex $x \in Y_i \setminus (V(M) \cup Q)$ with a neighbour $y \in Y_{i+1} \setminus V(M)$ then we can add the edge xy to M_i to obtain a contradiction. So we may assume that every vertex $x \in Y_i \setminus (V(M) \cup Q)$ has a neighbour in $Y_{i+1} \cap M_{i+1}$. Since $|Y_i \setminus (V(M) \cap Q)| > |Y_i| - d_i - d_{i-1} - |Q \cap Y_i| > d_{i+1}$, by the pigeonhole principle we may choose distinct vertices $x_i, x_i' \in Y_i \setminus (V(M) \cup Q)$ which share the same neighbour $y_i \in Y_{i+1} \cap V(M_{i+1})$. Let $z_i \in Y_{i+2}$ be the neighbour of y_i in M_{i+1}.

We now proceed iteratively as follows for each $j \in [\ell - 1]$ in turn: arbitrarily choose a vertex $x_{i+j} \in Y_{i+j} \setminus (V(M) \cup Q)$ and a neighbour $y_{i+j} \in Y_{i+j+1} \setminus (V(M_{i+j}) \setminus \{z_{i+j-1}\})$ of x_{i+j} ; this is possible since $|M_{i+j}| \leq d_{i+j}$ and $z_{i+j-1} \in V(M_{i+j})$ so there will be at least one neighbour available to choose. If $y_{i+j} \notin V(M_{i+j+1})$ then by adding the edge $x_i y_i$ to M_i and, for each $j' \in [j]$, replacing the edge $y_{i+j'-1} z_{i+j'-1}$ by the edge $x_{i+j'} y_{i+j'}$ in M_{i+j}, we obtain a family which contradicts the maximality of M. So we may assume that $y_{i+j} \in V(M_{i+j+1})$; let z_{i+j} be the neighbour of y_{i+j} in M_{i+j+1}.

If $y_{\ell-1} \neq x_i$, then by adding the edge $x_i y_i$ to M_i and, for each $j \in [\ell - 1]$, replacing the edge $y_{i+j-1} z_{i+j-1}$ by the edge $x_{i+j} y_{i+j}$ in M_{i+j}, we obtain a family which contradicts the maximality of M. On the other hand, if $y_{\ell-1} = x_i$ then the same is true with x_i' in place of x_i. In either case we conclude from the contradiction that actually we must have $\sum_{i \in [\ell]} |M_i| = \sum_{i \in [\ell]} d_i$, and so our chosen $\mathcal{M}$ satisfies the conditions of the proposition. □

Corollary 3.3.5. *Fix $k \geq 4$ and $\alpha \ll 1/k$, and suppose that $2k$ divides n. Let a* *graph G, vertex classes V_i for $i \in [k]$ and sets A_i, B_i, Z_i and X_i be as in Setup 3.3.1. Suppose also that $\delta^*(G) \geq (k+1)n/2k$, and that for each $i \in [2k]$, taking $d_i := |X_i| - (k-1)n/2k$, we have $0 \leq d_i \leq \alpha n$. Finally let Q be a set of at most αn vertices of G. Then there exists a matching M_i in $G[X_{i+k-1} \setminus Q, X_i]$ of size d_i for each $i \in [2k]$ such that the matchings M_i are pairwise vertex-disjoint.*

Proof. Observe first that each vertex $v \in X_{i+k-1}$ has at least $\delta^*(G) - (|V_i| - |X_i|) \geq d_i$ neighbours in X_i, and also note that the condition that $\alpha \ll 1/k$ ensures that $|Q \cap X_i| < |X_i| - 3\max_{i \in [2k]} d_i$ for each $i \in [2k]$. This is because

$$\begin{aligned} |Q \cap X_i| &= 1/2(|Q| + |X_i| - |Q \setminus X_i| - |X_i \setminus Q|) \\ &\leq \alpha n \leq (k-1)n/8k \leq (k-1)n/2k - 3\alpha n \\ &\leq |X_i| - 3\alpha n \leq |X_i| - 3\max_{i \in [2k]} d_i. \end{aligned}$$

If k is even, set $Y_i = X_{(k+1)i}$ for each $i \in [2k]$, and note that since $k+1$ and $2k$ are coprime, the sets $Y_1, \ldots, Y_{2k}$ are then precisely the sets $X_1, \ldots, X_{2k}$, just labelled differently but maintaining the same order. Our initial observation above then implies that each $x \in Y_i$ has at least d_i neighbours in Y_{i+1}, so we may apply Proposition 3.3.4 to find a family of pairwise vertex-disjoint matchings M'_i for $i \in [2k]$ such that M'_i is a matching of size d_i in $G[Y_i \setminus Q, Y_{i+1}]$; relabelling by taking $M_i := M'_{(k+1)i}$ then gives the required matchings.

If k is odd, set $Y_i^0 = X_{(k+1)i}$ and $Y_i^1 = X_{1+(k+1)i}$ for each $i \in [k]$, and note that since $\gcd(k+1, 2k) = 2$, the sets Y_i^a for $a \in \{0,1\}$ and $i \in [k]$ are again precisely the sets $X_1, \ldots, X_{2k}$, just labelled differently. For each $i \in [k]$ let $a(i) = i(k+1)$ and $b(i) = 1 + i(k+1)$; our initial observation then implies that each $x \in Y_i^0$ has at least $d_{a(i)}$ neighbours in Y_{i+1}^0, and each $x \in Y_i^1$ has at least $d_{b(i)}$ neighbours in Y_{i+1}^1. So applying Proposition 3.3.4 twice, first to the sets Y_i^0 for $i \in [k]$, then to the sets Y_i^1 for $i \in [k]$, yields families M_i^0 and M_i^1 of pairwise-disjoint matchings such that M_i^0 is a matching of size $d_{a(i)}$ in $G[Y_i^0 \setminus Q, Y_{i+1}^0]$, whilst M_i^1 is a matching of size $d_{b(i)}$ in $G[Y_i^1 \setminus Q, Y_{i+1}^1]$. Relabelling by taking $M_i := M'_{(2k+1)(i-1)}$ if i is odd and $M_i = M'_{(2k+1)i}$ if i is even then gives the required matchings. $\square$

Under the structure of Setup 3.3.1, a canonical cycle of index i is a copy of C_k with vertices in $Z_{c(i)}, X_{i+1}, X_{i+2}, \ldots, X_{i+k-1}$.

Proposition 3.3.6. *Let $1/n \ll \alpha \ll \beta \ll 1/k$. Let a graph G, vertex classes V_i for $i \in [k]$ and sets A_i, B_i, Z_i and X_i be as in Setup 3.3.1, and let Q be a set of at most βn 'forbidden' vertices of G, where in particular Q contains all vertices of G which are not α-good. Then the following statements hold.*

(i) *For each $i \in [2k]$ there is a canonical cycle in G of index i which does not include any vertex of Q.*

(ii) *Let $i \in [2k]$ and let $x \in X_i$ be an α-suitable vertex. Then there is a canonical cycle in G with index either $i-1$ or $i-2$ which contains x and does not include any vertex of Q other than x.*

(iii) *Let $i \in [2k]$ and let xy be an edge of G for which $y \in X_i$ is α-suitable and x is a good vertex in X_{i+k-1}. Then there exists a cycle of length k in G which includes the edge xy and has vertices in $X_i, X_{i+1}, \ldots, X_{i+k-1}$ where all vertices except perhaps x and y are not in Q.*

Proof. For (i) we can form such a cycle greedily. Choose a good vertex $x_1 \in X_{i+1} \setminus Q$, and then for each $j \in \{2, \ldots, k-1\}$ in turn choose a good vertex $x_j \in (X_{i+j} \cap N(x_{j-1})) \setminus Q$; in each case this is possible since x_{j-1} is α-good so has at least $(1-\alpha)|X_{i+j}| > \beta n \geq |Q|$ neighbours in X_{i+j}. To complete the cycle choose a vertex $y \in (Z_{c(i)} \cap N(x_{k-1}) \cap N(x_1)) \setminus Q$; this is possible since x_{k-1} and x_1 are both α-good so each has at least $(1-\alpha)|Z_{c(i)}|$ neighbours in Z_{i-1}, and it follows that $|Z_{c(i)} \cap N(x_{k-1}) \cap X_1| \geq |Z_{c(i)}| - \alpha n \geq |Q|$. We then have the desired cycle $x_1 x_2 \ldots x_{k-1} y$.

For (ii) we proceed similarly, except first we choose the neighbour of x in $V_{c(i-1)}$. Observe that either x has at least βn neighbours in X_{i-1} or at least βn neighbours in $Z_{c(i-1)}$. In the former case we form a canonical cycle with index $i-2$ which contains x, and in the latter case we form a canonical cycle with index $i-1$ which contains x.

The argument for (iii) is again similar: we greedily choose the vertices of the cycle in $X_{i+1}, \dots, X_{i+k-2}$ in turn. $\square$

Proposition 3.3.7. *Let a graph G, vertex classes V_i for $i \in [k]$ and sets A_i, B_i, Z_i and X_i be as in Setup 3.3.1. Suppose that $2k$ divides n, that $|A_i| = |B_i| = (k-1)n/2k$ for each $i \in [2k]$ and $|Z_i| = n/k$ for each $i \in [k]$. Suppose also that every vertex is $\frac{1}{4k}$-good. Then G contains a transversal C_k-factor.*

Proof. For each $i \in [k]$ arbitrarily partition Z_i into two parts Z'_i and Z'_{i+k} of equal size $n/2k$. Likewise, for each $i \in [2k]$ let $X_i^1, \dots, X_i^{k-1}$ be an arbitrary partition of X_i into $k-1$ parts of equal size $n/2k$. Each bipartite graph $G[X_i^j, X_{i+1}^{j+1}]$ then has minimum degree at least $n/4k$, and so admits a perfect matching. Taking the union of these perfect matchings, we obtain for each $i \in [2k]$ a perfect P_{k-2}-tiling $\mathcal{P}_i$ in $G[X_{i+1}^1, X_{i+2}^2, \dots, X_{i+k-1}^{k-1}]$. Let B_i be the bipartite graph whose vertex classes are Z'_i and $\mathcal{P}_i$, where $v \in Z'_i$ and $P \in \mathcal{P}_i$ are adjacent in B_i if v is a neighbour of both endvertices of P, that is, if v together with the vertices of P forms a copy of C_k in G. B_i then has minimum degree at least $n/4k$, and so admits a perfect matching. The corresponding copies of C_k then form a perfect C_k-tiling in $G[Z'_i, X_{i+1}^1, X_{i+2}^2, \dots, X_{i+k-1}^{k-1}]$, and combining these tilings for every $i \in [2k]$ gives a transversal C_k-factor in G, as required. $\square$

Finally we combine all of the previous results to prove Lemma 3.1.2.

Proof of Lemma 3.1.2. Choose constants α and β such that $\psi \ll \alpha \ll \beta \ll 1/k$. Let $m = \left\lceil 4\sqrt{\psi} n \right\rceil$. By Proposition 3.3.3 there exist sets A_i, B_i, Z_i and X_i for $i \in [k]$ meeting the conditions of Setup 3.3.1, and integers d_i with $|d_i| \leq m$ for $i \in [k]$, with the properties that

- $|A_i| = (k-1)n/2k + d_i + m$,
- $|B_i| = (k-1)n/2k - d_i + m$,
- $|Z_i| = n/k - 2m$,
- for each $i \in [2k]$, every vertex in X_i is α-suitable, and
- at most αn vertices are not α-good.

Here, the fourth conclusion holds because of the observation that a vertex being α-suitable implies that it is α'-suitable for each $\alpha' \leq \alpha$. This follows directly from the definition. For each $i \in [2k]$ let $c_i := |X_i| - (k-1)n/2k$, so that $0 \leq c_i \leq 2m$. Also let B consist of all vertices in G which are not α-good, so $|B| \leq \alpha n$. By Corollary 3.3.5 there exists a matching M which, for each $i \in [2k]$, contains precisely c_i edges with one vertex in $X_{i+k-1} \setminus Q$ and the other in X_i (and no edges other than these). Let M_i denote the edges for each i. Here, we can apply Corollary 3.3.5 because $\alpha \ll 1/k$.

Apply Proposition 3.3.6 (iii) iteratively to choose, for each $i \in [2k]$ and each edge $e \in M_i$, a copy C of C_k in G which includes the edge e and has vertices in $X_i, X_{i+1}, \dots, X_{i+k-1}$. In each of these applications of Proposition 3.3.6 we take the forbidden set Q to consist of all vertices in B, in M or in any previously-chosen cycles. Therefore, we can apply Proposition 3.3.6 since we have $|Q| \leq |B| + k \cdot |M| \leq \alpha n + 4k^2 m \leq \beta n$. In this way we ensure that all of the $\sum_{i \in [2k]} c_i$ chosen cycles are pairwise vertex-disjoint. Let $\mathcal{C}_1$ be the collection of cycles chosen in this step. Note that $|V(\mathcal{C}_1)| \leq k\,|M| \leq 4k^2 m$.

For each $i \in [2k]$ let $B_i := (B \cap X_i) \setminus V(\mathcal{C}_1)$, and write $b_i := |B_i|$. We next apply Proposition 3.3.6 (ii) iteratively to choose, for each $i \in [2k]$ and each vertex $x \in B_i$, a canonical cycle in G with index either $i-1$ or $i-2$ which contains x. In each of these applications of Proposition 3.3.6 we take the forbidden set Q to consist of all vertices in B or in $V(\mathcal{C}_1)$, or in any previously-chosen cycles. Therefore, we can apply Proposition 3.3.6 again because $|Q| \leq k \cdot |B| + |\mathcal{C}_1| \leq k\alpha n + 4k^2 m \leq \beta n$. This ensures that all of the $\sum_{i\in[2k]} b_i$ chosen cycles are pairwise vertex-disjoint. Let $\mathcal{C}_2$ be the collection of cycles chosen in this step. Note that $|V(\mathcal{C}_2)| \leq k\alpha n$.

For each $i \in [2k]$ we now apply Proposition 3.3.6 (i) iteratively to choose, for each $i \in [2k]$, a total of $(\sum_{j\in[2k]}(c_j + b_j)) - c_i - b_i$ canonical cycles in G of index i. In each of these applications of Proposition 3.3.6 we take the set Q to consist of all vertices in $V(\mathcal{C}_1) \cup V(\mathcal{C}_2)$ or in any previously-chosen cycles. Once again, we can apply Proposition 3.3.6 because $Q \leq k\alpha n + 4k^2 m + 8k^2 m + 4k^2\alpha n \leq \beta n$. This ensures that all of the chosen cycles are pairwise vertex-disjoint. Let $\mathcal{C}_3$ be the collection of cycles chosen in this step.

Let $\mathcal{C} := \mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3$ be the set of all cycles we have chosen, and write $t := |\mathcal{C}|$ and $n^* := n - t$. Note that $t = k(\sum_{j\in[2k]}(c_j + b_j))$. Let $X_i^* := X_i \setminus V(\mathcal{C})$ and $Z_i^* := Z_i \setminus V(\mathcal{C})$ for each i. For each i we then have

$$|X_i^*| = |X_i| - (k-1)t/k - c_i = (k-1)n^*/2k,$$

and $|Z_i^*| = n^* - |X_i^*| - |X_{i+k}^*| = n^*/k$. Moreover, since all vertices which were not good were covered by $\mathcal{C}$, all remaining vertices are 2α-good. By Proposition 3.3.7 then gives us a perfect C_k-tiling $\mathcal{C}'$ in $G^* := G \setminus \mathcal{C}$, whereupon $\mathcal{C} \cup \mathcal{C}'$ is a perfect C_k-tiling in G. □

3.4 Proof of Lemma 3.1.3

We now state the Regularity Lemma of Szemerédi [67]. Though this was originally conceived to solve a problem relating to arithmetic progressions in the integers, it has since seen numerous applications in results relating to embedding large subgraphs into graphs. Notably, an early application in the area of graph tilings is the proof by Komlós, Sárközy and Szemerédi [44] of a conjecture of Alon and Yuster [3], showing that for any graph F, there is some constant $C = C(F)$ such that for any n which is divisible by $|F|$, any n-vertex graph with minimum degree at least $(1 - \chi(F))n + C$

contains an F-factor. This was improved on by Kühn and Osthus [50] who showed that $\chi(F)$ could be replaced by a new quantity $\chi^*(F)$. Here, depending on some properties of the graph F, the quantity $\chi^*(F)$ is either equal to the critical chromatic number $\chi_{cr}(F)$, or the chromatic number $\chi(F)$. For brevity, we do not define these notions here. For more examples of applications of the regularity lemma to graph embedding results, see [47].

For a graph G and disjoint vertex sets $X, Y \subseteq V(G)$, we write $G[X, Y]$ for the bipartite subgraph of G with vertex classes X and Y comprising all edges of G containing both a vertex of X and a vertex of Y, and write $d_G(X, Y)$ for the density of this subgraph, that is, $d_G(X, Y) := \frac{e(G[X,Y])}{|X||Y|}$ (we omit the subscript G when it is clear from the context). Now let G be a bipartite graph with vertex classes A and B. We say that G is (d, ε)*-regular* if $d(X, Y) = d \pm \varepsilon$ for all sets $X \subseteq A$ and $Y \subseteq B$ with $|X| \geq \varepsilon|A|$ and $|Y| \geq \varepsilon|B|$. Sometimes we do not wish to specify the density exactly, in which case we say that G is ε-regular if there exists some $d \in [0, 1]$ for which G is (d, ε)-regular, and we say that G is $(\geq d, \varepsilon)$ regular if there exists $d' \in [d, 1]$ for which G is (d', ε)-regular.

Lemma 3.4.1 (Regularity Lemma [67])**.** *Suppose that $1/n \ll 1/T \ll 1/t, \varepsilon, 1/k$. Let G be a k-partite graph whose vertex classes $V_1, \ldots, V_k$ each have size n. Then there exist integers ℓ and m and a partition of V_i into parts $X^i_0, X^i_1, \ldots X^i_\ell$ for each $i \in [k]$ such that*

(i) $t \leq \ell \leq T$,

(ii) $|X^i_0| \leq \varepsilon n$ *for each* $i \in [k]$,

(iii) $|X^i_j| = m$ *for every* $i \in [k]$ *and* $j \in [\ell]$,

(iv) *for each* $i, i' \in [k]$ *with* $i \neq i'$ *and each* $j \in [\ell]$ *there are at most* $\varepsilon\ell$ *elements* $j' \in [\ell]$ *for which* $G[X^i_j, X^{i'}_{j'}]$ *is not* ε*-regular.*

Observe that we then have $n - \varepsilon n \leq m\ell \leq n$. We will commonly refer to the sets X^i_j for $i \in [k]$ and $j \in [\ell]$ obtained from the regularity lemma as clusters. For a given density parameter $d > 0$, once we have applied the regularity lemma to a graph G in the given form, we may form a d-reduced graph R whose vertices are the clusters X_j^i for $i \in [k]$ and $j \in [\ell]$ and which has an edge between X_j^i and $X_{j'}^{i'}$ if the subgraph $G[X_j^i, X_{j'}^{i'}]$ is $(\geq d, \varepsilon)$-regular; the edges of R then indicate where we can usefully work with regularity. Observe that the reduced graph R is naturally k-partite with vertex classes $U_i = \{X_j^i : j \in [\ell]\}$. Moreover, with respect to these vertex classes R inherits a degree condition from G; more precisely we have $\delta^*(R) \geq \frac{\ell}{n} \cdot \delta^*(G) - (d + 3\varepsilon)\ell$. Indeed, for any cluster X_j^i in U_i there are at least $m\delta^*(G)$ edges of G between X_j^i and V_{i+1}. At most $\varepsilon n \cdot m$ of these edges include a vertex of X_0^{i+1}, at most $\varepsilon \ell m^2 \leq \varepsilon nm$ are contained in pairs $G[X_j^i, X_{j'}^{i+1}]$ which are not ε-regular, and at most $(d+\varepsilon)m^2\ell \leq (d+\varepsilon)nm$ are contained in pairs $G[X_j^i, X_{j'}^{i+1}]$ with $d(X_j^i, X_{j'}^{i+1}) \leq (d+\varepsilon)$ (which is an upper bound for the density of a pair which is ε-regular but not $(\geq d, \varepsilon)$-regular). We conclude that the number of edges between X_j^i and V_{i+1} which are contained in pairs $G[X_j^i, X_{j'}^{i+1}]$ which are $(\geq d, \varepsilon)$-regular is at least $m\delta^*(G) - 2\varepsilon nm - (d+\varepsilon)nm = mn(\delta^*(G)/n - (d+3\varepsilon))$. Since each pair contains at most $m^2 \leq mn/\ell$ edges we conclude that X_j^i has at least $mn(\delta^*(G)/n - (d+3\varepsilon))/(mn/\ell) = \frac{\ell}{n}\delta^*(G) - (d+3\varepsilon)\ell$ neighbours in U_{i+1}; an identical argument with $i-1$ in place of $i+1$ shows that X_j^i has at least $\frac{\ell}{n}\delta^*(G) - (d+3\varepsilon)\ell$ neighbours in U_{i-1} as well, giving the claimed degree condition.

The key properties of regularity for our purposes are expressed by the next two lemmas; both are standard results but we include short proofs for completeness. The first, often called the slicing lemma, states that regularity is inherited (with weaker parameters) by subgraphs induced by not-too-small subsets of each cluster.

Lemma 3.4.2 (Slicing Lemma)**.** *Fix $d, \varepsilon, \alpha \in [0,1]$, and let G be a (d,ε)-regular bipartite graph with vertex classes A and B. For all sets $X \subseteq A$ and $Y \subseteq B$ with $|X| \geq \alpha|A|$ and $|Y| \geq \alpha|B|$ the induced subgraph $G[X,Y]$ is $(d, \varepsilon/\alpha)$-regular.*

Proof. For all sets $X' \subseteq X$ and $Y' \subseteq Y$ with $|X'| \geq (\varepsilon/\alpha)|X|$ and $|Y'| \geq (\varepsilon/\alpha)|Y|$ we have $|X'| \geq \varepsilon|A|$ and $|Y'| \geq \varepsilon|B|$. By regularity of G it follows that $d(X', Y') = d \pm \varepsilon$, so $d(X', Y') = d \pm \varepsilon/\alpha$, as required. $\square$

The second lemma states that if we have k clusters arranged in cyclical fashion, and each consecutive pair of clusters forms a dense regular pair, then we can find a copy of C_k with one vertex in each cluster. This is a vastly simplified form of the Counting Lemma, with allows us to estimate the number of partition-respecting copies of any small k-partite subgraph in G, but it suffices for our purposes.

Lemma 3.4.3. *Let G be a k-partite graph with vertex classes $X_1, \dots, X_k$ in which $G[X_i, X_{i+1}]$ is $(\geq d, \varepsilon)$-regular for every $i \in [k]$ (with addition on the indices taken modulo k). If $2\varepsilon < d \leq 1$, then G contains a transverse copy of C_k.*

Proof. Observe that the regularity condition implies that for each $i \in [k]$ fewer than $\varepsilon|X_i|$ vertices in X_i have at most $\varepsilon|X_{i+1}|$ neighbours in X_{i+1}, and fewer than $\varepsilon|X_{i+1}|$ vertices in X_{i+1} have at most $\varepsilon|X_i|$ neighbours in X_i. So we may choose a vertex $x_1 \in X_1$ with $|N(x_1) \cap X_2| > \varepsilon|X_2|$ and $|N(x_1) \cap X_k| > \varepsilon|X_k|$. Similarly, for each $2 \leq j \leq k-2$ in turn we may choose $x_j \in X_j \cap N(x_{j-1})$ with $|N(x_j) \cap X_{j+1}| > \varepsilon|X_{j+1}|$. Now write $S := N(x_1) \cap X_k$. Since $|S| \geq \varepsilon|X_k|$, the regularity condition ensures that fewer than $\varepsilon|X_{k-1}|$ vertices in X_{k-1} have at most $\varepsilon|S|$ neighbours in S. So we may choose $x_{k-1} \in N(x_{k-2}) \cap X_{k-1}$ with $|N(x_{k-1}) \cap S| \geq \varepsilon|S| > 0$, and then take x_k to be any neighbour of x_{k-1} in S, to obtain the desired cycle $x_1 x_2 \dots x_k x_1$. $\square$

We now prove Lemma 3.1.3 given Theorem 3.1.4 and Lemma 3.1.5. Haxell and Rödl [34] showed that for any H for sufficiently large n, whenever an n-vertex graph contains a perfect fractional H-factor, it also contains an 'almost perfect' H-tiling. We derive an an analogous result for transversal H-factors in multipartite graphs using similar methods.

Proof of Lemma 3.1.3. Introduce new constants with $1/n \ll 1/T \ll 1/t, \varepsilon \ll d \ll \sigma \ll \gamma \ll \psi' \ll \psi, 1/k$. Let G be a balanced k-partite graph whose vertex classes $V_1, \dots, V_k$ each have size n and which satisfies $\delta^*(G) \geq (\frac{1}{2} + \frac{1}{2k} - \gamma)n$. Apply

Lemma 3.1.5 to obtain an integer $z \leq \sigma n$ and an absorbing set $A \subseteq V(G)$ with $|A \cap V_i| = z$ for each $i \in [k]$ such that if $G - A$ has a transversal C_k-tiling of size at least $n - z - \sigma^2 n$, then G contains a transversal C_k-factor.

Apply the regularity lemma (Lemma 3.4.1) to $G - A$ (with vertex classes $V_i \setminus A$ for each $i \in [k]$) to obtain integers ℓ and m and, for each $i \in [k]$, a partition of $V_i \setminus A$ into parts $X_0^i, X_1^1, \dots X_\ell^i$, such that

1. $t \leq \ell \leq T$,
2. $|X_0^i| \leq \varepsilon(n - z) \leq \varepsilon n$ for each $i \in [k]$,
3. $|X_j^i| = m$ for every $i \in [k]$ and $j \in [\ell]$,
4. for each $i, i' \in [k]$ with $i \neq i'$ and each $j \in [\ell]$ there are at most $\varepsilon \ell$ elements $j' \in [\ell]$ for which $G[X_j^i, X_{j'}^{i'}]$ is not ε-regular.

For each $i \in [k]$ define $U_i := \{X_j^i : j \in [\ell]\}$, and let R be the d-reduced graph of G, as defined shortly after Lemma 3.4.1, so $X_j^i X_{j'}^{i'}$ is an edge of R if $G[X_j^i, X_{j'}^{i'}]$ is $(\geq d, \varepsilon)$-regular. As noted there, the reduced graph R is then k-partite with vertex classes $U_1, \dots, U_k$, each of which has size ℓ, and moreover we have $\delta^*(R) \geq \frac{\ell}{n-z} \cdot \delta^*(G - A) - (d + 3\varepsilon)\ell \geq (\frac{1}{2} + \frac{1}{2k} - \gamma - \sigma)\ell - (d + 3\varepsilon)\ell \geq (\frac{1}{2} + \frac{1}{2k} - 2\gamma)\ell$. We may therefore apply Theorem 3.1.4 to deduce that R either contains a perfect fractional C_k-tiling or is ψ'-extremal.

Suppose first that R contains a perfect fractional transversal C_k-tiling. Let $\mathcal{C}(R)$ denote the set of transverse copies of C_k in R, and for each $C \in \mathcal{C}(R)$ let $w_C \in [0, 1]$ be the weight of C in this tiling. Also for each $X \in V(R)$ let $\mathcal{C}(X) := \{C \in \mathcal{C}(R) : X \in V(C)\}$, so by the definition of a perfect fractional tiling we have $\sum_{C \in \mathcal{C}(X)} w_C = 1$ for each $X \in V(R)$. We now greedily choose, for each $C \in \mathcal{C}(R)$ in turn, a collection of $\lfloor (1 - d) w_C m \rfloor$ transverse copies of C_k in G, each with one vertex in each cluster in C. Moreover we do this so that the chosen copies of C_k in G are pairwise vertex-disjoint, meaning that they form a transversal C_k-tiling $\mathcal{C}$ in

$G - A$. We write $V(\mathcal{C}) := \bigcup_{C \in \mathcal{C}} V(C)$. To see that it is possible to form $\mathcal{C}$ in this way, observe that for each $X \in V(R)$ we have

$$|V(\mathcal{C}) \cap X| = \sum_{C \in \mathcal{C}(X)} \lfloor (1-d) w_C m \rfloor \leq (1-d)m \sum_{C \in \mathcal{C}(X)} w_C = (1-d)m.$$

This means that every time we come to choose a copy of C_k in G, with one vertex in each cluster in some cycle $C \in \mathcal{C}(R)$, we know that at least dm vertices of each cluster of C have not been used by previously-chosen cycles. So if we let $Y_1, \ldots, Y_k$ be the clusters of C (with $Y_i \in U_i$ for each $i \in [k]$), and for each $i \in [k]$ let Y_i' be the set of as-yet-unused vertices in Y_i, then $G[Y_i, Y_{i+1}]$ is $(\geq d, \varepsilon)$-regular by definition of R, so $G[Y_i', Y_{i+1}']$ is $(\geq d, \varepsilon/d)$-regular by Lemma 3.4.2. We may therefore use Lemma 3.4.3 to choose a transverse copy of C_k in G with one vertex in each set Y_i', as required. Using the fact that $|\mathcal{C}(R)| \leq \ell^k$, we find that $\mathcal{C}$ has size at least

$$\begin{aligned} |\mathcal{C}| = \sum_{C \in \mathcal{C}(R)} \lfloor (1-d) w_C m \rfloor &\geq \left(\sum_{C \in \mathcal{C}(R)} (1-d) w_C m \right) - \ell^k \\ &\geq (1-d)\ell m - \ell^k \geq (1-d)(1-\varepsilon)(n-z) - \sigma^2 n/2 \geq n - z - \sigma^2 n, \end{aligned}$$

where in the third inequality we used the properties of our application of the regularity lemma which imply that $\ell m \geq (1-\varepsilon)(n-z)$ and $\ell \leq T$, and our assumption that $1/n \ll 1/T$. By choice of A and the existence of $\mathcal{C}$ we conclude that G contains a transversal C_k-factor.

Now suppose instead that R is ψ'-extremal. This means that there exist disjoint subsets $A_i', B_i' \subseteq U_i$ with $|A_i'| = |B_i'| = \lfloor (\frac{1}{2} - \frac{1}{2k})\ell \rfloor$ for each $i \in [k]$ such that $d_R(A_1', A_k') \leq \psi'$, $d_R(B_1', B_k') \leq \psi'$, and for each $i \in [k-1]$ we have $d_R(A_i', B_{i+1}') \leq \psi'$ and $d_R(B_i', A_{i+1}') \leq \psi'$. Observe that for each $i \in [k]$ we have

$$\left| \bigcup_{X \in A_i'} X \right| = m|A_i'| = m \left\lfloor \left(\frac{1}{2} - \frac{1}{2k} \right) \ell \right\rfloor \leq \left\lfloor \left(\frac{1}{2} - \frac{1}{2k} \right) m\ell \right\rfloor \leq \left\lfloor \left(\frac{1}{2} - \frac{1}{2k} \right) n \right\rfloor,$$

and likewise we have $|\bigcup_{X\in B'_i} X| \le \lfloor(\frac{1}{2}-\frac{1}{2k})n\rfloor$, so we may choose subsets $A_i, B_i \subseteq V_i$ with $|A_i| = |B_i| = \lfloor(\frac{1}{2}-\frac{1}{2k})n\rfloor$ such that $\bigcup_{X\in A'_i} X \subseteq A_i$ and $\bigcup_{X\in B'_i} X \subseteq B_i$. Moreover, for each $i \in [k]$ we have

$$\left|\bigcup_{X\in A'_i} X\right| = m\left\lfloor\left(\frac{1}{2}-\frac{1}{2k}\right)\ell\right\rfloor \ge \left(\frac{1}{2}-\frac{1}{2k}\right)\ell m - m \ge \left(\frac{1}{2}-\frac{1}{2k}\right)(1-\varepsilon)(n-z) - \frac{n}{t} \ge \left(\frac{1}{2}-\frac{1}{2k}\right)n - 2\sigma n,$$

so $|A_i \setminus \bigcup_{X\in A'_i} X| \le 2\sigma n$, and likewise we have $|B_i \setminus \bigcup_{X\in B'_i} X| \le 2\sigma n$. So taking a crude bound, the number of edges between A_1 and A_k is at most $\psi'\ell^2m^2 + (d+\varepsilon)\ell^2m^2 + \varepsilon\ell^2m^2 + 4\sigma n^2 \le \psi|A_1||A_k|$, so $d_G(A_1, A_k) \le \psi$. Similar calculations show that $d_G(B_1, B_k) \le \psi$, and that for each $i \in [k-1]$ we have $d_G(A_i, B_{i+1}) \le \psi$ and $d_G(B_i, A_{i+1}) \le \psi$, so the sets A_i and B_i witness that G is ψ-extremal. □

3.5 Finding a perfect fractional C_k-tiling

The aim of this section is to prove Theorem 3.1.4, which we recall below.

Theorem 3.1.4. *Suppose that $1/n \ll \gamma \ll \psi, 1/k$, and let G be a balanced k-partite graph whose vertex classes each have size n. If $\delta^*(G) \ge (\frac{1}{2}+\frac{1}{2k})n - \gamma n$, then either G contains a perfect fractional C_k-tiling or G is ψ-extremal.*

Recall that a copy of C_k in a k-partite graph G is *transverse* if it contains exactly one vertex from each vertex-class of G. For a graph H on at most k vertices, we say that a copy of H is *partition-respecting* in a k-partite graph G if it contains at most one vertex from each vertex class. Thus, in some sense this generalises the notion of being transverse to graphs with fewer than k vertices. Unless otherwise specified, all cycles we consider in this section will be transversal cycles, and all paths will be partition-respecting. Furthermore, whenever the vertex set of a cycle is indexed by $[k]$, we assume that the vertices indexed by i and $i+1$ are adjacent on the cycle, for each $i \in [k]$. Similarly whenever the vertex set of a path is indexed by $[\ell]$, we assume that the vertices indexed by 1 and ℓ are the endvertices and for each $i \in [\ell-1]$, the vertices indexed by i and $i+1$ are adjacent on the path.

In order to illustrate the proof method, we first state and give a brief outline of the proof of the following theorem, which has a slightly weaker degree condition, and which can be used to give an alternative proof of the result of Ergemlidze and Molla [24].

Lemma 3.5.1. *Let $1/n \ll \gamma \ll 1/k \ll 1$. Let G be a k-partite graph with vertex classes $V_1, \dots, V_k$, each of size n, and $\delta^*(G) \geq (\frac{1}{2} + \frac{1}{2k} + \gamma)n$. Then G contains a perfect fractional C_k-tiling.*

Proof outline of Lemma 3.5.1. Let G be as above and suppose that G contains no perfect fractional C_k-tiling. Then by Fact 3.2.7, $\mathbf{1} \notin PC(\chi(T)\colon T \text{ is a copy of } C_k \text{ in } G)$. Therefore, by Lemma 3.2.8 (Farkas' lemma), there is a vector $\boldsymbol{a} \in \mathbb{R}^{kn}$ such that for each copy T of C_k in G, $\boldsymbol{a} \cdot \chi(T) \geq 0$, but $\boldsymbol{a} \cdot \mathbf{1} < 0$. For each $i \in [k]$, label the vertices of V_i by $v_{i,1}, \dots, v_{i,n}$ such that $\boldsymbol{a} \cdot \chi(v_{i,j}) \leq \boldsymbol{a} \cdot \chi(v_{i,j'})$ for each $j \leq j'$. Partition V_i into two parts $V_{i,T}$ and $V_{i,B}$, where $V_{i,T} = \{v_{i,1}, \dots, v_{i,\frac{k-1}{k}n}\}$ and $V_{i,B} = \{v_{i,\frac{k-1}{k}n+1} \dots v_{i,n}\}$.

Let $G_T = G[\bigcup_{i \in [k]} V_{i,T}]$ and let $n_T = \frac{(k-1)n}{k}$. Then $\delta^*(G_T) \geq (\frac{1}{2} + \gamma)n_T$. Therefore, for each $i \in [k]$, Hall's condition implies that we can find a perfect matching between $V_{i,T}$ and $V_{i+1,T}$. This implies the existence of a perfect fractional P_{k-2}-tiling in G_T, in which each copy of P_{k-2} has weight 0 or $1/(k-1)$. We prove this later in Lemma 3.5.3. More importantly, let $\mathcal{P}$ be the collection of paths assigned a non-zero weight in this fractional path tiling. Note that each such path intersects with exactly $k-1$ vertex classes. Split the collection $\mathcal{P}$ into k classes $\mathcal{P}_1, \dots, \mathcal{P}_k$ such that $\mathcal{P}_i$ contains all paths in $\mathcal{P}$ which do not intersect with V_i. There are $(k-1)n/k$ paths in $\mathcal{P}_i$ for each $i \in [k]$. Split this into n/k equal parts $\mathcal{P}_{i,1}, \dots, \mathcal{P}_{i,n/k}$ each of size $k-1$, and label the paths in $\mathcal{P}_{i,j}$ by $P_{i,j,\ell}$ for $\ell \in [k-1]$. Now for each $i \in [k]$, $j \in [n/k]$, $\ell \in [k-1]$, let $T_{i,j,\ell}$ be the set of vertices in $P_{i,j,\ell}$ together with the vertex $v_{i,(k-1)n/k+j}$. If $\mathcal{T}$ is the collection of all these sets $T_{i,j,\ell}$, then observe that each vertex in $V(G)$ appears in $k-1$ paths in $\mathcal{P}$ and therefore in $k-1$ sets in $\mathcal{T}$. Therefore, $\chi(T) = (k-1) \cdot \mathbf{1}$.

Also, for each $T_{i,j,\ell} \in \mathcal{T}$, there is a cycle $C_{i,j,\ell}$ in G such that $\boldsymbol{a} \cdot \chi(C_{i,j,\ell}) \leq$

$\boldsymbol{a} \cdot \chi(T_{i,j,\ell})$. This is because the endvertices of the path $P_{i,j,\ell}$ have at least $n/k + 2\gamma n$ common neighbours in V_i, and therefore, there is at least one such common neighbour with smaller index than $v_{i,(k-1)n/k+j}$. Therefore,

$$0 \leq \sum_{i \in [k], j \in [n/k], \ell \in [k-1]} \boldsymbol{a} \cdot \chi(C_{i,j,\ell}) \leq \sum_{i \in [k], j \in [n/k], \ell \in [k-1]} \boldsymbol{a} \cdot \chi(T_{i,j,\ell}) \cdot = (k-1) \cdot \boldsymbol{a} \cdot \mathbf{1} < 0,$$

a contradiction. Therefore, G contains a perfect fractional C_k-tiling. □

Our proof works similarly to the proof outlined above, where we begin by assuming there is no perfect fractional C_k-tiling and thus applying Farkas' lemma to find an ordering of the vertices. However, there are two major difficulties which arise when adapting the proof outlined above to our setting which has a weaker degree condition.

The first difficulty is finding a perfect fractional P_{k-2}-tiling in G_T. In the proof of Lemma 3.5.1, Hall's condition and the minimum degree immediately imply the existence of a perfect matching between $V_{i,T}$ and $V_{i+1,T}$ for each $i \in [k]$, and this immediately allows us to find a perfect fractional P_{k-2}-tiling in G_T. This is no longer the case with the degree condition in Theorem 3.1.4. A significant portion of this section is dedicated to determining when exactly we can find a perfect fractional path tiling in G_T. We make a distinction in cases depending on whether or not there is a robust pair of vertex classes in G_T. In particular we prove that if there is a robust pair $(V_{i,T}, V_{i+1,T})$ in G_T, then G_T contains a perfect fractional P_{k-2}-tiling, namely, we prove the following.

Lemma 3.5.2. *Let $1/n \ll \gamma \ll \nu \ll \tau \ll 1/k$. Let G be a k-partite graph with vertex classes $V_1, \dots, V_k$, each of size n, and $\delta^*(G) \geq (\frac{1}{2} - \gamma)n$. Suppose there is some $i^* \in [k]$ such that $G[V_{i^*}, V_{i^*+1}]$ is a robust (ν, τ)-expander. Then G contains a perfect fractional P_{k-2}-tiling.*

When there is no such robust pair, Lemma 3.2.5 gives us a partition of $(V_{i,T}, V_{i+1,T})$ into two graphs which are 'close to' complete bipartite graphs. In particular, each vertex class $V_{i,T}$ is partitioned twice. This time, we the case distinction we make is depending on how well aligned these two partitions are. Informally, we say these partitions are well aligned if there is some labelling of the partitions $A_1 \cup A_2$ and $B_1 \cup B_2$ such that A_i and B_i differ only in a small number of vertices. Otherwise, we say they are not well aligned. It turns out that when there is some i^* such that the two partitions of $V_{i^*,T}$ are not well aligned, we can find a perfect fractional path tiling in G_T. On the other hand, if these partitions are 'aligned' for each $i \in [k]$, then we show that either G_T contains a perfect fractional path tiling, or the original graph G is ψ-extremal.

The other key component of the proof of Lemma 3.5.1 which is difficult to adapt to our setting is the step in which we use the path tiling we have found in order to find a collection $\mathcal{T}$ of sets that cover the vertex set of G equally, and that each dominate a cycle in G. In the proof of Lemma 3.5.1, we can do this directly by pairing each path P with a vertex from $V_{i,B}$ such that $P \cap V_i = \emptyset$ to obtain each set T. As long as we do this in a balanced way (i.e. covering each vertex equally), this suffices in Lemma 3.5.1 since the degree condition ensures that the endvertices of P have a common neighbour in $V_{i,T}$, so there is a cycle C which satisfies $\boldsymbol{a} \cdot \chi(C) \leq \boldsymbol{a} \cdot \chi(T)$. For Theorem 3.1.4 however, the degree condition only guarantees that the endvertices of P have at least $n/k - 2\gamma n$ common neighbours in V_i. Therefore, we can only cover $(k-1)(n/k - 2\gamma n)$ vertices in each part $V_{i,T}$ and $n/k - 2\gamma n$ vertices in each $V_{i,B}$ in this way.

In order to overcome this, we first need to find $2k\gamma n$ cycles of length k which use no vertices from V_B. Again, by considering various cases depending on the structure of G_T, we show that these cycles can always be found unless G is close to extremal. Then for each part V_i, we choose $2\gamma n$ of these cycles which have not yet been chosen and match each one with a vertex from $\{v_{i,n/k+1}, \dots, v_{i,n/k+2\gamma n}\}$ to get sets T'. Obtain T from T' by deleting the vertex which lies in $V_{i,T} \cap T'$. These sets each dominate one of the cycles we have just found. Furthermore, if G' is the graph obtained by deleting these, if we can then find a perfect fractional P_{k-2}-tiling in

$V_T \cap V(G')$, this allows us to proceed as before to find a perfect fractional C_k-tiling in G.

The rest of the section is organised as follows. In Section 3.5.2, we prove Lemma 3.5.2. Then in Section 3.5.2, we discuss what happens when there is no robust pair. Finally in Section 3.5.3, we combine all the results to prove Theorem 3.1.4.

3.5.1 The robust expander case

The aim of this subsection is to prove Lemma 3.5.2. We begin by proving the following lemma which tells us that a perfect matching between each pair $V_{i,T}$ and $V_{i+1,T}$ gives a perfect fractional P_{k-2}-tiling.

Lemma 3.5.3. *Let $k \geq 3$, and $n \in \mathbb{N}$. Let G be a k-partite graph with vertex classes $V_1, \dots, V_k$, each of size n. If there is a perfect matching between V_i and V_{i+1} for every $i \in [k]$ (with indices taken modulo k), then G contains a perfect fractional partition-respecting P_{k-2}-tiling, in which each partition-respecting copy of P_{k-2} has weight 0 or $1/(k-1)$.*

Proof. We prove this directly. Find a perfect matching between V_i and V_{i+1} for each $i \in [k]$. Let H be the subgraph of G which contains only edges from this matching. Let $\mathcal{P}$ be the collection of all partition-respecting copies of P_{k-2} in G. We define the fractional P_{k-2}-tiling to be the function $f\colon \mathcal{P} \to [0,1]$ such that $f(P) = 1/(k-1)$ for each $P \in \mathcal{P}$ such that P lies in H, and $f(P) = 0$ otherwise. Clearly, f is well defined and each path is assigned a weight in $[0,1]$. Note that each vertex v of G is contained in exactly $k-1$ paths in $\mathcal{P}$ which also lie in H, (that is, one path containing v in each possible position along the path). Therefore, $\sum_{P \in \mathcal{P}, v \in P} f(P) = 1$, for each vertex $v \in V(G)$, as required. □

In order to prove Lemma 3.5.2, it would therefore be enough to find a perfect matching between V_i and V_{i+1} for each $i \in [k]$. However, this might not always be possible. For example, $G[V_i, V_{i+1}]$ could be the union of two almost complete biparitite graphs which have parts of unequal size, and therefore, there is no perfect matching between these parts. However, what we can do instead is to first remove a set of paths which, in some sense, balance the parts of G. That is, whenever there is a pair (V_i, V_{i+1}) which is close to the union of two almost complete bipartite graphs with *imbalance* in the part sizes, we remove a collection of paths which correct the imbalance in this pair while not affecting imbalance in any other pair. Therefore, once we have done this for each $i \in [k]$, this leaves us to find a perfect matching in the remaining graph and then apply Lemma 3.5.3. We now prove some lemmas which allow us to correct these imbalances.

3.5.1.1 Correcting imbalances

We begin by stating the following useful proposition.

Proposition 3.5.4. *Let $0 \leq \beta \leq 1/2$ and $n \in \mathbb{N}$. Let S be a set of size n. Suppose $A_1 \cup A_2$ and $B_1 \cup B_2$ are two partitions of S, and $(1/2-\beta)n \leq |A_i|, |B_i| \leq (1/2+\beta)n$ for each $i \in [2]$. Then $||A_i \cap B_j| - |A_{i'} \cap B_{j'}|| \leq 2\beta n$ for each $i, i', j, j' \in [2]$ with $i \neq i'$ and $j \neq j'$.*

Furthermore, if n is even, then for any $i, i', j, j' \in [2]$ with $i \neq i'$ and $j \neq j'$, we have that $||A_i \cap B_j| - |A_{i'} \cap B_{j'}||$ is odd if and only if $||A_{i'} \cap B_j| - |A_i \cap B_{j'}||$ is odd.

Proof. Observe that $|A_i \cap B_j| + |A_i \cap B_{j'}| = |A_i|$ and $|A_i \cap B_j| + |A_{i'} \cap B_j| = |B_j|$ for each $i, i', j, j' \in [2]$ with $i \neq i'$ and $j \neq j'$. Then

$$\begin{aligned} ||A_i \cap B_j| - |A_{i'} \cap B_{j'}|| &= ||A_i \cap B_j| + |A_i \cap B_{j'}| - |A_i \cap B_{j'}| - |A_{i'} \cap B_{j'}|| \\ &= \left||A_i| - \left|B'_j\right|\right| \leq 2\beta n, \end{aligned}$$

as required.

Furthermore, suppose $||A_i \cap B_j| - |A_{i'} \cap B_{j'}||$ is odd, but $||A_{i'} \cap B_j| - |A_i \cap B_{j'}||$ is even. Then $||A_i \cap B_j| - |A_{i'} \cap B_{j'}| + |A_{i'} \cap B_j| - |A_i \cap B_{j'}||$ must be odd, and this is equal to $||B_j| - |B_{j'}||$. Therefore, one of B_j and $B_{j'}$ must have odd size, and the other must have even size. This implies that $|B_j| + |B_{j'}| = n$ must be odd, a contradiction. $\square$

The following lemma says that if $G[V_i, V_{i+1}]$ is a robust pair, then we can use this to show that any vertex which is not in V_i or V_{i+1} is contained in plenty of cycles of length k.

Lemma 3.5.5. *Let $1/n \ll \gamma \ll \nu \ll \tau \ll 1/k$. Let G be a k-partite graph with parts $V_1, \ldots, V_k$ each of size n and $\delta^*(G) \geq (\frac{1}{2} - \gamma)n$. Suppose there is some $i \in [k]$ such that $G[V_i, V_{i+1}]$ is a robust (ν, τ)-expander. Then for every $j \in [k]$ with $j \neq i, i+1$, any vertex v in V_j is contained in at least $(\nu - 2\gamma)n$ transversal cycles which intersect only at the vertex v.*

Proof. By relabelling if necessary, suppose that $G[V_1, V_2]$ is a robust (ν, τ)-expander. Let $v \in V_j$ for some $j \in \{3, \ldots, k\}$. Let $W_{j-1} = N(v) \cap V_{j-1}$ and (unless $j - 1 = 2$) for each $i \in [j-2 \downarrow 2]$, define $W_i = N(W_{i+1}) \cap V_i$. Similarly, let $W_{j+1} = N(v) \cap V_{j+1}$ and (unless $j + 2 = 1$) for each $i \in [j+2 \uparrow k]$, let $W_i = N(W_{i-1}) \cap V_i$. We now define $W_1 = N(W_k) \cap RN(W_2, V_1) \cap V_1$. Note that $|W_i| \geq (1/2 - \gamma)n$ for each $i \in \{2, \ldots, j-2, j+2, \ldots, k\}$, and therefore, $|W_1| \geq (1/2-\gamma)n + (1/2+\nu-\gamma)n - 1 = (\nu - 2\gamma)n$. Therefore, we can greedily choose cycles which begin at v and pass through each W_i, for $i \in \{1, \ldots, j-1, j+1, \ldots, k\}$. $\square$

In the next lemma, we show that for any small set of vertices in G, we can find a collection of partition-respecting paths of length $k - 2$ which cover this set and which hit each vertex class equally.

Lemma 3.5.6. *Let $1/n \ll \gamma, \alpha \ll 1/k$. Let G be a k-partite graph with parts $V_1, \ldots, V_k$ each of size n and $\delta^*(G) \geq (\frac{1}{2} - \gamma)n$. Let X be any set of αn vertices in V_{i^*} for some $i^* \in [k]$. Then we can find a collection of vertex-disjoint partition-respecting paths of length $k - 2$ which covers X and contains exactly $(k-1)|X|$ vertices in each class.*

Proof. Take X to be as above. By the degree condition, each vertex in X has at least $(1/2 - \gamma)n$ neighbours in V_{i^*+1}, and each of these neighbours have at least $(1/2 - \gamma)n$ neighbours in V_{i^*+2} and so on. So, greedily, we find $|X|$ vertex-disjoint paths of length $k-2$ which have initial vertex in X. Now for each $i \in [i^*+1 \uparrow i^*-1]$ in turn, take a set $X_i \subseteq V_i$ with $|X_i| = |X|$ which does not intersect any previously found paths. Then similarly, greedily find a collection of vertex-disjoint paths of length $k-2$ which have initial vertex in X_i and do not intersect any previously found paths. We can always make this greedy choice since we use at most $k|X| \leq k\alpha n$ vertices from each set V_i and therefore there are always at least $(1/2 - \gamma - k\alpha)n$ choices available for the next vertex. Furthermore, each path with initial vertex in X_i intersects every class other than V_{i-1}. Therefore, when this process is finished, there are $(k-1)|X|$ paths which intersect V_i for each $i \in [k]$. $\square$

The next lemma says that we can find plenty of vertex-disjoint paths of length $k-2$ in G which remain in prescribed sets, provided these prescribed sets satisfy certain properties. Since the statement is fairly technical, we first describe the statement more informally. When we think of the bipartite graphs given by $G[V_i, V_j]$, either this pair is a robust (ν, τ)-expander, or it is not. In this lemma, we will assume that any time this is not a robust pair, it gives rise to a high-degree (α, β)-bipartition. Recall from Definition 3.2.4 that this is a partition of the vertex class V_i into V_{i,L_1} and V_{i,L_2} and a partition of the vertex class V_{i+1} into V_{i+1,R_1} and V_{i+1,R_2}, such that each of these subparts has size around $n/2$, and for each $j \in [2]$, there is a high minimum degree in $G[V_{i,L_j}, V_{i+1,R_j}]$. The '$L$' and '$R$' in the subscript here represent whether it is the 'left' or 'right' part of the partition. Indeed, for each $i \in [k]$, if the pair (V_i, V_{i+1}) is a non-(ν, τ)-robust pair, then we say V_i is the *left part* in a non-robust pair, and V_{i+1} is the *right part* in a non-robust pair. Recall that $\mathcal{L}_{\nu,\tau}(G)$ is the set of all indices i such that V_i is the left part in a non-(ν, τ)-robust pair, and $\mathcal{R}_{\nu,\tau}(G)$ is the analogous set for right parts. In general it need not be the case that a non-robust pair has a high-degree (α, β)-bipartition. However, we will see later that we can obtain this by moving vertices between parts and perhaps removing some paths and cycles. Therefore, for the sake of this lemma, assume this is the case.

Whenever a part is both the left part of a non-robust pair, and also the right part of a non-robust pair, we say it is *both parts* in a non-robust pair, and recall that the set of all such parts is $\mathcal{B}_{\nu,\tau}(G)$. When a part lies in $\mathcal{B}_{\nu,\tau}(G)$, it will have two different high-degree (α, β)-bipartitions. These partitions can be well-aligned or not well-aligned. However, we will later see that by removing a collection of cycles and paths of length $k-2$, we can arrive at a situation where either these partitions are perfectly well-aligned or they differ in at least linearly many vertices. Thus, we restrict to this situation in the following lemma. Thus, depending on the number of non-robust pairs that a part V_i belongs to, it will be partitioned into between 0 and 4 subparts, where each subpart has at least linear size, ηn.

Then the lemma says that for any such subpart U_1, there is some sequence of subparts $U_2, U_3, \ldots, U_k$ such that for each $s \in [k]$, there are plenty of vertex-disjoint partition-respecting paths of length $k-2$ which have their j-th vertex in U_{s+j-1} for each $j \in [k-1]$. In other words, if we restrict to $G[U_1, \ldots, U_k]$, then within this graph for each $s \in [k]$, there are at least ηn vertex-disjoint partition-respecting paths of length $k-2$ with initial vertex in U_s (such that these are also partition-respecting paths in G). More formally, we have the following statement.

Lemma 3.5.7. *Let $1/n \ll \gamma \ll \beta \ll \eta \ll \alpha \ll \nu \ll \tau \ll 1/k$. Let G be a k-partite graph with parts $V_1, \ldots, V_k$ each of size n and $\delta^*(G) \geq (\frac{1}{2} - \gamma)n$. For each $i \in \mathcal{L}_{\nu,\tau}(G)$, let $((V_{i,L_1}, V_{i,L_2}), (V_{i+1,R_1}, V_{i+1,R_2}))$ be a high-degree (α, β)-bipartition of (V_i, V_{i+1}). For each $i \in \mathcal{B}_{\nu,\tau}(G)$, for $j, \ell \in [2]$, let $V_{i,L_j,R_\ell} = V_{i,L_j} \cap V_{i,R_\ell}$ and suppose that for each $j', \ell' \in [2]$ with $j \neq j'$ and $\ell \neq \ell'$, either $\left|V_{i,L_j,R_\ell}\right| = \left|V_{i,L_{j'},R_{\ell'}}\right| = 0$, or $\left|V_{i,L_j,R_\ell}\right|, \left|V_{i,L_{j'},R_{\ell'}}\right| \geq \eta n$. For each $i \in \mathcal{L}_{\nu,\tau}(G) \setminus \mathcal{B}_{\nu,\tau}(G)$, for $j, j' \in [2]$ with $j \neq j'$, let $V_{i,L_j,R_j} = V_{i,L_j}$ and let $V_{i,L_j,R_{j'}} = \emptyset$. For each $i \in \mathcal{R}_{\nu,\tau}(G) \setminus \mathcal{B}_{\nu,\tau}(G)$, for $\ell, \ell' \in [2]$ with $\ell \neq \ell'$, let $V_{i,L_\ell,R_\ell} = V_{i,R_\ell}$ and let $V_{i,L_\ell,R_{\ell'}} = \emptyset$.*

Let $r \in \mathcal{L}_{\nu,\tau}(G) \cup \mathcal{R}_{\nu,\tau}(G)$, and let $\diamond, \circ \in [2]$ be such that $V_{i,L_\diamond,R_\circ} \neq \emptyset$. Let

$\mathcal{U} = \{U_1, \dots, U_k\}$ *be a collection of sets satisfying the following.*

(i) $U_r = V_{i,L_\diamond,R_\circ}$.

(ii) *For each* $i \in \mathcal{L}_{\nu,\tau}(G) \cup \mathcal{R}_{\nu,\tau}(G)$, $U_i = V_{i,L_j,R_\ell}$ *for some* $j, \ell \in [2]$.

(iii) *For each* $i \in [k] \setminus (\mathcal{L}_{\nu,\tau}(G) \cup \mathcal{R}_{\nu,\tau}(G))$, $U_i = V_i$.

(iv) *For each* $i \in \mathcal{L}_{\nu,\tau}(G)$, *each vertex in* U_i *has at most* αn *non-neighbours in* U_{i+1} *and each vertex in* U_{i+1} *has at most* αn *non-neighbours in* U_i.

(v) *For each* $i \notin \mathcal{L}_{\nu,\tau}(G)$, $|RN_\nu(U_i) \cap U_{i+1}| \geq (\nu - 2\beta)n$ *and* $|RN_\nu(U_{i+1}) \cap U_i| \geq (\nu - 2\beta)n$.

(vi) *For each* $i \in \mathcal{B}_{\nu,\tau}(G) \setminus \{r\}$, $|U_i| \geq n/8$. *If* $r \in \mathcal{B}_{\nu,\tau}(G)$ *then* $|U_r| \geq \eta n$.

Then for any $s \in [k]$, *we can find at least* ηn *vertex-disjoint paths of length* $k-2$ *with initial vertex in* U_s *and which lie entirely in* U_i *for* $i \in [s \uparrow s-2]$.

Proof. To prove this, we first find subsets $W_i^{(s)} \subseteq U_i$ which satisfy some useful properties. We then use these to find the paths that we need. Fix $s \in [k]$ as above. For the remainder of the the proof, we will omit the superscript, and simply write W_i to mean $W_i^{(s)}$. There are two cases to consider, where the approach is almost the same in both cases, with a distinction depending on whether or not we require paths passing through U_r. We will also drop the subscripts ν, τ whenever they are clear from context.

Case 1: $s \neq r+1$.

Let $W_r = U_r$. For each $i \in [r-1 \downarrow s]$, suppose we have found W_{i+1}, and define W_i as follows. If $i \in \mathcal{L}(G)$, let $W_i = U_i$. Else, define $W_i = RN_\nu(W_{i+1}) \cap U_i$. Similarly, for each $i \in [r+1 \uparrow s-2]$, suppose we have found W_{i-1}. Define W_i as follows. If $i \in \mathcal{R}(G)$, then let $W_i = U_i$. Otherwise let $W_i = RN_\nu(W_{i-1}) \cap U_i$. In the next series of claims, we will prove certain properties satisfied by these sets W_i which will help us find the paths we need. In the first of these, we will show that the sizes of these sets W_i are sufficiently large. Note that in each of the claims in this proof, we are split into two cases depending on whether $i \in \{s \uparrow r-1\}$ or $i \in \{r \uparrow s-2\}$. We will usually only discuss the first of these, and the second will follow by the same method after making the necessary changes, namely by replacing "$i+1$" by "$i-1$" and swapping the terms $\mathcal{L}(G)$ and $\mathcal{R}(G)$. In this case, we say that the second case follows by symmetry.

Claim 3.5.8. *The following hold:*

(A) $|W_r| \geq \eta n$. *Furthermore, if* $r \notin \mathcal{B}(G)$, *then* $|W_r| \geq (1/2-\beta)n$.

(B) *For each* $i \in \{s \uparrow r-1\}$*:*

  (i) *if* $i \in \mathcal{B}(G)$, *then* $|W_i| \geq n/8$,

  (ii) *if* $i \in \mathcal{L}(G) \setminus \mathcal{B}(G)$, *then* $|W_i| \geq (1/2-\beta)n$,

  (iii) *if* $i \in \mathcal{R}(G) \setminus \mathcal{B}(G)$, *then* $|W_i| \geq (\nu-2\beta)n$,

  (iv) *if* $i \in [k] \setminus (\mathcal{L}(G) \cup \mathcal{R}(G))$, *then* $|W_i| \geq (1/2+\nu-\beta)n$.

(C) *For each* $i \in \{r+1 \uparrow s-2\}$*:*

  (i) *if* $i \in \mathcal{B}(G)$, *then* $|W_i| \geq n/8$,

  (ii) *if* $i \in \mathcal{R}(G) \setminus \mathcal{B}(G)$, *then* $|W_i| \geq (1/2-\beta)n$,

  (iii) *if* $i \in \mathcal{L}(G) \setminus \mathcal{B}(G)$, *then* $|W_i| \geq (\nu-2\beta)n$,

  (iv) *if* $i \in [k] \setminus (\mathcal{L}(G) \cup \mathcal{R}(G))$, *then* $|W_i| \geq (1/2+\nu-\beta)n$.

Proof of Claim 3.5.8. First, note that $W_r = U_r = V_{r,L_j,R_\ell}$ for some $j, \ell \in \{1,2\}$. Therefore, $|W_i| \geq \eta n$ by the hypotheses of the lemma. Furthermore, if $r \notin \mathcal{B}(G)$, then $W_r = U_r = V_{r,L_j,R_j}$ for some $j \in \{1,2\}$, and $\left|V_{r,L_j,R_j}\right| \geq (1/2-\beta)n$. Thus, this proves (A).

We now prove (B) by induction. For each $i \in [r-1 \downarrow s]$, suppose the claim holds for each $i' \in \{r \downarrow i+1\}$. First, if $i \in \mathcal{B}(G)$, then $W_i = U_i = V_{i,L_{j_1},R_{\ell_1}}$ for some

$j_1, \ell_1 \in \{1,2\}$, where for $j_2, \ell_2 \in \{1,2\}$, $W_{i+1} = V_{i,L_{j_2},R_{j_1}}$ and $\left|V_{i,L_{j_1},R_{\ell_1}}\right| \geq \left|V_{i,L_{j_1},R_{\ell_2}}\right|$. In particular, as $\left|V_{i,L_{j_1},R_{\ell_1}}\right| + \left|V_{i,L_{j_1},R_{\ell_2}}\right| \geq (1/2-\beta)n$, we know $|W_i| \geq (1/2-\beta)n/2 \geq n/8$, proving (B) (i).

If $i \in \mathcal{L}(G) \setminus \mathcal{B}(G)$, then $W_i = U_i = V_{i,L_{j_1},R_{j_1}}$ for some $j_1 \in [2]$. Furthermore, since $i \notin \mathcal{B}(G)$, we know that $V_{i,L_{j_1},R_{j_2}} = V_{i,L_{j_2},R_{j_1}} = \emptyset$ for $j_2 \in [2]$ with $j_2 \neq j_1$. Therefore, $|W_i| \geq (1/2-\beta)n$, proving (B) (ii).

If $i \in \mathcal{R}(G) \setminus \mathcal{B}(G)$, then as before, $|U_i| \geq (1/2-\beta)n$. Also, $i+1 \in (\mathcal{L}(G) \setminus \mathcal{B}(G)) \cup ([k] \setminus (\mathcal{L}(G) \cup \mathcal{R}(G)))$. Thus by the induction hypothesis, $|W_{i+1}| \geq (1/2-\beta)n$. Since $G[V_i, V_{i+1}]$ is a robust (ν,τ)-expander, $|RN_\nu(W_{i+1}) \cap V_i| \geq |W_{i+1}| + \nu n \geq (1/2-\beta+\nu)n$. Therefore, $|W_i| = |RN_\nu(W_{i+1}) \cap U_i| \geq (1/2-\beta+\nu)n + (1/2-\beta)n - n = (\nu - 2\beta)n$, proving (B) (iii).

Finally, if $i \in [k] \setminus (\mathcal{L}(G) \cup \mathcal{R}(G))$, note first that $U_i = V_i$. Furthermore, $i+1 \in (\mathcal{L}(G) \setminus \mathcal{B}(G)) \cup ([k] \setminus (\mathcal{L}(G) \cup \mathcal{R}(G)))$, so by the same argument as in (B) (iii), $|W_i| = |RN_\nu(W_{i+1}) \cap U_i| = |RN_\nu(W_{i+1}) \cap V_i| \geq (1/2-\beta+\nu)n$, which proves (B) (iv), and therefore concludes the proof of (B).

The proof for (C) follows similarly to the proof for (B) by symmetry. $\square$

Generally speaking, in the following two claims, we show that each set W_i satisfies one of two properties, involving sending a small matching into a neighbouring set. More precisely, focusing on $i \in \{s \uparrow r-1\}$, in the first claim, we show that some of these sets W_i satisfy the property that any subset of W_i of the appropriate size 'sends a perfect matching' into its neighbouring set W_{i+1}. In the second claim, we show that the remaining sets W_i satisfy the property that if we take two subsets of the appropriate size of W_i and W_{i+1} respectively, then we can find a matching between these which covers one of these subsets. We now formalise this.

Claim 3.5.9. *The following holds.*

(A) *For each $i \in \{s \uparrow r-1\}$, if $i \notin \mathcal{L}(G)$, then for any set $X \subseteq W_i$ with $|X| \leq$*

$(\nu - 2\beta)n$*, there is a set* $Y \subseteq W_{i+1}$ *with* $|Y| = |X|$ *such that there is a perfect matching between* X *and* Y*.*

(B) *For each* $i \in \{r+1 \uparrow s-2\}$*, if* $i \notin \mathcal{R}(G)$*, then for any set* $X \subseteq W_i$ *with* $|X| \leq (\nu - 2\beta)n$*, there is a set* $Y \subseteq W_{i-1}$ *with* $|Y| = |X|$ *such that there is a perfect matching between* X *and* Y*.*

Proof of Claim 3.5.9. First, fix some $i \in \{s \uparrow r-1\}$ such that $i \notin \mathcal{L}(G)$, and let $X \subseteq W_i$ with $|X| \leq (\nu - 2\beta)n$. Then $W_i \subseteq RN_\nu(W_{i+1})$, so in particular, each vertex in W_i has at least νn neighbours in W_{i+1}. So, fix an arbitrary ordering of the vertices of X. Then in turn for each $v \in X$, choose a $u_v \in N(v) \cap W_{i+1}$ which has not yet been chosen, and take $Y = \bigcup_{v \in X}\{u_v\}$. We can always do this since by Claim 3.5.8 (A) and (B), for each $i \in \{s \uparrow r-1\}$, whenever $i \notin \mathcal{L}(G)$, $|W_{i+1}| \geq (1/2 - \beta)n \geq (\nu - 2\beta)n$. This proves Claim 3.5.9 (A). The proof of Claim 3.5.9 (B) follows by symmetry. □

Claim 3.5.10. *The following holds.*

(A) *For any* $i \in \{s \uparrow r-1\}$ *such that* $i \in \mathcal{L}(G)$*, for any set* $X \subseteq W_i$ *and* $Y \subseteq W_{i+1}$ *such that* $|X| \geq |Y| + \alpha n$*, there is a matching between* X *and* Y *which covers* Y*.*

(B) *For any* $i \in \{r+1 \uparrow s-2\}$ *such that* $i \in \mathcal{R}(G)$*, for any set* $X \subseteq W_i$ *and* $Y \subseteq W_{i-1}$ *such that* $|X| \geq |Y| + \alpha n$*, there is a matching between* X *and* Y *which covers* Y*.*

Proof of Claim 3.5.10. Fix some $i \in \{s \uparrow r-1\}$ such that $i \in \mathcal{L}(G)$. Let $X \subseteq W_i$ and $Y \subseteq W_{i+1}$ with $|X| \geq |Y| + \alpha n$. Recall that $W_i = U_i = V_{i,L_{j_1},R_{\ell_1}}$ and $W_{i+1} \subseteq U_{i+1} = V_{i+1,L_{j_2},R_{j_1}}$ for $j_1, j_2, \ell_1 \in \{1,2\}$. Each vertex of W_{i+1} has at most $\alpha \left|V_{i,L_{j_1}}\right| \leq \alpha n$ non-neighbours in $V_{i,L_{j_1}}$, and therefore also at most αn non-neighbours in any subset of $V_{i,L_{j_1}}$, in particular, X. Thus we fix an arbitrary ordering of the vertices in Y. Then for each vertex $v \in Y$ in turn, greedily select an unchosen vertex $u_v \in N(v) \cap X$ to find a matching between X and Y which covers Y. It only remains to check that we can always choose X and Y as suggested. Indeed we can do this since by Claim 3.5.8, we know that $|W_i| \geq \alpha n$ for each $i \in \{s \uparrow r-1\}$, and this proves (A). The proof of (B) follows by symmetry. $\square$

Thus, we have shown that each set W_i satisfies one of two properties, each of which involve finding matchings from fixed subsets of W_i. In the next claim, we show that we can tie these matchings together and start building paths. Before stating the next claim, we require the following definitions. We say that $i \in [k]$ is a *turning point* if either $i \in \{s \uparrow r-1\}$ and $i \in \mathcal{R}(G)$, or $i \in \{r+1 \uparrow s-2\}$, and $i \in \mathcal{L}(G)$. We say that i and i' are *consecutive turning points* if i, i' are both turning points, and

- either: $i, i' \in \{s \uparrow r-1\}$, and i appears before i' in the ordering given by $[s \uparrow r-1]$, and for each $i'' \in \{i \uparrow i'\} \setminus \{i, i'\}$, i'' is not a turning point,
- or: $i, i' \in \{r+1 \uparrow s-2\}$, and i appears before i' in the ordering given by $[s-2 \downarrow r+1]$, and for each $i'' \in \{i \downarrow i'\} \setminus \{i, i'\}$, i'' is not a turning point.

Claim 3.5.11. *The following holds.*

(A) *If t and t' are consecutive turning points in $\{s \uparrow r-1\}$, then between any set $X_t \subseteq W_t$ and $X_{t'} \subseteq W_{t'}$ with $\alpha n \leq |X_t| = |X_{t'}| \leq (\nu - 2\beta)n$, we can find $|X_t| - \alpha n$ vertex-disjoint partition-respecting paths which remain in W_i for $i \in \{t \uparrow t'\}$.*

(B) *If t and t' are consecutive turning points in $\{s-2 \downarrow r+1\}$, then between any set $X_t \subseteq W_t$ and $X_{t'} \subseteq W_{t'}$ with $\alpha n \leq |X_t| = |X_{t'}| \leq (\nu - 2\beta)n$, we can find $|X_t| - \alpha n$ vertex-disjoint partition-respecting paths which remain in W_i for $i \in \{t \uparrow t'\}$.*

Proof of Claim 3.5.11. Let t and t' be consecutive turning points in $\{s \uparrow r-1\}$. Then $t, t' \in \mathcal{R}(G)$, but for each $i \in \{t \uparrow t'\} \setminus \{t, t'\}$, $i \notin \mathcal{R}(G)$. Take a set $X_t \subseteq W_t$

and a set $X_{t'} \subseteq W_{t'}$ with $\alpha n \leq |X_t| = |X_{t'}| \leq (\nu - 2\beta)n$. This is possible since by Claim 3.5.8, $|W_i| \geq (\nu - 2\beta)n$ for each $i \in \{s \uparrow r-1\}$.

First, assume $t' > t+1$. Now for each $i \in [t+1 \uparrow t'-1]$ in turn, we find sets $X_i \subseteq W_i$ such that there is a perfect matching between X_{i-1} and X_i as follows. Suppose we have found X_{i-1}. Since t and t' are consecutive turning points, we know that $i \notin \mathcal{R}(G)$. Therefore, it follows that $i-1 \notin \mathcal{L}(G)$. So by Claim 3.5.9, we can find a set $X_i \subseteq W_i$ with $|X_i| = |X_{i-1}|$ such that there is a perfect matching between X_{i-1} and X_i. Again, this is possible since $|W_i| \geq (\nu - 2\beta)n$ for each $i \in \{s \uparrow r-1\}$. By repeating this, we find a set $X_{t'-1}$ such that by combining the matchings we have found, we find $|X_t|$ vertex-disjoint paths between X_t and X_{t-1}. Now, since $t' \in \mathcal{R}(G)$, $t'-1 \in \mathcal{L}(G)$. Let $X'_{t'} \subseteq X_{t'}$ be such that $|X'_{t'}| = |X_t| - \alpha n$. Then by Claim 3.5.10, we can find a matching between $X_{t'-1}$ and $X'_{t'}$ which covers $X'_{t'}$. Indeed, by combining the matchings once again, we obtain $|X_t| - \alpha n$ vertex-disjoint paths between X_t and X_{t-1} which remain in W_i for $i \in \{t \uparrow t'\}$.

Now assume $t' = t+1$. Let $X'_{t'} \subseteq X_{t'}$ with $|X'_{t'}| = |X_t| - \alpha n$. Since $t' \in \mathcal{R}(G)$, we know that $t \in \mathcal{L}(G)$. So by Claim 3.5.10, there is a matching between X_t and $X'_{t'}$ which covers $X'_{t'}$. In other words, there are $|X_t| - \alpha n$ vertex-disjoint paths between X_t and $X_{t'}$ which remain in W_i for $i \in \{t \uparrow t'\}$. This proves (A). The proof of (B), follows by symmetry. $\square$

It only remains now to combine the claims above to prove Case 1. We actually show that there are at least ηn vertex-disjoint paths with initial vertex in W_s and which remain in W_i for $i \in [k] \setminus \{s-1\}$. This suffices since $W_i \subseteq U_i$ for each $i \in [k] \setminus \{s-1\}$. Broadly speaking, the approach is to tie together the paths we found in Claim 3.5.11 and then finish these off if necessary by using Claim 3.5.9 and Claim 3.5.10. More precisely, in (I), for each $i \in [s \uparrow r-1]$ in turn, we want to find a set $X_i \subseteq W_i$ such that there is a large matching between X_{i-1} and X_i. Then in (II), for each $i \in [s-2 \downarrow r+1]$, we want to find a set $X_i \subseteq W_i$ such that there is a large matching between X_i and X_{i+1}. Then finally in (III) we want to tie these together by finding a set $X_r \subseteq W_r$ such that there is a large matching between X_r and X_{r-1}, and also a large matching between X_r and X_{r+1}. By tying all these matchings together we obtain many vertex-disjoint partition-respecting paths of length $k-2$ which have initial vertex in W_s.

(I) If $s = r$, then proceed to (II). Otherwise, let $X_s \subseteq W_s$ such that $|X_s| = (\nu - 2\beta)n$. This is possible by Claim 3.5.8. If $s = r-1$, proceed to (II). Otherwise, let ℓ_1 be the number of turning points in $\{s+1 \uparrow r-1\}$, and suppose first $\ell_1 = 0$. Therefore, for each $i \in \{s+1 \uparrow r-1\}$, $i \notin \mathcal{R}(G)$, and so $i-1 \notin \mathcal{L}(G)$. For each $i \in [s+1 \uparrow r-1]$, suppose we have found $X_{i-1} \subseteq W_{i-1}$ of size $(\nu - 2\beta)n$. Then by Claim 3.5.9, we can find a set $X_i \subseteq W_i$ such that $|X_i| = |X_{i-1}| = (\nu - 2\beta)n$ and there is a perfect matching between X_{i-1} and X_i, and by following these matchings along, we find $(\nu - 2\beta)n$ vertex-disjoint paths between X_s and X_{r-1}.

On the other hand, suppose $\ell_1 \geq 1$. Let $T = (t_1, \ldots, t_{\ell_1})$ be the list of these turning points such that each turning point in $\{s+1 \uparrow r-1\}$ appears in T exactly once, and such that for each $i' \in [\ell_1 - 1]$, $t_{i'}$ and $t_{i'+1}$ are consecutive turning points in $\{s+1 \uparrow r-1\}$.

Then if $s+1 = t_1$, take a subset $X_{t_1} \subseteq W_{t_1}$ of size $(\nu - 2\beta - \alpha)n$. Since $s+1 = t_1$, we know $s \in \mathcal{L}(G)$. So by Claim 3.5.10 we can find a matching between X_s and X_{t_1} which covers X_{t_1}. Else, if $s+1 \neq t_1$, then for each $i \in \{s+1 \uparrow t_1 - 1\}$, $i \notin \mathcal{R}(G)$, and so $i-1 \notin \mathcal{L}(G)$. So suppose for each $i \in [s+1 \uparrow t_1 - 1]$, we have found X_{i-1} of size $(\nu - 2\beta)n$. Then by Claim 3.5.9, we find a set $X_i \subseteq W_i$ of size $(\nu - 2\beta)n$ such that there is a perfect matching between X_{i-1} and X_i. Then as $t_1 \in \mathcal{R}(G)$, we know $t_1 - 1 \in \mathcal{L}(G)$. So, by Claim 3.5.10, we can find a set $X_{t_1} \subseteq W_{t_1}$ of size $(\nu - 2\beta - \alpha)n$ such that there is a matching between X_{t_1-1} and X_{t_1} which covers X_{t_1}.

If $\ell_1 = 1$, then we now have $(\nu - 2\beta - \alpha)n$ vertex-disjoint paths between X_s

and $X_{t_{\ell_1}}$. Otherwise, for each $i \in [2 \uparrow \ell_1]$, suppose we have found $X_{t_{i-1}}$ with $\left|X_{t_{i-1}}\right| = (\nu - 2\beta - (i-1)\alpha)n$, and take a subset $X_{t_i} \subseteq W_{t_i}$ with $|X_{t_1}| = (\nu - 2\beta - i\alpha)n$. Then by Claim 3.5.11, we can find $(\nu - 2\beta - i\alpha)n$ vertex-disjoint partition-respecting paths between $X_{t_{i-1}}$ and X_{t_i} which remain in W_j. Thus, for any value of ℓ_1, we have found $(\nu - 2\beta - \ell_1\alpha)n$ vertex-disjoint partition-respecting paths between X_s and $X_{t_{\ell_1}}$.

Then finally, if $t_{\ell_1} = r-1$, proceed to (II). Otherwise, for each $i \in \{t_{\ell_1}+1 \uparrow r-1\}$, note that $i \notin \mathcal{R}(G)$, and so $i-1 \notin \mathcal{L}(G)$. Therefore, for each $i \in [t_{\ell_1}+1 \uparrow r-1]$, suppose we have found $X_{i-1} \subseteq W_i$ of size $(\nu - 2\beta - \ell_1\alpha)n$. By Claim 3.5.9, we find $X_i \subseteq W_i$ of size $(\nu - 2\beta - \ell_1\alpha)n$ such that there is a perfect matching between X_{i-1} and X_i. By following this along, we find a set $X_{r-1} \subseteq W_{r-1}$ of size $(\nu - 2\beta - \ell_1\alpha)n \geq (\nu - k\alpha)n$ such that there are $(\nu - 2\beta - \ell_1\alpha)n$ vertex-disjoint paths between X_s and X_{r-1} which remain in W_i for $i \in \{s \uparrow r-1\}$.

(II) If $s - 2 = r$, then we proceed to (III). Otherwise, by symmetry, this step will follow similarly to (I), and we can find $(\nu - k\alpha)n$ vertex-disjoint paths between X_{s-2} and X_{r+1} which remain in W_i for $i \in \{s-2 \downarrow r+1\}$.

(III) Now to finish off, suppose first $r = s$. If $r \in \mathcal{L}(G)$, take a subset $X_r \subseteq W_r$ with $|X_r| = \eta n$. Then as $|X_r| \leq |X_{r+1}| - \alpha n$, we can apply Claim 3.5.10 to find a matching between X_r and X_{r+1} which covers X_r. Instead, if $r \notin \mathcal{L}(G)$, then we can apply Claim 3.5.9 to find a set $X_r \subseteq W_r$ such that there is a perfect matching between X_r and X_{r+1}. On each occasion, we have found at least ηn vertex-disjoint paths between X_r and X_{s-2} which lie in W_i for $i \in \{r \uparrow s-2\}$.

The case when $r = s-2$ is symmetrical, and we find at least ηn vertex-disjoint partition-respecting paths between X_s and X_r which lie in W_i for $i \in \{s \uparrow r\}$.

Now suppose $r \neq s$ and $r \neq s-2$. We know that $r \in \mathcal{L}(G) \cup \mathcal{R}(G)$. If $r \in \mathcal{B}(G)$, take a set $X_r \subseteq W_r$ such that $|X_r| = \eta n$. Then as $r \in \mathcal{R}(G)$, $r-1 \in \mathcal{L}(G)$. So by Claim 3.5.10, as $|X_r| \leq |X_{r-1}| - \alpha n$, we can find a matching between X_{r-1} and X_r which covers X_r. Also, as $r \in \mathcal{L}(G)$, $r+1 \in \mathcal{R}(G)$, and so again by

Claim 3.5.10, as $|X_r| \leq |X_{r+1}| - \alpha n$, we can find a matching between X_r and X_{r+1} which covers X_r. By combining these, we find ηn vertex-disjoint paths between X_s and X_{s-2} which lie in W_i for $i \in \{s \uparrow s-2\}$.

Instead, if $r \in \mathcal{L}(G) \setminus \mathcal{B}(G)$, then $r-1 \notin \mathcal{L}(G)$. So, by Claim 3.5.9, there is a set $X_r' \subseteq W_r$ with $|X_r'| = |X_{r-1}|$ such that there is a perfect matching between X_r' and X_{r-1}. Then take $X_r \subseteq X_r'$ with $|X_r| = \eta n$ (this is possible as $r \notin \mathcal{B}(G)$, and so $|W_r| \geq (1/2 - \beta)n$ by Claim 3.5.8). As $|X_r| \leq |X_{r+1}| - 1$, and since $r+1 \in \mathcal{R}(G)$, by Claim 3.5.10, we can find a matching between X_r and X_{r+1} which covers X_r. Thus, we obtain ηn vertex-disjoint paths between X_s and X_{s-2} which lie in W_i for $i \in \{s \uparrow s-2\}$. The case when $r \in \mathcal{R}(G) \setminus \mathcal{B}(G)$, is symmetrical and this concludes the proof of Case 1.

Case 2: $s = r + 1$.

The proof in this case is similar to the proof of Case 1, but a little easier. The slight difference is in how we find the sets W_i. These sets will then satisfy exactly the same properties as in Claims 3.5.8, 3.5.9 and 3.5.10, and we can use these properties to find paths exactly as in Case 1. Due to the similarities, we omit the latter part of the proof, and simply highlight how we find the sets $W_i^{(s)}$ (we omit the superscript for ease of notation).

Begin by letting $W_s = U_s$. Then for each $i \in [s+1 \uparrow s-2]$, suppose we have found W_{i-1}. If $i \in \mathcal{L}(G)$, let $W_i = U_i$. Otherwise, let $W_i = RN_\nu(W_{i-1}) \cap U_i$.

Thus, in each case, we find collection of at least ηn vertex-disjoint paths of length $k-2$ with initial vertex in W_s which lie in W_i for $i \in \{s \uparrow s-2\}$, which concludes the proof of the lemma. $\square$

Before stating and proving the next lemma, we give the following definition.

Definition 3.5.12. *Let G be a k-partite graph with vertex-classes $V_1, \ldots, V_k$, each of size n. For $i \in [k]$, suppose that $V_i^L = V_{i,L_1} \cup V_{i,L_2}$ and $V_i^R = V_{i,R_1} \cup V_{i,R_2}$ are two* *partitions of the vertex class V_i. Then we define the* joint partition $J(V_i^L, V_i^R)$ *to be the partition with vertex classes* $V_{i,L_j,R_\ell} \coloneqq V_{i,L_j} \cap V_{i,R_\ell}$ *for each* $j, \ell \in [2]$*. For each* $j, j', \ell, \ell' \in [2]$ *with* $j \neq j'$ *and* $\ell \neq \ell'$*, define the following.*

(i) *Say that* V_{i,L_1} *and* V_{i,L_2} *are* opposite subparts of V_i^L*. Similarly, say that* V_{i,R_1} *and* V_{i,R_2} *are* opposite subparts of V_i^R.

(ii) *Say that* V_{i,L_j,R_ℓ} *and* $V_{i,L_{j'},R_{\ell'}}$ *are* diagonal subparts of $J(V_i^L, V_i^R)$.

(iii) *Define* $\mathrm{diff}_{i,L}^G \coloneqq ||V_{i,L_1}| - |V_{i,L_2}||$ *and* $\mathrm{diff}_{i,R}^G \coloneqq ||V_{i,R_1}| - |V_{i,R_2}||$.

(iv) *Define* $\mathrm{diff}_{i,j,\ell}^G \coloneqq \left|\left|V_{i,L_j,R_\ell}\right| - \left|V_{i,L_{j'},R_{\ell'}}\right|\right|$.

(v) *Define* $\mathrm{diff}_i^G(J(V_i^L, V_i^R)) = \max\{\mathrm{diff}_{i,j,\ell}^G, \mathrm{diff}_{i,j',\ell}^G\}$.

Recall that $\mathcal{N}_{\nu,\tau}(G)$ is the set of all indices $i \in [k]$ such that (V_i, V_{i+1}) is a (ν,τ)-robust pair in G and $\mathcal{B}_{\nu,\tau}$ is the set of indices $i \in [k]$ such that neither (V_i, V_{i+1}) nor (V_{i-1}, V_i) is a (ν,τ)-robust pair in G. Lemma 3.2.5 tells us that whenever a pair (V_i, V_{i+1}) is not a (ν,τ)-robust pair, we can find an (α,β)-bipartition of it. We will see later that we can modify this slightly so that this is a high-degree (α,β)-bipartition, and furthermore, whenever a part is partitioned twice by this, the resulting joint partition contains at least linear sized subparts.

Given these assumptions, the next lemma tells us that if there is some $r \in \mathcal{L}(G) \cup \mathcal{R}(G)$ such that

- if $r \in \mathcal{B}(G)$ and the difference in size between some pair of diagonal subparts in V_r is at least 2, or
- if $r \notin \mathcal{B}(G)$ and the difference in size between opposite subparts of V_r is at least 2,

then we can find a constant sized collection of paths of length $k-2$ such that if we remove this collection, then within the resulting graph the aforementioned difference in size is reduced by exactly 2. Meanwhile, for every other pair of opposite or diagonal subparts, the difference in size in the original graph is the same as the difference in size in the new graph. Thus, the next lemma says that we can reduce the imbalance in r by 2 while maintaining the imbalance elsewhere.

Lemma 3.5.13. *Let $1/n \ll \gamma \ll \beta \ll \eta \ll \alpha \ll \nu \ll \tau \ll 1/k$. Let G be a k-partite graph with parts $V_1, \dots, V_k$ each of size n and $\delta^*(G) \geq (1/2 - \gamma)n$. Let $i^* \in [k]$ be such that $i^* \in \mathcal{N}_{\nu,\tau}(G)$. For each $i \notin \mathcal{N}_{\nu,\tau}(G)$, let $((V_{i,L_1}, V_{i,L_2}), (V_{i+1,R_1}, V_{i+1,R_2}))$ be a high-degree (α, β)-bipartition of (V_i, V_{i+1}). For each $i \in \mathcal{B}(G)$, for $j, \ell \in [2]$, let $V_{i,L_j,R_\ell} = V_{i,L_j} \cap V_{i,R_\ell}$ and suppose that for each $j', \ell' \in [2]$ with $j \neq j'$ and $\ell \neq \ell'$, either $\left|V_{i,L_j,R_\ell}\right| = \left|V_{i,L_{j'},R_{\ell'}}\right| = 0$, or $\left|V_{i,L_j,R_\ell}\right|, \left|V_{i,L_{j'},R_{\ell'}}\right| \geq \eta n$. For each $i \in \mathcal{L}(G)$, let $\mathrm{diff}^G_{i,L}$ and $\mathrm{diff}^G_{i+1,R}$ be as in Definition 3.5.12. For each $i \in \mathcal{B}(G)$ and $j, \ell \in [2]$, let $\mathrm{diff}^G_{i,j,\ell}$ be as in Definition 3.5.12.*

Suppose there exist $r \in [k]$ and $\diamond, \diamond', \circ, \circ' \in \{1, 2\}$ with $\diamond \neq \diamond'$ and $\circ \neq \circ'$ such that $\mathrm{diff}^G_{r,\diamond,\circ} \geq 2$. Then we can find a collection $\mathcal{Q}$ of vertex-disjoint partition-respecting paths of length $k-2$ in G such that $|V_i \cap V(\mathcal{Q})| = 2k-2$ for each $i \in [k]$, and if $G' = G - \mathcal{Q}$, then

(i) $\mathrm{diff}^{G'}_{r,\diamond,\circ} = \mathrm{diff}^G_{r,\diamond,\circ} - 2 = \mathrm{diff}^G_{r,\diamond',\circ'} - 2 = \mathrm{diff}^{G'}_{r,\diamond',\circ'}$,

(ii) *for each $i, j, \ell \in (\mathcal{B}(G) \times [2] \times [2]) \setminus \{(r, \diamond, \circ), (r, \diamond', \circ')\}$, we have* $\mathrm{diff}^{G'}_{i,j,\ell} = \mathrm{diff}^G_{i,j,\ell}$,

(iii) *for each $i \in \mathcal{L}(G) \setminus \mathcal{B}(G)$,* $\mathrm{diff}^{G'}_{i,L} = \mathrm{diff}^G_{i,L}$, *and*

(iv) *for each $i \in \mathcal{R}(G) \setminus \mathcal{B}(G)$,* $\mathrm{diff}^{G'}_{i,R} = \mathrm{diff}^G_{i,R}$.

Proof. For each $i \in \mathcal{L}(G) \setminus \mathcal{B}(G)$, for $j, j' \in [2]$ with $j \neq j'$, let $V_{i,L_j,R_j} = V_{i,L_j}$ and let $V_{i,L_j,R_{j'}} = \emptyset$. For each $i \in \mathcal{R}(G) \setminus \mathcal{B}(G)$, for $\ell, \ell' \in [2]$ with $\ell \neq \ell'$, let $V_{i,L_\ell,R_\ell} = V_{i,R_\ell}$ and let $V_{i,L_\ell,R_{\ell'}} = \emptyset$. Let $i^*, r, \diamond, \diamond', \circ, \circ'$ be as in the lemma statement. The proof will proceed as follows. We begin by finding some sets $U_{i,j}$ which satisfy the conditions in Lemma 3.5.7 to find many paths of length $k-2$ within these sets $U_{i,j}$. We then choose some paths to satisfy the desired properties. We are choosing a constant number of paths among linearly many, so we can do this greedily. We begin by defining the sets $U_{i,j}$.

If $|V_{r,L_\diamond,R_\circ}| - \left|V_{r,L_{\diamond'},R_{\circ'}}\right| > 0$, let $U_{r,1} = V_{r,L_\diamond,R_\circ}$ and $U_{r,2} = V_{r,L_{\diamond'},R_{\circ'}}$. Else, let $U_{r,1} = V_{r,L_{\diamond'},R_{\circ'}}$ and $U_{r,2} = V_{r,L_\diamond,R_\circ}$. Essentially, $U_{r,1}$ is the larger of the subparts in question, and $U_{r,2}$ is the smaller one. Note that $|U_{r,1}| \geq |U_{r,2}| + 2$. Now we find sets $U_{i,j}$ for each $i \in [k]$ and $j \in [2]$, which will guide the paths we eventually find. We define these iteratively as follows.

Unless $r = i^* + 1$, for each $i \in [r-1 \downarrow i^*+1]$, suppose we have found $U_{i+1,j}$ for $j \in \{1,2\}$. If $i+1 \in \mathcal{L}(G) \cup \mathcal{R}(G)$ then assume that $U_{i+1,1} = V_{i+1,L_{j_1},R_{\ell_1}}$ for some $j_1, \ell_1 \in [2]$, and assume that $U_{i+1,2} = V_{i+1,L_{j_2},R_{\ell_2}}$ for $j_2, \ell_2 \in [2]$ with $j_1 \neq j_2$ and $\ell_1 \neq \ell_2$. Else, if $i+1 \in [k] \setminus (\mathcal{L}(G) \cup \mathcal{R}(G))$, then assume $U_{i+1,j} = V_{i+1}$ for each $j \in \{1,2\}$. Then find $U_{i,j}$ as follows.

- If $i \in \mathcal{L}(G)$, then let $U_{i,1} = V_{i,L_{\ell_1},R_{\ell_3}}$ where $\ell_3 \in \{1,2\}$ satisfies that for $\ell_4 \in [2]$ with $\ell_3 \neq \ell_4$, $\left|V_{i,L_{\ell_1},R_{\ell_3}}\right| \geq \left|V_{i,L_{\ell_1},R_{\ell_4}}\right|$. Then set $U_{i,2} = V_{i,L_{\ell_2},R_{\ell_4}}$. Thus in this case, $U_{i,j}$ and $U_{i+1,j}$ have a high minimum degree between them for each $j \in [2]$.

- If $i \in \mathcal{R}(G) \setminus \mathcal{B}(G)$, then take $U_{i,1} = V_{i,L_{j_3},R_{j_3}}$ where $j_3 \in \{1,2\}$. Then set $U_{i,2} = V_{i,L_{j_4},R_{j_4}}$ with $j_4 \in \{1,2\}$ so that $j_3 \neq j_4$. In particular, $U_{i,1}, U_{i,2} \neq \emptyset$.

- If $i \in [k] \setminus (\mathcal{L}(G) \cup \mathcal{R}(G))$, then take $U_{i,1} = U_{i,2} = V_i$.

Thus we obtain $U_{i,j}$ for each $i \in \{i^*+1 \uparrow r\}$ and each $j \in [2]$. Similarly, unless $r = i^*$, we find $U_{i,j}$ for each $i \in [r+1 \uparrow i^*]$ and $j \in [2]$ as follows. Suppose we have found $U_{i-1,j}$ for $j \in \{1,2\}$. If $i-1 \in \mathcal{L}(G) \cup \mathcal{R}(G)$ then assume that $U_{i-1,1} = V_{i-1,L_{j_1},R_{\ell_1}}$ for some $j_1, \ell_1 \in [2]$, and assume that $U_{i-1,2} = V_{i-1,L_{j_2},R_{\ell_2}}$ for $j_2, \ell_2 \in [2]$ with $j_1 \neq j_2$ and $\ell_1 \neq \ell_2$. Else, if $i-1 \in [k] \setminus (\mathcal{L}(G) \cup \mathcal{R}(G))$, then assume $U_{i-1,j} = V_{i-1}$ for each $j \in \{1,2\}$. Then find $U_{i,j}$ as follows.

- If $i \in \mathcal{R}(G)$, then let $U_{i,1} = V_{i,L_{j_3},R_{j_1}}$, where $j_3 \in [2]$ is such that for $j_4 \in [2]$ with $j_3 \neq j_4$, $\left|V_{i,L_{j_3},R_{j_1}}\right| \geq \left|V_{i,L_{j_4},R_{j_1}}\right|$. Then set $U_{i,2} = V_{i,L_{j_4},R_{j_2}}$.

- If $i \in \mathcal{L}(G) \setminus \mathcal{B}(G)$, then let $U_{i,1} = V_{i,L_{j_3},R_{j_3}}$ for $j_3 \in [2]$ and let $U_{i,2} = V_{i,j_4,j_4}$ for $j_4 \in [2]$ with $j_3 \neq j_4$.
- If $i \in [k] \setminus (\mathcal{L}(G) \cup \mathcal{R}(G))$, then let $U_{i,1} = U_{i,2} = V_i$.

Thus, we have found $U_{i,j}$ for each $i \in [k]$, $j \in [2]$. Let $\mathcal{U}_j = \{U_{1,j}, \dots, U_{k,j}\}$. Note that each $U_{i,j}$ is some subpart in the partition of V_i. Furthermore, for each $j \in [2]$ the collection $\mathcal{U}_j$ satisfies the conditions of Lemma 3.5.7, and so for any $s \in [k]$, we can find a collection $\mathcal{Q}_{s,j}$ of ηn vertex-disjoint paths of length $k-2$ which have initial vertex in $U_{s,j}$ and which remain in the sets $U_{i,j}$ for $i \in \{s \uparrow s-2\}$.

In the remainder of the proof, we begin by choosing one special path, and then using the paths in $\mathcal{Q}_{i,j}$ to choose a collection of vertex-disjoint paths which satisfy the conclusions of the lemma. The first step is to find a path of length $k-2$ which has initial vertex in $U_{i^*,1}$ and each other vertex lies in $U_{i,2}$ for $i \in \{i^*+1 \uparrow i^*-2\}$. We do this in a similar way to how we found the paths in Lemma 3.5.7. We first find sets W'_i which will guide where this path lies. First, if $r = i^*-1$, then let $W'_1 = U_{i^*,1}$. For each $i \in [i^*+1 \uparrow i^*-2]$, suppose we have found W'_{i-1}. If $i \in \mathcal{R}(G)$, let $W'_i = U_{i,2}$. Otherwise, let $W'_i = RN_\nu(W'_{i-1}) \cap U_{i,2}$. If instead $r \neq i^*-1$, then let $W'_r = U_{r,2}$. Then unless $r = i^*+1$, for each $i \in [r-1 \downarrow i^*+1]$, let $W'_i = U_{i,2}$ if $i \in \mathcal{L}(G)$, and let $W'_i = RN_\nu(W'_{i+1}) \cap U_{i,2}$ if $i \notin \mathcal{L}(G)$. Similarly, for each $i \in [r-1 \uparrow i^*-2]$, unless $r = i^*-2$, let $W'_i = U_{i,2}$ if $i \in \mathcal{R}(G)$, and let $W'_i = RN_\nu(W_{i-1}) \cap U_{i,2}$ if $i \notin \mathcal{R}(G)$. Then let $W'_1 = RN_\nu(W_2) \cap U_{1,1}$. We then show that we can find a path of length $k-2$ which lies in W'_i for $i \in [k-1]$, that is, the following claim holds.

Claim 3.5.14. *We can find a path of length $k-2$ with initial vertex in W'_1 such that each vertex of this path lies in W'_i for $i \in \{1 \uparrow k-1\}$.*

We omit the proof of this claim, but note that this follows in exactly the same way as the proof of Lemma 3.5.7.

Let P^* be the path found in Claim 3.5.14, and add P^* to the collection $\mathcal{Q}$. Now we choose vertex-disjoint paths, to form the rest of the collection $\mathcal{Q}$, and show that this satisfies the conclusions of the lemma. We say that a path in $\mathcal{Q}$ *hits* $U_{i,j}$ if there is a vertex of the path which belongs to $U_{i,j}$. Since these paths are partition-respecting, each path in $\mathcal{Q}$ can only hit $U_{i,j}$ at most once. For each $i \in [k]$ and $j \in [2]$, let $H(i,j)$ be a function counting the number of paths in $\mathcal{Q}$ which hit $U_{i,j}$.

Claim 3.5.15. *In order to show that removing $\mathcal{Q}$ only reduces the difference between parts by 2 in V_r and leaves the difference unchanged elsewhere, it suffices to show that $H(r,1) = H(r,2) + 2$ and for each $i \in [k] \setminus \{r\}$, $H(i,1) = H(i,2)$.*

Proof of Claim 3.5.15. Let $G' = G - \mathcal{Q}$, and for each $i \in [k]$ and $j \in [2]$, let $U'_{i,j} = U_{i,j} \cap V(G')$. Suppose $H(r,1) = H(r,2) + 2$ and for each $i \in [k] \setminus \{r\}$, $H(i,1) = H(i,2)$. For any $i \in [k]$ and $j, j', \ell, \ell' \in [2]$ such that $j \neq j'$ and $\ell \neq \ell'$, if there is no $s \in [2]$ such that $U_{i,s} = V_{i,j,\ell}$, then there is no $s' \in [2]$ such that $U_{i,s'} = V_{i,j',\ell'}$. Therefore, $\text{diff}^{G'}_{i,j,\ell} = \text{diff}^{G}_{i,j,\ell}$, since the paths in $\mathcal{Q}$ are only chosen from sets $U_{i',t}$ for $i' \in [k]$ and $t \in [2]$.

On the other hand for any $i \in [k]$, $j, j', \ell, \ell' \in [2]$ with $j \neq j'$ and $\ell \neq \ell'$, such that $V_{i,j.\ell} = U_{i,s}$ and $V_{i,j',\ell'} = U_{i,s'}$ for some $s, s' \in [2]$ with $s \neq s'$, we have

$$\begin{aligned}\text{diff}^{G'}_{i,j,\ell} &= ||V_{i,j,\ell} \cap V(G')| - |V_{i,j',\ell'} \cap V(G')|| = ||U_{i,1} \cap V(G')| - |U_{i,2} \cap V(G')|| \\ &= |(|U_{i,1}| - H(i,1)) - (|U_{i,2}| - H(i,2))| = \text{diff}^{G}_{i,j,\ell} - (H(i,1) - H(i,2)).\end{aligned}$$

Therefore, $\text{diff}^{G'}_{r,\diamond,\circ} = \text{diff}^{G}_{r,\diamond,\circ} - (H(r,1) - H(r,2)) = \text{diff}^{G}_{r,\diamond,\circ} - 2$. Meanwhile for $i \in [k] \setminus \{r\}$, $j, j', \ell, \ell' \in [2]$ with $j \neq j'$ and $\ell \neq \ell'$, such that $V_{i,j,\ell} = U_{i,s}$ and $V_{i,j',\ell'} = U_{i,s'}$ for some $s, s' \in [2]$ with $s \neq s'$, we have $\text{diff}^{G'}_{i,j,\ell} = \text{diff}^{G}_{i,j,\ell}$, as required. □

We now find the collection $\mathcal{Q}$ and show that indeed, $H(r,1) = H(r,2) + 2$ and for each $i \in [k] \setminus \{r\}$, $H(i,1) = H(i,2)$, which will conclude the proof. Recall that $P^* \in \mathcal{Q}$. We add $2k - 3$ more paths to $\mathcal{Q}$ in total, one at a time, each time making sure that we choose only paths whose vertices are unused, and so that the resulting collection $\mathcal{Q}$ is a collection of vertex-disjoint paths. Note that we can do this since we will choose paths from $\mathcal{Q}_{i,j}$ for some $i \in [k]$ and $j \in [2]$, and each such collection satisfies $|\mathcal{Q}_{i,j}| \geq \eta n \geq 2k^2$. For technical reasons, we split this into three cases to consider, but the ideas in each are essentially the same. In each case below, we assume that any new paths chosen are vertex-disjoint from any paths already in the collection $\mathcal{Q}$.

Case A: $r = i^*$ For each $i \in [i^* + 1 \uparrow i^* - 1]$, for each $j \in [1 \uparrow 2]$ choose a path $Q_{i,j} \in \mathcal{Q}_{i,j}$ and add this to $\mathcal{Q}$. Then add a path $Q_{i^*,1} \in \mathcal{Q}_{i^*,1}$ to $\mathcal{Q}$. Now consider $H(i^*, 1)$. We note that for $i \in [k] \setminus \{i^* + 1\}$, the path $Q_{i,1}$ hits $U_{i^*,1}$. Also, P^* hits $U_{i^*,1}$. As these are the only paths to hit $U_{i^*,1}$, we know $H(i^*, 1) = k$. Meanwhile, for each $i \in [k] \setminus \{i^*, i^* + 1\}$, the path $Q_{i,2}$ hits $U_{i^*,2}$, and these are the only paths in $\mathcal{Q}$ which hit $U_{i^*,2}$, so $H(i^*, 2) = k - 2$. Now for any $i \in [k] \setminus \{i^*\}$, for each $s \in [k] \setminus \{i + 1\}$, the path $Q_{s,1}$ hits $U_{i,1}$ and these are the only paths to hit $U_{i,1}$, so $H(i, 1) = k - 1$. On the other hand, for any $i \in [k] \setminus \{1\}$, for each $s \in [k] \setminus \{1, i + 1\}$, the path $Q_{s,1}$ hits $U_{i,2}$ and so does the path P^*, and these are the only paths to hit $U_{i,2}$. So, $H(i, 2) = k - 1$. Thus, $H(i^*, 1) = H(i^*, 2) + 2$, and $H(i, 1) = H(i, j)$ for each $i \in [k] \setminus \{i^*\}$, so by Claim 3.5.15, we are done.

Case B: $r = i^* - 1$ Choose two distinct paths $Q_{i^*+1,1}$ and $Q^*_{i^*+1,1}$ in $\mathcal{Q}_{i^*+1,1}$ and add these to $\mathcal{Q}$. Then for each $i \in [k] \setminus \{i^*, i^* + 1\}$ choose a previously unchosen path $Q_{i,1}$ in $\mathcal{Q}_{i,1}$ and add this to $\mathcal{Q}$. For each $i \in [k] \setminus \{i^* + 1\}$, choose a path $Q_{i,2}$ in $\mathcal{Q}_{i,2}$ and add this to $\mathcal{Q}$.

Then $H(i^* - 1, 1) = k$ and $H(i^* - 1, 2) = k - 2$. Indeed, this is because the paths hitting $U_{i^*-1,1}$ are the paths $Q_{i,1}$ for $i \in [k] \setminus \{i^*\}$, together with the path $Q^*_{i^*+1,1}$. On the other hand, the paths hitting $U_{i^*-1,2}$ are the paths $Q_{i,2}$ for $i \in [k] \setminus \{i^*, i^* + 1\}$. Meanwhile, for each $i \in [k] \setminus \{i^* - 1\}$, $H(i, 1) = H(i, 2) = k - 1$. Indeed, the paths hitting $U_{i^*,1}$ are the paths $Q_{i,1}$ for $i \in [k] \setminus \{i^*, i^* + 1\}$ together with the path P^*. The paths hitting $U_{i^*,2}$ are the paths $Q_{i,2}$ for $i \in [k] \setminus \{i^* + 1\}$. This proves $H(i^*, 1) = H(i^*, 2) = k - 1$.

For $i \in [k] \setminus \{i^*-1, i^*\}$, the paths hitting $U_{i,1}$ are the paths $Q_{i',1}$ for $i' \in [k] \setminus \{i, i^*-1\}$ together with the path $Q^*_{i^*,1}$. On the other hand, the paths hitting $U_{i,2}$ are the paths $Q_{i',2}$ for $i' \in [k] \setminus \{i^*, i\}$ together with the path P^*. Therefore, $H(i,1) = H(i,2) = k-1$.

Case C: $r \notin \{i^*-1, i^*\}$ Choose two distinct paths $Q_{i^*+1,1}$ and $Q^*_{i^*+1,1}$ in $\mathcal{Q}_{i^*+1,1}$ and add these to $\mathcal{Q}$. Also, choose two distinct paths $Q_{r+1,2}$ and $Q^*_{r+1,2}$ and add these to $\mathcal{Q}$. Then for each $i \in [k] \setminus \{i^*+1, r+1\}$, choose an unselected path $Q_{i,1}$ in $\mathcal{Q}_{i,1}$ and add this to $\mathcal{Q}$. Then for each $i \in [k] \setminus \{i^*, i^*+1, r+1\}$, choose an unselected path $Q_{i,2}$ in $\mathcal{Q}_{i,2}$ and add this to $\mathcal{Q}$.

Then $H(r,1) = k$, while $H(r,2) = k-2$. This is because for each $i \in [k] \setminus \{r+1\}$, the paths $Q_{i,1}$ hit $U_{r,1}$, and so does the path $Q^*_{i^*+1,1}$. Meanwhile, for $i \in [k] \setminus \{i^*, i^*+1, r+1\}$, the paths $Q_{i,2}$ hit $U_{r,2}$, and so does the path P^*, a total of $k-2$ paths. Meanwhile, $H(i,1) = H(i,2) = k-1$ for each $i \in [k] \setminus \{r\}$. For $H_{i^*,1}$, the paths that hit $U_{i^*,1}$ are the paths $Q_{i',1}$ for $i' \in [k] \setminus \{i^*+1, r+1\}$, together with the path P^*. For $H(i^*,2)$, the paths hitting $U_{1,2}$ are the paths $Q_{i,2}$ for $i \in [k] \setminus \{i^*, i^*+1\}$ together with the path $Q^*_{r+1,2}$.

For each $i \in [k] \setminus \{i^*, r\}$, the paths that hit $U_{i,1}$ are the paths $Q_{i',1}$ for $i' \in [k] \setminus \{i+1, r+1\}$, together with the path $Q^*_{i^*+1,1}$. Meanwhile, the paths that hit $U_{i,2}$ are the paths $Q_{i',2}$ for $i' \in [k] \setminus \{i^*, i^*+1, i+1\}$, together with the paths P^* and $Q^*_{r+1,2}$. Therefore, $H(i,1) = H(i,2) = k-1$, concluding the proof. □

In the next lemma, we now apply Lemma 3.5.13 repeatedly, and remove a collection of paths of length $k-2$ such that in the resulting graph G', each pair of classes which was robust in G is still robust in G' (albeit with different constants), and each pair of classes which was not robust in G now has a bipartition into two almost-complete balanced bipartite graphs.

Lemma 3.5.16. *Let $1/n \ll \gamma \ll \beta \ll \eta \ll \alpha \ll \nu \ll \tau \ll 1/k$. Let G be a k-partite graph with parts $V_1, \ldots, V_k$ each of size n and $\delta^*(G) \geq (1/2-\gamma)n$. Let $i^* \in [k]$*

be such that $G[V_{i^*} \cup V_{i^*+1}]$ is a robust (ν,τ)-expander. For each $i \notin \mathcal{N}_{\nu,\tau}(G)$, let $((V_{i,L_1}, V_{i,L_2}), (V_{i+1,R_1}, V_{i+1,R_2}))$ be a high-degree (α,β)-bipartition of (V_i, V_{i+1}). For each $i \in \mathcal{B}_{\nu,\tau}(G)$, for each $j,\ell \in [2]$, let $\left|V_{i,L_j} \cap V_{i,R_\ell}\right| \geq \eta n$.

Then we can find a collection $\mathcal{Q}$ of vertex-disjoint paths of length $k-2$ and cycles of length k in G such that if $H = G - \mathcal{Q}$, then $H[(V_{i^*} \cup V_{i^*+1}) \cap V(H)]$ is a robust $(\nu/2,\tau)$-expander, and for each $i \notin \mathcal{N}_{\nu/2,\tau}(H)$, there is an $(2\alpha, 0)$-bipartition of $(V_i \cap V(H), V_{i+1} \cap V(H))$.

Proof. Let G be as above. By relabelling if necessary, assume $1 \in \mathcal{N}(G)$, so $G[V_1, V_2]$ is a robust (ν,τ)-expander. Now for each $i \in \mathcal{B}_{\nu,\tau}(G)$, for each $j, j', \ell, \ell' \in [2]$ with $j \neq j'$ and $\ell \neq \ell'$, let $V_{i,L_j,R_\ell} = V_{i,L_j} \cap V_{i,R_\ell}$. For $i \in \mathcal{L}(G) \setminus \mathcal{B}(G)$ and $j, j' \in [2]$ with $j \neq j'$, let $V_{i,L_j,R_j} = V_{i,L_j}$ and $V_{i,L_j,R_{j'}} = \emptyset$. For $i \in \mathcal{R}(G) \setminus \mathcal{B}(G)$, for $\ell, \ell' \in [2]$, let $V_{i,L_\ell,R_\ell} = V_{i,R_\ell}$ and let $V_{i,L_\ell,R_{\ell'}} = \emptyset$. Now, for each $i \in \mathcal{L}(G)$, let $\mathrm{diff}^G_{i,L}$ be as in Definition 3.5.12. For each $i \in \mathcal{R}(G)$, let $\mathrm{diff}^G_{i,R}$ be as in Definition 3.5.12. For each $i \in \mathcal{L}(G) \cup \mathcal{R}(G)$, for $j, j', \ell, \ell' \in [2]$ with $j \neq j'$ and $\ell \neq \ell'$, we let $\mathrm{diff}^G_{i,j,\ell}$ be as in Definition 3.5.12. In each case, we omit the graph from the superscript when it is clear from context.

We would like to remove paths and cycles to obtain a graph where the difference between the sizes of opposite subparts is zero everywhere. We claim that it suffices to make sure that the difference between the sizes of each pair of diagonal subparts is zero. Indeed, if this is the case, then for each $i \in \mathcal{L}(G)$, and each $j, j', \ell, \ell' \in [2]$ with $j \neq j'$ and $\ell \neq \ell'$, we have that $\left|V_{i,L_j,R_\ell}\right| = \left|V_{i,L_{j'},R_{\ell'}}\right|$. Then $|V_{i,L_1}| = |V_{i,L_1,R_1}| + |V_{i,L_1,R_2}| = |V_{i,L_2,R_2}| + |V_{i,L_2,R_1}| = |V_{i,L_2}|$, so $\mathrm{diff}_{i,L} = 0$. Similarly, we can show that for $i \in \mathcal{R}(G)$, if $\mathrm{diff}_{i,j,\ell} = 0$ for each $j, \ell \in [2]$, then $\mathrm{diff}_{i,R} = 0$. Therefore, in this proof, we will be working towards making $\mathrm{diff}_{i,j,\ell} = 0$ for all i, j, ℓ for which the quantity is defined.

If n is not even, then let C_0 be a cycle of length k in G, which exists by Lemma 3.5.5 and add this to the collection $\mathcal{Q}$. Remove this cycle from G to get G', so $G' = G - C_0$. Then let $n' := |V_i \cap V(G')| = n - 1$, and so n' is even. Furthermore, we can choose constants $\gamma', \alpha', \beta', \eta', \nu'$ such that $\gamma \ll \gamma' \ll \beta \ll \beta' \ll \eta' \ll \eta \ll \alpha \ll \alpha' \ll \nu' \ll \nu$

and G' satisfies the hypotheses of the lemma with these constants. If n is even, then let $G' = G$ and again, n' is even and G' satisfies all the hypotheses of the lemma with the constants above. For ease of notation, we assume n is even.

Now we want to make sure the difference in size between diagonal subparts is even for each part, since we have Lemma 3.5.13 which reduces the difference by 2 for a single part. Notice that as n is even, we know that for each $i \in (\mathcal{L}(G) \cup \mathcal{R}(G)) \setminus \mathcal{B}(G)$, $\mathrm{diff}_{i,j,\ell}$ is even for each $j, \ell \in [2]$. This is because when n is even, then if we split any set of size n into two subsets of sizes a and b respectively, then either both a and b are even or they are both odd, and in either case, $|a - b|$ is even.

Thus, the only case when a pair of diagonal subparts could have odd difference is when $i \in \mathcal{B}(G)$. By Proposition 3.5.4, $\mathrm{diff}_{i,j,\ell}$ is odd if and only if $\mathrm{diff}_{i,j,\ell'}$ is also odd for $j, \ell, \ell' \in [2]$ with $\ell \neq \ell'$. Let $Odd(G) \coloneqq \{i \in [k] \colon \text{there exist } j, \ell \in [2] \text{ such that } \mathrm{diff}_{i,j,\ell} \text{ is odd.}\}$. Suppose $|Odd(G)| = b$ and label the elements of $Odd(G) = \{i_1, \ldots, i_b\}$. For each $s \in [b]$, we will find vertex-disjoint cycles of length k $C_{s,1}$ and $C_{s,2}$ and a sequence of subgraphs $G_0 = G \supset G_1 \supset \ldots \supset G_b$ so that $Odd(G_b) = \emptyset$. Choose constants $\gamma_i, \beta_i, \eta_i, \alpha_i, \nu_i$ for each $i \in [b]$ such that $\gamma \ll \gamma_1 \ll \ldots \ll \gamma_b \ll \beta \ll \beta_1 \ll \ldots \ll \beta_b \ll \eta_b \ll \ldots \ll \eta_1 \ll \eta$, $\alpha \ll \alpha_1 \ll \ldots \ll \alpha_b \ll \nu_b \ll \ldots \ll \nu_1 \ll \nu$. Suppose for some $s \in [b]$, that we have found G_{s-1} and that G_{s-1} with $n_{s-1} \coloneqq |V_i \cap V(G_{s-1})| = n - 2(s-1)$, and that G_{s-1} satisfies the hypotheses in the lemma statement with parameters $\gamma_{s-1}, \beta_{s-1}, \eta_{s-1}, \alpha_{s-1}, \nu_{s-1}, \tau$. Suppose further that $Odd(G_{s-1}) \subset Odd(G)$ and $i_1, \ldots, i_{s-1} \notin Odd(G_{s-1})$.

Then let $V^{s-1}_{i,L_j,R_\ell} = V_{i,L_j,R_\ell} \cap G_{s-1}$. Again for simplicity of notation, we omit the $s-1$ in the superscript below. Define $U_{i_s,1} = V_{i_s,L_1,R_1}$ and $U_{i_s,2} = V_{i_s,L_2,R_1}$. Then for each $i \in [i_s - 1 \downarrow 2]$, suppose we have found $U_{i+1,1}$, and that if $i+1 \in \mathcal{L}(G) \cup \mathcal{R}(G)$, let $U_{i+1,1} = V_{i+1,L_{j_1},R_{\ell_1}}$ for $j_1, \ell_1 \in [2]$. Else, let $U_{i+1,1} = V_i$. Then

- if $i \in \mathcal{L}(G)$, then let $U_{i,1} = V_{i,L_{\ell_1},R_{\ell_2}}$ where $\ell_2 \in [2]$ is such that for $\ell_3 \in [2]$ with $\ell_2 \neq \ell_3$, $\left|V_{i,L_{\ell_1},R_{\ell_2}}\right| \geq \left|V_{i,L_{\ell_1},R_{\ell_3}}\right|$,
- if $i \in \mathcal{R}(G) \setminus \mathcal{B}(G)$, let $U_{i,1} = V_{i,L_{j_2},R_{j_2}}$ for any $j_2 \in [2]$,

- if $i \in [k] \setminus (\mathcal{L}(G) \cup \mathcal{R}(G))$, let $U_{i,1} = V_i$.

Similarly, for each $i \in [i_s + 1 \uparrow 1]$, suppose we have found $U_{i-1,1}$ and that if $i-1 \in \mathcal{L}(G) \cup \mathcal{R}(G)$, let $U_{i-1,1} = V_{i-1,L_{j_1},R_{\ell_1}}$ for $j_1, \ell_1 \in [2]$. Else, let $U_{i-1,1} = V_{i-1}$. Then

- if $i \in \mathcal{R}(G)$, then let $U_{i,1} = V_{i,L_{j_2},R_{j_1}}$ where $j_2 \in [2]$ is such that for $j_3 \in [2]$ with $j_2 \neq j_3$, $\left|V_{i,L_{j_2},R_{j_1}}\right| \geq \left|V_{i,L_{j_3},R_{j_1}}\right|$,
- if $i \in \mathcal{L}(G) \setminus \mathcal{B}(G)$, let $U_{i,1} = V_{i,L_{\ell_2},R_{\ell_2}}$ for any $\ell_2 \in [2]$,
- if $i \in [k] \setminus (\mathcal{L}(G) \cup \mathcal{R}(G))$, let $U_{i,1} = V_i$.

Then define $U_{i,2}$ to be the diagonal part to $U_{i,1}$ for each $i \in [k] \setminus \{i_s\}$. Then by Lemma 3.5.7 we know there is a collection $\mathcal{R}$ of $(\nu_{s-1} - 2\alpha_{s-1})n_{s-1}$ vertex-disjoint paths between $U_{i_s+1,1}$ and $U_{i_s-1,1}$ which lie in $U_{i,1}$ for $i \in [i_s + 1 \uparrow i_s - 1]$. Let $X_i = U_{i,1} \cap \mathcal{R}$. Then any vertex $v \in U_{i_s,1}$ has at most $\alpha_{s-1}n_{s-1}$ non-neighbours in X_{i_s-1} and at most $\alpha_{s-1}n_{s-1}$ non-neighbours in X_{i_s+1}, and therefore we can certainly find $(\nu - 4\alpha)n$ cycles containing v. We can do the same thing for any $w \in U_{i_s,2}$, and so in particular, we can find two vertex-disjoint cycles $C_{s,1}$ and $C_{s,2}$ such that for each $j \in [2]$, $C_{s,j}$ lies entirely in $U_{i,j}$ for $i \in [k]$. Furthermore, let $G_s = G_{s-1} - C_{s,1} - C_{s,2}$. Then $\mathrm{diff}^{G_s}_{i_s,1,1} = \mathrm{diff}^{G_{s-1}}_{i_s,1,1} - 1$ and $\mathrm{diff}^{G_s}_{i_s,1,2} = \mathrm{diff}^{G_{s-1}}_{i_s,1,2} - 1$, so as both of these quantities were odd, they are now even, and so in particular, $i_s \notin \mathit{Odd}(G)$. Furthermore, for each $i \in [k] \setminus \{i_s\}$, if $j, \ell \in [2]$ are such that $V_{i,L_j,R_\ell} = U_{i,1}$ and $V_{i,L_j,R_\ell} = U_{i,2}$, then $\mathrm{diff}^{G_s}_{i,j,\ell} = \mathrm{diff}^{G_{s-1}}_{i,j,\ell} - 2$. If $j, j', \ell, \ell' \in [2]$ with $j \neq j'$ and $\ell \neq \ell'$ are such that $V_{i,L_j,R_\ell} = U_{i,1}$ and $V_{i,L_{j'},R_{\ell'}} = U_{i,2}$, then $\mathrm{diff}^{G_s}_{i,j,\ell} = \mathrm{diff}^{G_{s-1}}_{i,j,\ell}$ and for each other $j, \ell \in [2]$, $\mathrm{diff}^{G_s}_{i,j,\ell} = \mathrm{diff}^{G_{s-1}}_{i,j,\ell}$. Therefore, the parity of the difference between the diagonal parts becomes even for i_s, and the difference remains unchanged for each other $i \in [k]$. Therefore, $\mathit{Odd}(G_s) = \mathit{Odd}(G_{s-1}) \setminus \{i_s\}$. Furthermore, G_s satisfies the conditions in the lemma with constants $\gamma_s, \beta_s, \eta_s, \alpha_s, \nu_s, \tau$.

Let $\mathcal{Q}^{(1)} = \bigcup_{i \in [b]} (C_{i,1} \cup C_{i,2})$. Let $H = G_b$. Then $H = G - \mathcal{Q}^{(1)}$. Now we know that the difference between each pair of diagonal parts has even parity. We will now repeatedly apply Lemma 3.5.13 to reduce this to 0 for each pair of diagonal parts. In particular, we find a sequence of subgraphs $H_0 = H \supset H_1 \supset \ldots \supset H_k$ as follows. Again, choose constants such that $\gamma_b \ll \gamma'_1 \ll \ldots \ll \gamma'_k \ll \beta_b \ll \beta'_1 \ll \ldots \ll \beta'_k \ll \eta'_k \ll \ldots \ll \eta'_1 \ll \eta_b \ll \alpha_b \ll \alpha'_1 \ll \ldots \ll \alpha'_k \ll \nu'_k \ll \ldots \ll \nu'_1 \ll \nu_b$.

For each $i \in [k]$, if $i \notin \mathcal{L}(G) \cup \mathcal{R}(G)$, then let $H_i = H_{i-1}$. If $\mathrm{diff}_{i,j,\ell}^{H_{i-1}} = 0$ for each $j, \ell \in [2]$, then let $H_i = H_{i-1}$. Otherwise, if there is some $j, \ell \in [2]$ such that $\mathrm{diff}_{i,j,\ell}^{H_{i-1}} > 0$, then in particular we know $\mathrm{diff}_{i,j,\ell}^{H_{i-1}} \geq 2$ as the parity is even. Then for each such j, ℓ, apply Lemma 3.5.13 $\mathrm{diff}_{i,j,\ell}^{H_{i-1}}/2$ times to find a collection of vertex-disjoint paths $\mathcal{P}_i$ such that if $H_i = H_{i-1} - \mathcal{P}_i$, then H_i satisfies the conditions in the lemma with constants $\gamma'_i, \beta'_i, \eta'_i, \alpha'_i, \nu'_i, \tau$. Furthermore, for each $j, j', \ell, \ell' \in [2]$ with $j \neq j'$, $\mathrm{diff}_{i,j,\ell}^{H_i} = 0 = \mathrm{diff}_{i,j',\ell}^{H_i}$, while $\mathrm{diff}_{i',j'',\ell''}^{H_i} = k_{i',j'',\ell''}^{H_{i-1}}$ for each $i' \in [k] \setminus \{i\}$, $j'', \ell'' \in [2]$.

Let $\mathcal{Q}$ be the collection of all paths and cycles found during this process. Note that during this process, we remove the same number of vertices from V_i for each $i \in [k]$, and in particular at most $\beta'_k n$ vertices. Therefore by Lemma 3.2.6, the resulting graph H satisfies that any pair which was in $\mathcal{N}_{\nu,\tau}(G)$ is now in $\mathcal{N}_{\nu/2,\tau}(H)$ and for each $i \notin \mathcal{N}_{\nu/2,\tau}(H)$, there is an $(2\alpha, 0)$-bipartition of (V_i^H, V_{i+1}^H), as required. □

The next lemma now allows us to take any graph in which each pair is either (ν, τ)-robust or it has an $(\alpha, 0)$-bipartition and use this to find a perfect fractional P_{k-2}-tiling.

Lemma 3.5.17. *Let $1/n \ll \gamma \ll \alpha \ll \nu \ll \tau \ll 1/k$. Let G be k-partite graph with parts $V_1, \ldots, V_k$ each of size n and $\delta^*(G) \geq (1/2 - \gamma)n$. Let $i^* \in [k]$ be such that (V_{i^*}, V_{i^*+1}) is a robust (ν, τ)-expander. For each $i \in [k]$ with $i \notin \mathcal{N}_{\nu,\tau}(G)$, let $((V_{i,L_1}, V_{i,L_2}), (V_{i+1,R_1}, V_{i+1,R_2}))$ be an $(\alpha, 0)$-bipartition of (V_i, V_{i+1}). Then G contains a perfect fractional P_{k-2}-tiling in which each copy of P_{k-2} has weight 0 or $1/(k-1)$.*

Proof. For each $i \in [k]$, we do the following. If (V_i, V_{i+1}) is a robust (ν, τ)-expander, then use Lemma 3.2.3 to find a perfect matching between V_i and V_{i+1}. Otherwise, we know there is an $(\alpha, 0)$-bipartition $((V_{i,L_1}, V_{i,L_2}), (V_{i+1,R_1}, V_{i+1,R_2}))$ of (V_i, V_{i+1}). By definition, this means that the partition of V_i into $V_{i,L_1} \cup V_{i,L_2}$ and of V_{i+1} into $V_{i+1,R_1} \cup V_{i+1,R_2}$ is such that

(i) $\left|V_{i,L_j}\right| = |V_{i+1,R_\ell}| = n/2$ for each $j, \ell \in [2]$, and

(ii) for each $j \in [2]$, every vertex in V_{i,L_j} has at least $(1-\alpha)n/2$ neighbours in V_{i+1,R_j} and vice versa.

We can verify that Hall's condition is satisfied and find a perfect matching between V_{i,L_j} and V_{i+1,R_j} for each $j \in [2]$, and therefore we can find a perfect matching between V_i and V_{i+1}. Thus we find perfect matchings between V_i and V_{i+1} for each $i \in [k]$. Therefore, we can apply Lemma 3.5.3 to find a perfect fractional P_{k-2}-tiling in G in which each copy of P_{k-2} has weight 0 or $1/(k-1)$. $\square$

3.5.1.2 Proof of Lemma 3.5.2

We can now apply the results we have seen to prove Lemma 3.5.2. Recall that we want to show that, given that there is some pair of parts (V_{i^*}, V_{i^*+1}) in G which is a (ν, τ)-robust pair, we can find a perfect fractional P_{k-2}-tiling in G. The proof of this result proceeds as follows. First for each pair (V_i, V_{i+1}) which is not robust, we find an (α, β)-bipartition as given by Lemma 3.2.5, that is, a split of (V_i, V_{i+1}) into two subgraphs A and B, each of which are close to being complete bipartite graphs, though there may be some vertices with low degree in A and B, and similarly there may be some slight imbalances in the part sizes. The first step is to find paths which hit each V_i evenly and which cover any of the vertices which have low degree where they should see almost everything. We can do this by applying Lemma 3.5.6. This will leave only vertices which see almost everything where they should.

Now suppose that there are some i such that both (V_{i-1}, V_i) and (V_i, V_{i+1}) are non-robust pairs. In this case, V_i will be partitioned twice, into 4 subparts. The second step is to cover any subparts which have small size. We do this by finding paths of length $k-2$ which cover this. Again, this can be done by applying Lemma 3.5.6. This will now leave only vertices which see almost everything where they should, and only large subparts. Furthermore, by doing this carefully, we can ensure that the difference in size between any pair of diagonal subparts is small compared to the sizes of these parts.

Then we are in a position to apply Lemma 3.5.16 which gives a collection of paths of length $k-2$ and cycles of length k that we can remove in order to make the graph 'balanced', that is, to make it so that whenever there is a pair (V_i, V_{i+1}) which is not a (ν', τ)-robust pair, there is an $(\alpha', 0)$-bipartition of this pair. In particular, the guaranteed robust pair is used for 'balancing' in Lemma 3.5.16. To the resulting graph, we then apply Lemma 3.5.17 to find a perfect fractional P_{k-2}-tiling in the remaining graph. The final step is to combine all the paths and cycles we have found in previous steps to obtain a fractional P_{k-2} tiling of the graph itself. We formalise this below.

Proof of Lemma 3.5.2. Let G be as above. Choose constants $\gamma \ll \nu_1 \ll \beta_1 \ll \alpha_1 \ll \nu_2 \ll \ldots \ll \nu_k \ll \beta_k \ll \alpha_k \ll \nu_{k+1} = \nu$. By Lemma 3.2.1, we know that if any graph H is a robust (ν_j, τ)-expander, then it is a robust (ν_{j-1}, τ)-expander for all $j \in \{2, \ldots, k+1\}$. By the Pigeonhole principle, there must be some $j \in [k]$ such that $\mathcal{N}_{\nu_j,\tau}(G) = \mathcal{N}_{\nu_{j+1},\tau}(G)$. Let $\nu' := \nu_j$, $\nu := \nu_{j+1}$, $\alpha := \alpha_j$ and $\beta := \beta_j$. We know by the hypothesis of the theorem that $|\mathcal{N}(G)| \geq 1$, since in particular, $i^* \in \mathcal{N}(G)$. By relabelling if necessary, suppose that $1 \in \mathcal{N}(G)$. We have two cases to consider.

First, assume that $\mathcal{N}(G) = [k]$. Then by Lemma 3.5.17, since each part has the same size and (V_i, V_{i+1}) is a (ν, τ)-robust pair for each $i \in [k]$, there is a perfect fractional P_{k-2}-tiling. Therefore, we may assume that there is at least one $i \in [k]$ such that $i \notin \mathcal{N}(G)$, that is, (V_i, V_{i+1}) is not a (ν, τ)-robust pair.

For each $i \notin \mathcal{N}(G)$, we let $((V_{i,L_1}, V_{i,L_2}), (V_{i+1,R_1}, V_{i+1,R_2}))$ be an (α, β)-bipartition of (V_i, V_{i+1}). Indeed, this exists by Lemma 3.2.5 since (V_i, V_{i+1}) is not a (ν', τ)-robust pair. Let $s \in [k]$ be such that (V_s, V_{s+1}) is a non-(ν', τ)-robust pair in G. Then for

$(i,X),(i',X') \in \{(s,L),(s+1,R)\}$ with $(i,X) \neq (i',X')$ and $j \in \{1,2\}$, we say a vertex $v \in V_{i,X_j}$ is *σ-partition-respecting* if $\left|N(v) \cap V_{i',X'_j}\right| \geq (1-\sigma)\left|V_{i',X'_j}\right|$. Otherwise, we say v is *σ-non-partition-respecting.* By the definition of an (α,β)-bipartition, for any $i \in [k]$, there can be at most $4\beta n$ vertices in V_i which are α-non-partition-respecting. This is because each V_i can belong to at most two non-(ν',τ)-robust pairs, and each such pair can create up to $2\beta n$ vertices which are α-non-partition-respecting. For each $i \in [k]$, let B_i be the set of α-non-partition-respecting vertices in V_i.

We will first find a collection $\mathcal{Q}^{(1)}$ of vertex-disjoint transversal paths of length $k-2$ such that $G' \coloneqq G[V(G) \setminus V(\mathcal{Q}^{(1)})]$ contains no σ'-non-partition-respecting vertices, for some σ'. Set $G_0 \coloneqq G$. We will find a sequence of graphs $G_0 \supset G_1 \supset \ldots \supset G_k = G'$ as follows. For each $i \in [1 \uparrow k]$, suppose we have found G_{i-1}, and let $V_j^{(i-1)} = V_j \cap V(G_{i-1})$ for each $j \in [k]$. Let $n_{i-1} = \left|V_j^{(i-1)}\right| = n - 4(k-1)\beta(i-1)n$ for each $j \in [k]$. Also suppose

$$\delta^*(G_{i-1}) \geq \left(\frac{1}{2} - \gamma - 4(k-1)\beta(i-1)\right)n \geq \left(\frac{1}{2} - \gamma - 4(k-1)\beta(i-1)\right)n_{i-1}.$$

Now let $X_i \subseteq V_i$ be such that $(B_i \cap V(G_{i-1})) \subseteq X_i$ and $|X_i| = 4\beta n$. Then by Lemma 3.5.6, we can find a collection $\mathcal{Q}_i^{(1)}$ of transversal paths of length $k-2$ which cover X_i and which collectively contain exactly $(k-1)|X_i| = 4(k-1)\beta n$ vertices from $V_j^{(i-1)}$ for each $j \in [k]$.

Let $G_i = G_{i-1}[V(G_{i-1}) \setminus V(\mathcal{Q}_i)]$, and let $V_j^{(i)} = V_j \cap V(G_i)$ for $j \in [k]$. Then $n_i = \left|V_j^{(i)}\right| = n_{i-1} - 4(k-1)\beta n = n - 4(k-1)\beta i n$. Furthermore,

$$\begin{aligned}\delta^*(G_i) \geq \delta^*(G_{i-1}) - 4(k-1)\beta n &\geq \left(\frac{1}{2} - \gamma - 4(k-1)\beta i\right)n \\ &\geq \left(\frac{1}{2} - \gamma - 4(k-1)\beta i\right)n_i.\end{aligned}$$

When this process has finished, let $G' \coloneqq G_k$, and for each $j \in [k]$, let $V'_j \coloneqq V_j^{(k)}$. Choose constants γ' and β' such that $\beta \ll \gamma' \ll \beta' \ll \alpha$. Then $n' \coloneqq n_k = (1-4(k-1)k\beta)n$, and $\delta^*(G') \geq (1/2 - \gamma - 4(k-1)k\beta)n \geq (1/2-\gamma')n'$. Let

$\mathcal{Q}^{(1)} = \bigcup_{i\in[k]} \mathcal{Q}_i^{(1)}$.

We have removed at most $\gamma' n$ vertices from each part. So, for each $i \in \mathcal{N}_{\nu,\tau}(G)$, we know by Lemma 3.2.6 that $i \in \mathcal{N}_{\nu-\gamma',\tau}(G')$. Now for each $i \notin \mathcal{N}_{\nu-\gamma',\tau}(G')$, we know $i \notin \mathcal{N}_{\nu,\tau}(G)$. Therefore, for any $i \notin \mathcal{N}_{\nu-\gamma',\tau}(G')$, for each $j \in [2]$, let $V'_{i,L_j} = V_{i,L_j} \cap V(G')$, and let $V'_{i+1,R_j} = V_{i+1,R_j} \cap V(G')$. Then

$$\begin{aligned}\left(\frac{1}{2}-\beta'\right)n' \le \left(\frac{1}{2}-\beta-4(k-1)k\beta\right)n \le \left|V'_{i,L_j}\right|, \left|V'_{i+1,R_j}\right| \le \left(\frac{1}{2}+\beta\right)n \\ \le \left(\frac{1}{2}+\beta'\right)n'.\end{aligned}$$

Furthermore, for any $v \in V'_{i,L_j}$, we know that $\left|N_G(v, V_{i+1,R_j})\right| \ge (1-\alpha)\left|V_{i+1,R_j}\right|$. Therefore,

$$\left|N_{G'}(v, V'_{i+1,R_j})\right| \ge (1-\alpha)\left|V_{i+1,R_j}\right| - 4(k-1)k\beta n \ge (1-2\alpha)\left|V'_{i+1,R_j}\right|,$$

and the reverse holds for each vertex in V'_{i+1,R_j} by symmetry. Therefore in particular, for each $i \notin \mathcal{N}_{\nu-\gamma',\tau}(G')$, (V_i, V_{i+1}) has a high-degree $(2\alpha,\beta')$-bipartition $((V'_{i,L_1}, V'_{i,L_2}),(V'_{i+1,R_1}, V'_{i+1,R_2}))$.

Now for each $i \in \mathcal{B}_{\nu-\gamma',\tau}(G')$, for each $j,\ell \in [2]$, let $V'_{i,L_j,R_\ell} = V'_{i,L_j} \cap V'_{i,R_\ell}$. For each $i \in \mathcal{L}_{\nu-\gamma',\tau}(G') \setminus \mathcal{B}_{\nu-\gamma',\tau}(G')$, for $j,j' \in [2]$ with $j \neq j'$, let $V'_{i,L_j,R_j} = V'_{i,L_j}$ and let $V'_{i,L_j,R_{j'}} = \emptyset$. For each $i \in \mathcal{R}_{\nu-\gamma',\tau}(G') \setminus \mathcal{B}_{\nu-\gamma',\tau}(G')$, for each $\ell,\ell' \in [2]$ with $\ell \neq \ell'$, let $V'_{i,L_\ell,R_\ell} = V'_{i,R_\ell}$ and let $V'_{i,L_\ell,R_{\ell'}} = \emptyset$. For each $i \in \mathcal{L}(G) \cup \mathcal{R}(G)$, for each $j,j',\ell,\ell' \in [2]$ with $j \neq j'$ and $\ell \neq \ell'$, we say V'_{i,L_j,R_ℓ} and $V'_{i,L_{j'},R_{\ell'}}$ are *diagonal parts*. By Proposition 3.5.4, we know that the difference between the size of two diagonal parts can be at most $2\beta' n'$.

We now want to remove any subparts which are small in size together with the parts which are diagonal to these. We first do the following. Let $\beta' \ll \eta_1 \ll \eta_2 \ll \ldots \ll \eta_{2k+2} \ll \alpha$. As there can be at most four subparts in each part, at most two of these can be 'small' and the difference between the size of diagonal parts must also be 'small', by the Pigeonhole principle there is some $r \in [2k+1]$ such that for each $i \in \mathcal{L}_{\nu,\tau}(G)$ and each $j, j', \ell, \ell' \in [2]$ such that $j \neq j'$ and $\ell \neq \ell'$,

- either $\left|V'_{i,L_j,R_\ell}\right|, \left|V'_{i,L_{j'},R_{\ell'}}\right| \leq \eta_r n'$,
- or $\left|V'_{i,L_j,R_\ell}\right|, \left|V'_{i,L_{j'},R_{\ell'}}\right| \geq \eta_{r+1} n'$.

Let $\eta \coloneqq \eta_r$ and $\eta' \coloneqq \eta_{r+1}$. Let $\mathcal{J}_\eta(G') = \{i \in [k] \colon$ there exist $j, \ell \in [2]$ such that $0 < \left|V_{i,L_j,R_\ell}\right| \leq \eta n'\}$, and suppose $|\mathcal{J}_\eta(G')| = \ell^*$. Let $\mathcal{J}_\eta(G') = \{i_1, \ldots, i_{\ell^*}\}$. We will find a collection of cycles $\mathcal{Q}^{(2)}$ such that if $H = G' - \mathcal{Q}$, then $\mathcal{J}_{\eta'/2}(H) = \emptyset$. That is, in the resulting graph, every subpart has size at least $\eta n'$.

Indeed, let $H_0 = G'$. We find a sequence of graphs $H_0 \supset H_1 \supset \ldots \supset H_{\ell^*}$ inductively as follows. Let $s \in [l^*]$ and suppose we have found H_{s-1}. Let $j_{s,1}, j_{s,2}, \ell_{s,1}, \ell_{s,2} \in [2]$ be such that $j_{s,1} \neq j_{s,2}$ and $\ell_{s,1} \neq \ell_{s,2}$, and $0 < \left|V'_{i_s,j_{s,1},\ell_{s,1}}\right|, \left|V'_{i_s,j_{s,2},\ell_{s,2}}\right| \leq \eta n'$. That is, choose a subpart which has size smaller than $\eta n'$. Let $Y_{i_s} \subseteq V_{i_s}$ be such that $(V'_{i_s,j_{s,1},\ell_{s,1}} \cup V'_{i_s,j_{s,2},\ell_{s,2}}) \subseteq Y_{i_s}$ and $|Y_{i_s}| = 2\eta n'$. That is, Y_{i_s} is a collection of vertices which covers both the subpart we chose and also the subpart which is diagonally opposite from this part. Then by Lemma 3.5.6, we can find a collection $\mathcal{Q}_s^{(2)}$ of transversal paths of length $k-2$ which cover Y_{i_s} and which collectively hit $V_j \cap V(H_{s-1})$ exactly $k-1$ times for each $j \in [k]$. Then let $H_s = H_{s-1}[V(H_{s-1}) \setminus V(\mathcal{Q}_s^{(2)})]$. Then $m_s \coloneqq |V_i \cap V(H_s)| = (1 - 2(k-1)s\eta)n'$ and $\delta^*(H) \geq (1/2 - \gamma' - 2(k-1)s\eta)n'$. Note that we can apply Lemma 3.5.6 at each stage here, because $m_{s-1} = V'_i \cap V(H_{s-1}) = (1 - 2(k-1)(s-1)\eta)n' \geq (1 - 2k^2\eta)n'$ and $\delta^*(H_{s-1}) \geq (1/2 - \gamma' - 2(k-1)(s-1)\eta)n' \geq (1/2 - \eta')m_{s-1}$ for $\eta' \gg \eta, \gamma'$.

When this process is complete, let $\mathcal{Q}^{(2)} = \bigcup_{i \in [\ell^*]} \mathcal{Q}_i^{(2)}$. Let $H \coloneqq H_{\ell^*}$ and $m \coloneqq m_{\ell^*}$. Choose constants γ'' and β'' such that $\eta \ll \gamma'' \ll \beta'' \ll \eta'$. Then $m = (1 - 2(k-1)\ell^*\eta)n' \geq (1 - \gamma'')n'$ and

$$\delta^*(H) \geq \left(\frac{1}{2} - \gamma' - 2(k-1)\ell^*\eta\right) n' \geq \left(\frac{1}{2} - \gamma''\right) m.$$

As we have removed at most $\gamma'' n'$ vertices from each part, then by Lemma 3.2.6, for each $i \in \mathcal{N}_{\nu-\gamma',\tau}(G')$, we know that $i \in \mathcal{N}_{\nu-2\gamma'',\tau}(H)$. For each $i \notin \mathcal{N}_{\nu-2\gamma'',\tau}(H)$, we know $i \notin \mathcal{N}_{\nu-\gamma',\tau}(G')$. So for each $i \notin \mathcal{N}_{\nu-2\gamma'',\tau}(H)$, for each $j \in [2]$, let $V^H_{i,L_j} = V_{i,L_j} \cap V(H)$, and $V^H_{i+1,R_j} = V_{i+1,R_j} \cap V(H)$. Then

$$\left(\frac{1}{2} - \beta''\right) m \leq \left(\frac{1}{2} - \beta'\right) n' - \gamma'' n' \leq \left|V^H_{i,L_j}\right|, \left|V^H_{i+1,R_j}\right| \leq \left(\frac{1}{2} + \beta'\right) n' \leq \left(\frac{1}{2} + \beta''\right) m.$$

Furthermore, for any $v \in V^H_{i,L_j}$, we know that $\left|N_{G'}(v) \cap V'_{i+1,R_j}\right| \geq (1-2\alpha)\left|V'_{i+1,R_j}\right|$. Therefore,

$$\left|N_H(v) \cap V^H_{i+1,R_j}\right| \geq (1-2\alpha)\left|V'_{i+1,R_j}\right| - \gamma'' n' \geq (1-3\alpha)\left|V^H_{i+1,R_j}\right|,$$

and the reverse holds for each vertex in V^H_{i+1,R_j} by symmetry. Therefore, each $i \notin \mathcal{N}_{\nu-2\gamma'',\tau}(H)$, (V^H_i, V^H_{i+1}) has a high-degree $(3\alpha, \beta'')$-bipartition $((V^H_{i,L_1}, V^H_{i,L_2}), (V^H_{i+1,R_1}, V^H_{i+1,R_2}))$. Furthermore, for any $i \in \mathcal{B}_{\nu-\gamma'',\tau}(H)$, for each $j, j', \ell, \ell' \in [2]$ with $j \neq j'$ and $\ell \neq \ell'$, if $V^H_{i,L_j,R_\ell} \neq \emptyset$ then $\left|V'_{i,L_j,R_\ell}\right| \geq \eta' n'$, and so $\left|V^H_{i,L_j,R_\ell}\right| \geq (\eta' - \gamma'')n' \geq \eta' m/2$.

Therefore, as $\eta' \gg \beta''$, we can apply Lemma 3.5.16 to find a collection of vertex-disjoint transversal paths $\mathcal{Q}^{(3)}$ such that if $H' := H[V(H) \setminus V(\mathcal{Q}^{(3)})]$, then there is some $i \in [k]$ such that $H'[V^{H'}_i \cup V^{H'}_{i+1}]$ is a robust $((\nu - 2\gamma'')/2, \tau)$-expander and for each $i' \in [k]$ such that $H'[V^{H'}_{i'} \cup V^{H'}_{i'+1}]$ is not a robust $((\nu - 2\gamma'')/2, \tau)$-expander, there is a balanced $(6\alpha, 0)$-bipartition of $(V^{H'}_{i'}, V^{H'}_{i'+1})$. Thus, we can apply Lemma 3.5.17 to find a perfect fractional path tiling in H'. In other words, there is a function f which assigns a weight in $[0,1]$ to each path of length $k-2$ in H' such that the weight at any vertex of H' is exactly 1. It now remains to find a perfect fractional path tiling in G. Indeed let $\mathcal{P}$ be the set of all paths of length $k-2$ in G. We define the function $g : \mathcal{P} \to [0,1]$ as follows.

$$g(P) = \begin{cases} 1, \text{ if } P \in \mathcal{Q}^{(i)} \text{ for some } i \in [3] \\ 1/(k-1), \text{ if } P \text{ is contained in some cycle } C \in \mathcal{Q}^{(i)} \text{ for some } i \in [3] \\ f(P), \text{ if } V(P) \subseteq V(H') \\ 0 \text{ otherwise.} \end{cases}$$

Then clearly the function is well defined since each path in $\mathcal{P}$ has been assigned some weight in $[0,1]$. To prove that this is a perfect fractional path tiling, we need to check that the weight at each vertex is exactly 1. Indeed, for any $v \in V(H')$, clearly $v \notin \mathcal{Q}^{(i)}$ for any $i \in [3]$. Therefore, since f is a perfect fractional path tiling of $V(H')$, we know that the weight at v must be 1. Now for any $v \notin V(H')$, we know that v must belong to some path in $Q^{(i)}$ for some $i \in [3]$. As this is a collection of vertex-disjoint paths and cycles of length k, we know if v belongs to a path in $\mathcal{Q}^{(1)} \cup \mathcal{Q}^{(2)} \cup \mathcal{Q}^{(3)}$, then it can only belong to one such path and therefore the weight at vertex v is exactly 1. On the other hand, if v belongs to a cycle in $\mathcal{Q}^{(1)} \cup \mathcal{Q}^{(2)} \cup \mathcal{Q}^{(3)}$, then it must belong to exactly $k-1$ paths, and so the weight at v is exactly 1 yet again, concluding the proof. □

3.5.2 No robust expanders

In the following lemma, we show that if G is a k-partitite graph with parts $V_1, \ldots, V_k$ each of size n and $\delta^*(G) \geq (1/2 - \gamma)n$, and if G is close to the almost-complete blow up of the union of 2 C_ks with parts of slightly unbalanced size, then G contains a perfect fractional P_{k-2}-tiling. We also show that if G is close to the almost-complete blow-up of C_{2k} with parts of slightly unbalanced size *and* k is even, then G also contains a perfect fractional P_{k-2}-tiling.

Lemma 3.5.18. *Let k be an even integer. Let $1/n \ll \gamma \ll \beta \ll \alpha \ll 1/k$. Let G be a k-partite graph with parts $V_1, \ldots, V_k$, each of size n, and $\delta^*(G) \geq (1/2 - \gamma)n$.*

For each $i \in [k]$, suppose (A_i, B_i) is a partition of V_i such that for each $i \in [k-1]$, $((A_i, B_i), (A_{i+1}, B_{i+1}))$ is a high-degree (α, β)-bipartition of (V_i, V_{i+1}). Suppose one of the following holds.

(i) *$((A_k, B_k), (A_1, B_1))$ is a high-degree (α, β)-bipartition of (V_k, V_1).*

(ii) *$((A_k, B_k), (B_1, A_1))$ is a high-degree (α, β)-bipartition of (V_k, V_1) and k is even, and $||A_i| - |B_i||$ is even for each $i \in [k]$.*

Then G contains a perfect fractional P_{k-2}-tiling.

Proof. Let G be as above. Suppose first that (i) holds. Let $c(A) = \min_{i\in[k]} |A_i|$ and let $c(B) = \min_{i\in[k]} |B_i|$. For each $i \in [k]$, let $f(i, A) = |A_i| - c(A)$, and let $f(i, B) = |B_i| - c(B)$. Let $f(A) = \sum_{i\in[k]} f(i, A)$ and $f(B) = \sum_{i\in[k]} f(i, B)$. Since each A_i and B_i belongs to an (α, β)-bipartition, we know that $f(A), f(B) \leq 2k\beta n$. Note that by the density condition, we can ensure that for each $i \in [k]$ and $X \in \{A, B\}$, there are plenty of vertex-disjoint paths of length $k - 2$ which have initial vertex in X_i and which remain in $X_{i'}$ for $i' \in [i \uparrow i - 2]$.

For each $i \in [k]$, for each $X \in \{A, B\}$, we do the following. For each $r \in [f(i, X)]$ and $j \in [k] \setminus \{i + 1\}$, choose a path of length $k - 2$ which has initial vertex in X_j which is disjoint from any other path chosen, and add this to the collection $\mathcal{Q}^{(1)}_{i,X}$. Note that the collection $\mathcal{Q}^{(1)}_{i,X}$ hits X_i exactly $(k - 1)f(i, X)$ times and hits $X_{i'}$ exactly $(k - 2)f(i, X)$ times for each $i' \in [k] \setminus \{i\}$. Let $\mathcal{Q}^{(1)} = \bigcup_{i\in[k],X\in\{A,B\}} \mathcal{Q}^{(1)}_{i,X}$. Let $G' = G - \mathcal{Q}^{(1)}$, and for each $i \in k]$ and $X \in \{A, B\}$, let $X'_i = X_i \cap V(G')$. Then

$$\begin{aligned}
|X'_i| &= |X_i| - \left|V(\mathcal{Q}^{(1)} \cap V(X_i)\right| \\
&= c(X) + f(i, X) - (k - 1)f(i, X) - \sum_{i'\in[k]\setminus\{i\}} (k - 2)f(i', X) \\
&= c(X) - (k - 2)f(X).
\end{aligned}$$

Therefore, $|X'_i| = \left|X'_{i+1}\right| \geq (1/2 - 2k^2\beta)n$ for each $i \in [k]$ and $X \in \{A, B\}$. Furthermore, let α' be such that $\alpha \ll \alpha' \ll 1$. Each vertex in X'_i has at least $(1 - \alpha')\left|X'_{i+1}\right|$

neighbours in X'_{i+1}, and vice versa. Therefore, by Hall's condition, we can find a perfect matching between X'_i and X'_{i+1}. Therefore, by Lemma 3.5.3, there is a perfect fractional path tiling in G'. To obtain a perfect fractional path tiling in G, we give each path which lies in G' the weight given by the perfect fractional path tiling of G'. If P lies in $\mathcal{Q}^{(1)}$, we give it a weight of 1. Otherwise, we give it a weight of 0. This is a perfect fractional P_{k-2}-tiling of G.

Now suppose that (i) does not hold but (ii) does hold. In this case, let $g(i) = |A_i| - |B_i|$. Since n is even. $|g(i)|$ is always even. For each $i \in [k]$ in turn, we do the following. If $g(i) = 0$, then we proceed to $i+1$. Otherwise, let $X, Y \in \{A, B\}$ with $X \neq Y$ be such that $|X_i| \leq |Y_i|$. then relabel the parts of G as $D_1, \dots, D_{2k}$ so that $D_j \coloneqq X_{i+j-1}$ for each $j \in [k]$ and $D_j \coloneqq Y_{i+j-k-1}$ for each $j \in \{k+1, \dots, 2k\}$. Note that for any $j \in [2k]$ (with indices taken modulo $2k$) each vertex in D_j has at least $\alpha\,|D_{j+1}|$ neighbours in D_{j+1}. Therefore, in particular, we can find plenty of vertex-disjoint paths of length $k-2$ which have initial vertex in D_j and which pass through the sets $D_{j'}$ for $j' \in \{j, \dots, j+k-2\}$. We will find a collection $\mathcal{Q}_i^{(2)}$ of vertex-disjoint paths of length $k-2$ which hits D_1 exactly $(k-2)\cdot g(i)$ times, D_{k+1} exactly $k \cdot g(i)$ times and hits D_j exactly $(k-1)\cdot g(i)$ times for each $j \in [2k] \setminus \{1, k+1\}$. Taking $\mathcal{Q}^{(2)} = \bigcup_{i\in[k]} \mathcal{Q}_i^{(2)}$, the net effect of removing $\mathcal{Q}^{(2)}$ from G will be that $|A_i| = |B_i|$ for each $i \in [k]$, and also that $|A_i| = |A_{i+1}|$ for each $i \in [k]$ (indices taken modulo k). Then the remainder of the graph will contain a perfect matching between A_i and A_{i+1} for each $i \in [k]$, and in particular, also contain a perfect fractional P_{k-2}-tiling, and then combining this with the paths in $\mathcal{Q}^{(1)}$ as before, we find a perfect fractional P_{k-2}-tiling in G. Thus, it suffices to find the collection of paths $\mathcal{Q}^{(2)}$. We do this as follows, at each stage making sure we choose paths that are vertex-disjoint from any previously chosen paths.

For $j \in \{2, k+2\}$, let $\mathcal{Q}_{i,j}^{(2)}$ be a collection of $g(i)$ vertex disjoint paths with initial vertex in D_j and add this to $\mathcal{Q}_i^{(2)}$. Now for each $j \in [k/2]$, choose a collection of $2g(i)$ vertex-disjoint paths $\mathcal{Q}_{i,1+2j}^{(2)}$ with initial vertex in D_{1+2j} and add these to $\mathcal{Q}_i^{(2)}$. Also, for each $j \in [k/2-1]$, choose $2g(i)$ vertex-disjoint paths $\mathcal{Q}^{(2)}_{i,k+2+2j}$ with initial vertex in D_{k+2+2j} and add these to $\mathcal{Q}^{(2)}_i$. Now we examine how many times D_j is hit by the collection $\mathcal{Q}^{(2)}_i$ for each $j \in [2k]$. We note that if j is odd then for $3 \leq j \leq k+1$, and if j is even then for $k+3 \leq j \leq 2k$, there are $2g(i)$ paths in $\mathcal{Q}^{(2)}_i$ with initial vertex in D_j.

Now consider D_1. This is hit by any paths with initial vertex in $D_{k+3}, \ldots, D_{2k}, D_1$. Note that between $k+3$ and $2k$, there are exactly $k/2-1$ odd integers. Therefore, D_1 is hit exactly $2(k/2-1)g(i) = (k-2)g(i)$ times. On the other hand, any paths with initial vertex in $D_3, \ldots, D_{k+1}$ will hit D_{k+1}, and there are exactly $k/2$ odd integers between 3 and $k+1$, so D_{k+1} is hit $k \cdot g(i)$ times. Meanwhile, for any $j \in [2k] \setminus \{1, k+1\}$, D_j is hit by any path in $\mathcal{Q}^{(2)}_i$ with initial vertex in $D_{j+1-\ell}$ for $\ell \in [k-1]$, and indices taken modulo $2k$. Let $H(j) = \{j+1-\ell \colon \ell \in [k-1]\}$. Then for any $j \in [2k] \setminus \{1, k+1\}$, $H(j)$ contains exactly $k/2-1$ indices j' such that there are $2g(i)$ paths in $\mathcal{Q}^{(2)}_i$ with initial vertex in $D_{j'}$, and exactly 1 index j' such that there is one path in $\mathcal{Q}^{(2)}_i$ with initial vertex in $D_{j'}$. Therefore, D_j is hit exactly $(2(k/2-1)+1) \cdot g(i) = (k-1)g(i)$ times for each $j \in [k] \setminus \{1, k+1\}$, as required, proving the lemma. $\square$

3.5.3 Proof of Theorem 3.1.4

It only remains to use the lemmas in this section to prove Theorem 3.1.4, that is, the following.

Theorem 3.1.4. *Suppose that $1/n \ll \gamma \ll \psi, 1/k$, and let G be a balanced k-partite graph whose vertex classes each have size n. If $\delta^*(G) \geq (\frac{1}{2} + \frac{1}{2k})n - \gamma n$, then either G contains a perfect fractional C_k-tiling or G is ψ-extremal.*

We begin by outlining the proof. The overall technique relies on Farkas' lemma, as we have seen in the proof outline of Lemma 3.5.1 at the beginning of this section. Therefore, we begin by applying Farkas' lemma to find a vector $\boldsymbol{a}$ which gives a weight to each vertex. We label the vertices in each part so that the 'lighter' vertices have smaller index and are at the 'top', and 'heavier' vertices have larger index and are at the 'bottom'. Rather than explicitly finding a perfect fractional C_k-tiling, the problem is now reduced to finding a collection of sets of size k such that each of these sets 'dominate' a cycle in the graph (that is, for each set, there is a cycle in the graph which is lighter than the set) and such that collectively, these k-sets cover each vertex the same number of times.

We then separate the lightest $(k-1)n/k$ vertices in each part V_i as the set $V_{i,T}$, and the heaviest n/k vertices in each V_i as $V_{i,B}$. The degree condition implies that if $G_T := G[\bigcup_{i\in[k]} V_{i,T}]$ and $n_T = (k-1)n/k$, then $\delta^*(G_T) \geq (1/2-\gamma)n_T$. Furthermore, for each $i \in [k]$, if we choose any vertex from V_i and any vertex from V_{i+2}, they have at least $n/k - 2\gamma n$ neighbours in V_{i+1}, and therefore, they have at least one neighbour which is lighter than any vertex in $V_{i+1,B}$ other than the lightest $2\gamma n$ vertices of $V_{i+1,B}$. The proof splits into multiple cases now depending on the behaviour of G_T.

The first case we consider in Claim 3.5.19 is the case when there is some $i \in [k]$ such that the graph of $G_T[V_i, V_{i+1}]$ is a (ν, τ)-robust expander. In this case, we first show we can find $2k\gamma n$ cycles which are dominated by a collection of sets covering the lightest $2k\gamma n$ vertices in $V_{i,B}$. Then we apply Lemma 3.5.2 to find a perfect fractional P_{k-2}-tiling in G_T. Then we use this to find a collection of sets in G which each dominate a cycle, thus giving a contradiction and therefore proving the existence of a perfect fractional C_k-tiling in G.

The remainder of the proof is dedicated to the case when there is no $i \in [k]$ such that $G_T[V_i, V_{i+1}]$ is a (ν, τ)-robust expander. Lemma 3.2.5 tells us that each pair $G_T[V_i, V_{i+1}]$ must therefore be 'close to' two almost complete bipartite graphs, where the size of each part in this bipartition is around $n/2$. Note that each vertex class $V_{i,T}$ receives two different bipartitions by Lemma 3.2.5. In Claim 3.5.20 we consider what happens when there is some i^* such that the two different bipartitions of $V_{i^*,T}$ do not 'nearly line up'. That is, suppose $A_1 \cup A_2$ and $B_1 \cup B_2$ are two bipartitions of

$V_{i^*,T}$. In Claim 3.5.20, we show that if $|A_j \cap B_\ell| \geq \beta' n$ for each $j, \ell \in [2]$, then we find a perfect fractional P_{k-2} tiling in G_T. As before, this implies the existence of a perfect fractional C_k-tiling in G.

Finally, we assume that the two different bipartitions 'nearly line up' in each part $V_{i,T}$. In this case, we show that if we cannot find $\beta' n$ cycles in G_T, then G is certainly ψ-extremal. If we can find at least $\beta' n$ cycles in G, then if k is even, we can find a perfect fractional C_k-tiling in G. If k is odd, then this is no longer guaranteed, but we show that either G contains a perfect fractional C_k-tiling, or it is ψ-extremal. We now formalise this below.

Proof of Theorem 3.1.4. Choose constants ν, τ such that $\gamma \ll \nu \ll \tau \ll \psi, 1/k$.

Suppose G contains no perfect fractional transversal C_k-tiling. From now on, assume that all cycles of length k are transverse and all paths are partition-respecting. By Fact 3.2.7, we know that $\mathbf{1} \notin PC(\chi(C) \colon C \in \mathcal{C})$. Thus, by Farkas' lemma, there is some $\boldsymbol{a} \in \mathbb{R}^{kn}$ such that $\boldsymbol{a} \cdot \mathbf{1} < 0$ but $\boldsymbol{a} \cdot \chi(C) \geq 0$ for each $C \in \mathcal{C}$.

For each $i \in [k]$, label the vertices of V_i by $v_{i,1}, \ldots, v_{i,n}$, and let $a_{i,1}, \ldots, a_{i,n}$ be the corresponding coordinates of $\boldsymbol{a}$, where labels are chosen so that $a_{i,1} \leq \ldots \leq a_{i,n}$. We say that a set $S = \{v_{i_1,j_1}, \ldots, v_{i_t,j_t}\}$ *dominates* a set $S' = \{v_{i_1,\ell_1}, \ldots, v_{i_t,\ell_t}\}$ if $j_s \geq \ell_s$ for each $s \in [t]$ (that is, if the corresponding indices are larger). The general idea of the proof is to show that if the graph contains certain structure, X, say, then we can find a collection $\mathcal{T}$ of sets such that each set $T \in \mathcal{T}$ dominates some cycle $C_T \in \mathcal{C}$ and such that $\sum_{T \in \mathcal{T}} \chi(T) = b \cdot \mathbf{1}$ for some $b > 0$. If we can do this, then

$$0 \leq \sum_{T \in \mathcal{T}} \boldsymbol{a} \cdot \chi(C_T) \leq \sum_{T \in \mathcal{T}} \boldsymbol{a} \cdot \chi(T) = b \cdot (\boldsymbol{a} \cdot \mathbf{1}) < 0, \tag{3.5.1}$$

which is a contradiction. Therefore, we use this to conclude the graph cannot have structure X. We do this for the first time in Claim 3.5.19. We then repeatedly do this to gain more and more information about the structure of the graph, and finally, use this to conclude that if we cannot find a contradiction to the existence of a perfect fractional C_k-tiling, the resulting graph must be close to extremal.

To begin, for each $i \in [k]$, partition V_i into two sets $V_{i,T}$ and $V_{i,B}$, where $V_{i,T} = \{v_{i,1}, \ldots, v_{i,\frac{k-1}{k}n}\}$ and $V_{i,B} = \{v_{i,\frac{k-1}{k}n+1} \ldots v_{i,n}\}$. Here, T and B represent 'top' and 'bottom', respectively, since if the vertices are represented graphically with indices in increasing order from top to bottom, then $V_{i,T}$ contains the 'top' $(k-1)n/k$ vertices of V_i, and $V_{i,B}$ contains the bottom $(k-1)n/k$ vertices of V_i. To provide intuition through the proof, we will often use the words 'top' and 'bottom' for this graphical representation. Let $V_T = \bigcup_{i \in [k]} V_{i,T}$ and $V_B = \bigcup_{i \in [k]} V_{i,B}$. Let $n_T = (k-1)n/k$ and $n_B = n/k$.

Claim 3.5.19. *Suppose that there is some $i^* \in [k]$ such that $G[V_{i^*,T}, V_{i^*+1,T}]$ is a robust (ν, τ)-expander. Then G contains a perfect fractional C_k-tiling.*

Proof of Claim 3.5.19. We know that for each $i \in [k]$, $\delta(G[V_{i,T}, V_{i+1,T}]) \geq (1/2-\gamma)n$. Furthermore, any pair of vertices $v \in V_i$ and $w \in V_{i+2}$ have at least $n/k-2\gamma n$ common neighbours in V_{i+1}.

We begin by finding $2k\gamma n$ vertex-disjoint transversal cycles of length 5 in $G[V_T]$. We can do this by selecting $2k\gamma n$ vertices in $V_{i^*+2,T}$. By applying Lemma 3.5.5, since the pair $(V_{i^*,T}, V_{i^*+1,T})$ is a (ν, τ)-robust pair in $G[V_T]$, each chosen vertex in $V_{i^*+2,T}$ is contained in at least $(\nu - 2\gamma)n$ cycles of length k in $G[V_T]$ which intersect only at the chosen vertex. Thus, for each chosen vertex, we can greedily choose a cycle of length k such that all cycles chosen are vertex-disjoint. Let $\mathcal{F}$ be the collection of cycles found in this step. These cycles will be used to cover the top $2\gamma n$ vertices of $V_{i,B}$.

Partition $\mathcal{F}$ into k parts of equal size, $\mathcal{F}_1, \ldots, \mathcal{F}_k$. For each $i \in [k]$, label the cycles in $\mathcal{F}_i$ by $F_{i,1}, \ldots, F_{i,2\gamma n}$. For each cycle $F_{i,j}$, we construct a new set $T_{i,j}$, where $T_{i,j}$ contains each vertex of $F_{i,j}$ which does not belong to V_i together with the vertex $v_{i,\frac{k-1}{k}n+j}$. In particular, $T_{i,j}$ dominates $F_{i,j}$ for each $i \in [k]$ and $j \in [2\gamma n]$.

Let $T = \bigcup_{i \in [k], j \in [2\gamma n]} T_{i,j}$. Let $G' = G[V(G) \setminus T]$. Let $V'_T = V_T \setminus T$ and $V'_B = V_B \setminus T$. Let $V'_{i,T} = V_{i,T} \setminus T$ and $V_{i,B} = V_{i,B} \setminus T$ for each $i \in [k]$. Then for each $i \in [k]$, we have removed exactly $2\frac{k-1}{k}\gamma n$ vertices from $V_{i,T}$ to obtain $V'_{i,T}$. So, for each $i \in [k]$,

$$n'_T := \left|V'_{i,T}\right| = \frac{k-1}{k}n - \frac{2(k-1)}{k}\gamma n = \frac{(1-2\gamma)(k-1)}{k}n.$$

For any vertex $v \in V_{i,T}$, we have removed at most $4(k-1)\gamma n/k$ vertices from its neighbourhood in $V_{i+1,T} \cup V_{i-1,T}$. Therefore,

$$\begin{aligned}\delta(G[V'_{i,T}, V'_{i+1,T}]) &\geq \left(\frac{1}{2} - \gamma\right) n_T - 4 \cdot \frac{k-1}{k}\gamma n = \left(\frac{1}{2} - \gamma\right) n_T - 4\gamma n_T \\ &= \left(\frac{1}{2} - 5\gamma\right) n_T.\end{aligned}$$

Furthermore, $G[V_{i^*,T}, V_{i^*+1,T}]$ was a robust (ν, τ)-expander. As, for each $i \in [k]$, we have removed $2(k-1)\gamma n/k = 2\gamma \left|V_{i,T}\right|$ vertices from $V_{i,T}$ to obtain $V'_{i,T}$, Lemma 3.2.6 implies that $G'[V'_{i^*,T}, V'_{i^*+1,T}]$ is a robust $(\nu - 2\gamma, \tau)$-expander. Therefore, we can apply Lemma 3.5.2 to find a perfect fractional partition-respecting P_{k-2}-tiling in $G'[V'_T]$. By definition, this means we find a function f which assigns a weight in $[0, 1]$ to each partition-respecting path of length $k-2$ in $G'[V'_T]$ in such a way that if $\mathcal{P}'$ is the collection of partition-respecting paths in $G'[V'_T]$, then for each vertex $v \in V'_T$, $\sum_{P \in \mathcal{P}'\colon v \in V(P)} f(P) = 1$. Let $\mathcal{P}$ be the collection of paths which are assigned a non-zero weight by f. Note that $\sum_{P \in \mathcal{P}} f(P) = \left|V_T\right|' / (k-1)$, since each path contributes towards the weight of $(k-1)$ vertices of V'_T and since the weight at each vertex is 1.

Now partition $\mathcal{P}$ into k parts $\mathcal{P}_1, \ldots, \mathcal{P}_k$, so that $\mathcal{P}_i = \{P \in \mathcal{P} \colon P \cap V'_{i,T} = \emptyset\}$, that is, $\mathcal{P}$ is the collection of paths in $\mathcal{P}$ which do not intersect $V'_{i,T}$. This does indeed partition $\mathcal{P}$, since each path in $\mathcal{P}$ is partition-respecting and has length $k-2$, and so there is exactly one class $V'_{i,T}$ with which this path does not intersect. Label the paths in $\mathcal{P}_i$ by $P_{i,1}, \ldots, P_{i,\ell_i}$, where $\ell_i = |\mathcal{P}_i|$. Then for any $i \in [k]$,

$$\left|V'_{i,T}\right| = \sum_{v \in V'_{i,T}} \sum_{P \in \mathcal{P} \colon v \in P} f(P) = \sum_{P \in \mathcal{P}} f(P) - \sum_{P \in \mathcal{P}_i} f(P).$$

$$\implies \sum_{P \in \mathcal{P}_i} f(P) = \frac{|V'_T|}{k-1} - \left|V'_{i,T}\right| = \frac{k(1-2\gamma)(k-1)n}{k(k-1)} - \frac{(1-2\gamma)(k-1)n}{k}$$
$$= \frac{(1-2\gamma)n}{k}.$$

For each $i \in [k]$ and $j \in [\ell_i]$, say $U(P_{i,j})$ is the set of common neighbours of the endvertices of $P_{i,j}$ in V_i. Note that this is now in the original graph G. Then $|U(P_{i,j})| \geq (1-2\gamma)n/k$. Let $u(P_{i,j}) \in U(P_{i,j})$ be such that $\boldsymbol{a} \cdot \chi(u(P_{i,j})) \leq \boldsymbol{a} \cdot \chi(v)$ for each $v \in U(P_{i,j})$, that is, $u(P_{i,j})$ is the vertex in $U(P_{i,j})$ which is assigned the lowest weight by $\boldsymbol{a}$. Let $Q_{i,j}$ be the cycle formed by taking $V(P_{i,j})$ together with $u(P_{i,j})$. Let $\mathcal{Q}_i = \{Q_{i,j} \colon j \in [\ell_i]\}$. Let $\mathcal{Q} = \bigcup_{i \in [k]} \mathcal{Q}_i$.

Now for each $P_{i,j}$, we construct $(1-2\gamma)n/k$ sets of vertices $S_{i,j,1}, \ldots, S_{i,j,(1-2\gamma)n/k}$ such that for each $r \in [(1-2\gamma)n/k]$, $S_{i,j,r} = V(P_{i,j}) \cup \{v_{i,\frac{k-1}{k}n+2\gamma n+r}\}$. Essentially, we construct the sets $S_{i,j,r}$ by assigning a collection of paths $P_{i,j}$ to each vertex in $V'_{i,B}$. Note that $S_{i,j,r}$ dominates $Q_{i,j}$ for each $r \in [(1-2\gamma)n/k]$. For each $i \in [k], j \in [\ell_i], r \in [(1-2\gamma)n/k]$, let

- $g(S_{i,j,r}) = f(P_{i,j})/((1-2\gamma)(n/k))$,
- $\mathcal{S}_{i,j} = \{S_{i,j,r'} \colon r' \in [(1-2\gamma)n/k]\}$
- $\mathcal{S}_i = \bigcup_{j' \in [\ell_i]} \mathcal{S}_{i,j'}$, and
- $\mathcal{S} = \bigcup_{i' \in [k]} \mathcal{S}_i$.

Then for each $v \in V'_{i,T}$,

$$\sum_{S \in \mathcal{S} \colon v \in S} g(S) = \sum_{P \in \mathcal{P} \colon v \in P} \frac{(1-2\gamma)(n/k) f(P)}{(1-2\gamma)(n/k)} = \sum_{P \in \mathcal{P} \colon v \in P} f(P) = 1.$$

On the other hand, for each $v \in V'_{i,B}$,

$$\sum_{S \in \mathcal{S}\colon v \in S} g(S) = \sum_{S \in \mathcal{S}, P \in \mathcal{P}_i\colon V(P_i) \subseteq S} g(S) = \sum_{P \in \mathcal{P}_i} \frac{f(P)}{(1-2\gamma)(n/k)} = 1$$

Here, the first equality holds because for each vertex in $V'_{i,B}$, the sets $S_{i,j,r}$ are formed by taking this vertex together with some path in $\mathcal{P}_i$, and each of the paths in $\mathcal{P}_i$ form a set S with exactly one vertex in $V'_{i,B}$. Now,

$$\begin{aligned}
&\boldsymbol{a} \cdot \left(\left(\sum_{i \in [k]} \sum_{j \in [\ell_i]} \sum_{r \in [(1-2\gamma)n/k]} g(S_{i,j,r}) \cdot \chi(S_{i,j,r}) \right) + \left(\sum_{i \in [k]} \sum_{j \in [2\gamma n]} \chi(T_{i,j}) \right) \right) \\
&\geq \left(\sum_{i \in [k]} \sum_{j \in [\ell_i]} \boldsymbol{a} \cdot \chi(Q_{i,j}) \sum_{r \in [(1-2\gamma)n/k]} g(S_{i,j,r}) \right) + \left(\sum_{i \in [k]} \sum_{j \in [2\gamma n]} \boldsymbol{a} \cdot \chi(F_{i,j}) \right) \\
&\geq \left(\sum_{i \in [k]} \sum_{j \in [\ell_i]} \boldsymbol{a} \cdot \chi(Q_{i,j}) f(P_{i,j}) \right) + \left(\sum_{i \in [k]} \sum_{j \in [2\gamma n]} \boldsymbol{a} \cdot \chi(F_{i,j}) \right) \geq 0.
\end{aligned}$$

Here, $\boldsymbol{a} \cdot \chi(Q_{i,j}), \boldsymbol{a} \cdot \chi(F_{i,j}) \geq 0$ by the definition of $\boldsymbol{a}$ and since each $Q_{i,j}$ and $F_{i,j}$ is a transversal cycle, and $f(P_{i,j}) > 0$ because these paths $P_{i,j}$ were chosen to be the paths which were assigned a non-zero value by f. On the other hand,

$$\begin{aligned}
&\boldsymbol{a} \cdot \left(\left(\sum_{i \in [k]} \sum_{j \in [\ell_i]} \sum_{r \in [(1-2\gamma)n/k]} g(S_{i,j,r}) \cdot \chi(S_{i,j,r}) \right) + \left(\sum_{i \in [k]} \sum_{j \in [2\gamma n]} \chi(T_{i,j}) \right) \right) \\
&\leq \boldsymbol{a} \cdot \left(\left(\sum_{i \in [k]} \sum_{j \in [\ell_i]} \sum_{r \in [(1-2\gamma)n/k]} g(S_{i,j,r}) \cdot \left(\sum_{v \in S_{i,j,r}} \chi(v) \right) \right) + \left(\sum_{i \in [k]} \sum_{j \in [2\gamma n]} \sum_{v \in T_{i,j}} \chi(v) \right) \right) \\
&\leq \boldsymbol{a} \cdot \left(\sum_{S \in \mathcal{S}, v \in S} g(S) \cdot \chi(v) + \sum_{T \in \mathcal{T}, v \in T} \chi(v) \right) \\
&\leq \boldsymbol{a} \cdot (\chi(V'_T) + \chi\,(V'_B) + \chi(V_T \setminus V'_T) + \chi(V_B \setminus V'_B)) = \ \boldsymbol{a} \cdot \mathbf{1} < 0.
\end{aligned}$$

This is a contradiction, which proves the claim. □

Therefore, when there exists an $i^* \in [k]$ such that $G[V_{i^*,T}, V_{i^*+1,T}]$ is a robust (ν, τ)-expander, then there is a perfect fractional transversal C_k-tiling. So, we may assume there is no such i^*.

Choose constants α, β, β' such that $\nu \ll \beta \ll \alpha \ll \tau$ and $\beta \ll \beta' \ll \psi, 1/k$. Apply Lemma 3.2.5 to find an (α, β)-bipartition $((V_{i,L_1}, V_{i,L_2}), (V_{i+1,R_1}, V_{i+1,R_2}))$ of $(V_{i,T}, V_{i+1,T})$ for each $i \in [k]$. For each $i \in [k]$ and $j, j', \ell, \ell' \in [2]$ with $j \neq j'$ and $\ell \neq \ell'$, let $V_{i,L_j,R_\ell} = V_{i,L_j} \cap V_{i,R_\ell}$ and let $\text{diff}^G_{i,j,\ell}$ be as in Definition 3.5.12. Then note that $\text{diff}^G_{i,j,\ell} \leq 2\beta n_T$ for each $i \in [k]$ where $n_T = |V_{i,T}|$.

Claim 3.5.20. *Suppose there is some $s \in [k]$ such that $\left|V_{s,L_j,R_\ell}\right| \geq \beta' n_T$ for each $j, \ell \in [2]$. Then G contains a perfect fractional C_k-tiling.*

Proof of Claim 3.5.20. Take $U_{s+1,1} = V_{s+1,L_j,R_\ell}$ with $j, \ell \in [2]$, where each set is chosen so that $\left|V_{s+1,L_j,R_\ell}\right| \geq n_T/8$. Then for $j', \ell' \in [2]$ with $j \neq j'$ and $\ell \neq \ell'$, let $U_{s+1,2} = V_{s+1,L_{j'},R_{\ell'}}$. Then for each $i \in [s+2 \uparrow s-1]$, if $U_{i-1,1} = V_{i-1,L_{j_1},R_{\ell_1}}$, then set $U_{i,1} = V_{i,L_{j_2},R_{j_1}}$ for $j_2 \in [2]$ such that $\left|V_{i,L_{j_2},R_{j_1}}\right| \geq n_T/8$. Then for $j_3, \ell_2 \in [2]$ such that $j_2 \neq j_3$ and $j_1 \neq \ell_2$, we let $U_{i,2} = V_{i,L_{j_3},R_{\ell_2}}$. It remains to find $U_{s,1}$ and $U_{s,2}$. For each $i \in \{s+1, s-1\}$ and $r \in [2]$, suppose that $U_{i,r} = V_{i,L_{j_i},R_{\ell_i}}$ for $j_i, \ell_i \in [2]$. Then set $U_{s,r} = V_{s,L_{j_2},R_{\ell_k}}$.

Note that for each $i \in [k]$ and $r \in [2]$, any vertex in $U_{i,r}$ has at most αn_T non-neighbours in $U_{i+1,r}$ and at most αn_T non-neighbours in $U_{i-1,r}$. If we take any set of $n_T/8 - \alpha n_T$ vertices in $U_{s+1,r}$, we can greedily find a matching between $U_{s+1,r}$ and $U_{s+2,r}$ of size $n_T/8 - \alpha n_T$. Repeat this for each $i \in [k]$ to find a collection of $n_T/8 - \alpha n_T$ vertex-disjoint paths of length $k-2$ which have initial vertex in $U_{s+1,r}$ and which remain in $U_{i,r}$ for $i \in [k] \setminus \{s\}$. Let $W_{i,r}$ be the vertices of these paths which intersect with $U_{i,r}$. Now $U_{s,r}$ has size at least $\beta' n_T$. Consider any vertex in $U_{s,r}$. We know this has at most αn_T non-neighbours in $W_{s+1,r}$ and at most αn_T non-neighbours in $W_{s-1,r}$. Therefore for each $r \in [2]$, greedily, we can find a collection $\mathcal{Q}^{(1)}_r$ of $\beta' n_T$ vertex-disjoint cycles of length k which remain in $U_{i,r}$ for $i \in [k]$. The point of these cycles is the same as the $2k\gamma n$ cycles we found in Claim 3.5.19. In other words, these cycles will be dominated by sets which cover the top $2\gamma n$ vertices of $V_{i,B}$ for each $i \in [k]$. We will also possibly use one of these cycles to make sure that the number of vertices in each class is even. We then use these to fix imbalances between sizes of subparts in the bipartitions, so that in the resulting graph we can find perfect matchings between each pair of vertex classes. These perfect matchings will then give us a perfect fractional P_{k-2}-tiling in V_T, which we can use to find a perfect fractional C_k-tiling in G. Therefore in order to do this, we require this collection of cycles to be sufficiently large, and in particular, this works because $\beta \ll \beta' \ll 1/k$.

Now do the following. For each $r \in [2]$, take $k\gamma n$ cycles of length k in $\mathcal{Q}_r^{(1)}$, and let this be the set $\mathcal{Q}_r^{(2)}$. Partition this into k parts of size γn (note that we assume that $\gamma n \in \mathbb{N}$), and label the parts $\mathcal{Q}_{i,r}^{(2)}$ for $i \in [k]$, and label the elements of this by $Q_{i,r,1}^{(2)}, \dots, Q_{i,r,\gamma n}^{(2)}$. Then for each $i \in [k]$, let $T_{i,r,j}^{(1)}$ be the set of vertices of $Q_{i,r,j}^{(2)}$ which do not intersect V_i, together with the vertex $v_{i,(k-1)n/k+j}$. Let $\mathcal{T}^{(1)}$ be the collection of all these sets $T_{i,r,j}^{(1)}$. Also, if n_T is odd, then let $T^{(2)}$ be a cycle in $\mathcal{Q}^{(1)}$ which has not yet been chosen.

Let $G' = G - V(\mathcal{T}^{(1)} \cup T^{(2)})$. Let $V_i' = V_i \cap V(G')$, $V_{i,B}' = V_{i,B} \cap V(G')$ and $V_{i,T}' = V_{i,T} \cap V(G')$ for each $i \in [k]$. Let $n' = |V_i'|$ and let $n_T' = \left|V_{i,T}'\right|$. Note that n_T' is even.

Now for each $i \in [k]$ in turn, do the following. Suppose $\mathrm{diff}_{i,1,1}^{G'}$ is odd, then by Proposition 3.5.4, we know that $\mathrm{diff}_{i,1,2}^{G'}$ is also odd. Let $T_{i,1}^{(3)}$ be a set containing exactly one vertex from the larger of V_{i,L_1,R_1}' and V_{i,L_2,R_2}', together with exactly one previously unchosen vertex from $V_{i',B}'$ for each $i' \in [k] \setminus \{i\}$. Let $T_{i,2}^{(3)}$ be a set containing exactly one vertex from the larger of V_{i,L_1,R_2}' and V_{i,L_2,R_1}', together with exactly one previously unchosen vertex from $V_{i',B}'$ for each $i' \in [k] \setminus \{i\}$. Choose a previously unchosen cycle from $\mathcal{Q}_1^{(1)}$ and let $T_{i,3}^{(3)}$ be the set containing each vertex from this cycle which does not intersect with $V_{i,T}'$, together with a previously unchosen vertex from $V_{i,B}'$. Similarly, choose a previously unchosen cycle from $\mathcal{Q}_2^{(1)}$ and let $T_{i,4}^{(3)}$ be the set containing each vertex from this cycle which does not intersect with $V_{i,T}'$, together with a previously unchosen vertex from $V_{i,B}'$. Note that for each $r \in [4]$, there is a cycle $C_{i,r}^{(3)}$ in G such that $\boldsymbol{a} \cdot \chi(C_{i,r}^{(3)}) \leq \boldsymbol{a} \cdot \chi(T_{i,r}^{(3)})$. For $r = 1$ and $r = 2$, this is by the minimum degree condition. For $r = 3$ and $r = 4$, these are the cycles we found in $\mathcal{Q}^{(1)}$.

Let $\mathcal{T}^{(3)}$ be the collection of sets $T_{i,j}^{(3)}$ for each $i \in [k]$ and $j \in [4]$. Then let $H = G' - V(\mathcal{T}^{(3)})$. As before, let $V_i^H = V_i \cap V(H)$, $V_{i,B}^H = V_{i,B} \cap V(H)$ and $V_{i,T}^H = V_{i,T} \cap V(H)$ for each $i \in [k]$. Let $m = \left|V_i^H\right|$ and let $m_T = \left|V_{i,T}^H\right|$. Note m_T is still even since the collection $\mathcal{T}^{(3)}$ intersects $V'_{i,T}$ an even number of times for each $i \in [k]$. Furthermore, $\mathrm{diff}_{i,j,\ell}^H$ is even for each $i \in [k]$ and $j, \ell \in [2]$.

Now do the following for each $i \in [k]$ in turn. If $\mathrm{diff}_{i,1,1}^H = 0$, then let $\mathcal{T}_{i,1}^{(4)} = \emptyset$. Else, if $\mathrm{diff}_{i,1,1}^H \geq 2$, then let $j_1, j_2 \in [2]$ with $j_1 \neq j_2$ be such that $\left|V_{i,L_{j_1},R_{j_1}}^H\right| \geq \left|V_{i,L_{j_2},R_{j_2}}^H\right|$. For each $j \in [\mathrm{diff}_{i,1,1}^H]$, let $T_{i,j}^{(4)}$ be the set containing one previously unchosen vertex in $V_{i,L_{j_1},R_{j_1}}^H$ together with exactly one previously unchosen vertex in $V_{i',B}^H$ for each $i' \in [k] \setminus \{i\}$, and let $\mathcal{T}_{i,1}^{(4)}$ be the collection of all these sets $T_{i,j}^{(4)}$. Next, if $\mathrm{diff}_{i,1,2}^H = 0$, then let $\mathcal{T}_{i,2}^{(4)} = \emptyset$. Else, if $\mathrm{diff}_{i,1,2}^H \geq 2$, then let $j_3, j_4 \in [2]$ with $j_3 \neq j_4$ be such that $\left|V_{i,L_{j_3},R_{j_4}}^H\right| \geq \left|V_{i,L_{j_4},R_{j_3}}^H\right|$. For each $j \in [\mathrm{diff}_{i,1,2}^H]$, let $T_{i,j}^{(4)}$ be the set containing one previously unchosen vertex in $V_{i,L_{j_3},R_{j_4}}^H$ together with exactly one previously unchosen vertex in $V_{i',B}^H$ for each $i' \in [k] \setminus \{i\}$, and let $\mathcal{T}_{i,2}^{(4)}$ be the collection of all these sets $T_{i,j}^{(4)}$. Let $\mathcal{T}^{(4)} = \bigcup_{i \in [k], j \in [2]} \mathcal{T}_{i,j}^{(4)}$. Note that for each $T \in \mathcal{T}^{(4)}$, there is some cycle C_T of length k in G which satisfies $\boldsymbol{a} \cdot \chi(C_T) \leq \boldsymbol{a} \cdot \chi(T)$ by the degree condition.

Now if both $\mathrm{diff}_{i,1,1}^H = 0$ and $\mathrm{diff}_{i,1,2}^H = 0$, then let $\mathcal{T}_i^{(5)} = \emptyset$. Otherwise, observe that $(\mathrm{diff}_{i,1,1}^H + \mathrm{diff}_{i,1,2}^H)/2$ is an integer, since both terms in the sum are even. Therefore, for each $r \in [2]$, for each $s \in [(\mathrm{diff}_{i,1,1}^H + \mathrm{diff}_{i,1,2}^H)/2]$, choose a previously unchosen cycle from $\mathcal{Q}_r^{(1)}$ and let $T_{i,r,s}^{(5)}$ be the set containing each vertex from this cycle which does not intersect with V_i^H, together with a previously unchosen vertex in $V_{i,B}^H$. Let $\mathcal{T}_{i,r}^{(5)}$ be the collection of all these cycles over all $s \in [(\mathrm{diff}_{i,1,1}^H + \mathrm{diff}_{i,1,2}^H)/2]$. Let $\mathcal{T}^{(5)} = \bigcup_{i \in [k], r \in [2]} \mathcal{T}_{i,r}^{(5)}$. Again, for each $T \in \mathcal{T}^{(4)}$, there is some cycle C_T of length k in G which satisfies $\boldsymbol{a} \cdot \chi(C_T) \leq \boldsymbol{a} \cdot \chi(T)$, namely the cycle from $\mathcal{Q}_r^{(1)}$ which was chosen when forming the set T.

Now let $H' = H - \mathcal{T}^{(4)} - \mathcal{T}^{(5)}$. In particular, $H' = G - \mathcal{T}^{(1)} - \mathcal{T}^{(2)} - \mathcal{T}^{(3)} - \mathcal{T}^{(4)} - \mathcal{T}^{(5)}$.

For any fixed $i \in [k]$, the following holds.

- $\left|V_{i,T} \cap V(\mathcal{T}^{(1)})\right| = 2(k-1)\gamma n$,
- $\left|V_{i,T} \cap V(\mathcal{T}^{(2)})\right| = 1$ if n_T is odd,
- $\left|V_{i,T} \cap V(\mathcal{T}^{(2)})\right| = 0$ if n_T is even,
- $\left|V_{i,T} \cap V(\mathcal{T}^{(3)})\right| = 2\ell^* \leq 2k$, where ℓ^* is the number of parts in which there are two differences which are odd,
- $\left|V_{i,T} \cap V(\mathcal{T}^{(4)})\right| = \mathrm{diff}_{i,1,1} + \mathrm{diff}_{i,1,2} \leq 4\beta n$,
- $\left|V_{i,T} \cap V(\mathcal{T}^{(5)})\right| = \sum_{i' \in [k] \setminus \{i\}} \mathrm{diff}_{i',1,1} + \mathrm{diff}_{i',1,2} \leq 4(k-1)\beta n$.

Thus overall, for each $i, i' \in [k]$, $\left|V_{i,T}^{H'}\right| = \left|V_{i',T}^{H'}\right| \geq (k-1)n/k - 2\alpha n$, and so in particular, each vertex in $V_{i,L_j,R_\ell}^{H'}$ has at most $2\alpha n$ non-neighbours in $V_{i+1,L_{j'},R_j}$, and vice versa for each $i \in [k]$ and $j, j', \ell \in [2]$. We can check the same quantities for $V_B^{H'}$. Indeed, for any $i \in [k]$, the following holds.

- $\left|V_{i,B} \cap V(\mathcal{T}^{(1)})\right| = 2\gamma n$,
- $\left|V_{i,B} \cap V(\mathcal{T}^{(2)})\right| = 0$,
- $\left|V_{i,B} \cap V(\mathcal{T}^{(3)})\right| = 2\ell^* \leq 2k$, where ℓ^* is the number of parts in which there are two differences which are odd,
- $\left|V_{i,B} \cap V(\mathcal{T}^{(4)})\right| = \sum_{i' \in [k] \setminus \{i\}} \mathrm{diff}_{i',1,1} + \mathrm{diff}_{i',1,2} \leq 4(k-1)\beta n$,
- $\left|V_{i,B} \cap V(\mathcal{T}^{(5)})\right| = \mathrm{diff}_{i,1,1} + \mathrm{diff}_{i,1,2} \leq 4\beta n$.

Therefore, similarly, for each $i, i' \in [k]$, $\left|V_{i,B}^{H'}\right| = \left|V_{i',T}^{H'}\right| \geq n/k - 2\alpha n$. More precisely, we consider two cases. First, in the case when n_T is even,

$$\begin{aligned}(k-1)\left|V_{i,B}^{H'}\right| &= (k-1)\left(\frac{n}{k}-2\gamma n-2\ell^*-\sum_{i'\in[k]}\text{diff}_{i',1,1}+\text{diff}_{i',L_1,R_2}\right)\\ &= \left|V_{i,T}^{H'}\right|-(k-2)\left(2\gamma n-2\ell^*-\sum_{i'\in[k]}\text{diff}_{i',1,1}+\text{diff}_{i',1,2}\right)\\ &\leq \left|V_{i,T}^{H'}\right|-7\beta n.\end{aligned}$$

Hence, we choose $(k-2)(2\gamma n-2\ell^*-\sum_{i'\in[k]}\text{diff}_{i',1,1}+\text{diff}_{i',1,2})/2$ previously unchosen cycles from $\mathcal{Q}_r^{(1)}$ for each $r\in[2]$, and take the collection of these cycles to be $\mathcal{T}^{(6)}$. Indeed, we can do this because each term is even and because $\gamma n\in\mathbb{N}$. Furthermore, for each $i\in[k]$, $j,\ell\in[2]$, we have $\text{diff}_{i,j,\ell}^{H'}=0$. If H'' is the graph $H'-\mathcal{T}^{(6)}$, then for each $i\in[k]$, $j,\ell\in[2]$, we have $\text{diff}_{i,j,\ell}^{H''}=0$. So, we can find a perfect fractional P_{k-2}-tiling in $H''[V_T]$. The proof then proceeds similarly to the proof of Claim 3.5.19. We omit the details here, but the point is that the various collections of cycles found dominate sets of vertices in the graph, such that each vertex $v\in V(G)$ is contained in these sets with equal weight. In this way, as in the proof of Claim 3.5.19, we can find a contradiction to Fact 3.2.7, and therefore, show that G must contain a perfect fractional C_k-tiling.

On the other hand if n_T is odd, the proof proceeds almost identically to the previous case, other than that we additionally choose $\lfloor(k-2)/2\rfloor$ more previously unchosen cycles from $\mathcal{Q}_r^{(1)}$ for each $r\in[2]$ to add to $\mathcal{T}^{(6)}$. If $k-2$ is odd, we add an additional previously unchosen cycle from $\mathcal{Q}_1^{(1)}$ to add to $\mathcal{T}^{(6)}$. Then take $H''=H'-\mathcal{T}^{(6)}$. Then if $k-2$ is even, the argument follows in exactly the same way as for the case when n_T is even, as for each $i\in[k]$, $j,\ell\in[2]$, we have $\text{diff}_{i,j,\ell}^{H''}=0$. On the other hand, if $k-2$ is odd, then for each $i\in[k]$, there exists $j,j',\ell,\ell'\in[2]$ with $j\neq j'$ and $\ell\neq\ell'$ such that $\text{diff}_{i,j,\ell}^{H''}=0$ but $\text{diff}_{i,j',\ell'}^{H''}=1$. More precisely, if $\left|V_{i,L_j,R_{\ell'}}^{H''}\right|=\left|V_{i,L_{j'},R_\ell}^{H''}\right|-1$, then $\left|V_{i+1,L_{j''},R_j}^{H''}\right|=\left|V_{i+1,L_{\ell''},R_{j'}}\right|-1$ for $\ell'',j''\in[2]$ with $\ell''\neq j''$. Meanwhile, $\left|V_{i,L_j,R_\ell}^{H''}\right|=\left|V_{i,L_{j'},R_{\ell'}}^{H''}\right|$. This implies that if $\left|V_{i,L_j}^{H''}\right|=\left|V_{i,R_j}\right|$ for each $i \in [k]$, $j \in [2]$ (this quantity is either $\lfloor m''/2 \rfloor$ or $\lfloor m''/2 \rfloor + 1$). Thus, by the degree condition, we can still find a perfect matching between $V_{i,T}^{H''}$ and $V_{i+1,T}^{H''}$ for each $i \in [k]$, and therefore we can proceed as in the case when n_T is even. Thus, we find a perfect fractional C_k-tiling in G, proving the claim. $\square$

This claim showed that whenever we have a part $V_{s,T}$ which is partitioned twice in such a way that in the joint partition, each part is 'sufficiently large', then this provides enough flexibility for us to fix the imbalances between parts, and allow us to find a perfect fractional C_k-tiling. Thus, we deduce that for each $i \in [k]$, when $V_{i,T}$ is partitioned twice, these partitions are 'well-aligned'. Intuitively, this suggests that $G[V_T]$ resembles either the union of two blow-ups of C_k (**Case 1**), or the blow-up of a C_{2k} (**Case 2**). In the first case, we show that G contains a C_k-factor. In the second case, we show that G is close to extremal. We now formalise this below.

For each $i \in [k]$, there exist $j_1, j_2, \ell_1, \ell_2 \in [2]$ with $j_1 \neq j_2$ and $\ell_1 \neq \ell_2$ such that $\left|V_{i,L_{j_1},R_{\ell_1}}\right|, \left|V_{i,L_{j_2},R_{\ell_2}}\right| \leq \beta' n_T$ and $\left|V_{i,L_{j_1},R_{\ell_2}}\right|, \left|V_{i,L_{j_2},R_{\ell_1}}\right| \geq (1/2 - \beta - \beta')n_T$. Let $U_{1,1} = V_{1,L_j,R_\ell}$ where $j, \ell \in [2]$ are such that $\left|V_{2,L_j,R_\ell}\right| \geq (1/2 - \beta - \beta')n_T$. Then let $U_{1,2} = V_{1,L_{j'},R_{\ell'}}$ where $j', \ell' \in [2]$ are such that $j \neq j'$ and $\ell \neq \ell'$. Then for each $i \in [2 \uparrow k]$, for each $r \in [2]$, suppose $U_{i-1,r} = V_{i-1,L_j,R_\ell}$ for some $j, \ell \in [2]$. Let $U_{i,r} = V_{i,L_{j'},R_j}$ for $j' \in [2]$ such that $\left|V_{i,L_{j'},R_j}\right| \geq (1/2 - \beta - \beta')n$. Suppose that $j_1, j_2, \ell_1, \ell_2 \in [2]$ are such that $j_1 \neq j_2$ and $\ell_1 \neq \ell_2$, and $U_{k,1} = V_{k,L_{j_1},R_{\ell_1}}$, and $U_{k,2} = V_{k,L_{j_2},R_{\ell_2}}$. Let $U'_{1,1} = V_{1,L_{j_3},R_{j_1}}$ and $U'_{1,2} = V_{1,L_{j_4},R_{j_2}}$ where $j_3, j_4 \in [2]$ with $j_3 \neq j_4$ and $\left|U'_{1,1}\right|, \left|U'_{1,2}\right| \geq (1/2 - \beta - \beta')n_T$. Now we have two cases to consider.

Case 1: $U'_{1,j} = U_{1,j}$ for each $j \in [2]$. In this case, we first find $2k\gamma n$ cycles of length k in $G[V_T]$. To do this, we use the property that each vertex in $U_{i,1}$ has at most αn_T non-neighbours in $U_{i+1,1}$. Let $\mathcal{R}$ be the set of these cycles. Partition this into k parts of size γn, and label the parts $\mathcal{R}_i$ for $i \in [k]$, and label the elements of this by $R_{i,1}, \ldots, R_{i,\gamma n}$. Then for each $i \in [k]$, let $T_{i,j}$ be the set of vertices of $R_{i,j}$ which do not intersect V_i, together with the vertex $v_{i,(k-1)n/k+j}$. Let $\mathcal{T}$ be the collection of all these sets $T_{i,r,j}$. Then in particular, for each $T \in \mathcal{T}$, there is a cycle C_T in G which satisfies $\boldsymbol{a} \cdot \chi(C_T) \leq \boldsymbol{a} \cdot \chi(T)$.

Let $G' = G - \mathcal{T}$, and let $n' = n - 2k\gamma n$. Then apply Lemma 3.5.18 to find a perfect fractional P_{k-2}-tiling in G'_T. Then using the same methods as in Claim 3.5.19 and Claim 3.5.20, we can use this to find a perfect fractional C_k-tiling in G', and combining this with the sets $\mathcal{T}$, we can find a perfect fractional C_k-tiling in G.

Case 2: $\boldsymbol{U'_{1,1} = U_{1,2}}$ and $\boldsymbol{U'_{1,2} = U_{1,1}}$. We want to show that either G is ψ-extremal or we can find a collection of $2k\gamma n$ cycles, together with a perfect fractional path tiling in G. We do the following. For each $i \in [k]$, let $A_i \cup B_i \cup Z_i$ be a partition of V_i so that $|A_i|, |B_i| = (k-1)n/2k$ and $|Z_i| = n/k$, and furthermore such that $|A_i \cap U_{i,1}| \geq (1/2 - 2\beta')(k-1)n/k$, $|B_i \cap U_{i,2}| \geq (1/2 - 2\beta')(k-1)n/k$ and $Z_i = V_{i,B}$. Let $\alpha \ll \alpha' \ll \psi$. First, observe that for each $i \in [k-1]$,

$$d(A_i, A_{i+1}) = \frac{(1-\alpha')\,|A_i| \cdot (1-\alpha') \cdot (1-\alpha')\,|A_{i+1}|}{|A_i|\,|A_{i+1}|} \geq 1 - \psi.$$

Similarly, $d(B_i, B_{i+1}), d(A_k, B_1), d(B_k, A_1) \geq 1 - \psi$. Suppose $d(A_s, B_{s+1}) \leq \beta'$ and $d(B_s, A_{s+1}) \leq \beta'$ for each $s \in [k-1]$ and also $d(A_k, A_1) \leq \beta'$ and $d(B_k, B_1) \leq \beta'$. Then the first four criteria in Definition 3.1.1 are immediately satisfied. Now suppose that k is even. Then in order to check that the graph is ψ-close to extremal, it remains only to check that $d(Z_i, A_{i+1})$, $d(A_i, Z_{i+1})$, $d(Z_i, B_{i+1})$, $d(B_i, Z_{i+1}) \geq 1 - \psi$ for each $i \in [k]$. Indeed, for $i \in [k-1]$, note that $d(A_i, V_{i+1}) < \beta'$. That is,

$$\frac{|E(A_i, B_{i+1})|}{|A_i|\,|B_{i+1}|} < \beta'.$$

Meanwhile, by the minimum degree condition, $|E(A_i, V_{i+1})| \geq |A_i|\,(k+1)n/2k$.

Therefore,

$$\begin{aligned}|E(A_i, Z_{i+1})| &= |E(A_i, V_{i+1})| - |E(A_i, A_{i+1})| - |E(A_i, B_{i+1})| \\ &\geq |A_i| \frac{(k+1)n}{2k} - |A_i| |A_{i+1}| - \beta' |A_i| |B_{i+1}| \\ &= |A_i| \left(\frac{n}{k} - \frac{\beta' \cdot (k-1)}{2k} \right) \geq |A_i| |Z_{i+1}| (1 - \psi).\end{aligned}$$

Here, the final inequality holds since $\psi \gg \beta'$. This implies that $d(A_i, Z_{i+1}) \geq 1 - \psi$. By repeating this argument for each of $d(Z_i, A_{i+1})$, $d(Z_i, B_{i+1})$ and $d(B_i, Z_{i+1})$, we get that the fifth criterion in Definition 3.1.1 holds, and therefore, when k is even, G is ψ-close to extremal. The same argument can be used to prove the sixth criterion in Definition 3.1.1 when k is odd, and therefore prove that G is ψ-close to extremal.

Thus, it remains only to consider the case when there is some $s \in [k-1]$ such that $d(A_s, B_{s+1}) \geq \beta'$, or the case when there is some $s \in [k-1]$ such that $d(B_s, A_{s+1}) \geq \beta'$. By symmetry, it suffices only to check the first of these. That is, we consider the case when there are at least $\beta' |A_s| |B_{s+1}|$ edges between A_s and B_{s+1}. Then in particular, this implies that we can find a matching of size $\beta'(k-1)n/2k$ between A_s and B_{s+1}. Using the fact that for each $i \in [k-1]$ there is a high density of edges between B_i and B_{i+1}, between A_i and A_{i+1} and between B_k and A_1, we can find at least $\beta' n$ cycles which pass through the sets $A_1, \ldots, A_s, B_{s+1}, \ldots, B_k$. First remove $2k\gamma n$ cycles from this, to cover up the top vertices of V_B, as before. Let G' be the graph obtained after removing the necessary vertices. So, $V'_{i,T} = V_{i,T} \cap V(G')$, and $n'_T = \left| V'_{i,T} \right|$ for any $i \in [k]$. Also for each $i \in [k]$ and $j, \ell \in [2]$, define $V'_{i,L_j} = V_{i,L_j} \cap V(G')$ $V'_{i,R_j} = V_{i,R_j} \cap V(G')$ Note that in G', $|V_{i,T}| \geq (1/2 - \beta)n_T - 2(k-1)\gamma n \geq (1/2 - 2\beta)n'_T$. Now the proof splits into two cases, depending on whether or not k is even.

Claim 3.5.21. *If k is even and $d(A_s, B_{s+1}) \geq \beta'$, then G contains a perfect fractional C_k-tiling.*

Proof of Claim 3.5.21. Suppose first that k is even. We show that we can find a perfect fractional P_{k-2}-tiling in G. As we have done before, for each $i \in [k-1]$, let X_i be the set of vertices in V'_{i,L_j} which have fewer than $(1-2\alpha)\left|V'_{i+1,R_j}\right|$ neighbours in V'_{i+1,R_j}, together with the vertices in V'_{i,R_j} which have fewer than $(1-2\alpha)\left|V'_{i-1,L_j}\right|$ neighbours in V'_{i-1,L_j} for each $j \in [2]$. Also, for each $i \in [k]$, let Y_i be the set of vertices in $V_{i,L_j,R_\ell} \cup V_{i,L_{j'},R_{\ell'}}$, where $j, j', \ell, \ell' \in [2]$ are such that $j \neq j'$ and $\ell \neq \ell'$, and $\left|V_{i,L_j,R_\ell}\right| \leq \beta' n_T$. Now for each $i \in [k]$, we take a set of vertices $Z'_i \subseteq V'_{i,T}$ such that $X_i \cup Y_i \subseteq Z'_i$, and such that $|Z'_i| = 10\beta' n_T$. Then apply Lemma 3.5.6 to find a collection $\mathcal{Q}_i^{(1)}$ of paths which cover Z'_i and which intersect each vertex class exactly $(k-1)|Z_i|$ times. Remove this and repeat for $i+1$. If H is the graph obtained at the end of this process, then let $V_T^H = V'_T \cap V(H)$ and similarly label $V_i \cap V(H)$ as V_i^H. Note that $V_{i,T}^H = \left|V_{i',T}^H\right|$ for each $i, i' \in [k]$. So, let $|V_{i,T}| = m$. If m is not even, first find an additional cycle which only remains in V_T^H and then in the remainder of H, apply Lemma 3.5.18 to find a perfect fractional P_{k-2}-tiling in $H[V_T^H]$. Combined with the paths in $\mathcal{Q}^{(1)}$ this gives a perfect fractional P_{k-2}-tiling of $G'[V'_T]$. Therefore, as before, we can complete this to a perfect fractional C_k-tiling of G. $\square$

Claim 3.5.22. *If k is odd, and both $d(A_s, B_{s+1}) \geq \beta'$ and $d(B_s, A_{s+1}) \geq \beta'$, then G contains a perfecct fractional C_k-tiling.*

We omit the full details in proving Claim 3.5.22, but the idea is as follows. The claim implies that both $d(U_{s,1}, U_{s+1,2}) \geq \beta'/2$ and $d(U_{s,2}, U_{s+1,1}) \geq \beta'/2$. In this case, we can find at least $\beta' n/2$ cycles which pass through $U_{1,1}, \ldots, U_{s,1}, U_{s+1,2}, \ldots, U_{k,2}$, and also $\beta' n/2$ cycles which pass through $U_{1,2}, \ldots, U_{s,2}, U_{s+1,1}, \ldots, U_{k,1}$. We can then as before remove the smaller subparts and non-partition-respecting vertices in each V_i and then find a perfect fractional P_{k-2}-tiling in the remainder of the graph by balancing using similar methods as Lemma 3.5.18.

Therefore, we may assume that if k is odd, then for each $i \in [k-1]$, whenever $d(A_i, B_{i+1}) \geq \beta'$, we have $d(B_i, A_{i+1}) \leq \beta'$. Therefore, by the degree condition, this implies that for each $i \in [k]$, there exists $X, Y \in \{A, B\}$ such that $d(X_i, Z_{i+1}), d(Y_i, Z_{i-1}) \geq 1 - \psi$, and therefore G satisfies Definition 3.1.1 in the case when k is odd, and so G is ψ-extremal. Note that we can do something similar when there exists $s \in [k-1]$ such that $d(B_s, A_{s+1}) \geq \beta'$, or when there exists $X \in \{A, B\}$ such that $d(X_k, X_1) \geq \beta'$.

Therefore, this concludes the proof of the result.

□

CHAPTER 4

REMOVING INDUCED POWERS OF CYCLES FROM A GRAPH

The aim of this chapter is to prove Theorem 1.2.1, which we recall.

Theorem 1.2.1. *Let $t \geq 1$ and $h \geq 4t(2t+1)$ be integers, with $(t+1) \mid h$. Let $c_0 = \lfloor (\lfloor h/t \rfloor + 1)/3 \rfloor$, $\ell_0 = \lceil h/(2t+1) \rceil$, and let $p_0 = t/(c_0\ell_0 - c_0 - \ell_0 + t + 1)$. Then for all $p \in [p_0, 1/\lceil h/(2t+1) \rceil]$, we have that*

$$\mathrm{ed}_{\mathrm{Forb}(C_h^t)}(p) = \frac{p(1-p)}{t + \left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1 \right) p}.$$

This chapter is set out as follows. In Section 4.1, we describe some general constructions which can be used to calculate $\mathrm{dist}(G(n,p), \mathcal{H})$. We then use these constructions to prove the main result in Section 4.2.

4.1 Coloured regularity graphs

In this section, we describe a structure which defines a set of rules for editing a graph to make it satisfy some hereditary property.

A *coloured regularity graph (CRG)* is a complete graph K on k vertices, in which each vertex is coloured either black or white, and each edge is coloured black, white or grey. Formally, we say that we can partition the vertex set $V(K)$ into two sets $\mathrm{VB}(K)$ and $\mathrm{VW}(K)$, of black and white vertices respectively, and we can partition the edge set $E(K)$ into $\mathrm{EB}(K)$, $\mathrm{EW}(K)$, and $\mathrm{EG}(K)$ of black, white, and grey edges respectively. For ease of notation, when the CRG K is clear from context, we write VB to mean $\mathrm{VB}(K)$, and analogously for VW, EB, EW, and EG. We say a CRG K' is a *sub-CRG* of a CRG K if K' can be obtained by deleting vertices of K.

We say a graph H *embeds* in a CRG K, and we write $H \mapsto K$, if there exists some function $\phi\colon V(H) \longrightarrow V(K)$ which, for all $u, v \in V(K)$, satisfies the following conditions.

(i) If $uv \in E(H)$, we have either $\phi(u) = \phi(v) \in \mathrm{VB}$, or $\phi(u)\phi(v) \in \mathrm{EB} \cup \mathrm{EG}$.

(ii) If $uv \notin E(H)$, we have either $\phi(u) = \phi(v) \in \mathrm{VW}$, or $\phi(u)\phi(v) \in \mathrm{EW} \cup \mathrm{EG}$.

For our purposes, due to a key property of CRGs, it is most interesting to consider all those CRGs into which H does not embed. Specifically, suppose we find a CRG K into which a graph H does not embed. Then any graph H' which contains H as an induced subgraph will also not embed in K. So, if there is a graph G which embeds into K, then crucially, G cannot contain H as an induced subgraph, implying that $G \in \mathrm{Forb}(H)$. Thus we find a very clear relationship between the class of CRGs into which a graph H does not embed, and the family $\mathrm{Forb}(H)$.

Recall that $\mathcal{F}(\mathcal{H})$ is the family of all forbidden subgraphs of $\mathcal{H}$. Then for any hereditary property $\mathcal{H}$, we let

$$\mathcal{K}(\mathcal{H}) = \{K\colon H \not\mapsto K \text{ for all } H \in \mathcal{F}(\mathcal{H})\}\,.$$

Let $K \in \mathcal{K}(\mathcal{H})$ for some $\mathcal{H}$, and suppose G does not belong to $\mathcal{H}$. We can view K as a set of rules by which to edit G, in order for it to belong to $\mathcal{H}$. We begin by partitioning $V(G)$ into k parts $V_1, \ldots, V_k$ such that each part V_i corresponds to a distinct vertex v_i of K. The optimal sizes of these parts are to be determined, but for the purpose of understanding the method, we may assume that these are all equal. Then we edit (that is, we add or remove edges) according the following rules.

(i) If $v_i \in \mathrm{VB}$, we add all edges with both endpoints in V_i.

(ii) If $v_i \in \mathrm{VW}$, we remove all edges with both endpoints in V_i.

(iii) If $v_i v_j \in \mathrm{EB}$, we add all edges with one endpoint in V_i and the other in V_j.

(iv) If $v_i v_j \in \mathrm{EW}$, we remove all edges with one endpoint in V_i and the other in V_j.

Let G' be the graph obtained from G by carrying out these edits. Then G' embeds into K, and so as we have observed, $G' \in \mathcal{H}$. We will see that there will be some such CRG such that the edit distance can be determined by applying these rules to edit the graph $G(n,p)$ and then counting the expected number of edge changes which would be required.

4.1.1 Measuring the edits defined by a CRG

Suppose we are given a CRG $K \in \mathcal{K}(\mathcal{H})$, and we would like to count the expected proportion of edges of $G(n,p)$ which would be changed with respect to these rules, for some fixed p. We will define a quadratic program $g_K(p)$ which counts exactly this quantity. In order to define $g_K(p)$, we first define the matrix $\mathbf{M}_K(p)$. We label the vertices of K by $v_1, \ldots, v_k$, and let $\mathbf{M}_K(p)$ be a $k \times k$ matrix whose entries are given by

$$[\mathbf{M}_K(p)]_{ij} = \begin{cases} p & \text{if either } i = j \text{ and } v_i \in \mathrm{VW}, \text{ or } i \neq j \text{ and } v_i v_j \in \mathrm{EW}, \\ 1 - p & \text{if either } i = j \text{ and } v_i \in \mathrm{VB}, \text{ or } i \neq j \text{ and } v_i v_j \in \mathrm{EB}, \\ 0 & \text{if } i \neq j \text{ and } v_i v_j \in \mathrm{EG}. \end{cases}$$

Then we define the quadratic program

$$g_K(p) = \begin{cases} \min & \mathbf{x}^T \mathbf{M}_K(p) \mathbf{x} \\ \text{s.t.} & \mathbf{x} \cdot \mathbf{1} = 1 \\ & \mathbf{x} \geq \mathbf{0}. \end{cases} \tag{4.1.1}$$

Note that the vector $\mathbf{x}$ which has exactly one entry equal to 1 and all other entries equal to 0 is a feasible solution to this program. Thus, there is some optimal vector $\mathbf{x}^*$ which attains the minimum in the program above, and so $g_K(p)$ has a solution for every CRG K. Furthermore, for any given CRG K, we can easily find the value of $g_K(p)$ using the method of Lagrange multipliers.

The vector $\mathbf{x}^*$ depends on the matrix $\mathbf{M}_K(p)$, which captures information about the adjacencies in K. Recall that every CRG K defines a partition of the random graph $G(n,p)$, where the vertices of K represent parts in this partition. The vector $\mathbf{x}$ assigns a weight to every vertex of K. In particular, for any $v \in K$, the weight $\mathbf{x}(v)$ corresponds to the proportion of vertices of $G(n,p)$ which lie in the part corresponding to v, and the optimal weight vector $\mathbf{x}^*$ gives the assignment of vertices of $G(n,p)$ to parts in a way which minimises the expected proportion of edge changes required. Thus, the function $g_K(p)$ measures exactly the expected proportion of edge changes of $G(n,p)$ which the CRG K defines.

The following result of Alon and Stav [2] suggests that for any hereditary property $\mathcal{H}$, we can use the function $g_K(p)$ to determine the edit distance function.

Theorem 4.1.1 (Alon and Stav [2])**.** *Let $\mathcal{H}$ be a hereditary property. Then for all $p \in [0,1]$,*

$$\mathrm{ed}_{\mathcal{H}}(p) = \inf_{K \in \mathcal{K}(\mathcal{H})} g_K(p).$$

Marchant and Thomason [52] later showed that there is in fact a CRG which attains the infimum in Theorem 4.1.1, that is, they showed the following.

Theorem 4.1.2 (Marchant and Thomason [52])**.** *Let $\mathcal{H}$ be a hereditary property. Then for all $p \in [0,1]$,*

$$\mathrm{ed}_{\mathcal{H}}(p) = \min_{K \in \mathcal{K}(\mathcal{H})} g_K(p).$$

This is a key result in this area, since for any hereditary property $\mathcal{H}$, the problem of determining the edit distance function is reduced to instead finding the CRG $K \in \mathcal{K}(\mathcal{H})$ which attains the minimum in Theorem 4.1.2.

An implication of Theorem 4.1.2 is that for any $K \in \mathcal{K}(\mathcal{H})$, the function $g_K(p)$ provides an upper bound to the edit distance function. Thus, rather than examining all CRGs in $\mathcal{K}(\mathcal{H})$, we begin by only examining those CRGs which have all grey edges, and use these to obtain an upper bound to $\mathrm{ed}_{\mathcal{H}}(p)$.

We denote by $K(r, s)$ the CRG on $r + s$ vertices, which has r white vertices, s black vertices, and all its edges are grey. For any hereditary property $\mathcal{H}$, we define the *clique spectrum* to be

$$\Gamma(\mathcal{H}) = \left\{(r, s) \in \mathbb{Z}^2_{\geq 0} \colon H \not\mapsto K(r, s) \text{ for all } H \in \mathcal{F}(\mathcal{H})\right\}.$$

The important property of $\Gamma(\mathcal{H})$ which we will be using is its monotonicity. That is, if $(r, s) \in \Gamma(\mathcal{H})$, then for all $0 \leq r' \leq r$, $0 \leq s' \leq s$, we have $(r', s') \in \Gamma(\mathcal{H})$. This follows immediately from the definition, and gives rise to important elements of $\Gamma(\mathcal{H})$ known as *extreme points*, which are pairs $(r, s) \in \Gamma(\mathcal{H})$ such that $(r+1, s), (r, s+1) \notin \Gamma(\mathcal{H})$. We denote by $\Gamma^*(\mathcal{H})$ the set of extreme points of $\Gamma(\mathcal{H})$.

We state the following useful lemma, which allows us to observe another useful property of these grey edge CRGs. Let K be a CRG. We say a sub-CRG K' of K is a *component* if every edge leaving K' is grey, that is, if for all $v \in V(K')$ and all $w \in V(K \setminus K')$, we have that $vw \in \mathrm{EG}(K)$. So, every CRG has a 'decomposition' into components, that is, a partition of the vertex set such that all edges leaving the sub-CRG induced on any part in this partition are grey. Then we can state the following lemma, a result of work by Martin [53].

Lemma 4.1.3 (Martin [53])**.** *Let K be a CRG with components $K^{(1)}, \ldots, K^{(\ell)}$. Then*

$$(g_K(p))^{-1} = \sum_{i=1}^{\ell} (g_{K^{(i)}}(p))^{-1}.$$

We can now state the following useful result, which suggests that for any grey-edge CRG $K(r, s)$, it is sufficient to know r and s in order to calculate the value of $g_{K(r,s)}(p)$.

Lemma 4.1.4 (Martin [53]).

$$g_{K(r,s)}(p) = \frac{p(1-p)}{r(1-p)+sp}.$$

For any pair $(r,s) \in \Gamma^*(\mathcal{H})$, we have $K(r,s) \in \mathcal{K}(\mathcal{H})$. By minimising over all the grey edge CRGs in $\mathcal{K}(\mathcal{H})$, we obtain an upper bound for $\mathrm{ed}_{\mathcal{H}}(p)$. Formally, we define

$$\gamma_{\mathcal{H}}(p) = \min\left\{g_{K(r,s)}(p) \colon (r,s) \in \Gamma(\mathcal{H})\right\} = \min\left\{\frac{p(1-p)}{r(1-p)+sp} \colon (r,s) \in \Gamma(\mathcal{H})\right\}.$$

Then $\gamma_{\mathcal{H}}(p) \geq \mathrm{ed}_{\mathcal{H}}(p)$. Furthermore, suppose that $(r,s) \in \Gamma^*(\mathcal{H})$. Then for all $0 \leq r' \leq r$, $0 \leq s' \leq s$ we have $g_{K(r,s)}(p) \leq g_{K(r',s')}(p)$. Thus when calculating $\gamma_{\mathcal{H}}(p)$, it suffices to consider only those pairs (r,s) which are extreme points of the clique spectrum.

The advantage of this is that the value of $\gamma_{\mathcal{H}}(p)$ is determinable for any hereditary property, and thus this upper bound is easier to calculate than directly minimising the function $g_K(p)$ over all $K \in K(\mathcal{H})$. Through the course of this paper, we may refer to the CRG which 'attains $\gamma_{\mathcal{H}}(p)$' or 'attains $\mathrm{ed}_{\mathcal{H}}(p)$' for some value of p, by which we mean the CRG K for which $g_K(p) = \gamma_{\mathcal{H}}(p)$, or $g_K(p) = \mathrm{ed}_{\mathcal{H}}(p)$, respectively. We will also define the following special type of CRG. This was originally introduced by Peck [64].

Definition 4.1.5. *For a hereditary property $\mathcal{H}$, we say that a CRG K is a* candidate CRG *for $\mathcal{H}$ if $K \in \mathcal{K}(\mathcal{H})$ and $g_K(p) < \gamma_{\mathcal{H}}(p)$.*

If the hereditary property $\mathcal{H}$ is clear from context, we omit the phrase 'for $\mathcal{H}$' from Definition 4.1.5.

4.1.2 The p-core CRGs and symmetrisation

As we have seen previously, the edit distance function $\mathrm{ed}_{\mathcal{H}}(p)$ can be determined by finding the CRG $K \in \mathcal{K}(\mathcal{H})$ which minimises $g_K(p)$. We say a CRG K is *p-core* if

$g_K(p) < g_{K'}(p)$ for any sub-CRG K' of K. It is clear from the definition of these structures that any CRG which minimises $g_K(p)$ must be p-core. Marchant and Thomason [52] identified the following useful classification of these structures.

Theorem 4.1.6 (Marchant-Thomason [52])**.** *Let K be a p-core CRG. Then the following holds.*

(i) *If $p = 1/2$, then all edges of K are grey.*

(ii) *If $p < 1/2$, then $EB = \emptyset$ and there are no white edges incident to white vertices.*

(iii) *If $p > 1/2$, then $EW = \emptyset$ and there are no black edges incident to black vertices.*

We consider again the quadratic program $g_K(p)$ defined in (4.1.1). Recall that $\mathbf{x}^*$ is the vector which attains the optimum. For ease of notation, we will omit the $*$ and assume that $\mathbf{x}$ is the optimal vector. Marchant and Thomason [52] showed that if K is a p-core CRG, then this optimal vector which solves the quadratic program in (4.1.1) is in fact unique, and moreover that this optimal vector $\mathbf{x}$ contains no zero entries.

Now the vector $\mathbf{x}$ assigns to each vertex $v \in V(K)$ a weight $\mathbf{x}(v)$. We let $d_B(v)$ ($d_W(v)$, $d_G(v)$ respectively) denote the sum of weights of vertices in the neighbourhood of v which are adjacent to v via a black (white, grey respectively) edge. Martin [53] found the following bounds on the quantity $d_G(v)$ using symmetrisation techniques.

Lemma 4.1.7 (Martin [53])**.** *Let $p \in (0, 1/2]$ and let K be a p-core CRG with optimal weight function $\mathbf{x}$. Then $\mathbf{x}(v) = g_K(p)/p$ for all $v \in \mathrm{VW}(K)$. Moreover, for all $v \in \mathrm{VB}(K)$, we have*

$$d_G(v) = \frac{p - g_K(p)}{p} + \frac{1-2p}{p}\mathbf{x}(v).$$

Lemma 4.1.7 can be thought of as a symmetrisation lemma, and tells us that the weight is distributed evenly among all vertices in $\mathrm{VW}(K)$ by the vector $\mathbf{x}$. This gives a way of determining the value of $\mathbf{x}$ at any vertex. We can use this to state the following useful lemma.

Lemma 4.1.8 (Martin [53]). *Let $p \in (0, 1/2]$ and let K be a p-core CRG with optimal weight function $\mathbf{x}$. Then for all $v \in \mathrm{VB}(K)$, we have $\mathbf{x}(v) \le g_K(p)/(1-p)$.*

4.2 Determining $\mathbf{ed}_{\mathbf{Forb}(C_h^t)}(p)$

In this section we aim to determine $\mathrm{ed}_{\mathrm{Forb}(C_h^t)}(p)$. Recall that Berikkyzy, Martin and Peck [10] determined this for all p in the case when $(t+1) \nmid h$ and for $p \in [1/\lceil h/(2t+1)\rceil, 1]$ in the case when $(t+1) \mid h$. Therefore, we will be focusing on determining this function for $0 \le p \le 1/\lceil h/(2t+1)\rceil$ in the case when $(t+1) \mid h$.

For the remainder of this section, we let $\mathcal{H} = \mathrm{Forb}(C_h^t)$, for integers h and t.

4.2.1 Preliminaries

We begin by stating the value of $\gamma_{\mathcal{H}}(p)$, which was determined by Berikkyzy, Martin and Peck [10].

Lemma 4.2.1 (Berikkyzy, Martin and Peck [10]). *Let $t \ge 1$ and $h \ge \max\{t(t+1), 4\}$ be integers. Then for all $p \in [0, 1]$, the following holds.*

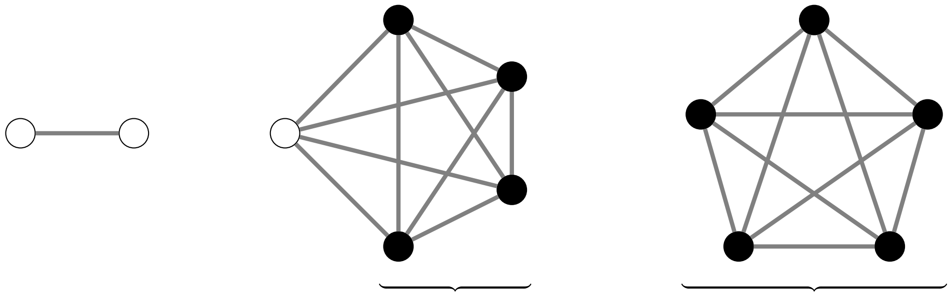
(i) *If $(t+1) \nmid h$, then*

$$\gamma_{\mathcal{H}}(p) = \min_{r \in \{0,1,\ldots,t\}} \left\{ \min\left\{ \frac{p}{t+1}, \frac{p(1-p)}{r(1-p) + \left(\left\lceil \frac{h}{t+r+1} \right\rceil - 1\right)p} \right\} \right\}.$$

(ii) *If $(t+1) \mid h$ then*

$$\gamma_{\mathcal{H}}(p) = \min_{r \in \{0,1,\ldots,t\}} \left\{ \frac{p(1-p)}{r(1-p) + \left(\left\lceil \frac{h}{t+r+1} \right\rceil - 1\right)p} \right\}.$$

In Fig. 4.1, we see the structures which correspond to $\Gamma^*(\mathrm{Forb}(C_h))$ and by applying Lemma 4.1.4, we obtain the values of $g_K(p)$ for each of these CRGs K.



$\lceil h/3 \rceil - 1$ black vertices $\lceil h/2 \rceil - 1$ black vertices

$K(2,0)$ (for odd h) $K(1, \lceil h/3 \rceil - 1)$ $K(0, \lceil h/2 \rceil - 1)$

Figure 4.1: The structures which correspond to $\Gamma^*(\mathrm{Forb}(C_h))$.

This gives $\gamma_{\mathrm{Forb}(C_h)}(p)$, which is exactly equal to Lemma 4.2.1 in the case $t = 1$. We recall the notion of a candidate CRG as in Definition 4.1.5, that is, consider a CRG $K \in \mathcal{K}(\mathcal{H})$ such that $g_K(p) < \gamma_{\mathcal{H}}(p)$ for some p. By Theorem 1.2.2, we know that in order for such a CRG to exist, we must have $(t+1) \mid h$ and $p \leq 1/\lceil h/(2t+1) \rceil$. We will determine some characteristics for such a candidate CRG.

By taking sub-CRGs, we may assume that K is p-core. Thus, in particular by Theorem 4.1.6, in the case when $p \leq 1/\lceil h/(2t+1) \rceil$, we may assume that all edges of K are grey or white, and that there are no white edges incident to white vertices. Here, we are assuming that $1/\lceil h/(2t+1) \rceil \leq 1/2$, which is true provided that

$$h \geq 2(2t+1).$$

So, in particular, we have this lower bound on h throughout. Recall that VW is the set of white vertices of K, and VB is the set of black vertices of K. Let $K_B = K[\mathrm{VB}]$. The following lemma gives some structural conditions on K.

Lemma 4.2.2 (Berikkyzy, Martin and Peck [10]). *Let $p \in (0, 1/2]$ and let $t \geq 1$ and $h \geq 2t+2$ be integers. Let $K \in \mathcal{K}(\mathcal{H})$ be p-core with r white vertices.*

(i) *If $r \in \{0, \ldots, t-1\}$ and $h \geq t(t-1)$, then K_B contains no grey cycle of length in $\left\{\left\lceil \frac{h}{t+r+1} \right\rceil, \ldots, \left\lfloor \frac{h}{t} \right\rfloor\right\}$.*

(ii) *If* $r = t$, *then* $|V(K_B)| \leq \left\lceil \frac{h}{2t+1} \right\rceil - 1$.

(iii) *If* $r \geq t+1$, *then* $(t+1) \nmid h$ *and* $V(K_B) = \emptyset$.

We will determine further characteristics of such a candidate CRG in the case that $(t+1) \mid h$ and $p < 1/\lceil h/(2t+1) \rceil$, with the aim of finding a contradiction and thus disproving its existence. We begin by determining the value of $\gamma_{\mathcal{H}}(p)$ for small p.

Lemma 4.2.3. *Let* $t \geq 1$ *and* $h \geq t(2t+1)$ *be integers with* $(t+1) \mid h$. *Then for* $p \leq 1/\lceil h/(2t+1) \rceil$, *we have*

$$\gamma_{\mathcal{H}}(p) = \frac{p(1-p)}{t(1-p) + \left(\left\lceil \frac{h}{2t+1} \right\rceil - 1 \right) p}.$$

Proof. We would like to show that when $p \leq 1/\lceil h/(2t+1) \rceil$, for every $r \in \{0, \ldots, t-1\}$, we have

$$\frac{p(1-p)}{r(1-p) + \left(\left\lceil \frac{h}{t+r+1} \right\rceil - 1 \right) p} \geq \frac{p(1-p)}{t(1-p) + \left(\left\lceil \frac{h}{2t+1} \right\rceil - 1 \right) p},$$

that is, the minimum in Lemma 4.2.1 (ii) is attained when $r = t$. By rearranging, we get that the inequality above is equivalent to showing

$$t - r + \left(\left\lceil \frac{h}{2t+1} \right\rceil - \left\lceil \frac{h}{t+r+1} \right\rceil - t + r \right) p \geq 0. \tag{4.2.1}$$

Now, note that

$$\begin{aligned} \left\lceil \frac{h}{2t+1} \right\rceil - \left\lceil \frac{h}{t+r+1} \right\rceil - t + r &\leq \frac{h}{2t+1} + 1 - \frac{h}{t+r+1} - t + r \\ &= \frac{-h(t-r)}{(2t+1)(t+r+1)} - (t-r-1) < 0. \end{aligned}$$

Therefore, this function is decreasing, and so the function in (4.2.1) is minimised at the upper bound on p, that is, when $p = 1/\lceil h/(2t+1) \rceil$. Combined with the fact that $r \leq t-1$ and by applying the crude upper and lower bounds on the ceiling function, we have

$$t-r+\left(\left\lceil\frac{h}{2t+1}\right\rceil-\left\lceil\frac{h}{t+r+1}\right\rceil-t+r\right)p \geq t-r+1-\frac{\left\lceil\frac{h}{t+r+1}\right\rceil}{\left\lceil\frac{h}{2t+1}\right\rceil}-\frac{t-r}{\left\lceil\frac{h}{2t+1}\right\rceil}$$
$$\geq t-r+1-\frac{h(2t+1)-(2t+1)(t+r+1)(t-r-1)}{h(t+r+1)} \tag{4.2.2}$$

We require (4.2.2) to be at least 0 in order for (4.2.1) to hold. In other words, by rearranging (4.2.2), we show that we require

$$h \geq \frac{(t+r+1)(t-r-1)(2t+1)}{(t+r)(t-r)} \tag{4.2.3}$$

Note that since $0 \leq r \leq t-1$,

$$\frac{(t+r+1)(t-r-1)(2t+1)}{(t+r)(t-r)} \leq \frac{(2t+1)^2(t-1)}{2t-1}.$$

On the other hand, $h \geq t(2t+1)$ and therefore, since

$$\begin{aligned} t(2t+1)-\frac{(2t+1)^2(t-1)}{2t-1} &= \frac{(2t^2+t)(2t-1)-(2t+1)^2(t-1)}{2t-1} \\ &= \frac{4t^3+2t^2-2t^2-t-(4t^3+4t^2+t-4t^2-4t-1)}{2t-1} \\ &= \frac{2t+1}{2t-1} > 0, \end{aligned}$$

whenever $t \geq 1$. Therefore, whenever $h \geq t(2t+1)$, (4.2.2) is at least 0, and therefore, (4.2.1) holds, and thus the lemma holds. □

Recall that for any pair of vertices $v, w \in V(K)$, $\mathbf{x}(v)$ is the weight assigned to vertex v by optimal vector $\mathbf{x}$ found by the quadratic program $g_K(p)$, $d_G(v)$ is the sum of the weights of all vertices incident to v via a grey edge, and $\deg_G(v)$ is the number of vertices adjacent to V via a grey edge. We define $d_G^W(v)$ be the sum of the weights of white vertices adjacent to v via a grey edge, and let $d_G^B(v)$ be the analogous quantity for black vertices. Additionally, let $\deg_G^B(v)$ be the number of black vertices adjacent to v via a grey edge. We further define $d_G(v,w)$ to be the weight of the common grey neighbourhood of v and w, and extend the definitions given above to common neighbourhoods. Recall also that for a set of vertices S, $\mathbf{x}(S)$ is the sum of weights of vertices in that set. In the following proposition, we state some important properties which the vertices of any candidate CRG K must satisfy.

Proposition 4.2.4. *Let $t \geq 1$ and $h \geq t(2t+1)$ be integers with $(t+1) \mid h$. Let K be a p-core candidate CRG, for $p \leq 1/\lceil h/(2t+1) \rceil$ with $g_K(p) = \gamma_{\mathcal{H}}(p) - \varepsilon$ for some $\varepsilon > 0$. Suppose that K has r white vertices, for $r \in \{0, \dots, t-1\}$. Then for any $v \in VB(K)$, the following hold.*

(i) $d_G(v) = \dfrac{t-1+\left(\left\lceil \frac{h}{2t+1} \right\rceil - t\right)p}{t+\left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right)p} + \dfrac{\varepsilon}{p} + \dfrac{1-2p}{p}\mathbf{x}(v).$

(ii) $d_G^B(v) = \dfrac{t-r-1+\left(\left\lceil \frac{h}{2t+1} \right\rceil - t + r\right)p}{t+\left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right)p} + \dfrac{(r+1)\varepsilon}{p} + \dfrac{1-2p}{p}\mathbf{x}(v).$

(iii) $\mathbf{x}(v) \leq \dfrac{p}{t+\left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right)p} - \dfrac{\varepsilon}{1-p}.$

(iv) $\deg_G^B(v) > \left(\left\lceil \frac{h}{2t+1} \right\rceil - 1\right)(t-r).$

Proof. Let $v \in \mathrm{VB}(K)$. To prove (i), we apply Lemma 4.1.7 to get that

$$\begin{aligned} d_G(v) &= \frac{p - g_K(p)}{p} + \frac{1-2p}{p}\mathbf{x}(v) = 1 - \frac{\gamma_{\mathcal{H}}(p) - \varepsilon}{p} + \frac{1-2p}{p}\mathbf{x}(v) \\ &= 1 - \frac{1-p}{t+\left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right)p} + \frac{\varepsilon}{p} + \frac{1-2p}{p}\mathbf{x}(v) \\ &= \frac{t-1+\left(\left\lceil \frac{h}{2t+1} \right\rceil - t\right)p}{t+\left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right)p} + \frac{\varepsilon}{p} + \frac{1-2p}{p}\mathbf{x}(v). \end{aligned}$$

Here, the third equality holds by applying Lemma 4.2.3, since $h \geq t(2t+1)$ and $p \leq 1/\lceil h/(2t+1) \rceil$.

To prove (ii), we note that since K is p-core for $p \leq 1/2$, every edge incident to a white vertex in K is grey by Theorem 4.1.6. Furthermore, by Lemma 4.1.7, for any white vertex $u \in \mathrm{VW}(K)$, we have

$$\mathbf{x}(u) = \frac{g_K(p)}{p} = \frac{1-p}{t + \left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right) p} - \frac{\varepsilon}{p}. \tag{4.2.4}$$

Here again the second equality holds by an application of Lemma 4.2.3. Then $d_G^B(v) = d_G(v) - d_G^W(v) = d_G(v) - \mathbf{x}(\mathrm{VW})$, and we obtain the result by observing that $\mathbf{x}(\mathrm{VW})$ is r times the bound given in (4.2.4), and subtracting this from the bound on $d_G(v)$ given in (i).

To prove (iii), we note that $\mathbf{x}(\mathrm{VB}) \geq \mathbf{x}(v) + d_G^B(v)$. On the other hand, we also have $\mathbf{x}(\mathrm{VB}) = 1 - \mathbf{x}(\mathrm{VW})$. Thus, combining these gives

$$\begin{aligned}
\mathbf{x}(v) &\leq 1 - \mathbf{x}(\mathrm{VW}) - d_G^B(v) \\
&= 1 - \frac{r(1-p)}{t + \left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right) p} + \frac{r\varepsilon}{p} - \frac{t - r - 1 + \left(\left\lceil \frac{h}{2t+1} \right\rceil - t + r\right) p}{t + \left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right) p} \\
&\quad - \frac{(r+1)\varepsilon}{p} - \frac{1-2p}{p}\mathbf{x}(v) \\
&= \frac{1-p}{t + \left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right) p} - \frac{\varepsilon}{p} - \frac{1-2p}{p}\mathbf{x}(v).
\end{aligned}$$

Rearranging and solving for $\mathbf{x}(v)$ gives the result.

Finally, to prove (iv), we have

$$\begin{aligned}
\deg_G^B(v) &\geq \left\lceil \frac{d_G^B(v)}{\max_{u \in \mathrm{VB}} \mathbf{x}(u)} \right\rceil > \frac{t - r - 1 + \left(\left\lceil \frac{h}{2t+1} \right\rceil - t + r\right) p}{p} \\
&\geq \left(\left\lceil \frac{h}{2t+1} \right\rceil - 1\right)(t - r),
\end{aligned}$$

where the final inequality holds because $p \leq 1/\left\lceil h/(2t+1) \right\rceil$.

□

Note that when applying this result, we will usually ignore the exact value of ε and rely only on the fact that $g_K(p) < \gamma_{\mathcal{H}}(p)$. Finally, we state the following fact, which follows immediately from Cases 2 and 3 in the proof of Theorem 3 in the work of Berikkyzy, Martin, and Peck [10]. This allows us to now restrict to the case when the candidate

Fact 4.2.5. *Let $p \in (0, 1/2]$ and let $t \geq 1$ and $h \geq 2t(t+1)+1$. If K is a p-core CRG with $g_K(p) < \gamma_{\mathcal{H}}(p)$, that is, a candidate CRG, then K has exactly $t-1$ white vertices.*

4.2.2 Existence of long cycles

Before stating and proving the key lemmas in this section, we define some useful notation relating to paths and cycles. We will define these in terms of labelled paths, and remark that the corresponding definitions also apply for cycles. We say the *length* of a path P is the number of edges in P, and we denote this by $|P|$. If P is a labelled path, say $P = v_1 \ldots v_\ell$, then the *successor* of a vertex v_i on P is the vertex v_{i+1}. The *predecessor* of v_i on P is the vertex v_{i-1}. For indices $1 \leq i < j \leq \ell$, the subpath $v_i\, P\, v_j$ is the path $v_i\, v_{i+1} \ldots v_{j-1}\, v_j$, that is, the path which has initial vertex v_i and final vertex v_j and follows the labelling of P. The path $v_j\, P^{-1}\, v_i$ is the path which has initial vertex v_j, final vertex v_i and follows the reverse of the labelling of P. We also recall that for a set $S \subseteq V(K)$, the value $\mathbf{x}(S)$ denotes the sum of weights of vertices in that set, that is, $\mathbf{x}(S) = \sum_{v \in S} \mathbf{x}(v)$. We now prove a lemma which shows that for sufficiently small values of p, any p-core candidate CRG which has $t-1$ white vertices contains a grey cycle of length at least $\lceil h/2t \rceil$ in the subgraph induced on the black vertices. Recall that $K_B = K[\mathrm{VB}]$.

Lemma 4.2.6. *Let $t \geq 1$, and $h \geq 4t(2t+1)\}$ be integers with $(t+1) \mid h$. Let K be a p-core candidate CRG for some $p \leq 1/\lceil h/(2t+1) \rceil$, and suppose that K contains $t-1$ white vertices. Then K_B contains a grey cycle of length at least $\lceil h/2t \rceil$.*

Proof. We would like to find a grey cycle of length at least $\lceil h/2t+1 \rceil$. Recall that all edges of K_B are either grey or white. Assume for a contradiction that the longest grey cycle in K_B has length at most $\lceil h/2t \rceil - 1$. By Proposition 4.2.4 (ii), for any vertex $v \in \mathrm{VB}$, we have that

$$d_G^B(v) > \frac{\left(\left\lceil \frac{h}{2t+1} \right\rceil - 1\right) p}{t + \left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right) p} + \frac{1-2p}{p} \mathbf{x}(v). \tag{4.2.5}$$

Let P be a longest grey path in K_B. Let $P = v_1 \dots v_\ell$ so that v_1 and v_ℓ are the endvertices of P and $v_i v_{i+1}$ are edges of P for each $i \in [\ell - 1]$. Note that by the maximality of P, each grey neighbour of v in VB must lie on P. Let $Q = \{v_2, \dots, v_m\}$ where m is the largest index in $[\ell]$ such that $m \leq \lceil h/2t \rceil$ (so $Q \cup \{v_1\}$ is the set of the first $m+1$ vertices of P). The aim is to show that v_1 must have a grey neighbour in VB which lies outside Q, and this will give us a cycle of length at least $\lceil h/2t \rceil$. So we assume that $|Q| = \lceil h/2t \rceil - 2$, that is $Q = \{v_2, \dots, v_{\lceil h/2t \rceil - 1}\}$.

Consider $b \in \{2, \dots, \lceil h/2t \rceil - 1\}$ such that $v_1 v_b$ is an edge. Then by applying a rotation, we can find a grey path $P_b = v_{b-1} P^{-1} v_1 v_b P v_\ell$ which has initial vertex v_{b-1} and has the same length as P. By the maximality of P, the path P_b must also be a longest grey path in K_B and so the grey neighbourhood in VB of v_{b-1} must also lie entirely on P_b. Furthermore, note that the set of vertices $Q \cup \{v_1\}$ are still the first $\lceil h/2t \rceil - 1$ vertices of the path P_b. Therefore, if v_{b-1} had a grey neighbour on P outside $Q \cup \{v_1\}$, we would find a cycle of length at least $\lceil h/2t \rceil$, as required. So, we may assume that the grey neighbourhood of v_{b-1} in VB lies entirely in the set $Q \cup \{v_1\}$. In particular, the set of the first $\lceil h/(2t) \rceil - 1$ vertices of P is the same as the set of the first $\lceil h/2t \rceil - 1$ vertices of P', albeit that they have been rearranged.

We know by Proposition 4.2.4 (iv) that v_1 has at least $\lceil h/(2t+1) \rceil$ grey neighbours in VB. Say v_1 has exactly $\lceil h/(2t+1) \rceil + y$ grey neighbours in VB for some $y \in \{0, \dots, \lceil h/2t \rceil - \lceil h/(2t+1) \rceil - 2\}$. Let Q' be the set of vertices in Q which are predecessors in P of grey neighbours of v_1, that is, the set of vertices $v_j \in V(P)$ and $v_1 v_{j+1}$ is a grey edge and $j \in Q$. Then $|Q'| = \lceil h/(2t+1) \rceil + y - 1$ (since v_1 itself is a predecessor of v_2 on P, but v_1 does not lie in Q). Let $Q'' = Q \setminus Q'$. Then $|Q''| = \lceil h/2t \rceil - \lceil h/(2t+1) \rceil - y - 1$. Now by the upper bound on the weight of a black vertex given in Proposition 4.2.4 (iii), we obtain that

$$\mathbf{x}(Q) \leq \frac{\left(\left\lceil \frac{h}{2t} \right\rceil - 2\right) p}{t + \left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right) p}. \tag{4.2.6}$$

On the other hand, since $\mathbf{x}(Q) \geq d_G^B(v_1)$, we have

$$\mathbf{x}(Q) > \frac{\left(\left\lceil \frac{h}{2t+1} \right\rceil - 1\right) p}{t + \left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right) p} + \frac{1-2p}{p} \mathbf{x}(v_1). \tag{4.2.7}$$

Furthermore, again by the upper bound on the weight of a black vertex, we have

$$\mathbf{x}(Q'') \leq \frac{\left(\left\lceil \frac{h}{2t} \right\rceil - \left\lceil \frac{h}{2t+1} \right\rceil - y - 1\right) p}{t + \left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right) p}. \tag{4.2.8}$$

Therefore, since $\mathbf{x}(Q') = \mathbf{x}(Q) - \mathbf{x}(Q'')$, we have

$$\mathbf{x}(Q') > \frac{\left(2\left\lceil \frac{h}{2t+1} \right\rceil - \left\lceil \frac{h}{2t} \right\rceil + y\right) p}{t + \left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right) p} + \frac{1-2p}{p} \mathbf{x}(v_1). \tag{4.2.9}$$

Therefore, by averaging over the weights of all vertices in Q', there exists some vertex $u \in Q'$ such that

$$\mathbf{x}(u) > \frac{\left(2\left\lceil \frac{h}{2t+1} \right\rceil - \left\lceil \frac{h}{2t} \right\rceil + y\right) p}{\left(\left\lceil \frac{h}{2t+1} \right\rceil + y - 1\right)\left(t + \left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right) p\right)} + \frac{1-2p}{\left(\left\lceil \frac{h}{2t+1} \right\rceil + y - 1\right) p} \mathbf{x}(v_1). \tag{4.2.10}$$

When $h \geq 4t(2t+1)$, we have that

$$\mathbf{x}(u) > \frac{\left(2\left\lceil \frac{h}{2t+1} \right\rceil - \left\lceil \frac{h}{2t} \right\rceil\right) p}{\left(\left\lceil \frac{h}{2t+1} \right\rceil - 1\right)\left(t + \left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right) p\right)} + \frac{1-2p}{\left(\left\lceil \frac{h}{2t} \right\rceil - 3\right) p} \mathbf{x}(v_1). \tag{4.2.11}$$

This bound holds because for $(x+y)/(x'+y) \geq x/x'$ whenever $x' \geq x$. So, if we let $x = 2\lceil h/(2t+1) \rceil - \lceil h/2t \rceil$ and $x' = \lceil h/(2t+1) \rceil - 1$, then it suffices to have $x \leq x'$, and in order for this to be true, it suffices to have $h \geq 4t(2t+1)$. Now note that

$$\mathbf{x}(Q \cup \{v_1\}) \leq \frac{\left(\left\lceil \frac{h}{2t} \right\rceil - 2\right) p}{t + \left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right) p} + \mathbf{x}(v_1). \tag{4.2.12}$$

On the other hand, since $u \in Q'$, it is the predecessor on P to some grey neighbour of v_1 and so as we have seen already, the entire grey neighbourhood of u in K_B must lie in $Q \cup \{v_1\}$. Therefore,

$$\begin{aligned}
&\mathbf{x}(Q \cup \{v_1\}) > \frac{\left(\left\lceil \frac{h}{2t+1} \right\rceil - 1\right) p}{t + \left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right) p} + \frac{1-p}{p}\mathbf{x}(v) \\
&> \frac{\left(\left\lceil \frac{h}{2t+1} \right\rceil - 1\right) p}{t + \left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right) p} + \frac{\left(2\left\lceil \frac{h}{2t+1} \right\rceil - \left\lceil \frac{h}{2t} \right\rceil\right)(1-p)}{\left(\left\lceil \frac{h}{2t+1} \right\rceil - 1\right)\left(t + \left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right) p\right)} \\
&\quad + \frac{(1-2p)(1-p)}{\left(\left\lceil \frac{h}{2t} \right\rceil - 3\right) p^2}\mathbf{x}(v_1) \\
&= \frac{2\left\lceil \frac{h}{2t+1} \right\rceil - \left\lceil \frac{h}{2t} \right\rceil + \left(\left\lceil \frac{h}{2t+1} \right\rceil^2 - 4\left\lceil \frac{h}{2t+1} \right\rceil + \left\lceil \frac{h}{2t} \right\rceil + 1\right) p}{\left(\left\lceil \frac{h}{2t+1} \right\rceil - 1\right)\left(t + \left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right) p\right)} + \frac{(1-2p)(1-p)}{\left(\left\lceil \frac{h}{2t} \right\rceil - 3\right) p^2}\mathbf{x}(v_1).
\end{aligned} \tag{4.2.13}$$

Now we prove the following claim.

Claim 4.2.7. *If $t \geq 1$, $h \geq 4(2t+1)$ and $p \leq 1/\lceil h/(2t+1) \rceil$, we have*

$$\frac{(1-2p)(1-p)}{\left(\left\lceil \frac{h}{2t} \right\rceil - 3\right) p^2} \geq 1.$$

Proof of Claim 4.2.7. We would like to show that when $p \leq 1/\lceil h/(2t+1) \rceil$, we have $(1-2p)(1-p)/p^2 \geq \lceil h/2t \rceil - 3$. It suffices to show that whenever $p \leq (2t+1)/h$, we have

$$\frac{1-3p}{p^2} \geq \frac{h}{2t} - 2.$$

Since the left side of the equation above is decreasing in p in the interval $(0, 2/3)$, it suffices to show that this holds for the upper bound on p, that is, it suffices to show

$$1 - 3\left(\frac{2t+1}{h}\right) \geq \left(\frac{2t+1}{h}\right)^2 \left(\frac{h}{2t} - 2\right). \tag{4.2.14}$$

Now suppose that $h = k(2t+1)$ for some $k > 0$. Then in order for the inequality in (4.2.14) to hold, we need

$$1 - \frac{3}{k} \geq \frac{1}{k^2}\left(\frac{k(2t+1)}{2t} - 2\right).$$

or equivalently, we need $k^2 - k(4 + 1/2t) + 2 \geq 0$. In particular, since $t \geq 1$, it suffices to have $k^2 - 4.5k + 2 \geq 0$. The left side of this inequality is a positive quadratic with roots at $k = 0.5$ and $k = 4$, so in particular, this holds whenever $k \geq 4$. Therefore, in particular, it suffices to have $h \geq 4(2t+1)$. □

Note that we have $h \geq 4t(2t+1) \geq 4(2t+1)$ and $p \leq 1/\left\lceil h/(2t+1)\right\rceil$. Therefore, by applying Claim 4.2.7 to (4.2.13), we get

$$\mathbf{x}(Q \cup \{v_1\}) > \frac{2\left\lceil \frac{h}{2t+1}\right\rceil - \left\lceil \frac{h}{2t}\right\rceil + \left(\left\lceil \frac{h}{2t+1}\right\rceil^2 - 4\left\lceil \frac{h}{2t+1}\right\rceil + \left\lceil \frac{h}{2t}\right\rceil + 1\right)p}{\left(\left\lceil \frac{h}{2t+1}\right\rceil - 1\right)\left(t + \left(\left\lceil \frac{h}{2t+1}\right\rceil - t - 1\right)p\right)} + \mathbf{x}(v_1). \tag{4.2.15}$$

Combining this with the upper bound on $\mathbf{x}(Q \cup \{v_1\})$ given in (4.2.12), and rearranging, we get that

$$p > \frac{2\left\lceil \frac{h}{2t+1}\right\rceil - \left\lceil \frac{h}{2t}\right\rceil}{\left\lceil \frac{h}{2t}\right\rceil \left\lceil \frac{h}{2t+1}\right\rceil - \left\lceil \frac{h}{2t+1}\right\rceil^2 - 2\left\lceil \frac{h}{2t}\right\rceil + 2\left\lceil \frac{h}{2t+1}\right\rceil + 1}.$$

However, when $h \geq 4t(2t+1)$, the inequality above contradicts the assumption that $p \leq 1/\left\lceil h/(2t+1)\right\rceil$. For $t \geq 2$, we can show this by the fact that $m \leq \lceil m \rceil \leq m + 1$. To show this for $t = 1$, this crude upper and lower bound is not quite good enough and we can instead show this by splitting into 3 cases depending on the value of h mod 3. Therefore, the black-vertex subgraph K_B must contain a grey cycle of length at least $\lceil h/2t \rceil$. $\square$

4.2.3 Using long cycles to find shorter cycles

In this section, we prove two key lemmas. First, we show that given a graph satisfying a particular condition on the degree, we can use 'long' cycles to find shorter cycles in the graph.

Lemma 4.2.8. *Let G be a graph, $m \geq 5$ a positive integer. Suppose that in any set of $\lfloor m/3 \rfloor$ vertices of G, there are some two which have a common neighbour. If G contains a cycle of length at least m, then it also contains a cycle of length between $\lceil m/2 \rceil$ and $m-1$.*

Proof. Let G be as above, and let C be a shortest cycle in G of length at least m. The idea of the proof is as follows. We carefully choose a set of $\lfloor m/3 \rfloor$ vertices of C. Then the hypothesis suggests that these have a common neighbour. We use this common neighbour to find a new cycle C' which has length strictly shorter than C which is still bounded below by $\lceil m/2 \rceil$. By minimality of C, we must have that C' has length at most $m-1$, thus proving the result. It remains only to choose the set of vertices, and show that we can always find a new cycle C'.

We begin by labelling the vertices of C by $u_1, \ldots, u_\ell$, so that $u_i u_{i+1}$ is an edge for each $i \in [\ell]$ (indices taken with addition modulo ℓ). Let M be the subset of vertices of C given by

$$M = \left\{ u_{3i-2} \colon i \in \left\{ 1, \ldots, \left\lfloor \frac{m}{3} \right\rfloor \right\} \right\}.$$

Since M contains exactly $\lfloor m/3 \rfloor$ vertices, we know by hypothesis that there exist two distinct vertices $u_i, u_j \in M$ which have a common neighbour v in G. Without loss of generality, we may assume $i < j$. By choice of M, the paths $u_i \, C \, u_j$ and $u_j \, C \, u_i$ have length at least 3. Hence, the path $u_i \, v \, u_j$ requires at least one edge which does not belong to the cycle C.

We have two cases to consider. Recall that in each case, we aim to find a new cycle C' such that $\lceil m/2 \rceil \leq |C'| \leq |C| - 1$.

Case 1: v lies outside the cycle C. Now note that $3 \leq j - i \leq m - 3$. If $j - i \geq \lceil m/2 \rceil - 2$ then let $C' = u_i \, C \, u_j \, v \, u_i$. Then $|C'| = (j - i) + 2$. Hence,

$$\left\lceil \frac{m}{2} \right\rceil = \left\lceil \frac{m}{2} \right\rceil - 2 + 2 \leq |C'| \leq |C| - 3 + 2 = |C| - 1,$$

as required. On the other hand, if $j - i \leq \lceil m/2 \rceil - 2$, then let $C' = u_i \, v \, u_j \, C \, u_i$. Then $|C'| = |C| - (j - i) + 2$ and

$$\left\lceil \frac{m}{2} \right\rceil \leq m - \left\lceil \frac{m}{2} \right\rceil + 2 + 2 \leq |C'| \leq |C| - 1,$$

as required. Thus, if the common neighbour lies outside C we are done.

Case 2: v lies on the cycle C. Since both the paths $u_i \, C \, u_j$ and $u_j \, C \, u_i$ have length at least 3, and the vertex v must lie on one of these paths, we may assume without loss of generality that both of the paths $P_1 = u_i \, C \, v$ and $P_2 = v \, C \, u_i$ have length at length at least 2. In particular, the edge $u_i \, v$ is a chord on C. Thus, either P_1 or P_2 have length at least $\lceil |C| / 2 \rceil$, say this is P_1 without loss of generality, and let $C' = P_1 \, u_i$. Then

$$\left\lceil \frac{m}{2} \right\rceil \leq \frac{|C|}{2} + 1 \leq |C'| \leq |C| - 1,$$

as required. Thus, if the common neighbour lies on the cycle, then we are done once again. $\square$

The following lemma gives a condition on the common neighbourhoods of vertices in a weighted graph. For any vertex $v \in V(K)$, let $N_G(v)$ be the set of grey neighbours of v in $V(K)$, that is,

$$N_G(v) = \{u \in V(K) \colon uv \in \mathrm{EG}\}.$$

Lemma 4.2.9. *Let K be a CRG with all black vertices and all edges grey or white. Let $m \geq 2$ be an integer. If $d_G(v) > 1/m$ for all $v \in V(K)$, then in any subset $M \subseteq V(K)$ of size at least m, there exist at least 2 vertices which have a common neighbour in their grey neighbourhood.*

Proof. Consider a set $M \subseteq V(K)$ with $|M| \geq m$. Suppose that every pair of vertices in M have a disjoint neighbourhood. Then

$$1 \geq \left| \bigcup_{v \in M} N_G(v) \right| = \sum_{v \in M} |N_G(v)| = \sum_{v \in M} d_G(v) > 1,$$

a contradiction. Here, the first inequality holds since the sum of weights of all vertices in K is exactly 1, the second equality holds by the assumption that every pair of vertices have a disjoint neighbourhood and the final inequality holds because each term of the sum has size at least $1/m$ by assumption. □

4.2.4 Proof of main result

It remains only to combine the lemmas we have seen in this section to prove the main result.

Proof of Theorem 1.1.2. Let $t \geq 1$ and $h \geq 4t(2t+1)$. Recall $c_0 = \lfloor (\lfloor h/t \rfloor + 1)/3 \rfloor$, $\ell_0 = \lceil h/(2t+1) \rceil$, and $p_0 = t/(c_0\ell_0 - c_0 - \ell_0 + t + 1)$, with $p \in [p_0, 1/\lceil h/(2t+1) \rceil]$.

Assume for a contradiction that there exists a p-core candidate CRG K. Suppose that K has r white vertices. By Fact 4.2.5, we $r = t - 1$. By Lemma 4.2.6, we know that K_B contains a grey cycle of length at least $\lceil h/(2t) \rceil$. By Lemma 4.2.2, K_B cannot contain a grey cycle of length ℓ for any $\ell \in \{\lceil h/(2t) \rceil, \ldots, \lfloor h/t \rfloor\}$. Combining these facts together, we may assume that K_B contains a grey cycle of length at least $\lfloor h/t \rfloor + 1$.

Now clearly, $\mathbf{x}(\mathrm{VB}) \leq 1$. Furthermore, by Proposition 4.2.4 (ii) as there are exactly $t-1$ white vertices, we know that for any vertex $v \in \mathrm{VB}$,

$$d_G^B(v) \geq \frac{\left(\left\lceil \frac{h}{2t+1} \right\rceil - 1\right) p}{t + \left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right) p} + \frac{1-2p}{p}\mathbf{x}(v). \tag{4.2.16}$$

Claim 4.2.10. *When $p \geq p_0$, we have that $d_G^B(v) > 1/c_0$.*

Proof of Claim 4.2.10. It suffices to show that when $p \geq p_0$,

$$\frac{\left(\left\lceil \frac{h}{2t+1} \right\rceil - 1\right) p}{t + \left(\left\lceil \frac{h}{2t+1} \right\rceil - t - 1\right) p} > \frac{1}{c_0},$$

as by (4.2.16), this implies the claim. This is equivalent to showing

$$\left(\left(\left\lceil \frac{h}{2t+1} \right\rceil - 1\right) c_0 - \left\lceil \frac{h}{2t+1} \right\rceil + t + 1\right) p > t. \tag{4.2.17}$$

Now note that

$$\begin{aligned}
&\left(\left\lceil \frac{h}{2t+1} \right\rceil - 1\right) \left\lfloor \frac{\left\lfloor \frac{h}{t} \right\rfloor + 1}{3} \right\rfloor - \left\lceil \frac{h}{2t+1} \right\rceil + t + 1 \\
&\geq \left(\frac{h}{2t+1} - 1\right)\left(\frac{h}{3t} - 1\right) - \frac{h}{2t+1} - 1 + t + 1 \\
&= \frac{h^2}{3t(2t+1)} - \frac{h}{3t} - \frac{h}{2t+1} + 1 + t \\
&= \frac{h^2 - h(5t+1) + 3t(t+1)(2t+1)}{3t(2t+1)}.
\end{aligned}$$

This is positive whenever $h^2 - h(5t+1) + 3t(t+1)(2t+1) \geq 0$. Indeed, this has roots at

$$t_0^{\pm} = \frac{(5t+1) \pm \sqrt{(5t-1)^2 - 12t(t+1)(2t+1)}}{2}.$$

For $t \geq 4$, the quantity $(5t-1)^2 - 12t(t+1)(2t+1)$ is negative, and so this is a positive quadratic with no roots, and so is always positive. On the other hand, for $t = 1$, this is a positive quadratic with larger root between 5 and 5.5, and so for $h \geq 4t(2t+1) \geq 6$, the equation is positive. For $t = 2$, this has larger root between

8 and 8.5 and so for $h \geq 4t(2t+1) \geq 8$, the equation is positive. Finally, for $t = 3$, this has larger root at 9, and therefore, for $h \geq 4t(2t+1) \geq 9$, the equation is always positive.

In particular, when $t \geq 1$ and $h \geq 4t(2t+1)$, the inequality $h^2 - h(5t+1) + 3t(t+1)(2t+1) \geq 0$ holds. Therefore, we can rearrange (4.2.17) and get that

$$p \geq \frac{t}{\left(\left\lceil \frac{h}{2t+1} \right\rceil - 1\right) \left\lceil \frac{\lfloor \frac{h}{t} \rfloor + 1}{3} \right\rceil - \left\lceil \frac{h}{2t+1} \right\rceil + t + 1} = p_0. \tag{4.2.18}$$

□

So by Claim 4.2.10, since $p \geq p_0$, any vertex $v \in \mathrm{VB}(K)$ satisfies $d_G^B(v) > \mathbf{x}(\mathrm{VB}) / \lceil (\lceil h/t \rceil + 1)/3 \rceil$. Therefore by Lemma 4.2.9, in any set of $\left\lceil \frac{\lfloor h/t \rfloor + 1}{3} \right\rceil$ vertices of K_B, there are some two which have a common grey neighbour in K_B. Thus, we can apply Lemma 4.2.8 to show that K_B must contain a grey cycle of length ℓ for some $\ell \in \{\lceil (\lfloor h/t \rfloor + 1)/2 \rceil, \ldots, \lfloor h/t \rfloor\}$. Note that $\lceil (\lfloor h/t \rfloor + 1)/2 \rceil \geq \lceil h/(2t) \rceil$. Indeed, if $h = x + yt$ for $x \in \{0, \ldots, t-1\}$, then

$$\left\lceil \frac{\lfloor \frac{h}{t} \rfloor + 1}{2} \right\rceil - \left\lceil \frac{h}{2t} \right\rceil = \left\lceil \frac{\lfloor \frac{x+yt}{t} \rfloor + 1}{2} \right\rceil - \left\lceil \frac{x+yt}{2t} \right\rceil = \left\lceil \frac{y}{2} + \frac{1}{2} \right\rceil - \left\lceil \frac{y}{2} + \frac{x}{2t} \right\rceil \geq 0.$$

Here, the second equality holds because of the definition of the floor function and the final inequality holds because $x/t < 1$. Thus, we have found a cycle of length in the range forbidden by Lemma 4.2.2, giving a contradiction to the assumption that such a K exists, and therefore, $\mathrm{ed}_{\mathrm{Forb}(C_h^t)}(p) = \gamma_{\mathrm{Forb}(C_h^t)}(p)$ for $p \in [p_0, 1/\lceil h/(2t+1) \rceil]$, concluding the proof. □

Therefore, we have proved that when $p \geq p_0$, Theorem 1.1.2 does indeed hold. In the next lemma, we check that $p_0 < 1/\lceil h/2t+1 \rceil$ for all $h \geq 4t(2t+1)$, and therefore the result does indeed extend on the range given in the work of Berikkyzy, Martin and Peck [10].

Lemma 4.2.11. *Let $t \geq 1$ and $h \geq 4t(2t+1)$. Then $p_0 \leq 1/\lceil h/(2t+1) \rceil$, that is,*

$$\frac{t}{\left(\left\lceil \frac{h}{2t+1} \right\rceil - 1\right) \left\lfloor \frac{\left\lfloor \frac{h}{t} \right\rfloor + 1}{3} \right\rfloor - \left\lceil \frac{h}{2t+1} \right\rceil + t + 1} < \frac{1}{\left\lceil \frac{h}{2t+1} \right\rceil}.$$

Proof. Let t and h be as in the lemma. It suffices to show that

$$\left\lceil \frac{h}{2t+1} \right\rceil \left\lfloor \frac{\left\lfloor \frac{h}{t} \right\rfloor + 1}{3} \right\rfloor - \left\lfloor \frac{\left\lfloor \frac{h}{t} \right\rfloor + 1}{3} \right\rfloor - (t+1) \left\lceil \frac{h}{2t+1} \right\rceil + t + 1 > 0. \tag{4.2.19}$$

Note that

$$\begin{aligned}
&\left\lceil \frac{h}{2t+1} \right\rceil \left\lfloor \frac{\left\lfloor \frac{h}{t} \right\rfloor + 1}{3} \right\rfloor - \left\lfloor \frac{\left\lfloor \frac{h}{t} \right\rfloor + 1}{3} \right\rfloor - (t+1) \left\lceil \frac{h}{2t+1} \right\rceil + t + 1 \\
&\geq \frac{h}{2t+1} \left(\frac{\frac{h}{t}}{3} - 1 \right) - \frac{\frac{h}{t} + 1}{3} - (t+1) \left(\frac{h}{2t+1} + 1 \right) + (t+1) \\
&= \frac{h^2}{3t(2t+1)} = \frac{h}{2t+1} = \frac{h+1}{3t} - \frac{(t+1)h}{2t+1} \\
&= \frac{h^2 - h(3t^2 + 8t + 1) - (2t+1)}{3t(2t+1)}.
\end{aligned}$$

As the denominator is always positive, this is positive when the numerator is positive. The numerator is a positive quadratic, and so is always positive when h is at least the larger root. So, if we can show that the larger root is at most $4t(2t+1)$, then we are done. Indeed, the larger root is at

$$\begin{aligned}
h_0 &= \frac{3t^2 + 8t + 1 + \sqrt{(3t^2 + 8t + 1)^2 + 4(2t+1)}}{2} \\
&= \frac{3t^2 + 8t + 1 + \sqrt{9t^4 + 48t^3 + 70t^2 + 24t + 5}}{2}.
\end{aligned}$$

For $t \geq 2$, we have

$$\begin{aligned}
\frac{3t^2 + 8t + 1 + \sqrt{9t^4 + 48t^3 + 70t^2 + 24t + 5}}{2} &< \frac{3t^2 + 8t + 1 + \sqrt{(13t^2 - 1)^2}}{2} \\
&= 4t(2t+1).
\end{aligned}$$

Therefore, the lemma holds for each $t \geq 2$. Meanwhile, for $t = 1$, we can directly calculate this. Indeed, note that if we substitute $t = 1$ into the left hand side (4.2.19), we get

$$\begin{aligned}\left\lceil \frac{h}{3} \right\rceil \left\lfloor \frac{h+1}{3} \right\rfloor - \left\lfloor \frac{h+1}{3} \right\rfloor - 2\left\lceil \frac{h}{3} \right\rceil + 2 &\geq \frac{h}{3}\left(\frac{h+1}{3} - 1\right) - \frac{h+1}{3} - 2\left(\frac{h}{3} + 1\right) + 2 \\ &= \frac{h^2 + h - 3}{9}.\end{aligned}$$

The numerator is a positive quadratic and is positive whenever h is at least the largest root. In particular, we need $h \geq -1/2 + \sqrt{13}/2$. Since we have $h \geq 4t(2t+1) = 12 = -1/2 + \sqrt{25^2}/2 > -1/2 + \sqrt{13}/2$, the equation is positive for all $h \geq 12$. Therefore, for each $t \geq 1$, for each $h \geq 4t(2t + 1)$, we have that (4.2.19) holds, concluding the proof. $\square$

BIBLIOGRAPHY

[1] S. Abbasi. “The solution of the El-Zahar problem”. PhD thesis. Rutgers University, 1998.

[2] N. Alon and U. Stav. What is the furthest graph from a hereditary property? *Random Structures & Algorithms* 33.1 (2008), pp. 87–104.

[3] N. Alon and R. Yuster. H-factors in dense graphs. *Journal of Combinatorial Theory Series B* 66.2 (1996), pp. 269–282.

[4] K. Appel and W. Haken. Every planar map is four colorable. I. Discharging. *Illinois Journal of Mathematics* 21.3 (1977), pp. 429–490.

[5] K. Appel, W. Haken, and J. Koch. Every planar map is four colorable. II. Reducibility. *Illinois Journal of Mathematics* 21.3 (1977), pp. 491–567.

[6] M. Axenovich, A. Kézdy, and R. Martin. On the editing distance of graphs. *J. Graph Theory* 58.2 (2008), pp. 123–138.

[7] M. Axenovich and R. Martin. Multicolor and directed edit distance. *Journal of Combinatorics* 2.4 (2011), pp. 525–556.

[8] J. Balogh and R. Martin. Edit distance and its computation. *Electronic Journal of Combinatorics* 15.1 (2008), p. 20.

[9] J. Balogh, A. Treglown, and A. Z. Wagner. Tilings in randomly perturbed dense graphs. *Combinatorics, Probability and Computing* 28.2 (2019), pp. 159–176.

[10] Z. Berikkyzy, R. Martin, and C. Peck. On the edit distance of powers of cycles. *Discrete Mathematics* 342.10 (2019), pp. 2804–2817.

[11] T. Bohman, A. Frieze, and R. Martin. How many random edges make a dense graph Hamiltonian? *Random Structures & Algorithms* 22.1 (2003), pp. 33–42.

[12] B. Bollobás. **Extremal Graph Theory**. Courier Corporation, 2004.

[13] J. Böttcher. Large-scale structures in random graphs. *Surveys in Combinatorics* 440.2017 (2017), p. 87.

[14] J. Böttcher, R. Montgomery, O. Parczyk, and Y. Person. Embedding spanning bounded degree graphs in randomly perturbed graphs. *Mathematika* 66.2 (2020), pp. 422–447.

[15] J. Böttcher, O. Parczyk, A. Sgueglia, and J. Skokan. Cycle factors in randomly perturbed graphs. *Procedia Computer Science* 195 (2021), pp. 404–411.

[16] J. Böttcher, M. Schacht, and A. Taraz. Proof of the bandwidth conjecture of Bollobás and Komlós. *Mathematische Annalen* 343.1 (2009), pp. 175–205.

[17] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems* 30.1-7 (1998), pp. 107–117.

[18] Y. Chang, J. Han, Y. Kohayakawa, P. Morris, and G. O. Mota. Factors in randomly perturbed hypergraphs. *Random Structures & Algorithms* 60.2 (2022), pp. 153–165.

[19] K. Corradi and A. Hajnal. On the maximal number of independent circuits in a graph. *Acta Mathematica Hungarica* 14.3-4 (1963), pp. 423–439.

[20] B. Csaba, J. Nagy-György, I. Levitt, and E. Szemerédi. "Tight bounds for embedding bounded degree trees". *Fete of Combinatorics and Computer Science*. Springer, 2010, pp. 95–137.

[21] L. DeBiasio, D. Kühn, T. Molla, D. Osthus, and A. Taylor. Arbitrary orientations of Hamilton cycles in digraphs. *SIAM Journal on Discrete Mathematics* 29.3 (2015), pp. 1553–1584.

[22] L. DeBiasio and T. Molla. Semi-degree threshold for anti-directed Hamiltonian cycles. *Electronic Journal of Combinatorics* P4.34 (2015).

[23] G. A. Dirac. Some theorems on abstract graphs. *Proceedings of the London Mathematical Society* 3.1 (1952), pp. 69–81.

[24] B. Ergemlidze and T. Molla. Transversal C_k-factors in subgraphs of the balanced blow-up of C_k. *Combinatorics, Probability and Computing* (2021), pp. 1–17.

[25] L. Euler. Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae* (1741), pp. 128–140.

[26] J. Farkas. Theorie der einfachen Ungleichungen. *Journal für die reine und angewandte Mathematik (Crelles Journal)* 1902.124 (1902), pp. 1–27.

[27] E. Fischer. Variants of the Hajnal–Szemerédi Theorem. *Journal of Graph Theory* 31.4 (1999), pp. 275–282.

[28] A. Ghouila-Houri. Une condition suffisante d'existence d'un circuit Hamiltonien. *C.R. Acad. Sci. Paris* 251.4 (1960), pp. 495–497.

[29] G. Grimmett and D. Stirzaker. **Probability and Random Processes**. Oxford University Press, 2001.

[30] A. Hajnal and E. Szemerédi. Proof of a conjecture of P. Erdős. *Combinatorial Theory and Its Applications* 2 (1970), pp. 601–623.

[31] P. Hall. On representatives of subsets. *Classic Papers in Combinatorics* (1987), pp. 58–62.

[32] J. Han, P. Morris, and A. Treglown. Tilings in randomly perturbed graphs: Bridging the gap between Hajnal-Szemerédi and Johansson-Kahn-Vu. *Random Structures & Algorithms* 58.3 (2021), pp. 480–516.

[33] J. Han and Y. Zhao. Hamiltonicity in randomly perturbed hypergraphs. *Journal of Combinatorial Theory, Series B* 144 (2020), pp. 14–31.

[34] P. E. Haxell and V. Rödl. Integer and fractional packings in dense graphs. *Combinatorica* 21.1 (2001), pp. 13–38.

[35] S. Janson, T. Łuczak, and A. Ruciński. **Random Graphs**. John Wiley & Sons, 2011.

[36] R. Johansson. Triangle-factors in a balanced blown-up triangle. *Discrete Mathematics* 211.1-3 (2000), pp. 249–254.

[37] F. Joos and J. Kim. Spanning trees in randomly perturbed graphs. *Random Structures & Algorithms* 56.1 (2020), pp. 169–219.

[38] A. Kathapurkar and R. Montgomery. Spanning trees in dense directed graphs. *Journal of Combinatorial Theory, Series B* 156 (2022), pp. 223–249.

[39] P. Keevash and R. Mycroft. A multipartite Hajnal–Szemerédi theorem. *Journal of Combinatorial Theory, Series B* 114 (2015), pp. 187–236.

[40] D. G. Kirkpatrick and P. Hell. On the complexity of general graph factor problems. *SIAM Journal on Computing* 12.3 (1983), pp. 601–609.

[41] J. Komlós, G. N. Sárközy, and E. Szemerédi. Proof of a packing conjecture of Bollobás. *Combinatorics, Probability and Computing* 4.3 (1995), pp. 241–255.

[42] J. Komlós, G. N. Sárközy, and E. Szemerédi. Proof of the Seymour conjecture for large graphs. *Annals of Combinatorics* 2.1 (1998), pp. 43–60.

[43] J. Komlós, G. N. Sárközy, and E. Szemerédi. Spanning trees in dense graphs. *Combinatorics, Probability and Computing* 10.5 (2001), p. 397.

[44] J. Komlós, G. N. Sárközy, and E. Szemerédi. Proof of the Alon-Yuster conjecture. *Discrete Mathematics* 235.1-3 (2001), pp. 255–269.

[45] M. Krivelevich, M. Kwan, and B. Sudakov. Bounded-degree spanning trees in randomly perturbed graphs. *SIAM Journal on Discrete Mathematics* 31.1 (2017), pp. 155–171.

[46] M. Krivelevich, M. Kwan, and B. Sudakov. Cycles and matchings in randomly perturbed digraphs and hypergraphs. *Combinatorics, Probability and Computing* 25.6 (2016), pp. 909–927.

[47] D. Kühn and D. Osthus. Embedding large subgraphs into dense graphs. *Surveys in Combinatorics* (2009), pp. 137–167.

[48] D. Kühn and D. Osthus. Hamilton decompositions of regular expanders: a proof of Kelly's conjecture for large tournaments. *Advances in Mathematics* 237 (2013), pp. 62–146.

[49] D. Kühn and D. Osthus. Hamilton decompositions of regular expanders: applications. *Journal of Combinatorial Theory Series B* 104 (2014), pp. 1–27.

[50] D. Kühn and D. Osthus. The minimum degree threshold for perfect graph packings. *Combinatorica* 29.1 (2009), pp. 65–107.

[51] A. Lo and K. Markström. A multipartite version of the Hajnal–Szemerédi theorem for graphs and hypergraphs. *Combinatorics, Probability and Computing* 22.1 (2013), pp. 97–111.

[52] E. Marchant and A. Thomason. "Extremal graphs and multigraphs with two weighted colours". *Fete of Combinatorics and Computer Science*. Vol. 20. Bolyai Soc. Math. Stud. János Bolyai Math. Soc., Budapest, 2010, pp. 239–286.

[53] R. Martin. The edit distance function and symmetrization. *Electronic Journal of Combinatorics* 20.3 (2013), p. 26.

[54] R. Martin. On the computation of edit distance functions. *Discrete Mathematics* 338.2 (2015), pp. 291–305.

[55] R. Martin. "The edit distance in graphs: Methods, results, and generalizations". *Recent Trends in Combinatorics*. Springer, 2016, pp. 31–62.

[56] R. Martin and T. McKay. On the edit distance from $K_{2,t}$-free graphs. *Journal of Graph Theory* 77.2 (2014), pp. 117–143.

[57] R. Martin and A. W. N. Riasanovsky. On the edit distance function of the random graph. *Combinatorics, Probability and Computing* 31.2 (2022), pp. 345–367.

[58] A. McDowell and R. Mycroft. Hamilton ℓ-cycles in randomly-perturbed hypergraphs. *arXiv preprint arXiv:1802.04242* (2018).

[59] R. Montgomery. Spanning trees in random graphs. *Advances in Mathematics* 356 (2019), p. 106793.

[60] R. Montgomery, A. Pokrovskiy, and B. Sudakov. Embedding rainbow trees with applications to graph labelling and decomposition. *Journal of the European Mathematical Society* 22.10 (2020), pp. 3101–3132.

[61] R. Mycroft and T. Naia. Spanning trees of dense directed graphs. *Electronic Notes in Theoretical Computer Science* 346 (2019), pp. 645–654.

[62] R. Mycroft and T. Naia. Trees and tree-like structures in dense digraphs. *arXiv preprint arXiv:2012.09201* (2020).

[63] M. Pavez-Signé, N. Sanhueza-Matamala, and M. Stein. Dirac-type conditions for spanning bounded-degree hypertrees. *arXiv preprint arXiv:2012.09824* (2020).

[64] C. Peck. “On the edit distance from a cycle- and squared cycle-free graph”. MA thesis. United States: Iowa State University, 2013.

[65] N. Robertson, D. Sanders, P. Seymour, and R. Thomas. The four-colour theorem. *Journal of Combinatorial Theory Series B* 70.1 (1997), pp. 2–44.

[66] V. Rödl, E. Szemerédi, and A. Ruciński. An approximate Dirac-type theorem for k-uniform hypergraphs. *Combinatorica* 28.2 (2008), pp. 229–260.

[67] E. Szemerédi. “Regular partitions of graphs”. *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976).* Vol. 260. Colloq. Internat. CNRS. CNRS, Paris, 1978, pp. 399–401.