FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

BACHELOR THESIS

Václav Rozhoň

Sufficient conditions for embedding trees

Department of Applied Mathematics

Supervisor of the bachelor thesis:	Mgr. Tereza Klimošová, Ph.D.
Consultant of the bachelor thesis:	Mgr. Diana Piguet, Ph.D.
Study programme:	Computer science
Study branch:	General computer science

Prague 2018

I declare that I carried out this bachelor thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In .... date .... signature of the author

Title: Sufficient conditions for embedding trees

Author: Václav Rozhoň

Department: Department of Applied Mathematics

Supervisor: Mgr. Tereza Klimošová, Ph.D., Department of Applied Mathematics

Consultant: Mgr. Diana Piguet, Ph.D., The Czech Academy of Sciences, Institute of Computer Science

Abstract:

We study sufficient degree conditions that force a host graph to contain a given class of trees. This setting involves some well-known problems from the area of extremal graph theory. The most famous one is the Erdős-Sós conjecture that asserts that every graph with average degree greater than $k-1$ contains any tree on $k+1$ vertices.

Our two main results are the following. We prove an approximate version of the Erdős-Sós conjecture for dense graphs and trees with sublinear maximum degree. We also study a natural refinement of the Loebl-Komlós-Sós conjecture and prove it is approximately true for dense graphs.

Both results are based on the so-called regularity method. The second mentioned result is a joint work with T. Klimošová and D. Piguet.

Keywords:

extremal graph theory, tree embedding, Loebl-Komlós-Sós conjecture, Erdős-Sós conjecture, regularity lemma

My great thanks go to my supervisor Tereza Klimošová, and my consultant Diana Piguet; their help was invaluable. I would also like to thank to Jan Hladký for careful reading of some parts of the thesis, and to Stephan Wagner, who gave the proof of Proposition 3.1.

I am very grateful to my family for their lasting support. I dedicate the thesis to my dear partner, Hanka. Without your kindness and devotion this would not be possible.

Contents

Chapter 1

Introduction

Typical problems in extremal graph theory ask, how many edges in a graph force it to contain a given subgraph. A classical example of a result in this area is Turán's Theorem, which determines the average degree that guarantees the containment of the complete graph K_r. A more complex example is the Erdős-Stone Theorem [ES46], which essentially determines the average degree condition guaranteeing that the host graph contains a fixed non-bipartite graph. On the other hand, for a general bipartite graph the problem is wide open.

In this thesis we study this question when the embedded graph is a tree. Which conditions force the host graph G to contain any tree with fixed number of vertices as a subgraph? The two classical conjectures in this area that we investigate are the Erdős-Sós conjecture and the Loebl-Komlós-Sós conjecture.

Conjecture 1.1 (The Erdős-Sós conjecture). *Every graph G with average degree $\overline{\deg}(G) > k - 1$ contains any tree on $k + 1$ vertices.*

Here $\overline{\deg}(G)$ means the average degree of G; similarly, we denote the minimum and the maximum degree of G by $\delta(G)$ and $\Delta(G)$, respectively. Observe that the conjecture is optimal, since a graph with average degree at most $k - 1$ may have only k vertices.

The conjecture trivially holds when the tree is a star and a classical result of Erdős and Gallai [EG59] proves that it also holds for paths. There are many other partial results concerning the celebrated conjecture. It has been verified for some special families of host graphs [BD96, SW97, BD07, WLL00, Dob02], special families of trees embedded [McL05, FS07, Fan13], or when the size of the host graph is only slightly larger than the size of the tree [Woz96, Tin10, GZ16]. Finally, a solution of this conjecture for large k, based on an extension of the regularity lemma, has been announced in the early 1990's by Ajtai, Komlós, Simonovits, and Szemerédi. This result will be published as a sequence of three papers [AKSSa, AKSSc, AKSSb].

Another well-known conjecture in this area is the Loebl-Komlós-Sós conjecture.

Conjecture 1.2 (The Loebl-Komlós-Sós conjecture). *Let G be a graph of order n. If at least $n/2$ vertices of G have degree at least k, then G contains every tree on $k + 1$ vertices.*

Note that the conjecture is again almost best possible. The degree k cannot be lowered due to the example of the star $K_{1,k}$. We have to assume that at least half of the vertices have high degree due to the following example. Consider a graph consisting of many disjoint copies of a graph on $k + 1$ vertices that we get from K_{k+1} by deleting all edges in a subset of vertices of size at least $(k+3)/2$. Such a graph does not contain a path on $k+1$ vertices.

This conjecture was also verified for some special classes of host graphs [Sof00, Dob02], or trees [BLW00, PS08]. It was proved for dense graphs by [HP16] and, independently by [Coo09], building on results from [PS12] and [Zha11]. Finally, the approximate version of this conjecture was proved in a series of four long papers [HKP+17a, HKP+17b, HKP+17c, HKP+17d] (see [HPS+15] for an overview).

The two main results of this thesis concern these two conjectures and both have a similar flavour. Firstly, both results are approximate (we get arbitrarily close to the desired result for large sizes of the host graph) and are nontrivial only if the size of the embedded tree is linear in the size of the host graph.

Secondly, both results concern the class of *skewed* trees, i.e., trees such that the size of one of their colour class is at most rk. This allows us to consider refinements of the conjectures above, in particular the conditions imposed on the host graph can be weakened depending on r.

The first of the two results that we prove in Chapter 3 is the following Erdős-Sós-like result. Roughly speaking, it states that one can embed a tree with k vertices and skew r in every large enough host graph with positive proportion of vertices of degree roughly k and with minimum degree roughly rk. We have to further assume that the degree of the tree is sublinear.

Theorem 3.8. *For any $r, \eta > 0$ there exists n_0 and $\gamma > 0$ such that the following holds. Let G be a graph of order $n > n_0$ and T a tree of order k with two colour classes T_1, T_2 such that $|T_1| \leq rk$ and $\Delta(T_2) \leq \gamma k$. If $\delta(G) \geq rk + \eta n$, and at least ηn vertices of G have degree at least $k + \eta n$, then G contains T.*

As we will later see, this result is interesting only if $k > \eta n/2$, otherwise there is a simple greedy way of embedding T in G. Hence we interpret this result as one for trees of size linear in the size of the host graph; only for such a class of trees this result is nontrivial. A simple consequence of Theorem 3.8 with $r = 1/2$ is that the Erdős-Sós conjecture holds approximately (with error term linear in n) for trees with sublinear maximum degree.

Theorem 3.10. *For any $\eta > 0$ there exists n_0 and $\gamma > 0$ such that for every $n > n_0$, any graph of order n with average degree $\overline{\deg}(G) \geq k + \eta n$ contains every tree on k vertices with maximum degree $\Delta(T) \leq \gamma k$.*

The theorem is again trivial if the size of the tree is not linear in the size of the host graph. Although this theorem is only a special case of the result of Ajtai, Komlós, Simonovits, and Szemerédi, we still believe that it is of interest, as its proof is straightforward and most probably substantially simpler than their proof.

The main result of Chapter 4 is that the natural extension of the Loebl-Komlós-Sós conjecture to skewed trees hold approximately for dense graphs (we call this *skew* Loebl-Komlós-Sós conjecture).

Theorem 4.4. *For any $0 < r \leq 1/2$ and $\eta > 0$ there exists $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$, any graph of order n with at least rn vertices of degree at least $k + \eta n$ contains every tree of order at most k such that the size of its smaller colour class is at most rk.*

This result is again nontrivial only for trees of size linear in n.

The following chapter of the thesis contain general techniques for embedding trees that are then used in subsequent two chapters to prove the two mentioned results together with several others.

The other results in this thesis include a proof that the Loebl-Komlós-Sós conjecture holds both for trees of diameter at most five (Theorem 4.2) and for paths (Theorem 5.1 and its algorithmic version Theorem 5.2). We also propose several conjectures and prove another Erdős-Sós-like result for trees of diameter at most four (Theorem 3.6).

The results in Chapters 3 and 4 will be published in a series of three papers. In the first paper we provide a proof of Theorem 3.8, in the second paper we prove Theorem 4.4, and in the last paper we prove Theorems 3.6 and 4.2. The last two papers are a result of joint work with Tereza Klimošová and Diana Piguet. My contribution is in all three cases proportional to the number of authors of the papers.

Chapter 2

Techniques

In this chapter we introduce several general results regarding embedding of trees that we will use in subsequent chapters. We start by introducing some terminology and proposing few simple structural results. Then we introducing the regularity method – a very efficient tool for embedding results in dense host graphs. This is the basis of our subsequent results for dense graphs that we prove in Chapters 3 and 4.

2.1 Terminology

Throughout the thesis we use mostly the standard notation. We list all non-usual terminology here.

All graphs in the thesis are simple and loopless. The *degree* $\deg(x)$ of a vertex x is the number of its neighbours. By $\deg(x, X)$ we denote the number of neighbours of x in the set X. The minimum and maximum degree of G are $\delta(G)$ and $\Delta(G)$, respectively. We denote the second largest degree of G by $\Delta_2(G)$, with possible equality $\Delta_2(G) = \Delta(G)$. Let G be a graph and let X, Y be disjoint subsets of its vertices. We define $E(X, Y)$ as the set of edges of G with one end in X and the other end in Y; we set $e(X, Y) = |E(X, Y)|$. The *density* of the pair (X, Y) is defined as $d(X, Y) = \frac{e(X,Y)}{|X||Y|}$. The average degree is $\overline{\deg}(X, Y) = e(X, Y)/|X| = |Y|d(X, Y)$. The length of the shortest path between two vertices u, v in G is denoted by $\mathrm{dist}_G(x, y)$. We also sometimes use a symbol $\mathcal{T}_k$ to denote the class of all trees on k vertices, and $\mathcal{T}_k^r$ to denote the class of tree on k vertices with skew r.

When speaking about the Loebl-Komlós-Sós conjecture, we will use the term (r, k)-LKS graph to denote a graph fulfilling conditions of the conjecture.

Definition 2.1 (LKS graphs). *An (r, k)-LKS graph is a non-empty graph that contains at least rn vertices of degree at least k for $0 < r \leq \frac{1}{2}$ and $k > 0$.*

Furthermore, in the context of the Loebl-Komlós-Sós conjecture we use the name L-vertices for the vertices of G with degree at least k. Similarly, S-vertices are vertices of G that are not L-vertices.

For a given cycle C denote by $\overrightarrow{C}$ and $\overleftarrow{C}$ its two orientations. For $v, w \in C$ we denote by $v\overrightarrow{C}w$ the path starting at v and following the orientation of $\overrightarrow{C}$ up to w. We also use symbols v^+ and v^- to denote the successor and the predecessor of v on $\overrightarrow{C}$. We shall use analogous notation for the other orientation of C as well as for oriented paths. When we, for example, write $\overrightarrow{P} = u\overrightarrow{P}v$, we say that the first vertex of the oriented path is u, while the last one is v.

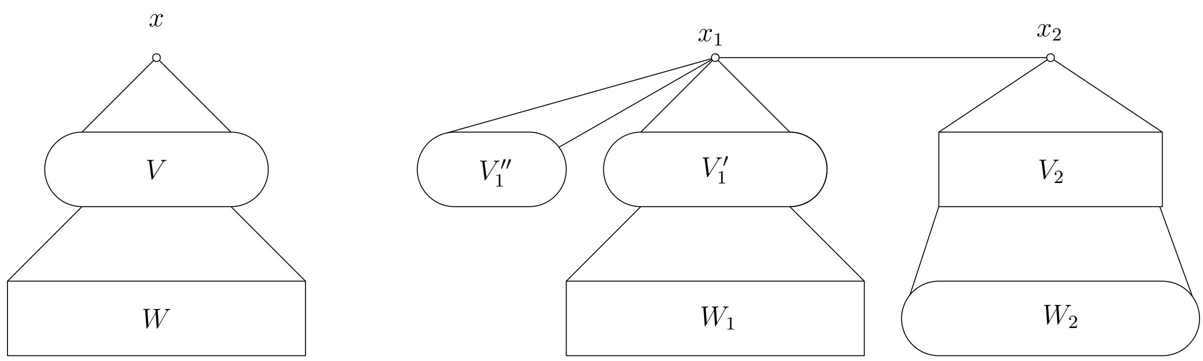

Figure 2.1: Trees of diameter four and five – notation.

We denote the diameter of T (length of the longest path in T) by $\mathrm{diam}(T)$. One can observe that all trees of diameter at most five have very simple structure (Figure 2.1). In each tree T of diameter at most four there is a vertex x such that all remaining vertices of T are at distance at most 2 from x. We denote by V the set of neighbours of x and define $W = N(V) \setminus \{x\}$. Similarly, in each nontrivial tree of diameter at most five there are two vertices x_1, x_2 and vertex sets V_1, V_2, W_1, W_2 such that $V_i = N(x_i) \setminus \{\{x_1\} \cup \{x_2\}\}$ and $W_i = N(V_i) \setminus \{x_i\}$. Moreover, we denote the set of leaf neighbours of x_1 by V_1'' and define $V_1' = V_1 \setminus V_1''$.

2.2 Structure of the host graph

At first note that there is a simple greedy algorithm for embedding a tree T in a host graph G: unless the whole T is embedded, choose a yet non-embedded vertex u with an embedded neighbour v and try to injectively extend the partial mapping φ of T to u by embedding u in the neighbourhood of $\varphi(v)$. If $\delta(G) \geq |T|$, we will be always able to injectively extend φ. We will use this observation several times later on.

It is not hard to see that, when proving the Loebl-Komlós-Sós conjecture, one may assume that the host graph does not contain any edges between its S-vertices. Another simple, yet important observation is that when one proves the Erdős-Sós conjecture, she may assume that the minimum degree of the host graph is at least $k/2$. For completeness, we give a proof here. Moreover, we prove two similar auxiliary lemmas (their versions for $r = 1/2$ are known).

Lemma 2.2 (Folklore). *Let G be a graph with $\overline{\deg}(G) > t$. Then it contains a non-empty subgraph H such that $\overline{\deg}(H) > t$ and $\delta(H) > t/2$.*

Proof. Let H be a minimal subgraph of G with $\overline{\deg}(H) > t$ and for contradiction assume that v is its vertex of degree less than or equal to $t/2$. We may erase v from H and the resulting non-empty graph H' contradicts its minimality, because

$$\overline{\deg}(H') = 2\frac{n\overline{\deg}(H)/2 - \deg(v)}{n-1} > 2\frac{nt/2 - t/2}{n-1} = t.$$

□

We continue with a similar observation about (r, k)-LKS graphs that will be later used in Chapter 5 to verify that the skew Loebl-Komlós-Sós conjecture holds for paths.

Lemma 2.3 (Lemma 5 in [BLW00] for $r = 1/2$)**.** *Let G be an (r,k)-LKS graph without edges between its S-vertices. Then it contains a non-empty (r,k)-LKS subgraph H such that for any subset $X \subseteq S(H)$ we have $|N(X)| > r|X|$.*

Proof. Let H be a minimal (r,k)-LKS subgraph of G. Suppose that we have $X \subseteq S(H)$ such that $|N(X)| \le r|X|$. Then we may erase X and the resulting non-empty graph H' contradicts the minimality of H, because we have

$$\frac{|L(H')|}{|V(H')|} \ge \frac{|L(H) \setminus N(X)|}{|V(H) \setminus X|} \ge \frac{|L(H)| - r|X|}{|V(H)| - |X|} \ge \frac{r|V(H)| - r|X|}{|V(H)| - |X|} = r,$$

where $V(H)$ and $V(H')$ denote the set of L-vertices of H and H', respectively. □

We employ a similar idea once more in the following lemma to show that every (r,k)-LKS graph contains a subgraph of average degree depending on r and k. A simple consequence of this result is that if we possess an $(r, k + \eta n)$-LKS graph such that $k \le \frac{1+r}{r}\eta n$, we may embed any tree on k vertices in the host graph by the greedy algorithm.

Lemma 2.4 (Theorem 5 in [Sof00] for $r = 1/2$)**.** *Each (r,k)-LKS graph G contains a subgraph $H \subseteq G$ such that $\overline{\deg}(H) \ge 2\frac{r}{1+r}k$.*

Proof. We will prove that the required property holds either for G itself or for $G[L]$ – the subgraph of G induced by its L-vertices.

Suppose that $2e(L)/|L| = \overline{\deg}(G[L]) < 2\frac{r}{1+r}k$. From this we get that

$$e(L) < \frac{r}{1+r}|L|k.$$

From the condition on the degree of L-vertices we have $2e(L) + e(L,S) \ge |L|k$, thus we have

$$e(L,S) \ge |L|k - 2e(L) > (1 - 2\frac{r}{1+r})|L|k.$$

For the average degree of G we finally have

$$\overline{\deg}(G) = \frac{2(e(L) + e(L,S))}{n} = \frac{2e(L) + e(L,S) + e(L,S)}{n} \ge \frac{|L|k + (1 - 2\frac{r}{1+r})|L|k}{n} = 2\frac{|L|k}{(1+r)n} \ge 2\frac{r}{1+r}k.$$

□

2.3 The regularity method

In this section we introduce the regularity method, a well-known technique that can be applied for embedding trees. The main idea behind the method is that we try to use the fact that it is generally easier to embed trees in random graphs, as their expansion properties can compensate for the lack of edges. Large dense graphs are behaving in a pseudorandom way (this is the regularity lemma), hence it is possible to successfully apply this idea to them.

2.3.1 The regularity lemma

We say that (X, Y) is an *ε-regular pair*, if for every $X' \subseteq X$ and $Y' \subseteq Y$, $|X'| \geq \varepsilon|X|$ and $|Y'| \geq \varepsilon|Y|$, $|d(X', Y') - d(X, Y)| \leq \varepsilon$.

Next well-known lemma states that subsets of regular pairs to some extend inherit the regularity of the whole pair. For more on the lemma, see e.g. [Dvo], or [Die97].

Lemma 2.5. *Let G be a graph and (X, Y) be an ε-regular pair of density d in G. Let $X' \subseteq X$ and $Y \subseteq Y$ such that $|X'| \geq \alpha|X|$ and $|Y'| \geq \alpha|Y|$. Then, (X', Y') is an ε'-regular pair of density at least $d - \varepsilon$, where $\varepsilon' = \max(\varepsilon/\alpha, 2\varepsilon)$.*

We say that a partition $\{V_0, V_1, \ldots, V_m\}$ of $V(G)$ is an *ε-regular partition*, if $|V_0| \leq \varepsilon|V(G)|$ all but at most εm^2 pairs (V_i, V_j), $1 \leq i < j \leq m$, are ε-regular. Each set of the partition is called cluster. We call the cluster V_0 the *garbage set*. We call a regular partition *equitable* if $|V_i| = |V_j|$ for every $1 \leq i < j \leq m$.

Lemma 2.6 (Szemerédi regularity lemma)**.** *For every $\varepsilon > 0$ there is n_0 and M such that every graph of size at least n_0 admits an ε-regular equitable partition $\{V_0, \ldots, V_m\}$ with $1/\varepsilon \leq m \leq M$.*

Given an ε-regular pair (X, Y), we call a vertex $x \in X$ *typical* with respect to a set $Y' \subseteq Y$ if $\deg(x, Y') \geq (d(X, Y) - \varepsilon)|Y'|$. Note that from the definition of regularity it follows that all but at most $\varepsilon|X|$ vertices of X are typical with respect to any subset of Y of size at least $\varepsilon|Y|$. This observation can be strengthened as follows.

Lemma 2.7 (Lemma 4 in [KPR18])**.** *Let $\{V_0, V_1, \ldots, V_m\}$ be an ε-regular partition of $V(G)$ and let $X = V_j$ for some $1 \leq j \leq m$. Then all but at most $\sqrt{\varepsilon}|X|$ vertices of a cluster X are typical w. r. t. all but at most $\sqrt{\varepsilon}m$ sets V_i, $i \in \{1, \ldots, m\} \setminus j$. In Chapter 4 we call such vertices of X* ultratypical.

2.3.2 Partitioning trees

Here we state a crucial lemma from [HKP+17d] that allows us to partition the tree in controllable number of small subtrees that we also call *microtrees*. These trees are neighbouring with a set of vertices of bounded size consisting of vertices that we call *seeds*. Moreover, we need to work separately with seeds from different colour classes of T. In the following definition, the set $W_A \cup W_B$ is the set of seeds of T and the set $\mathcal{D}_A \cup \mathcal{D}_B$ is the set of its microtrees.

Definition 2.8. *[HKP+17d, Definition 3.3] Let T be a tree on $k+1$ vertices. An ℓ-fine partition of T is a quadruple $(W_A, W_B, \mathcal{D}_A, \mathcal{D}_B)$, where $W_A, W_B \subseteq V(T)$ and $\mathcal{D}_A$ and $\mathcal{D}_B$ are families of subtrees of T such that*

1. *the three sets W_A, W_B and $\{V(K)\}_{K \in \mathcal{D}_A \cup \mathcal{D}_B}$ partition $V(T)$ (in particular, the trees in $K \in \mathcal{D}_A \cup \mathcal{D}_B$ are pairwise vertex disjoint),*
2. $\max\{|W_A|, |W_B|\} \leq 336k/\ell$,
3. *for $w_1, w_2 \in W_A \cup W_B$ their distance is odd if and only if one of them lies in W_A and the other one in W_B,*
4. *$|K| \leq \ell$ for every tree $K \in \mathcal{D}_A \cup \mathcal{D}_B$,*

5. *for each $K \in \mathcal{D}_A$ we have $N_T(V(K)) \setminus V(K) \subseteq W_A$. Similarly for each $K \in \mathcal{D}_B$ we have $N_T(V(K)) \setminus V(K) \subseteq W_B$.*

6. *$|N(V(K)) \cap (W_A \cup W_B)| \leq 2$ for each $K \in \mathcal{D}_A \cup \mathcal{D}_B$,*

7. *if $N(V(K)) \cap (W_A \cup W_B)$ contains two vertices z_1, z_2 for some $K \in \mathcal{D}_A \cup \mathcal{D}_B$, then $\mathrm{dist}_T(z_1, z_2) \geq 6$.*

We did not list all properties of ℓ-fine partition from [HKP+17d], only those we need.

Lemma 2.9. *[HP16, Lemma 5.3] Let T be a tree on $k+1$ vertices and let $\ell \in \mathbb{N}, \ell < k$. Then T has an ℓ-fine partition.*

In the subsequent applications we are always working with $\ell = \beta k$ for some small $\beta > 0$.

Observe that the structure that we work with is actually very similar to the structure of the tree of diameter five – the seeds W_A, W_B behave similarly to the two vertices x_1, x_2 and the four sets $V(\bigcup \mathcal{D}_A) \cap N(W_A), V(\bigcup \mathcal{D}_A) \setminus N(W_A), V(\bigcup \mathcal{D}_B) \cap N(W_B), V(\bigcup \mathcal{D}_B) \setminus N(W_B)$ are similar to the sets V_1, W_1, V_2, W_2. This is the reason, why we always aim to prove any embedding result at first for trees of small diameter (cf. Theorem 3.6 and Theorem 4.2). Although the class of such trees may not be of particular interest by itself, the proof guides us towards a general proof for all trees in the dense setting. In Chapter 3 devoted to local approach to Erdős-Sós conjecture, we are not able to prove a general result for trees of diameter five, only for trees of diameter four (Theorem 3.6). Similarly, we are not later able to prove a general result for all trees in the dense settings (such a result actually cannot hold there). This is the reason, why we turn our attention to trees of sublinear degree that still form a rather general class of trees. For each such tree we can additionally suppose that the seeds of its ℓ-fine partition are only in one colour class, thus the structure of this *one-sided* ℓ-fine partition resembles the structure of trees of diameter four.

Definition 2.10. *Let $T \in \mathcal{T}_{k+1}$ be a tree and T_1, T_2 its colour classes. Let $\Delta = \max_{v \in T_2} \deg(v)$. A one-sided ℓ-fine partition of T is a pair $(W, \mathcal{D})$, where $W \subseteq V(T_1)$ and $\mathcal{D} = \mathcal{D}' \sqcup \mathcal{D}''$ is a family of subtrees of T such that*

1. *the two sets W and $\{V(K)\}_{K \in \mathcal{D}}$ partition $V(T)$,*

2. *$|W| \leq 336k(1+\Delta)/\ell$,*

3. *$|K| \leq \ell$ for every tree $K \in \mathcal{D}$,*

4. *For each $K \in \mathcal{D}$ we have $N_T(V(K)) \setminus V(K) \subseteq W$.*

5. *We can split $\mathcal{D}$ into two subfamilies, $\mathcal{D} = \mathcal{D}' \sqcup \mathcal{D}''$, in such a way that all trees from $\mathcal{D}'$ have at most two neighbours $z_1, z_2 \in W$ such that $\mathrm{dist}_T(z_1, z_2) \geq 4$, while all of at most $336k/\ell$ trees from $\mathcal{D}''$ are singletons with at most Δ neighbours in W.*

Lemma 2.11. *Let $T \in \mathcal{T}_{k+1}$ and let $\ell \in \mathbb{N}, \ell < k$. Then T has a one-sided ℓ-fine partition.*

Proof. Let $(W_A, W_B, \mathcal{D}_A, \mathcal{D}_B)$ be an ℓ-fine partition of T. Suppose that $W_B \subseteq T_2$. Let $W = W_A \cup N(W_B)$ and define $\mathcal{D}$ as the set of trees of the forest $T \setminus W$. The conditions (1), (2), and (4) are clearly satisfied. Each vertex from W is now a singleton tree in $\mathcal{D}$. Define $\mathcal{D}''$ as the family of these singleton trees and set $\mathcal{D}' = \mathcal{D} \setminus \mathcal{D}''$. Each tree in $\mathcal{D}''$ clearly satisfies the conditions (3) and (5). Each tree from $\mathcal{D}'$ is either a tree from $\mathcal{D}_A$, or a subtree of a tree from $\mathcal{D}_B$, all such trees satisfy the condition (3). Finally recall that for each tree from $\mathcal{D}_A \cup \mathcal{D}_B$ with two neighbours z_1 and z_2 in $W_A \cup W_B$ we have $\mathrm{dist}_T(z_1, z_2) \geq 6$. Thus, all trees from $\mathcal{D}_A$ satisfy the condition (5). Each tree from $\mathcal{D}_B$ with two neighbours $z_1, z_2 \in W_B$ was split into one tree with two neighbours in W, such that their distance in T is at least 4, and maybe several other trees with only one neighbour in W. All such trees also satisfy (5). □

2.3.3 Embedding in regular pairs

In this section we present three embedding lemmas. The first will be used in Chapter 3 to embed the seeds of a one-sided partition, together with the set $\mathcal{D}''$, in vertices of two neighbouring clusters.

Proposition 2.12. *For any $d, \beta, \varepsilon > 0$, $\varepsilon \leq d^2/100$ there exist k_0 and $\gamma > 0$ such that the following holds.*

Let T be a tree of order $k \geq k_0$ and T_2 one of its colour classes such that $\Delta(T_2) \leq \gamma k$. Moreover, let $(W, \mathcal{D}), \mathcal{D} = \mathcal{D}' \sqcup \mathcal{D}''$ be its one-sided βk-fine partition. Let $\mathbf{v}_1$ and $\mathbf{v_2}$ be two clusters of vertices forming an ε-regular pair of density at least d. Suppose that $|\mathbf{v}_1| = |\mathbf{v}_2| \geq k/M_{L2.6}(\varepsilon)$, where $M_{L2.6}(\varepsilon)$ is the output of the regularity lemma (Lemma 2.6) with an input ε. Let $U \subseteq \mathbf{v}_1, |U| \leq 2\sqrt{\varepsilon}|\mathbf{v}_1|$. Then there is an injective mapping φ of $W \cup (\bigcup \mathcal{D}'')$ that embeds vertices of W in $\mathbf{v}_1 \setminus U$ and vertices of $\bigcup \mathcal{D}''$ in $\mathbf{v}_2$.

Proof. Choose $\gamma, k_0 > 0$ such that

$$\gamma = \frac{\beta d}{2000 M_{L2.6}(\varepsilon)},$$
$$k_0 = \frac{10}{\gamma}.$$

Note that in this case we have

$$\begin{aligned} \left|\bigcup D''\right| \leq |W| &\leq \frac{336(1+\gamma k)}{\beta} \leq \frac{500\gamma k}{\beta} \\ \boxed{\text{definition of } \gamma} \quad &= \frac{500\beta dk}{\beta \cdot 2000 M_{L2.6}(\varepsilon)} = \frac{dk}{4M_{L2.6}(\varepsilon)} \\ \boxed{|\mathbf{v}_1| \geq k/M_{L2.6}(\varepsilon)} \quad &\leq \frac{d}{4}|\mathbf{v}_1|. \end{aligned}$$

Take an arbitrary vertex $r \notin \bigcup \mathcal{D}''$ of T and root the tree at r. Order all vertices of $W \cup (\bigcup \mathcal{D}'')$ according to an order, in which they are visited by a depth-first search starting at r. Let $U' \subseteq \mathbf{v}_1 \cup \mathbf{v}_2$ be the set of vertices of $\mathbf{v}_1$ not typical to $\mathbf{v}_2$ together with vertices of $\mathbf{v}_2$ not typical to $\mathbf{v}_1$. We will provide an algorithm that gradually defines a partial embedding φ of the vertices of $W \cup (\bigcup \mathcal{D}'')$ such that $\varphi(W) \subseteq \mathbf{v}_1 \setminus (U \cup U')$ and $\varphi(\bigcup \mathcal{D}'') \subseteq \mathbf{v}_2 \setminus U'$.

We iterate over the sequence $x_1, x_2, x_3, \dots$ of vertices from $W \cup (\bigcup \mathcal{D}'')$, where the vertices are ordered by the depth-first search. In the i-th step we deal with the vertex $x = x_i$. At first we deal with the case $x \in W$.

Suppose that $y \in \bigcup \mathcal{D}''$ is the already embedded parent of x (if $y \notin \bigcup \mathcal{D}''$, our task is simpler). We want to embed x in an arbitrary neighbour of y in $\mathbf{v}_1 \setminus (U \cup \varphi(W) \cup U')$. To do so, it suffices to verify that $N(y) \setminus (U \cup \varphi(W) \cup U')$ is nonempty. This can be done with the help of the fact that $\varphi(y)$ is typical to $\mathbf{v}_1$ and together with our bound on $|W|$:

$$|N(y) \setminus (U \cup \varphi(W) \cup U')| \geq |\mathbf{v}_1|((d-\varepsilon) - 2\sqrt{\varepsilon} - \frac{d}{4} - \varepsilon) > 0.$$

Similarly, suppose that $x \in \bigcup \mathcal{D}''$. From the definition of $\mathcal{D}''$ we know that its parent y is certainly in W and $\varphi(y)$ is typical to $\mathbf{v}_2$. Now we similarly verify that

$$\left| N(y) \setminus \left(\varphi \left(\bigcup \mathcal{D}'' \right) \cup U' \right) \right| \geq |\mathbf{v}_2|((d-\varepsilon) - \frac{d}{4} - \varepsilon) > 0.$$

□

Next we state another proposition that will help us in Chapter 3 to embed small trees from a fine partition of T in the regular pairs of the host graph. The result is folklore.

Next, we state a similar proposition that enables us to embed small trees from a fine partition of T in the regular pairs of the host graph. The proposition is a variation on a folklore result and is similar to e.g. Lemma 5 in [KPR18].

Proposition 2.13. *For all $1 \geq d, \varepsilon > 0$ such that $\varepsilon < d^2/100$ there exists $\beta > 0$ such that the following holds.*

Let $\mathbf{v}_1, \mathbf{u}, \mathbf{v}$ be three clusters of vertices such that $\mathbf{v}_1\mathbf{u}$ and $\mathbf{u}\mathbf{v}$ are ε-regular pairs of density at least d. Let v_1, v_2 be two (not necessarily distinct) vertices of $\mathbf{v}_1$. Suppose that $|\mathbf{v}_1| = |\mathbf{u}| = |\mathbf{v}| \geq k/M_{L2.6}(\varepsilon)$. Let K be a tree of order at most βk and let x_1, x_2 be its two vertices from the same colour class of K such that if $v_1 \neq v_2$, then $x_1 \neq x_2$. Let U be a subset of vertices of $\mathbf{u} \cup \mathbf{v}$ such that $|\mathbf{u} \setminus U| \geq 4\sqrt{\varepsilon}|\mathbf{u}|$ and $|\mathbf{v} \setminus U| \geq 4\sqrt{\varepsilon}|\mathbf{v}|$. Moreover, suppose that either

1. *the vertices v_1, v_2 are typical to $\mathbf{u}$ and $\overline{\deg}(\mathbf{v}_1, \mathbf{u}) - |U \cap \mathbf{u}| \geq 4\sqrt{\varepsilon}|\mathbf{u}|$,*
2. *or we have $|N(v_i) \cap (\mathbf{u} \setminus U)| \geq 3\varepsilon|\mathbf{u}|$ for $i = 1, 2$.*

Then there is an injective mapping φ of K in $\mathbf{u} \cup \mathbf{v}$ such that $\varphi(K) \cap U = \emptyset$. Moreover, $\varphi(x_1)$ is a neighbour of v_1 and $\varphi(x_2)$ is a neighbour of v_2.

Proof. We show the proof for the harder case when $v_1 \neq v_2$. Choose

$$\beta = \varepsilon / M_{L2.6}(\varepsilon).$$

From this we get

$$|\mathbf{v}_1| \geq \frac{k}{M_{L2.6}(\varepsilon)} = \beta \cdot \frac{M_{L2.6}(\varepsilon)}{\varepsilon} \cdot \frac{k}{M_{L2.6}(\varepsilon)} = \frac{\beta k}{\varepsilon}.$$

Note that $\mathbf{u} \setminus U$ contains at least $3\varepsilon|\mathbf{u}|$ vertices, and similarly for $\mathbf{v}$. Hence there are at most $\varepsilon|\mathbf{u}|$ vertices in $\mathbf{u}$ that are not typical to $\mathbf{v} \setminus U$, and similarly for $\mathbf{v}$. We will use only typical vertices for embedding, so let U' denote the set U together with vertices not typical to $\mathbf{u} \setminus U$ or $\mathbf{v} \setminus U$, respectively. Observe that for each such vertex $u \in \mathbf{u}$ we have

$$\begin{aligned} |N(u) \cap (\mathbf{v} \setminus U')| &\geq (d-\varepsilon)|\mathbf{v} \setminus U| - \varepsilon|\mathbf{v}| \\ \boxed{|\mathbf{v} \setminus U| \geq 4\sqrt{\varepsilon}|\mathbf{v}|} \quad &\geq (d-\varepsilon)4\sqrt{\varepsilon}|\mathbf{v}| - \varepsilon|\mathbf{v}| \\ \boxed{d \gg \sqrt{\varepsilon}} \quad &\geq \sqrt{\varepsilon} \cdot 4\sqrt{\varepsilon}|\mathbf{v}| - \varepsilon|\mathbf{v}| \geq 2\varepsilon|\mathbf{v}| \\ \boxed{|\mathbf{v}| \geq \beta k/\varepsilon} \quad &\geq \varepsilon|\mathbf{v}| + \beta k \geq \varepsilon|\mathbf{v}| + |K|, \end{aligned}$$

and similar holds for any $u \in \mathbf{v}$. This means that during embedding we may always find a neighbour of u in $\mathbf{v} \setminus U'$ that was not yet used for embedding. The same applies for both vertices v_1, v_2. In the case (1) the vertices v_1, v_2 are typical to $\mathbf{u}$ and hence we have

$$\begin{aligned} |N(v_i) \cap (\mathbf{u} \setminus U')| &\geq (d(\mathbf{v}_1, \mathbf{u}) - \varepsilon)|\mathbf{u}| - |U' \cap \mathbf{u}| \\ &\geq \overline{\deg}(\mathbf{v}_1, \mathbf{u}) - |U \cap \mathbf{u}| - 2\varepsilon|\mathbf{u}| \\ \boxed{\overline{\deg}(\mathbf{v}_1, \mathbf{u}) - |U \cap \mathbf{u}| \geq 4\sqrt{\varepsilon}|\mathbf{u}|} \quad &\geq 4\sqrt{\varepsilon}|\mathbf{u}| - 2\varepsilon|\mathbf{u}| \geq 2\varepsilon|\mathbf{u}| \\ \boxed{|\mathbf{u}| \geq \beta k/\varepsilon} \quad &\geq \varepsilon|\mathbf{u}| + \beta k \geq \varepsilon|\mathbf{u}| + |K|, \end{aligned}$$

while in the case (2) we have

$$\begin{aligned} |N(v_i) \cap (\mathbf{u} \setminus U')| &\geq |N(v_i) \cap (\mathbf{u} \setminus U)| - \varepsilon|\mathbf{u}| \\ \boxed{|N(v_i) \cap (\mathbf{u} \setminus U)| \geq 3\varepsilon|\mathbf{u}|} \quad &\geq 2\varepsilon|\mathbf{u}| \\ \boxed{|\mathbf{u}| \geq \beta k/\varepsilon} \quad &\geq \varepsilon|\mathbf{u}| + \beta k \geq \varepsilon|\mathbf{u}| + |K|. \end{aligned}$$

We start by embedding the path $t_1 = x_1, t_2, \ldots, t_\ell = x_2$ connecting x_1 with x_2 in K. Embed x_1 in an arbitrary vertex of $\mathbf{u} \setminus U'$. For i going from 2 to $\ell - 2$ we always map t_i to a neighbour of $\varphi(t_{i-1})$ not lying in U'. Now we observe that both $N(v_2) \cap (\mathbf{u} \setminus U')$ and $N(t_{\ell-2}) \cap (\mathbf{v} \setminus U')$ have sizes at least $\varepsilon|\mathbf{v}_1|$, thus there is an edge connecting those two neighbourhoods. Map $t_{\ell-1}$ and t_ℓ in the two endpoints of the edge. The rest of the tree can be then embedded in the greedy manner. □

Finally, we state a very similar lemma that will be used in Chapter 4.

Lemma 2.14. *Let T be a tree with colour classes F_1 and F_2. Let $R \subseteq F_1$, $|R| \leq 2$ such that vertices of R do not have a common neighbour in T (if $|R| = 2$).*

Let $\varepsilon > 0$ and $\alpha > 2\varepsilon$. Let (X, Y) be an ε-regular pair in a graph G with density $d(X, Y) > 3\alpha$ such that $|F_1| \leq \varepsilon|X|$ and $|F_2| \leq \varepsilon|Y|$. Let $X' \subseteq X, Y' \subseteq Y$ be sets satisfying $|X'| > 2\frac{\varepsilon}{\alpha}|X|, |Y'| > 2\frac{\varepsilon}{\alpha}|Y|$.

Let φ be any injective mapping of vertices of R to vertices of X' with degree greater than $3\varepsilon|Y|$ in Y'. Then there exists extension of φ that is an injective homomorphism from T to (X, Y) satisfying $\varphi(F_1) \subseteq X'$ and $\varphi(F_2) \subseteq Y'$.

Proof. We embed vertices of $V(T) \setminus R$ into vertices of X' and Y' which are typical to Y' and X', respectively. Assume that we have already embedded some part of the tree in this way. We claim that every vertex of this partial embedding in X is incident with more than $\varepsilon|Y|$ vertices typical with respect to X' which have not been used for the partial embedding. Similarly, every vertex of the partial embedding in Y is incident with more than $\varepsilon|X|$ vertices typical with respect to Y', which have not been used for the partial embedding.

We give arguments only for vertices embedded into X, arguments for vertices embedded into Y are symmetric. For $\varphi(r) \in X$, $r \in R$, the claim follows from the fact that $\varphi(r)$ has more than $3\varepsilon|Y|$ neighbours in Y' and out of them, at most $\varepsilon|Y|$ are not typical with respect to X' and at most $\varepsilon|Y|$ have already been used for the partial embedding. Let $\varphi(v)$, $v \in V(T) \setminus R$ be a vertex of the partially constructed embedding and without loss of generality assume $\varphi(v) \in X'$. Since $\varphi(v)$ was chosen to be typical with respect to Y', it is adjacent to at least $(d-\varepsilon)|Y'|$ vertices of Y'. Again, out of these vertices, at most $\varepsilon|Y|$ are not typical with respect to X' and at most $\varepsilon|Y|$ have already been used for the partial embedding. Thus, $\varphi(v)$ is typical to at least $(d-\varepsilon)|Y'| - 2\varepsilon|Y| > ((d-\varepsilon)2\frac{\varepsilon}{\alpha} - 2\varepsilon)|Y|$. This is strictly greater than $\varepsilon|Y|$, since $d > 3\alpha$ and $\alpha > 2\varepsilon$.

It follows that if $|R| < 2$, we can construct embedding greedily.

If $|R| = 2$, $R = \{u, v\}$, we first embed vertices of a path connecting u and v, starting from u and embedding all but the last two internal vertices of a path into typical vertices, last embedded vertex being u'. Then we find an edge between the sets the set X'' of vertices of $N(u')$ which are typical to Y' and set Y'' of vertices of $N(v)$ which are typical to X'. Since, X'' and Y'' have size greater than $\varepsilon|X|$ and $\varepsilon|Y|$, respectively by our previous argument, from ε-regularity of (X, Y), it follows that there is an edge xy between X'' and Y''. We embed the last two internal vertices to x and y.

□

Chapter 3

A local approach to the Erdős-Sós conjecture

After one verifies that the Erdős-Sós conjecture is true both for trees of diameter at most three, and for paths (this was done already by Erdős and Gallai in 1959 [EG59]) one can observe that such trees can be embedded even in the case when the host graph contains a vertex of degree at least k and its minimum degree is at least $k/2$. This is trivial for trees of small diameter, while for the case of paths this follows from the mentioned proof of Erdős and Gallai.

While this local condition on the minimum and maximum degree of G, indeed, suffices both for these special cases, it already fails for trees of diameter four, as is demonstrated by the following example from [HRSW16]. Let T be a tree consisting of a vertex connected to centres of three stars on $k/3$ vertices and let G be a graph consisting of a vertex complete to either two cliques of size $k/2$, or $K_{k/2,k/2}$. Then $\Delta(G) \geq k$ and $\delta(G) \geq k/2$, but T is not contained in G (see Figure 3.1). This example shows that it would be naïve to try to prove the Erdős-Sós conjecture in the most general setting using only the local consequence of the bound on the average degree on the maximum and minimum degree of G. We will actually show in Section 3 that trees of diameter at most three and paths are special cases; with high probability, a random tree on $k+1$ vertices cannot be embedded in the host graph with two cliques from Figure 3.1.

Despite this fact, we devote this chapter to this *local* approach to the Erdős-Sós conjecture, showing for example, that it can be used to prove an approximate version of the Erdős-Sós conjecture for trees such that their size is linear in the size of the host

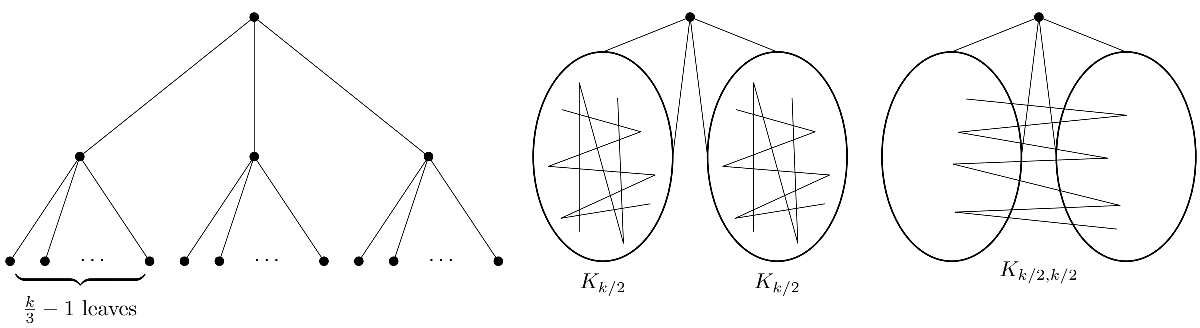

Figure 3.1: A tree on $k+1$ vertices and two host graphs of the same size showing that there are graphs with $\Delta(G) = k$ and $\delta(G) \geq k/2$ that do not contain a tree on $k+1$ edges. The example is taken from [HRSW16].

graph, while their maximum degree is sublinear. The idea to use only conditions on the minimum and maximum degree comes from the paper of Havet, Reed, Stein, and Wood [HRSW16].

We discuss the following natural questions.

1. Which trees can be embedded in any host graph satisfying $\Delta(G) \geq k$ and $\delta(G) \geq k/2$?

2. What is the smallest constant c_1 such that every graph with $\Delta(G) \geq k$ and $\delta(G) \geq c_1 k$ allows embedding of any tree with $k+1$ vertices? Strictly speaking, the smallest constant may not exist. On the other hand, setting $c_1 = 1$ clearly suffices.

3. Is there a minimal c_2 such that every graph with $\Delta(G) \geq c_2 k$ and $\delta(G) \geq k/2$ allows embedding of any tree with $k+1$ vertices?

4. What is the minimal number of vertices of degree at least k that a graph G with $\delta(G) \geq k/2$ has to contain, so that it then allows embedding of any T on $k+1$ vertices?

The second question was considered in the paper of Havet, Reed, Stein, and Wood [HRSW16], and we only state their results.

1) Restricting the class of embedded trees

We observe that the example graph with two cliques from Figure 3.1 actually provides a large class of trees on $k+1$ vertices that cannot be embedded in this graph.

Proposition 3.1 (S. Wagner, personal communication)**.** *For even k it holds that the probability that a random unlabelled tree of size $k+1$ can be embedded in the graph G consisting of a vertex complete to two cliques of size $k/2$ is in $\mathcal{O}(k^{-1/2})$.*

Proof. We at first classify trees on $k+1$ vertices that can be embedded in G. A vertex $u \in T$ is a centroid, if after removing it from T we obtain a family of trees such that each tree is of size at most $k/2$. Since the size of the graph is the same as the size of the tree that we embed, only a centroid of T can be embedded in the vertex of G complete to all other vertices. Since $k+1$ is odd, the centroid of the tree is unique. Hence T can be embedded if and only if the subtrees created after removing its centroid can be partitioned into two classes such that the number of vertices in each class is $k/2$. We call such trees *balanced*.

Let r_k be the number of unlabelled rooted trees with k vertices. A formula of Otter (see e.g. page 481 of [FS10]) states that $r_k = \Theta(k^{-3/2} \cdot B^k)$ for some positive constant B. Similarly, the number of unlabelled unrooted trees s_k is in $\Theta(k^{-5/2} \cdot B^k)$ for the same constant B (again page 481 of [FS10]).

Note that the number of balanced trees of order $k+1$ is at most $r_{k/2+1}^2$, since each such tree can be decomposed into two rooted trees with $k/2+1$ vertices each. Hence the number of balanced trees is in $\mathcal{O}(k^{-3}B^k)$. Comparing this with the sequence s_k, we conclude that the probability that a random unlabelled tree is balanced goes to 0 at a rate of at least $k^{-1/2}$.

□

2) Greater minimum degree

The second question was considered by Havet, Reed, Stein, and Wood in [HRSW16]. They conjectured the following:

Conjecture 3.2 (Conjecture 1.1 in [HRSW16])**.** *If G is a graph such that $\delta(G) \geq \lfloor 2k/3 \rfloor$ and $\Delta(G) \geq k$, then G allows embedding of any tree on $k+1$ vertices.*

As one can see from the example in Figure 3.1, this is tight. As an evidence for their conjecture, they prove its two following weakened variants. The first variant relaxes the condition on the maximum degree:

Theorem 3.3 (Theorem 1.2 in [HRSW16])**.** *There is a function g such that any graph G with $\delta(G) \geq \lfloor 2k/3 \rfloor$ and $\Delta(G) \geq g(k)$ allows embedding of any tree on $k+1$ vertices.*

The second weakening on the other hand shows that the constant c_1 from the second question is strictly smaller than 1.

Theorem 3.4 (Theorem 1.3 in [HRSW16])**.** *There is a constant $\varepsilon > 0$ such that if G is a graph with $\delta(G) \geq (1-\varepsilon)k$ and $\Delta(G) \geq k$, then G allows embedding of any tree on $k+1$ vertices.*

3) Greater maximum degree

The third question seems to be similar to the previous one. The example in Figure 3.1 shows that we have to take $\Delta(G) \geq 4k/3$. We conjecture that this is tight:

Conjecture 3.5. *If G is a graph such that $\delta(G) \geq k/2$ and $\Delta(G) \geq 4k/3$, then G allows embedding of any tree on $k+1$ vertices.*

If true, this conjecture would imply that the constant $2/3$ from Theorem 3.3 can be improved to $1/2$. We were able to verify the weakening of Conjecture 3.5 with $\Delta(G) \geq 4k/3$ replaced by $\Delta(G) \geq g(k)$ for some function g for trees of diameter at most four.

Theorem 3.6. *If G is a graph with $\delta(G) \geq k/2$ and $\Delta(G) \geq 2k^7$, then G allows embedding of any tree from on $k+1$ vertices of diameter at most four.*

Note that the Erdős-Sós conjecture was also verified for trees of diameter four in [McL05], but these two results are incomparable.

4) Many high degree vertices

Finally we consider the question of how many vertices of degree k a graph with G with $\delta(G) \geq k/2$ has to have so as to contain all trees on $k+1$ vertices. We propose the following conjecture.

Conjecture 3.7. *Every graph G on n vertices with $\delta(G) \geq k/2$ and at least $\frac{n}{2\sqrt{k}}$ vertices of degree at least k contains every tree of order $k+1$.*

Note that the fraction $\frac{1}{2\sqrt{k}}$ cannot be substantially improved due to the following example in the spirit of example from Figure 3.1.

Let k be an odd square and T be a tree of order $k+1$ consisting of a vertex connected to centres of $\sqrt{k}$ stars on $\sqrt{k}$ vertices. Let G be a graph consisting of two disjoint cliques of order $\frac{k-1}{2}$ and $\frac{k+1}{2}$, and an independent set of $\frac{\sqrt{k}-1}{2}$ vertices complete to both cliques. A simple calculation shows that the proportion of high degree vertices of G is

$$\frac{\frac{\sqrt{k}-1}{2}}{k+\frac{\sqrt{k}-1}{2}} < \frac{1}{2\sqrt{k}} .$$

Note that for any $c<1$ the left hand side is larger than $\frac{c}{2\sqrt{k}}$ for sufficiently large k. One can check that G does not contain T.

We prove a weakened variant of Conjecture 3.7. Specifically, we show that it is asymptotically true if the number of high degree vertices of G as well as the size of the tree T is linear in the size of G and, moreover, the maximum degree of T is sublinear. As we have already mentioned, we state a finer version of this result for skewed trees. Specifically, if we know that the skew of T is at most r, then G contains T even if its minimum degree is roughly rk.

Theorem 3.8. *For any $r,\eta>0$ there exist n_0 and $\gamma>0$ such that the following holds. Let G be a graph of order $n>n_0$ and T a tree of order k with two colour classes T_1,T_2 such that $|T_1|\le rk$ and $\Delta(T_2)\le\gamma k$. If $\delta(G)\ge rk+\eta n$, and at least ηn vertices of G have degree at least $k+\eta n$, then G contains T.*

We postpone the proof of this theorem to the last section of this chapter. As a special case for $r=1/2$, we get the following weakening of Conjecture 3.7.

Corollary 3.9. *For any $\eta>0$ there exist n_0 and $\gamma>0$ such that the following holds. Let G be a graph of order $n>n_0$ and T a tree of order k such that $\Delta(T)\le\gamma k$. If $\delta(G)\ge k/2+\eta n$, and at least ηn vertices of G have degree at least $k+\eta n$, then G contains T.*

Finally, Corollary 3.9 yields an approximate version of the Erdős-Sós conjecture for trees with sublinear degree.

Theorem 3.10. *For any $\eta>0$ there exist n_0 and $\gamma>0$ such that for every $n>n_0$, any graph of order n with average degree $\overline{\deg}(G)\ge k+\eta n$ contains every tree on k vertices with maximum degree $\Delta(T)\le\gamma k$.*

Proof. Let $\eta'=\eta/2$ and let G be a graph on $n\ge n_0=\frac{n_{0,C3.9}(\eta')}{\eta}$ vertices (here $n_{0,C3.9}(\eta')$ means the output of Corollary 3.9 with input η'). Suppose that $k\ge\eta n/2$.

We choose a subgraph $G'\subseteq G$ such that $\overline{\deg}(G')\ge k+\eta n$ and $\delta(G')\ge k/2+\eta n/2$. Hence we know that the size of G' is at least $k+\eta n\ge\eta n\ge n_{0,C3.9}$.

We claim that at least $\eta'|G'|$ vertices of G' have degree at least $k+\eta' n$ and hence we may apply Corollary 3.9. Otherwise, most of the vertices of G' have degree less than $k+\eta' n$ and we may compute that

$$\overline{\deg}(G')\le\eta'\cdot n+(1-\eta')\cdot(k+\eta' n)<\eta' n+(k+\eta' n)=k+2\eta' n\le k+\eta n,$$

a contradiction. □

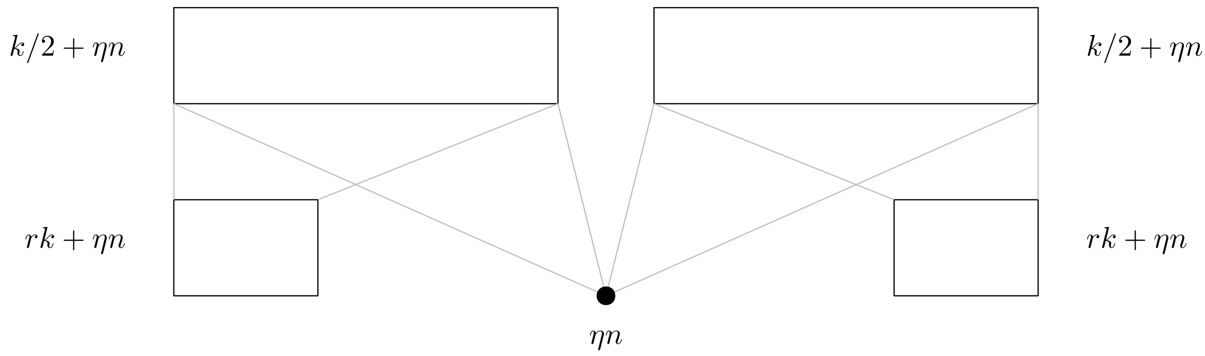

Figure 3.2: Example showing that the condition on bounded degree is needed in the statement of Theorem 3.8.

Let us state one more remark regarding Theorem 3.8. Although the result of Ajtai, Komlós, Simonovits, and Szemerédi [AKSSa, AKSSc, AKSSb] implies that the condition on the maximum degree $\Delta(T)$ in Theorem 3.10 is only an imperfection, it cannot be omitted in the statement of Theorem 3.8. We show that the theorem is false if we omit this condition.

Specifically, we show that for all $0 < r < 1/3$ there exists $\eta > 0$ such that the following is true. Let G be a graph on n vertices consisting of two disjoint copies of complete bipartite graphs with colour classes of sizes $rk + \eta n$ and $k/2 + \eta n$. Moreover, ηn additional vertices are complete to both larger colour classes of the two bipartite graphs (see Figure 3.2). Let T be a tree on k vertices consisting of a vertex x complete to centres of rk stars of sizes $\lfloor \frac{1}{r} \rfloor$ and $\lceil \frac{1}{r} \rceil$. The smaller colour class of T has size rk. Note that for fixed r the maximum degree of this smaller colour class of T is constant, though it is not true for the larger colour class, hence Theorem 3.8 does not apply. We claim that the tree T is not contained in G if we choose η sufficiently small. Suppose that there is an embedding of T in G. Since G is bipartite with one colour class of size at most $2rk + 3\eta n < (1-r)k$ if k is big enough and η sufficiently small, the vertex x must be embedded in the larger colour class. Out of $(1-r)k - 1$ leaves at least $(1-r)k - 1 - \eta n \cdot \lceil \frac{1}{r} \rceil > k/2 + \eta n$ have to be embedded in the same colour class as x, a contradiction.

Theorem 3.8 is thus an example of an asymptotic result that does not seem to have a natural exact strengthening. On the other hand, we believe that the assumption on the sublinear maximum degree in Corollary 3.9 can be dropped.

3.1 Proof of Theorem 3.6

In this subsection we prove Theorem 3.6. For trees of diameter four we use the notation from Section 2. The proof is reasonably straightforward, because due to the assumption that there is a vertex with huge degree, we have a lot of flexibility, if we place the central vertex $x \in T$ in the highest degree vertex of G. On the other hand, it does not always suffice to embed c in the highest degree vertex – as a counterexample consider G to be a complete bipartite graph with one partite of size $k/2$ and the other partite arbitrarily huge. If $|W| > k/2$, its central vertex has to be embeded in a low degree vertex. This harder case, however, does not occur when $|W| < k/2$. This will be crucial in the next subsection, where we prove the corresponding result for dense graphs, being guided by the approach for small diameter trees from this section.

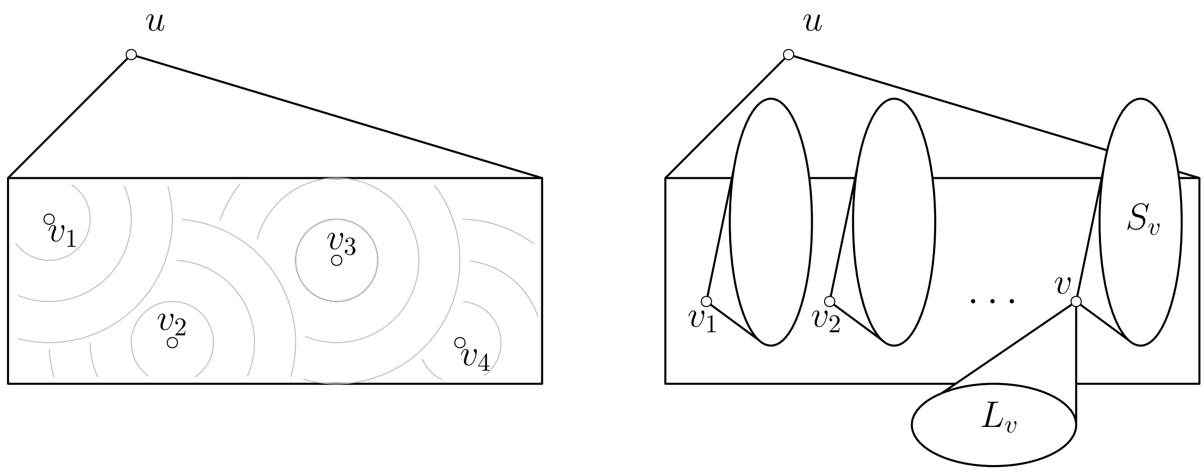

Figure 3.3: Two embedding configurations from Lemma 3.11 (left) and Theorem 3.6 (right).

Before proving Theorem 3.6 we propose the following lemma.

Lemma 3.11. *If $T \in \mathcal{T}_{k+1}$ and G is a graph with $\Delta(G) \geq 2k \cdot \Delta_2(G)^{\operatorname{diam}(T)-2}$ and $\delta(G) \geq k/2$, then G allows embedding of T.*

Proof. Let u be the vertex of degree Δ. We find a sequence of vertices $v_1, \dots, v_k$ in $N(u)$ such that for each $1 \leq i, j \leq k$, $i \neq j$, we have $\operatorname{dist}_{G\setminus\{u\}}(v_i, v_j) \geq \operatorname{diam}(T) - 1$. We find the desired set in k steps. In the i-th step we choose the vertex v_i and mark all vertices in G such that their distance to v_i in $G \setminus \{u\}$ is at most $\operatorname{diam}(T) - 2$. In each step we choose the new vertex only from the vertices that have not been marked yet (see Figure 3.3). Since the number of vertices with their distance to fixed vertex being precisely ℓ can be bounded by $\Delta(G')^\ell = \Delta_2(G)^\ell$, the number of marked vertices in each step is bounded by

$$\Delta_2(G)^0 + \Delta_2(G)^1 + \cdots + \Delta_2(G)^{\operatorname{diam}(T)-2} \leq 2\Delta_2(G)^{\operatorname{diam}(T)-2} .$$

Thus, as $(k-1) \cdot 2\Delta_2(G)^{\operatorname{diam}(T)-2} + 1 \leq \Delta(G)$, we may, indeed, find all the vertices $v_1, \dots, v_k$ using the described procedure.

The embedding of T is now straightforward. Let c be a centre of T, i.e., a vertex of T such that the subtrees $T_1, \dots, T_p$ of $T \setminus \{c\}$ are of sizes at most $(k+1)/2 - 1 \leq k/2$. If there are two centres, we choose any. We embed c in u and then gradually embed the subtrees $T_1, \dots, T_p$. We embed the root of each T_i in v_i and then proceed with embedding of the rest of T_i in $G \setminus \{u\}$ by the greedy method. This can be done for all the subtrees, because we know that $\delta(G \setminus \{u\}) \geq k/2 - 1 \geq |T_i| - 1$ and two overlapping trees T_i and T_j would imply that there is a path of length $\operatorname{diam}(T) - 1 + 2 = \operatorname{diam}(T) + 1$ in T, a contradiction. □

We now proceed with a proof of Theorem 3.6.

Proof. Let u_0 be the vertex of G of degree $\Delta(G)$. We define L as the set of vertices of degree at least k and let S be its complement. Invoking Lemma 3.11, we further assume that $\Delta_2(G) \geq k^3 + 1$ as otherwise we would have $2k \cdot (k^3)^{4-2} \leq \Delta(G)$. Let u be the vertex of $G \setminus \{u_0\}$ of degree $\Delta_2(G)$. Erase the edge uu_0 if present; now we assume that all of at least k^3 neighbours of u are S-vertices, otherwise we would get a smaller counterexample by deleting an edge between two L-vertices, neither of them being u_0.

Let x be the vertex of T such that the distance of all vertices of T to x is at most two. Let $T_1, \ldots, T_{|V|}, |T_1| \geq \cdots \geq |T_{|V|}|$ be the star subtrees with roots $y_1, \ldots, y_{|V|}$ that are children of x.

We now find a sequence of vertices $v_1, \ldots, v_k$ in $N(u)$ such that $(N(v_i) \cup v_i) \cap (N(v_j) \cup v_j) \cap S = \emptyset$. Similarly to Lemma 3.11, we do this in a simple step-by-step manner. In the ith step we choose any unmarked vertex from $N(u)$ and mark the vertex itself, its at most $k/2 - 1$ S-neighbours and at most k neigbours of each of these vertices. Since

$$(k-1)(1 + (k/2-1) + k(k/2-1)) + 1 \leq k^3 \leq |N(u)|,$$

we find all the vertices $v_1, \ldots, v_k$ by this procedure.

We now consider two cases depending on the skew of T. At first suppose that $|V| > |W|$ and, thus, $|W| \leq k/2 - 1$. We embed r in u (i.e., set $\varphi(r) = u$) and proceed with embedding its subtrees $T_1, \ldots, T_{|V|}$. In the i-th step we start by embedding r_i in v_i. Then we embed its leaf neighbours in the vertices from $N(v_i)$ that have not been used yet for embedding. As we know that no two vertices v_i, v_j are connected by an edge, we may, indeed, always do it, because $\deg(v_i) \geq k/2 \geq |W| + |\{r\}|$.

Further we assume that $|V| \leq |W|$. If we had that $|N(v_i) \cap S| \geq |T_1| - 1$ for all $1 \leq i \leq a_1$, we could embed T by setting $\varphi(x) = u$, $\varphi(y_i) = v_i$ for all i and finally embedding at most $|T_1| - 1$ leaf neighbours of all vertices y_i in their S-neighbourhood. Thus, we assume the existence of a vertex $v \in \{v_1, \ldots, v_k\}$ such that $|N(v) \cap S| \leq |T_1| - 2$. Set $L_v := N(v) \cap L$ and $S_v := N(v) \cap S$. Note that $|L_v| + |S_v| \geq k/2$ (see Figure 3.3).

We set $\varphi(x) = v$ and then proceed by step-by-step greedy embedding of subtrees $T_{|V|}, T_{|V|-1}, \ldots$ in $G \setminus L_v$. We can continue this process while it is for $|V| \geq \ell \geq 1$ the case that each vertex in $(\{v\} \cup N(v)) \setminus L_v$ has degree at least $|\{x\} \cup T_{|V|} \cup \cdots \cup T_\ell| - 1$ in $G \setminus L_v$, i.e., while it holds that

$$|T_{|V|}| + \cdots + |T_\ell| \leq k/2 - |L_v| \leq |S_v|.$$

In the following, let ℓ be the smallest number satisfying the inequality, i.e., we have embedded trees $T_\ell, \ldots, T_{|V|}$. Now we consider two cases. At first suppose that x has at least $|S_v|$ leaf neighbours. The preceding procedure than embeds the last $|S_v|$ leaf subtrees $T_{|V|-|S_v|+1}, \ldots, T_{|V|}$ in S_v. Observe that $|V| \leq |W|$ implies $|V| \leq k/2$. Hence, $|L_v| \geq k/2 - |S_v| \geq |V| - |S_v|$, thus we can embed the vertices $y_1, \ldots, y_{|V|-|S_v|}$ in L_v and finish with embedding their leaf neighbours by the greedy method.

In the second case we embed the vertex $y_{\ell-1}$ in $u \in L_v$. We know that $|T_{\ell-1}| + \cdots + |T_{|V|}| \geq |S_v| + 1$. Now it suffices to show that $|L_v| \geq \ell - 1$, because then we can embed all the vertices $y_1, \ldots, y_{\ell-2}$ in $L_v \setminus \{u\}$ and then finish by embedding their leaf neighbours in a greedy manner.

From the fact that all subtrees $T_1, \ldots, T_{\ell-2}$ have size at least two and $|T_1| \geq |S_v| + 2$ we conclude that

$$|T_2| + \cdots + |T_{\ell-2}| \geq k - (|S_v| + 2) - (|S_v| + 1) = k - 2|S_v| - 3.$$

Note that each tree $T_2, \ldots, T_{\ell-2}$ is of size at least two, thus

$$\ell - 3 \leq \frac{|T_2| + \cdots + |T_{\ell-2}|}{2} \leq \frac{k-3}{2} - |S_v|.$$

If k is odd, we have $|L_v| + |S_v| \geq \frac{k+1}{2}$, thus $\ell - 3 \leq \frac{k+1-4}{2} - |S_v| \leq |L_v| - 2$. If k is even, we have actually $\ell - 3 \leq \frac{k-4}{2} - |S_v|$, thus we also get $\ell - 3 \leq |L_v| - 2$. In either case it holds that $\ell - 1 \leq |L_v|$, as desired. □

Observe that the proof of Lemma 3.11 and Theorem 3.6 can be easily altered to give a proof of the following result.

Proposition 3.12. *If G is a graph with $\delta(G) \geq rk$, $r \leq 1/2$ and $\Delta(G) \geq 2k^7$, then G allows embedding of any tree from $\mathcal{T}_{k+1}^r$ of diametr at most four such that its smaller colour class contains the vertex x such that the distance of all other vertices of T from x is at most two.*

Indeed, it suffices to look only at configurations in which we embed the vertex x in the highest degree vertex. This is similar to the proof in the next section.

3.2 Proof of Theorem 3.8

In this section we prove Theorem 3.8. We split the proof into three parts. At first we preprocess the host graph by applying the regularity lemma and we partition the tree by applying Lemma 2.11. In the second part we find a suitable matching structure in the host graph. In the last part we embed the tree in the host graph.

Preprocessing the host graph and the tree

Fix η, r. Suppose that $\eta < 1$. Choose $d, \varepsilon, \beta, n_0$ such that

$$\begin{aligned}
d &= \frac{(\eta r)^2}{1000},\\
\varepsilon &= \frac{(\eta r d)^{20}}{10^{15}},\\
\beta &= \min\left(\beta_{P2.13}(d,\varepsilon,f), \frac{\eta d}{10^5 \cdot M_{L2.6}(\varepsilon)}\right),\\
\gamma &= \gamma_{P2.12}(d,\varepsilon,\beta),\\
n_0 &= \max\left(n_{0,L2.6}(\varepsilon), 2\frac{k_{0,P2.12}(d,f)}{\eta}, n_{0,P2.13}(d,\varepsilon,\beta)\right).
\end{aligned}$$

Let G be a fixed graph on $n \geq n_0$ vertices with at least ηn vertices of degree $k+\eta n$ and with $\delta(G) \geq rk + \eta n$. Suppose that $k \geq \eta n/2$. We apply the regularity lemma (Lemma 2.6) on G with $\varepsilon_{L2.6} = \varepsilon$ and obtain an ε-regular equitable partition $V_0, V_1, \ldots, V_m$ with $1/\varepsilon \leq m \leq M_{\mathrm{RL}}$ clusters. Each cluster has average degree at least $rk + \eta n$.

Erase all edges within sets V_i of the partition, between irregular pairs, and between pairs of density lower that d. We have erased at most $m \cdot \binom{n/m}{2} \leq \frac{n^2}{m} \leq \varepsilon n^2$ edges withing the sets V_i, at most $\varepsilon m^2 \cdot (n/m)^2 = \varepsilon n^2$ edges in irregular pairs, and at most $\binom{m}{2} \cdot d \cdot (n/m)^2 \leq d \cdot n^2$ edges in pairs of low density. Erase the garbage set V_0 and all of at most $\varepsilon n \cdot n$ incident edges. Note that we have erased at most $(3\varepsilon + d)n^2$ edges. We abuse the notation and still call the resulting graph G.

Note that the quantity $\sum_{1 \leq i \leq m} |V_i| \cdot \overline{\deg}(V_i)$ dropped down by at most $(6\varepsilon + 2d)n^2$. Thus there are at most $\sqrt{6\varepsilon + 2d} \cdot m$ clusters such that their average degree dropped down by more than $\sqrt{6\varepsilon + 2d} \cdot n$. Delete all such clusters and incident edges. We again call the resulting graph G. The average degree of each cluster of G that was not deleted at first dropped by at most $\sqrt{6\varepsilon + 2d} \cdot n$. Then we erased at most $\sqrt{6\varepsilon + 2d} \cdot m$ clusters, so now it is at least $rk + \eta n - 2 \cdot \sqrt{6\varepsilon + 2d} \cdot n > rk + \eta n/2$. Moreover, G contains at least $(\eta - \varepsilon - \sqrt{6\varepsilon + 2d})n \geq \eta n/2$ vertices of degree at least $k + \eta n - 2 \cdot \sqrt{6\varepsilon + 2d} \cdot n \geq k + \eta n/2$. Hence, there exists a cluster, without loss of generality it is V_1, such that the proportion of vertices of degree at least $k + \eta n/2$ in that cluster is at least $\eta/2 \geq \varepsilon$. If we denote by L this set of high degree vertices of V_1, we thus we have $\overline{\deg}(V_1, V_i) \geq \overline{\deg}(L, V_i) - \varepsilon|V_i|$ from regularity of each pair (V_1, V_i). This yields that $\overline{\deg}(V_1) \geq \overline{\deg}(L) - \varepsilon n \geq k + \eta n/3$.

The cluster graph $\mathbf{G}$ of G is a graph such that its vertex set are the clusters of G and there is an edge between two vertices of $\mathbf{G}$ if and only if there is a regular pair of density at least d between the corresponding two clusters in G. The weight of each edge $\mathbf{uv}$ is the average degree of $\mathbf{u}$ in $\mathbf{v}$. We use boldface font to denote the vertices and sets of vertices of $\mathbf{G}$. The vertex set of $\mathbf{G}$ is denoted $\mathbf{v}_1, \ldots, \mathbf{v}_m$, where each $\mathbf{v}_i$ corresponds to the cluster V_i of G.

After preprocessing the host graph we turn our attention to the tree T. Let T_1, T_2 be its colour classes such that $|T_1| \leq rk$ and $\Delta(T_2) \leq \gamma k$. We apply Lemma 2.11 with parameter $\ell_{L2.11} = \beta k$ and obtain its one-sided βk-fine partition $(W, \mathcal{D}), \mathcal{D} = \mathcal{D}' \sqcup \mathcal{D}''$ such that $|W| \leq 336(1 + \gamma k)/\beta$ and $|\bigcup \mathcal{D}''| \leq 336/\beta$. Moreover, for each $K \in \mathcal{D}'$ we have $|K| \leq \beta k$ and for each $K \in \mathcal{D}''$ we have $|K| = 1$. Also note that $W \subseteq T_1$.

Structure of the host graph

We now find a suitable structure in the cluster graph $\mathbf{G}$ that will be used for the embedding of T. It suffices to look at the cluster $\mathbf{v}_1$, that will serve for the embedding of the seeds of T, and its neighbourhood.

Let $\mathbf{M}$ a maximal matching in $N(\mathbf{v}_1)$. We will denote by $\mathbf{M}$ both the graph and its underlying vertex set. Suppose that $\mathbf{uv} \in \mathbf{M}$. Note that from the condition on maximality we get that there cannot be two vertices $\mathbf{x} \neq \mathbf{y} \in N(\mathbf{v}_1) \setminus \mathbf{M}$ such that both $\mathbf{xu}$ and $\mathbf{yv}$ are edges of $\mathbf{G}$. Thus there are two possibilities for each edge $\mathbf{uv}$; either only one of its endpoints have neighbours in $N(\mathbf{v}_1) \setminus \mathbf{M}$, or both of its endpoints have just one neighbour in $N(\mathbf{v}_1) \setminus \mathbf{M}$. We can get rid of the second special case as follows. For each vertex in $N(\mathbf{v}_1) \setminus \mathbf{M}$ we either delete it if it is a common neighbour of at least $\eta m/40$ matching pairs, or we delete all edges in at most $2 \cdot \eta m/40$ regular pairs connecting the vertex with these matching pairs. In this way we delete at most $40/\eta$ clusters and the degree of all remaining clusters of $\mathbf{G}$ drops down by at most $\eta m/20 \cdot |\mathbf{v}_1| + 40/\eta \cdot |\mathbf{v}_1| \leq (\eta/20 + 40\varepsilon/\eta) \cdot n \leq \eta n/10$. We abuse the notation and still call the resulting graph $\mathbf{G}$. The degree of $\mathbf{v}_1$ is at least $k + \eta n/3 - \eta n/10 \geq k + \eta n/5$ and the average degree of every cluster is similarly at least $rk + \eta n/5$. The matching $\mathbf{M}$ is still maximal in $N(\mathbf{v}_1)$. Moreover, we can split $\mathbf{M}$ into two colour classes, $\mathbf{M} = \mathbf{M}_1 \cup \mathbf{M}_2$, in such a way that only clusters from $\mathbf{M}_2$ have neighbours in $N(\mathbf{v}_1) \setminus \mathbf{M}$. Let $\mathbf{O}_1 = N(\mathbf{v}_1) \setminus \mathbf{M}$. Note that it is an independent set. Define $\mathbf{O}_2 = N(\mathbf{O}_1) \setminus \{\{\mathbf{v}_1\} \cup \mathbf{M}\}$. Note that $N(\mathbf{v}_1) \cap \mathbf{O}_2 = \emptyset$. All these sets are shown in Figure 3.4.

Embedding

We split the last part further into three subparts. At first we give an overview of the method that we use for the construction of the mapping φ. Then we formulate several

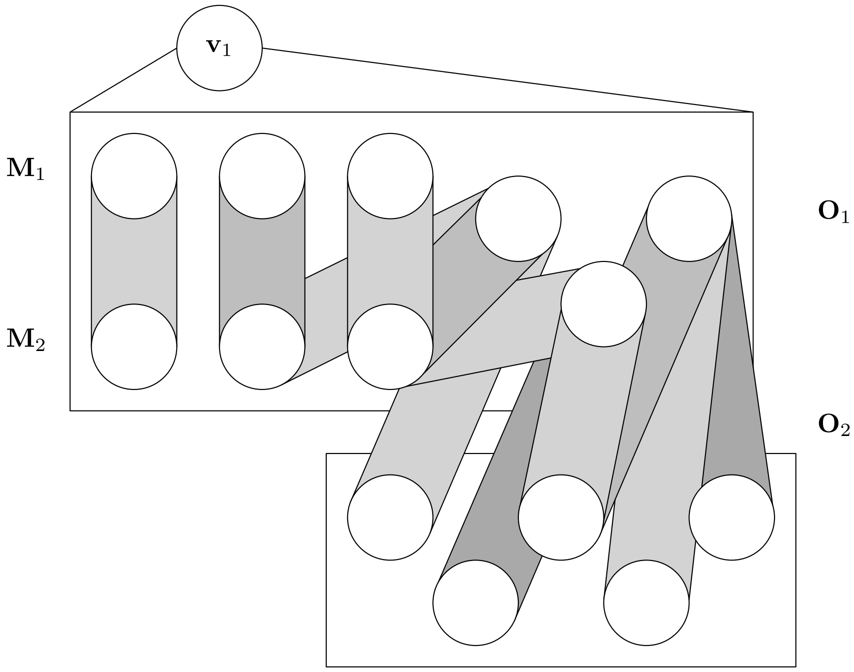

Figure 3.4: Cluster $\mathbf{v}_1$ and four sets of clusters $\mathbf{M}_1, \mathbf{M}_2, \mathbf{O}_1, \mathbf{O}_2$ that will be used for embedding. The regular pairs of different density are sketch by shades of grey (we omit pairs touching $\mathbf{v}_1$).

preparatory technical claims. In the last part we propose the embedding algorithm.

Overview

We gradually construct an injective mapping φ from T to G. In each step φ denotes the partial embedding that we already constructed. The idea behind the embedding process is very straightforward – we will try to embed microtrees of $\mathcal{D}$ inside the regular pairs in $\mathbf{M}$ and 'through' the vertices of $\mathbf{O}_1$. We will, however, have to overcome several technical difficulties.

One of the standard approaches of embedding trees, pursued, e.g., in [KPR18], is to start by embedding the seeds of T in vertices of two clusters (one for each colour class) such that the neighbourhood of these special clusters is sufficiently rich. Moreover, we embed the seeds in such vertices that are typical to almost all neighbouring clusters. We then split the microtrees in T into several subsets and embed these each subset of microtrees in some part of the neighbourhood of the special clusters. Here we take a different approach. We start in the same way by embedding the seeds W of T in a high degree cluster of G that we call $\mathbf{v}_1$. We then propose an algorithm that iterates over clusters in the neighbourhood of $\mathbf{v}_1$, each time finding two clusters that can be used for embedding of a microtree.

There are two main technical difficulties that we have to overcome. Recall that each seed is embedded in a vertex that is typical to almost all clusters. This means that when we choose a pair of clusters that we will use for embedding, we have to find a microtree that has not yet been embedded such that its adjacent seeds are embedded in vertices typical to the first cluster from the pair. We can ensure that there will be such microtree, unless the number of vertices that remain to be embedded, is very small, specifically $\sqrt[4]{\varepsilon}k$. To ensure that we can embed the whole tree T, we at first allocate a small fraction of vertices $F \subseteq \bigcup(\mathbf{M} \cup \mathbf{O})$ that we do not use for the embedding during the main embedding procedure. When only at most $\sqrt[4]{\varepsilon}k$ vertices remain to be embedded, we finally embed this small proportion of trees in the set F.

The second technical problem is that we cannot ensure that all the microtrees have the same skew. This complicate the main embedding procedure that would have been simpler in the case of microtrees with uniform skew. During the embedding procedure we behave against intuition and sometimes redefine an embedding of some microtrees.

Preparations

Note that there there may be at most $\sqrt{\varepsilon}|\mathbf{v}_1|$ vertices of $\mathbf{v}_1$ that are not typical to more than $\sqrt{\varepsilon}m$ clusters. Indeed, otherwise there would be at least $\varepsilon m|\mathbf{v}_1|$ pairs of a cluster and a vertex not typical to it, which in turn implies existence of a cluster such that more than $\varepsilon|\mathbf{v}_1|$ vertices are not typical to it. For each cluster $\mathbf{v} \in \mathbf{M}_1 \cup \mathbf{O}_1$ fix its arbitrary subset $F_\mathbf{v}$ of size $\lfloor \eta rd|\mathbf{v}|/300 \rfloor$. By the same reasoning there are at most $\sqrt{\varepsilon}|\mathbf{v}_1|$ vertices of $\mathbf{v}_1$ that are not typical to more than $\sqrt{\varepsilon}m$ sets $F_{\mathbf{v}_i}$.

We invoke Proposition 2.12 with parameters $d_{P2.12} = d$, $\beta_{P2.12} = \beta$, $\varepsilon_{P2.12} = \varepsilon$, and $f_{P2.12} = f$. We also choose $\mathbf{v}_{2,P2.12} = \mathbf{v}_2$ to be any cluster from the neighbourhood of $\mathbf{v}_{1,P2.12} = \mathbf{v}_1$. Finally, we define the set $U_{P2.12}$ to be the set of at most $2\sqrt{\varepsilon}|\mathbf{v}_1|$ vertices not typical to more than $\sqrt{\varepsilon}m$ neighbouring clusters $\mathbf{v}_i$, or their subsets $F_{\mathbf{v_i}}$. Note that due to our initial choice of constants all the conditions from the statement of the proposition are satisfied. Hence we embed the vertices of W in $\mathbf{v}_1$, while the vertices of $\bigcup \mathcal{D}''$ will be embedded in $\mathbf{v}_2$. Moreover, each vertex from W is typical to all but at most $\sqrt{\varepsilon}m$ clusters $\mathbf{v}_i$ and their fixed subsets $F_{\mathbf{v}_i}$ of size $\lfloor \eta rd|\mathbf{v}_1|/300 \rfloor$.

Note that each microtree $K \in \mathcal{D}'$ has at most two neighbours in W. We call a cluster $\mathbf{u} \neq \mathbf{v}_1$ *nice* with respect to $K \in \mathcal{D}'$, if all neighbours of K are embedded in vertices typical to $\mathbf{u}$. Note that each vertex from W was mapped to a vertex that is typical to all but at most $\sqrt{\varepsilon}m$ clusters, thus for each tree K there are at most $2\sqrt{\varepsilon}m$ clusters that are not nice to K. We will now, yet again, employ a simple doublecounting argument. This time we doublecount connections between microtrees from $\mathcal{D}'$ and clusters that are not nice to them; each such connection is weighted by the size of the tree. We get that there are at most $2\sqrt[4]{\varepsilon}m$ clusters such that if we take all trees such that the cluster is not nice to them, then the union of all such trees contains more than $\sqrt[4]{\varepsilon}k$ vertices. Delete all such clusters and if they are from $\mathbf{M}$, delete also their neighbours in $\mathbf{M}$. We also delete the cluster $\mathbf{v}_2$. Moreover, if it is the case that $\overline{\deg}(\mathbf{v}_1) > 2k$, we delete several pairs between $\mathbf{v}_1$ and the rest of $\mathbf{G}$ so as to achieve that $\overline{\deg}(\mathbf{v}_1) \leq 2k$. Observe that the average degree of each cluster is still at least

$$\begin{aligned} & rk + \eta n/10 - (4\sqrt[4]{\varepsilon}m + 2)|\mathbf{v_1}| \\ \boxed{m \geq 1/\varepsilon} \quad & \geq rk + \eta n/10 - (4\sqrt[4]{\varepsilon} + 2\varepsilon)n \\ \boxed{\varepsilon \ll \eta} \quad & \geq rk + \eta n/20. \end{aligned}$$

Similarly the degree of $\mathbf{v}_1$ is still at least $\overline{\deg}(\mathbf{v}_1) \geq k + \eta n/20$. We still call the new graph $\mathbf{G}$. We also know for each $\mathbf{u} \in N(\mathbf{v}_1)$ that the number of vertices in microtrees such that $\mathbf{u}$ is not nice to them is at most $\sqrt[4]{\varepsilon}k$.

Now we will define a small set $F \subseteq \bigcup(\mathbf{M} \cup \mathbf{O})$ that will be used at the end for embedding of several leftover microtrees with at most $\sqrt[4]{\varepsilon}k$ vertices.

Claim 3.13. *There is a set $F \subseteq \bigcup(\mathbf{M} \cup \mathbf{O})$ satisfying $|F| \leq \eta r\overline{\deg}(\mathbf{v}_1)/100$, $F_\mathbf{u} \subseteq F \cap \mathbf{u}$ for any $\mathbf{u} \in \mathbf{M}_1 \cup \mathbf{O}_1$ and $|F \cap \mathbf{u}| = |F \cap \mathbf{v}|$ for any $\mathbf{uv} \in \mathbf{M}$. Moreover, if we extend* *the partial mapping φ of T satisfying $\varphi(T) \cap F = \emptyset$ to all trees from $\mathcal{D}$ except of some $\bar{\mathcal{D}} \subseteq \mathcal{D}$ with $|\bigcup \bar{\mathcal{D}}| \leq \sqrt[4]{\varepsilon}k$, then we can injectively extend φ to the whole tree T.*

Proof. We define F as follows. For each $\mathbf{u} \in \mathbf{M}_1 \cup \mathbf{O}_1$ we begin by adding $F_\mathbf{u}$ to F. Then for each set $F_\mathbf{u}$ we find a set of the same size in some neighbour $\mathbf{v} \neq \mathbf{v}_1$ of $\mathbf{u}$ and also add this set to F. We call this set $G_\mathbf{u}$. For $\mathbf{uv} \in \mathbf{M}$ we take $G_\mathbf{u} = F_\mathbf{v}$. For $\mathbf{u} \in \mathbf{O}_1$ we find its neighbouring cluster in $\mathbf{O}_2 \cup \mathbf{M}_2$ with at least $\lfloor \eta r d |\mathbf{u}|/300 \rfloor$ vertices that were not yet added to F and we set $G_\mathbf{u}$ to be this set (we explain later, why we always find a suitable neighbouring cluster). In the case when $F_\mathbf{u} \in \mathbf{O}_1$, but $G_\mathbf{u} \in \mathbf{M}_2$, it is no longer true that $|F \cap \mathbf{u}'| = |F \cap \mathbf{v}'|$ for some matching edge $\mathbf{u}'\mathbf{v}' \in \mathbf{M}$. We again establish the condition by adding $\lfloor \eta r d |\mathbf{u}'|/300 \rfloor$ vertices from $\mathbf{u}'$ to F. This implies that

$$|F| \leq 3 \cdot \sum_{\mathbf{u} \in \mathbf{M}_1 \cup \mathbf{O}_1} \lfloor \eta r d |\mathbf{u}|/300 \rfloor \leq \eta r \overline{\deg}(\mathbf{v}_1)/100.$$

Now we explain, why each cluster $\mathbf{u} \in \mathbf{O}_1$ has a neighbour in $\mathbf{M}_2 \cup \mathbf{O}_2$ with at least $\lfloor \eta r d |\mathbf{u}|/300 \rfloor$ vertices that are not yet in F. Since we know that

$$\overline{\deg}(\mathbf{u}, \bigcup (\mathbf{M}_2 \cup \mathbf{O}_2)) \geq rk > 2|F| > 2\overline{\deg}(\mathbf{u}, F),$$

there is certainly a cluster $\mathbf{v} \in \mathbf{M}_2 \cup \mathbf{O}_2$ such that $\overline{\deg}(\mathbf{u}, \mathbf{v}) > 2\overline{\deg}(\mathbf{u}, F \cap \mathbf{v})$, thus $\overline{\deg}(\mathbf{u}, \mathbf{v} \setminus F) > \overline{\deg}(\mathbf{u}, \mathbf{v})/2 \geq d|\mathbf{v}|/2$, meaning that there is a subset of at least $d|\mathbf{v}|/2 > \lfloor \eta r d |\mathbf{v}|/300 \rfloor$ vertices in $\mathbf{v}$ that can be used to define $G_\mathbf{u}$.

Now we show how to embed any $\bar{\mathcal{D}}$ of small size in F. We define the embedding φ of all trees $K \in \bar{\mathcal{D}}$ in a step-by-step manner. Suppose that $\mathbf{u} \in \mathbf{M}_1 \cup \mathbf{O}_1$ and $G_\mathbf{u} \subseteq \mathbf{v}$. If the at most two neighbours z_1, z_2 of K in W are embedded in two vertices of $\mathbf{v}_1$ that are typical to set $F_\mathbf{u}$ and, moreover, $|\varphi(T) \cap F_\mathbf{u}| \leq \frac{d}{2}|F_\mathbf{u}|$ and $|\varphi(T) \cap G_\mathbf{u}| \leq \frac{d}{2}|G_\mathbf{u}|$, we can compute that for $i = 1, 2$ we have

$$|G_\mathbf{u} \setminus \varphi(T)| \geq (1 - \frac{d}{2})|G_\mathbf{u}| \geq 4\sqrt{\varepsilon}|\mathbf{v}|$$

and

$$\begin{aligned} |N(v_i) \cap (F_\mathbf{u} \setminus \varphi(T))| &\geq |N(v_i) \cap F_\mathbf{u}| - |\varphi(T) \cap F_\mathbf{u}| \\ \boxed{v_i \text{ is typical to } F_\mathbf{u}} \quad &\geq (d - \varepsilon)|F_\mathbf{u}| - |\varphi(T) \cap F_\mathbf{u}| \\ \boxed{\varepsilon \ll \eta r d^2,\ |\varphi(T) \cap F_\mathbf{u}| \leq d|F_\mathbf{u}|/2} \quad &\geq \frac{d}{3}|F_\mathbf{u}| \geq 3\varepsilon|\mathbf{u}|. \end{aligned}$$

Hence we can use Proposition 2.13 case (2) with parameters $U_{P2.13} = \bar{F} \cup \varphi(T)$, where $\bar{F}$ means the complement of F in our graph, $d_{P2.13} = d$, $\varepsilon_{P2.13} = \varepsilon$, $f_{P2.13} = f$, $\beta_{P2.13} = \beta$, $\mathbf{v}_{1,P2.13} = \mathbf{v}_1$, $\mathbf{u}_{P2.13} = \mathbf{u}$, $\mathbf{v}_{P2.13} = \mathbf{v}$, $K_{P2.13} = K$, $v_{1,P2.13} = \varphi(z_1)$, $v_{2,P2.13} = \varphi(z_2)$. The proposition then allows us to embed K.

Now it suffices to show that for any K we always find a suitable $\mathbf{u}$ such that $\varphi(z_1), \varphi(z_2)$ are typical to $F_\mathbf{u}$ and both $F_\mathbf{u}$ and $G_\mathbf{u}$ do not contain many embedded vertices of T. Recall that vertices $\varphi(z_1), \varphi(z_2)$ are typical to all but at most $\sqrt{\varepsilon}m$ sets $F_\mathbf{u}$. If we cannot use for embedding any other set $F_\mathbf{u}$ from remaining clusters of $\mathbf{M}_1 \cup \mathbf{O}_1$, it means that we have embedded at least $\lfloor \frac{d}{2} \cdot \eta r d |\mathbf{v}_1|/300 \rfloor$ vertices to this set $F_\mathbf{u}$, or we have embedded at least the same number of vertices in the appropriate set $G_\mathbf{u}$. This means that the number of vertices we have embedded is at least

$$\begin{aligned}\left(|\mathbf{M}_1 \cup \mathbf{O}_1| - 2\sqrt{\varepsilon}m\right) \cdot \left(\frac{d}{2} \cdot \lfloor \eta r d|\mathbf{v}_1|/300\rfloor\right) &\geq \left(\frac{|\mathbf{M} \cup \mathbf{O}_1|}{2} - 2\sqrt{\varepsilon}m\right) \cdot \frac{d^2 r\eta}{700}|\mathbf{v}_1| \\ \boxed{\varepsilon \ll d^{10}r^5\eta^5} \quad &\geq \left(\frac{\overline{\deg}(\mathbf{v}_1)}{2} - 2\sqrt{\varepsilon}m|\mathbf{v}_1|\right) \cdot \sqrt[5]{\varepsilon} \\ \boxed{m|\mathbf{v}_1| \leq n} \quad &\geq \left(\frac{k}{2} - 2\sqrt{\varepsilon}n\right) \cdot \sqrt[5]{\varepsilon} \\ \boxed{k \geq \eta n/2} \quad &\geq \left(\frac{1}{2} - \frac{4\sqrt{\varepsilon}}{\eta}\right)\sqrt[5]{\varepsilon}k \\ &> \sqrt[4]{\varepsilon}k,\end{aligned}$$

a contradiction. □

Embedding algorithm

So far we have embedded the set W in vertices of $\mathbf{v}_1$ that are typical to almost all clusters in the neighbourhood of $\mathbf{v}_1$. We also embedded the small set $\mathcal{D}''$. We invoke Claim 3.13 to get a small set F. Now we will gradually embed microtrees from $\mathcal{D}$ in $\bigcup(\mathbf{M} \cup \mathbf{O}) \setminus F$, until the number of vertices of microtrees that were not embedded yet is at most $\sqrt[4]{\varepsilon}k$. Then we embed the remaining parts of T in F using Claim 3.13. We will use the following notation for the sake of brevity.

Definition 3.14. *We say that a cluster* $\mathbf{u}$ *is full, if*

$$|\mathbf{u} \cap (\varphi(T) \cup F)| \geq |\mathbf{u}| - 4\sqrt{\varepsilon}\,|\mathbf{u}|\,.$$

We say that a cluster $\mathbf{u} \in N(\mathbf{v}_1)$ *is saturated, if*

$$|\mathbf{u} \cap (\varphi(T) \cup F)| \geq \overline{\deg}(\mathbf{v}_1, \mathbf{u}) - 4\sqrt{\varepsilon}\,|\mathbf{u}|\,.$$

We say that a matching edge $\mathbf{uv} \in \mathbf{M}$ *is saturated, if*

$$|(\mathbf{u} \cup \mathbf{v}) \cap (\varphi(T) \cup F)| \geq \overline{\deg}(\mathbf{v}_1, (\mathbf{u} \cup \mathbf{v})) - 8\sqrt{\varepsilon}\,|\mathbf{u}| - \beta k.$$

Note that every full cluster is also saturated. The intuition behind these definitions will be clear from the statements of the following claims.

Claim 3.15. *If* $\mathbf{u} \in N(\mathbf{v}_1)$ *is not saturated and* $\mathbf{v} \in N(\mathbf{u}) \setminus \{\mathbf{v}_1\}$ *is not full, then, unless* $|\mathrm{dom}(\varphi)| \geq k - \sqrt[4]{\varepsilon}k$*, we may injectively extend* φ *to some* $K \in \mathcal{D}$ *that was not yet embedded in such a way that* $\varphi(K \cap \mathcal{D}_1) \subseteq \mathbf{u}$*,* $\varphi(K \cap \mathcal{D}_2) \subseteq \mathbf{v}$*, and* $\varphi(K) \cap F = \emptyset$*.*

Proof. We have ensured that all trees of $\mathcal{D}$ such that $\mathbf{u}$ is not nice to them have at most $\sqrt[4]{\varepsilon}k$ vertices. Hence there is a yet non-embedded tree $K \in \mathcal{D}$ such that its at most two neighbours t_1, t_2 in W are embedded in vertices of $\mathbf{v}_1$ that are typical to $\mathbf{u}$. We may now apply Proposition 2.13 (1) with $d_{P2.13} := d, \varepsilon_{P2.13} = \varepsilon, \beta_{P2.13} = \beta, \mathbf{v}_{1,P2.13} = \mathbf{v}_1, \mathbf{u}_{P2.13} = \mathbf{u}, \mathbf{v}_{P2.13} = \mathbf{v}, K_{P2.13} = K, v_{i,P2.13} = \varphi(t_i), x_{i,P2.13} = N(t_i) \cap K, U_{P2.13} = \varphi(T) \cup F$. The proposition then allows us to extend injectively φ to K. □

Claim 3.16. *Let* φ *be a partial embedding of* T *in* G*.*

1. *There exists either an unsaturated vertex of* $\mathbf{O}_1$ *or an unsaturated edge of* $\mathbf{M}$.

2. *Suppose that* $\varphi(\mathcal{D}_1) \cap (\mathbf{M_2}) = \emptyset$ *and let* $\mathbf{u} \in \mathbf{O}$. *There exists a vertex in* $N(\mathbf{u}) \setminus \{\mathbf{v}_1\}$ *that is not full.*

Proof. 1. Suppose that for each edge $\mathbf{uv} \in \mathbf{M}$ we have

$$\begin{aligned}
|\,(\mathbf{u} \cup \mathbf{v}) \cap (\varphi(T) \cup F)\,| &\geq \overline{\deg}(\mathbf{v}_1, \mathbf{u} \cup \mathbf{v}) - 8\sqrt{\varepsilon}\,|\mathbf{u}| - \beta k \\
\boxed{\overline{\deg}(\mathbf{v}_1, \mathbf{u} \cup \mathbf{v}) \geq 2d} \quad &\geq \overline{\deg}(\mathbf{v}_1, \mathbf{u} \cup \mathbf{v}) \left(1 - \frac{8\sqrt{\varepsilon}\,|\mathbf{v}_1| + \beta k}{2d|\mathbf{v}_1|}\right) \\
\boxed{|\mathbf{v}_1| \geq n/M_{L2.6}(\varepsilon),\ k \leq n} \quad &\geq \overline{\deg}(\mathbf{v}_1, \mathbf{u} \cup \mathbf{v}) \left(1 - \frac{4\sqrt{\varepsilon}}{d} - \frac{\beta n}{2dn/M_{L2.6}(\varepsilon)}\right) \\
\boxed{\varepsilon \ll d,\ \beta \ll d/M_{L2.6}(\varepsilon)} \quad &\geq \overline{\deg}(\mathbf{v}_1, \mathbf{u} \cup \mathbf{v})(1 - \eta/100)
\end{aligned}$$

and similarly for each $\mathbf{u} \in \mathbf{O}_1$ we have

$$|\mathbf{u} \cap (\varphi(T) \cup F)| \geq \overline{\deg}(\mathbf{v}_1, \mathbf{u})(1 - \eta/100).$$

Hence we have

$$\begin{aligned}
\left|\bigcup(\mathbf{M} \cup \mathbf{O}_1) \cap (\varphi(T) \cup F)\right| &\geq \overline{\deg}(\mathbf{v}_1)(1 - \eta/100) \\
&= \eta\overline{\deg}(\mathbf{v}_1)/100 + \overline{\deg}(\mathbf{v}_1)(1 - \eta/50) \\
\boxed{\overline{\deg}(\mathbf{v}_1) \geq k + \eta k/20} \quad &\geq \eta\overline{\deg}(\mathbf{v}_1)/100 + (k + \eta k/20)(1 - \eta/50) \\
\boxed{|F| \leq \eta\overline{\deg}(\mathbf{v}_1)/100} \quad &> |F| + k,
\end{aligned}$$

a contradiction.

2. Similarly as in the previous case we can compute that we have embedded at least $|\mathbf{v}|(1 - \eta/100) \geq \overline{\deg}(\mathbf{u}, \mathbf{v})(1 - \eta/100)$ vertices into each full cluster $\mathbf{v}$. Since we know that $\overline{\deg}(\mathbf{v}_1) \leq 2k$ and $\overline{\deg}(\mathbf{u}) \geq rk + \eta k/20$, we thus we have

$$\begin{aligned}
\left|\bigcup(\mathbf{M}_2 \cup \mathbf{O}_2) \cap (\varphi(T \cap \mathcal{D}_2) \cup F)\right| &\geq \overline{\deg}(\mathbf{u})(1 - \eta/100) \\
&\geq \eta\overline{\deg}(\mathbf{u})/50 + \overline{\deg}(\mathbf{u})(1 - \eta/30) \\
&\geq \eta r\overline{\deg}(\mathbf{v}_1)/100 + (rk + \eta k/20)(1 - \eta/30) \\
&> |F| + rk \\
&\geq |F| + |\mathcal{D}_2|,
\end{aligned}$$

a contradiction.

□

We can now finish the proof of Theorem 3.8.

Proof. We will gradually embed microtrees from $\mathcal{D}'$ in $\bigcup(\mathbf{M} \cup \mathbf{O})$ in a specified manner using Claim 3.15, until $|\mathrm{dom}(\varphi)| \geq k - \sqrt[4]{\varepsilon}k$, or all edges of $\mathbf{M}$ and all vertices of $\mathbf{O}$ are saturated – from Claim 3.16 (1) we know that the latter actually cannot be true. When $|\mathrm{dom}(\varphi)| \geq k - \sqrt[4]{\varepsilon}k$, we finish by applying Claim 3.13 on our set F. We split the embedding procedure into three phases:

1. *Phase 1 – saturating the matching edges of* $\mathbf{M}$. In the first phase we embed gradually the microtrees of $\mathcal{D}'$ in the edges of $\mathbf{M}$ in such a way that for each $K \in \mathcal{D}'$ we have $\varphi(K \cap \mathcal{D}_1) \subseteq \mathbf{M}_1$. We run the process of applying Claim 3.15 for each edge $\mathbf{uv}$ until either $\mathbf{u} \in \mathbf{M_1}$ is saturated, $\mathbf{v} \in \mathbf{M_2}$ is full, or $|\text{dom}(\varphi)| \geq k - \sqrt[4]{\varepsilon}k$.

2. *Phase 2 – saturating the clusters in* $\mathbf{O}$. We repeatedly pick a cluster $v \in \mathbf{O}_1$ and then embed trees from $\mathcal{D}'$ in it by repeatedly applying Claim 3.15 in such a way that for each embedded K we have $\varphi(K \cap \mathcal{D}_1) \subseteq \mathbf{O}_1$ and $\varphi(K \cap \mathcal{D}_2) \subseteq \mathbf{M}_2 \cup \mathbf{O}_2$. Note that due to Claim 3.16 (2) the cluster $\mathbf{v}$ has always a neighbour that is not full and can be, thus, used for embedding. Hence we can apply this procedure until all clusters from $\mathbf{O}_1$ are saturated, or $|\text{dom}(\varphi)| \geq k - \sqrt[4]{\varepsilon}k$.

3. *Phase 3 – finalising the matching* $\mathbf{M}$. All clusters in $\mathbf{O}$ are now saturated. Our goal is now to show how to saturate the remaining edges of $\mathbf{M}$. This may not be possible with current φ as it is defined right now, since it could have for example happened that after the first phase we completely filled one cluster from a matching pair, while the other cluster remained almost empty. We solve this problem by potentially redefining the embedding of several microtrees that were embedded in $\mathbf{M}_1 \cup \mathbf{M_2}$ in Phase 1.

   Note that for each edge $\mathbf{uv} \in \mathbf{M}$, $\mathbf{u} \in \mathbf{M}_1$, it is true that either $\mathbf{u}$ is saturated, or $\mathbf{v}$ is full at the end of Phase 1. We deal with the first case in part (a). In the latter case we did not embed anything in $\mathbf{v}$ in Phase 2. We undefine embedding of all trees that were embedded in $\mathbf{uv}$ and saturate this edge in part (c).

   (a) If $\mathbf{u}$ is saturated, we repeatedly embed trees in $\mathbf{uv}$ in such a way that for each $K \in \mathcal{D}'$ we have $\varphi(K \cap \mathcal{D}_1) \subseteq \mathbf{v}$. We do this until either $\mathbf{u}$ is full, or $\mathbf{v}$ is saturated. In the latter case the whole edge is saturated. We deal with the first case in (b).

   (b) Suppose that $\mathbf{u}$ is full, but $\mathbf{v}$ is not saturated. Note that Claim 3.13 ensures that $|F \cap \mathbf{u}| = |F \cap \mathbf{v}|$. Hence it must be the case that $|\varphi(T) \cap \mathbf{u}| \geq |\varphi(T) \cap \mathbf{v}|$. Moreover, in Phase 2 we did not embedded trees in $\mathbf{u}$. This means that there exists a tree $K \in \mathcal{D}'$ that was embed in the matching edge $\mathbf{uv}$ in such a way that $|\varphi(K) \cap \mathbf{u}| \geq |\varphi(K) \cap \mathbf{v}|$. As long as it is true that $|(\varphi(T) \cup F) \cap \mathbf{u}| \geq |(\varphi(T) \cup F) \cap \mathbf{v}|$, we find any tree K with this property and we redefine its embedding. When this procedure ends, we have $\Big||\varphi(T) \cap \mathbf{u}| - |\varphi(T) \cap \mathbf{v}|\Big| \leq \beta k$, i.e., the embedding in the edge is *balanced*.

   (c) Finally it suffices to show how to saturate an edge $\mathbf{uv}$ fulfilling the balancing condition (note that if $\varphi(T) \cap \mathbf{uv} = \emptyset$, then the matching edge is certainly balanced). We again embed the microtrees in $\mathbf{uv}$ one after another. Unless one of the clusters is saturated, we choose to embed $K \in \mathcal{D}'$ in such a way that the colour class of K with less vertices is embedded in the cluster such that more of its vertices were already used for embedding of T. In this way we ensure that the balancing condition still holds.

       After one cluster, say $\mathbf{u}$, becomes saturated, we continue by embedding only in such a way that for each $K \in \mathcal{D}'$ we have $\varphi(K \cap \mathcal{D}_1) \subseteq \mathbf{v}$. We do this until either $\mathbf{v}$ becomes saturated, or $\mathbf{u}$ is full. In the first case the whole edge $\mathbf{uv}$ is clearly saturated. In the other case note that we have $|(\varphi(T) \cup F) \cap \mathbf{u}| \geq |\mathbf{u}| - 4\sqrt{\varepsilon}|\mathbf{u}| \geq \overline{\deg}(\mathbf{v}_1, \mathbf{v}) - 4\sqrt{\varepsilon}|\mathbf{v}|$ and $|(\varphi(T) \cup F) \cap \mathbf{v}| \geq \overline{\deg}(\mathbf{v}_1, \mathbf{u}) - 4\sqrt{\varepsilon}|\mathbf{u}| - \beta k$ due to our balancing condition. Hence the matching edge is saturated.

We described an algorithm that terminates when $|\mathrm{dom}(\varphi)| \geq k - \sqrt[4]{\varepsilon}k$, or all edges of $\mathbf{M}$ and all vertices of $\mathbf{O}$ are saturated. But the latter cannot happen due to Claim 3.16 (1). We finish by invoking Claim 3.13. $\square$

Chapter 4

The skew Loebl-Komlós-Sós conjecture for dense graphs

In this chapter we propose the skew version of the Loebl-Komlós-Sós conjecture. This was asked by Simonovits [personal communication].

Conjecture 4.1. *Any graph of order n with at least rn vertices of degree at least k contains every tree of order at most $k+1$ vertices with skew r.*

If true, the conjecture is best possible for the similar reason as the Loebl-Komlós-Sós conjecture. Indeed, given $r \in (0, 1/2]$, consider a graph consisting of a disjoint union of copies of a graph H with $k+1$ vertices consisting of a clique of order $\lfloor r(k+1) \rfloor - 1$, an independent set on the remaining vertices and the complete bipartite graph between the two sets (see Figure 4.1). Such a graph does not contain a path on $2\lfloor r(k+1) \rfloor$ vertices (or, to give an example of a tree of maximal order, a path on $2\lfloor r(k+1) \rfloor$ vertices with one end-vertex identified with the centre of a star with $k+1-2\lfloor r(k+1) \rfloor$ leaves).

A proof attempt of such a result starts by verifying that the conjecture holds for trees of diameter five. This is because (as we mentioned in Chapter 2) the structure of such trees is similar to the structure of the ℓ-fine partition of a general tree. We provide a proof of this result in the next section.

Theorem 4.2. *Let G be a graph on n vertices such that at least rn of its vertices have degree at least k. Then G allows embedding of any tree from $\mathcal{T}_{k+1}^{r}$ with diameter at most five.*

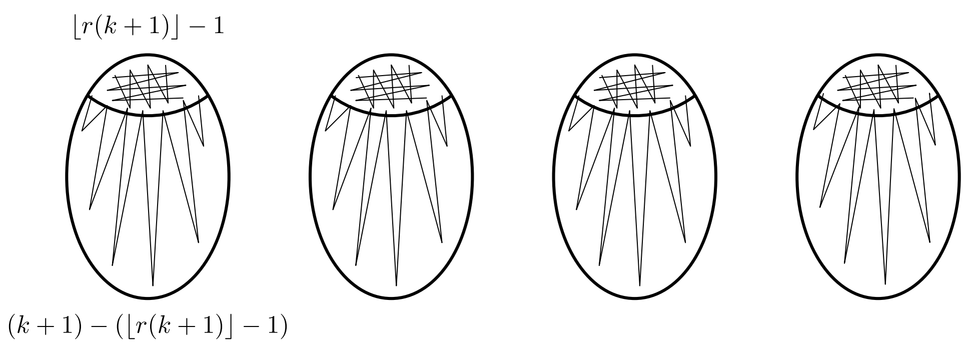

Figure 4.1: The graph showing the tightness of Conjecture 4.1 is a disjoint union of graphs of order $k+1$.

Note that this was already shown in [PS08] for the case $r = 1/2$.

We continue with the proof of a dense approximate version of the conjecture in the last section of this chapter. One can find a sketch of the proof in the extended abstract [KPR17].

Theorem 4.3. *For any $0 < r \leq 1/2$ and $\eta > 0$ there exists $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$, any graph of order n with at least rn vertices of degree at least $k + \eta n$ contains every tree of order at most k such that the size of its smaller colour class is at most rk.*

We will actually prove the following formulation of Theorem 4.3. The equivalence of the two statements follows from Lemma 2.4.

Theorem 4.4 (restated version of Theorem 4.3)**.** *For all $q, r, \eta > 0$ there is n_0 such that every graph on $n \geq n_0$ vertices with at least rn of its vertices having degree at least $(1 + \eta)k$ contains any tree from $\mathcal{T}_k^r$, if $k \geq qn$.*

This extends the main result of [PS12], which is a special case of Theorem 4.4 for $r = 1/2$. While we use and extend some of their techniques, our analysis is more complex. As in [PS12], we partition the tree into small rooted subtrees, which we then embed into regular pairs of the host graph. In order to connect those small rooted trees, we need two adjacent clusters with adequate average degree to those regular pairs, which typically will be represented by a matching in the cluster graph. Hence, we need a matching in the cluster graph that is as large as possible. To this end we use disbalanced regularity decomposition (see [HLT02]), placing large degree vertices into smaller clusters than the remaining vertices, hence covering as many low degree vertices as possible by this matching. We then consider several possible embedding configurations in the regularity decomposition, depending on the structure of the cluster graph, in particular depending on the properties of the adjacent clusters with suitable average degree to the optimal matching.

4.1 Proof of Theorem 4.2

Fix $T \in \mathcal{T}_{k+1}^r$ of diameter at most five; we will use the notation from Section 2.1. Note that we can without loss of generality assume that $r(k+1)$ is an integer, otherwise we may work with $r' < r$ for which it holds.

At first we decompose G into suitable subsets, i.e., find suitable structure for embedding. Define $L = \{v \in V(G) : \deg(v) \geq k\}$ and $S = V(G) \setminus L$. We define $S_1 = \{v \in S : \deg(v, G) \geq r(k+1)\}$ and set $S_0 = S \setminus S_1$. Further we define $L^* = \{v \in L : \deg(v, L) \geq r(k+1)\}$ and $N = (N(L^*) \cup N(S_1)) \cap (L \setminus L^*)$.

We will now describe two possible *configurations* in G and for both of these configurations we embed T in G. Then we show that at least one of these configurations has to appear in the host graph.

We will use the following version of the greedy method.

Lemma 4.5. *Let G be a graph and φ a partial embedding of $T \in \mathcal{T}_{k+1}$ in G such that the only non-embedded vertices of T are leaves. Moreover, suppose that $\deg(\varphi(u)) \geq k$ for any $u \in T$ with a non-embedded neighbour. Then φ can be injectively extended to the whole T.*

Proof. In each step we may arbitrarily embed any yet non-embedded vertex of the tree. □

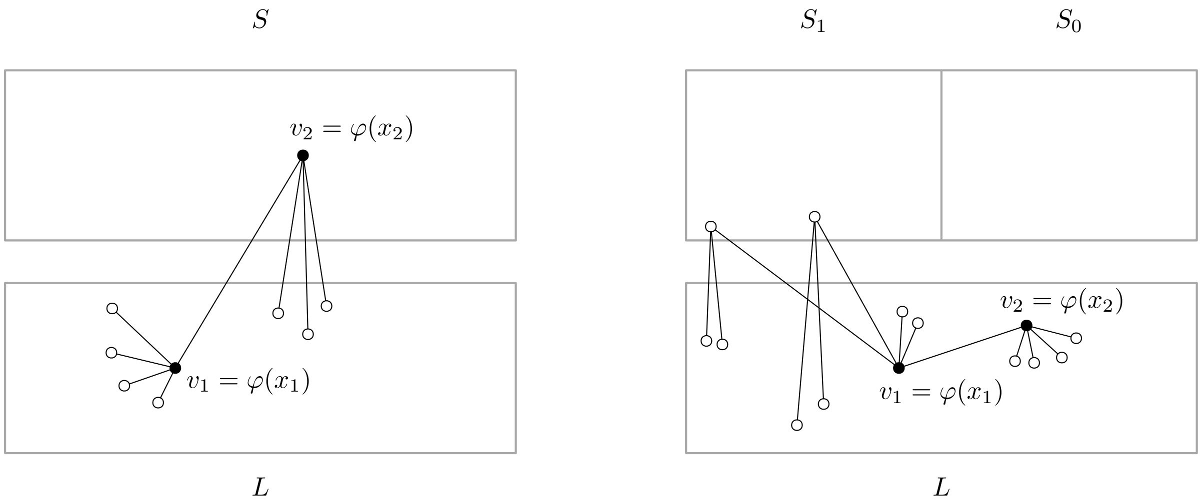

Figure 4.2: Two embedding configurations from Proposition 4.6 (left) and Proposition 4.7 (right). We finish both cases by applying Lemma 4.5.

A simple consequence of Lemma 4.5 is that if we define φ injectively on $\{x_1, x_2\} \cup V_1' \cup V_2$ such that $\varphi(\{x_1\} \cup V_1' \cup V_2) \subseteq L$, then φ can be injectively extended to the whole T.

We now propose the two configurations. In both configurations we assume that there are two neighbouring vertices of high degree and we prove that in that case we can embed T in G.

Proposition 4.6. *Suppose that there are two neighbours $v_1, v_2 \in G$ such that $v_1 \in L$, $\deg(v_1, L) \geq r(k+1)$ and $\deg(v_2, L) \geq |V_2| + 1$. Then T can be embedded in G.*

Proof. Set $\varphi(x_1) = v_1$ and $\varphi(x_2) = v_2$. Embed the vertices of V_2 arbitrarily in the L-neighbourhood of $\varphi(x_2)$. Then embed the vertices of V_1' arbitrarily in the L-neighbourhood of $\varphi(x_1)$. This can be done, since

$$\deg(\varphi(x_1)) \geq r(k+1) \geq |\{x_1\}| + |V_2| + |W_1| \geq |\{x_2\}| + |V_2| + |V_1'|.$$

We finish by using Lemma 4.5. □

Proposition 4.7. *Suppose that there are two neighbours $v_1, v_2 \in G$ such that $v_1 \in L$, $\deg(v_1, L \cup S_1) \geq |V_1'| + 1$ and $\deg(v_2, L) \geq r(k+1)$. Then T can be embedded in G.*

Proof. Set $\varphi(x_1) = v_1$ and $\varphi(x_2) = v_2$. Embed the vertices from V_1' arbitrarily in $N(\varphi(u_1)) \setminus \{v_2\}$, suppose that vertices $y_1, \ldots, y_t \in V_1'$ were embedded in S_1. We continue by embedding their leaf neighbours greedily. This can be done as for any $1 \leq i \leq t$ we can compute that the degree of each vertex $\varphi(y_i)$ is greater than or equal to the number of already embedded vertices of T plus the number of leaves to be embedded, i.e.,

$$\begin{aligned}\deg(\varphi(y_i)) &\geq r(k+1) \geq |\{x_1\} \cup W_1| \\ &\geq |\{x_1\}| + |W_1 \cap N(y_1 \cup \cdots \cup y_t)| + |\{y_{t+1} \cup \cdots \cup y_{|V_1'|}\}|,\end{aligned}$$

where we used in the last inequality the fact that every vertex from V_1' has at least one leaf neighbour in W_1. We follow by embedding the set V_2 in L. This can be done, since

$$\begin{aligned}\deg(\varphi(u_2)) &\geq r(k+1) \geq |\{x_1\} \cup W_1 \cup V_2| \\ &\geq |\{x_1\}| + |W_1 \cap N(y_1 \cup \cdots \cup y_t)| + |\{y_{t+1} \cup \cdots \cup y_{|V_1|}\}| + |V_2|.\end{aligned}$$

Finally we embed the rest of T by invoking Lemma 4.5. □

Now we may proceed with the proof of Theorem 4.2. It suffices to show that the configurations from Propositions 4.6 and 4.7 have to appear in each (r, k)-LKS graph.

Proof. Suppose that G is an (r, k)-LKS graph not containing the configurations from Propositions 4.6 and 4.7. Observe that the absence of the first configuration implies that for every $v \in N$ we have $\deg(v, L^*) \le |V_2|$. The absence of the second configuration implies that there are no edges between any two vertices $u \in L^*$ and $v \in L^* \cup S_1$, because $r(k+1) \ge |V_1'| + 1$. Moreover, for any $u \in N$ we have $\deg(u, L \cup S_1) \le |V_1'|$.

We now derive a contradiction with the assumption that G is an (r, k)-LKS graph. We start by bounding the numbers of edges $e(N, S_0)$. We have

$$e(N, S_0) \le |S_0|(r(k+1) - 1) - e(S_0, L^*) - e(S_0, L \setminus (L^* \cup N)), \tag{4.1}$$

since all vertices in S_0 have degree less than $r(k+1)$ (note that it is an integer). On the other hand, we have

$$e(N, S_0) \ge |N|(k - |V_1'|) \tag{4.2}$$

from the assumption that all vertices of N have at most $|V_1'|$ neighbours outside S_0. We continue by bounding the quantity $e(N, L^*)$. We have

$$e(N, L^*) \le N|V_2|, \tag{4.3}$$

because we assume that all vertices in N have at most $|V_2|$ neighbours in L^*. On the other hand it holds that

$$e(N, L^*) \ge |L^*|k - e(L^*, S_0), \tag{4.4}$$

since the vertices in L have degree at least k and we assume that vertices from L^* have no neighbours in $L^* \cup S_1$. After adding all four inequalities we get

$$|S_0|(r(k+1) - 1) - e(S_0, L^*) - e(S_0, L \setminus (L^* \cup N)) + N|V_2| \ge \\ |N|(k - |V_1'|) + |L^*|k - e(L^*, S_0)$$

and after rewriting and bounding $rk > r(k+1) - 1$ we have

$$|S_0|rk + |N|(|V_2| + |V_1'|) \ge (|N| + |L^*|)k + e(S_0, L \setminus (L^* \cup N)).$$

Now we observe that vertices in $L \setminus (L^* \cup N)$ have less than $r(k+1)$ neighbours in L and that they have no neighbours in S_1, hence their degree in S_0 is at least $k - r(k+1) + 1 > (1 - r)k$. Therefore we have

$$|S_0|rk + |N|(|V_2| + |V_1'|) \ge (|N| + |L^*|)k + |L \setminus (L^* \cup N)|(1 - r)k.$$

We use the bound $|V_1'| + |V_2| \le r(k+1) - 1 < rk$ to get

$$|S_0|rk + |N|rk \ge (|N| + |L^*|)k + |L \setminus (L^* \cup N)|(1 - r)k. \tag{4.5}$$

Moreover, the last three adjustments gave a strict inequality, if S_0, $L \setminus (L^* \cup N)$, or N were non-empty. But if all of these sets were empty, then also L^* is empty, because its vertices do not have neighbours in $L^* \cup S_1$, which would mean that G is empty. We have thus a strict inequality that can be divided to give

$$r|S_0| + r|N| > |N| + |L^*| + (1 - r)|L \setminus (L^* \cup N)|,$$

which implies

$$r|S_0| > (1 - r)(|N| + |L^*| + |L \setminus (L^* \cup N)|) = (1 - r)|L|,$$

a contradiction. □

4.2 Proof of Theorem 4.4

We start by fixing notation and proving some preparatory lemmas in section 4.2.1. Then we prove the theorem with the use of two crucial propositions, Proposition 4.10 and 4.11. In the next section we prove Proposition 4.10. Then we state several technical embedding lemmas in the next section 4.2.4. Finally, we prove Proposition 4.11 in the last section.

4.2.1 Preliminaries

We shall switch freely between a graph H and its corresponding cluster graph $\mathbf{H}$. For example $A \subseteq V(H)$ may as well denote a cluster in an original graph, as $A \in V(\mathbf{H})$ a vertex in the corresponding cluster graph. We shall freely use the term *clusters* in a cluster graph $\mathbf{H}$ to denote vertices of $\mathbf{H}$. If $\mathcal{S} \subseteq V(\mathbf{H})$ denotes a set of clusters, then $\bigcup \mathcal{S}$ denotes the corresponding union of vertices in the original graph H. If $A \in V(\mathbf{H})$ is a cluster and $\mathcal{S} \subseteq V(\mathbf{H})$ a set of clusters, then $\overline{\deg}(A, \mathcal{S})$ denotes the average degree of vertices in A to $\bigcup \mathcal{S}$ and $\overline{\deg}(A)$ stands short for $\overline{\deg}(A, V(\mathbf{H}))$.

We shall use the following notation. The class of all trees of order k is denoted as $\mathcal{T}_k$. For a graph G and two sets $A \in V(G)$ and $B \in V(G)$ let $G[A, B]$ denote the subgraph of G induced by all edges with one endpoint in A and the other in B.

Definition 4.8. *Let $r \leq 1/2$. We say that a graph H is an r-*skew LKS-graph with parameters $(k, \eta, \varepsilon, d)$ *if there exists a partition $\{L_1, \dots, L_{m_L}, S_1, \dots, S_{m_S}\}$ of $V(H)$ satisfying the following*

1. *$m_L \geq (1 + \eta)m_S$,*
2. *all sets L_i have the same size and all sets S_i have the same size,*
3. *$r|S_j| = (1 - r)|L_i|$ for all i, j,*
4. *each (L_i, L_j), $i, j \in [m_L]$ and each (L_i, S_j), $i \in [m_L], j \in [m_S]$ is an ε-regular pair of density either 0 or at least d,*
5. *there are no edges inside the sets and no edges between S_i and S_j for $i \neq j$,*
6. *average degree of vertices in each L_i is at least $(1 + \eta)k$.*

We call the sets L_i, $i \in [m_L]$, the L-clusters. Similarly, we call the sets S_i, $i \in [m_S]$, the S-clusters.

Let $\mathbf{H}$ be the graph with vertex set $\{L_1, \dots, L_{m_L}, S_1, \dots, S_{m_S}\}$ and with an edge (L_i, L_j), (L_i, S_j) whenever (L_i, L_j) or (L_i, S_j), respectively forms an ε-regular pair of positive density in H. Observe that for any edge (L_i, L_j) we have $\overline{\deg}(L_i, L_j) = \overline{\deg}(L_j, L_i)$, but for any edge (L_i, S_j) we have $r \cdot \overline{\deg}(L_i, S_j) = (1 - r) \cdot \overline{\deg}(S_j, L_i)$. We call $\mathbf{H}$ the r-skew LKS-cluster graph. We use a dot instead of an explicit parameter when the value of the parameter is not relevant in the given context.

Proposition 4.9. *Let H be an r-skewed LKS graph of order n with parameters $(\cdot, \cdot, \varepsilon, \cdot)$ and let $\mathbf{H}$ be its corresponding cluster graph.*

1. *Let C and D be an L-cluster and an S-cluster of $\mathbf{H}$, respectively. Then $|C| \leq n/|V(\mathbf{H})|$ and $|D| \leq \frac{n}{r|V(\mathbf{H})|}$.*

2. *If $v \in V(H)$ is an ultratypical vertex and $\mathcal{S} \subseteq V(\mathbf{H})$, then* $\deg(v, \bigcup \mathcal{S}) \geq \overline{\deg}(C, \mathcal{S}) - 2\sqrt{\varepsilon}n/r$, *where C is the cluster of $\mathbf{H}$ containing v.*

Proof.

1. The first inequality follows from the fact that the size of L-clusters is always at most the size of S-clusters. Then we compute $|D| = \frac{1-r}{r}|C| \leq \frac{n}{r|V(\mathbf{H})|}$.

2. If v is ultratypical, there are at most $\sqrt{\varepsilon}|V(\mathbf{H})|$ clusters D in $\mathbf{H}$ such that v is not typical to D. Denote by $\mathcal{D}$ the set of those clusters. Then by (1) we have $|\bigcup \mathcal{D}| \leq |\mathcal{D}| \cdot n/(r|V(\mathbf{H})|) \leq \sqrt{\varepsilon}n/r$. Then

$$
\begin{aligned}
\deg(v, \bigcup \mathcal{S}) &\geq \deg(v, \bigcup(\mathcal{S} \setminus \mathcal{D})) \\
&\geq \overline{\deg}(C, \bigcup(\mathcal{S} \setminus \mathcal{D})) - \varepsilon n \\
&\geq \overline{\deg}(C, \bigcup \mathcal{S}) - |\bigcup \mathcal{D}| - \varepsilon n \\
&\geq \overline{\deg}(C, \bigcup \mathcal{S}) - 2\sqrt{\varepsilon}n/r \; .
\end{aligned}
$$

□

Finally, we state two propositions that will be proved in Sections 4.2.3 and 4.2.5, respectively. The first proposition says that every LKS-graph contains one the four configurations, while the second proposition asserts that occurrence of these configurations implies containment of a given tree. Note that the first proposition is concerned only with the structure of the cluster graph, not the underlying graph, and could be stated in terms of weighted graphs instead.

Proposition 4.10. *Let H be a r'-skew LKS-graph $\mathbf{H}$ with parameters $(k, \eta, \cdot, \cdot)$ and let $\mathbf{H}$ be the corresponding cluster graph. We denote by $\mathcal{L}$ and $\mathcal{S}$, respectively, its set of L-clusters and S-clusters, respectively. For any numbers $a_1, a_2, b_1, b_2 \in \mathbb{N}_0$ with $a_2 + b_1 = \tilde{r}k$, $\tilde{r} \leq r'$, there is a matching $\mathbf{M}$ in $\mathbf{H}[\mathcal{L}, \mathcal{S}]$ and two adjacent clusters $X, Y \in V(\mathbf{H})$ such that, setting $\mathcal{S}_M = \mathcal{S} \cap V(\mathbf{M})$ and $\mathcal{S}_1 = \{Z \in \mathcal{S} : \overline{\deg}(Z) \geq (\tilde{r} + r'\eta)k\} \setminus \mathcal{S}_{\mathcal{M}}$, one of the four following configurations occurs.*

A) $\overline{\deg}(X, \mathcal{S}_1 \cup \mathcal{S}_M) \geq a_2 \cdot (1 - \tilde{r})/\tilde{r} + \eta k/4$, *and* $\overline{\deg}(Y, \mathcal{L}) \geq \tilde{r}k + \eta k/4$,

B) $\tilde{r}a_1 > (1 - \tilde{r})a_2$, $\overline{\deg}(X, \mathcal{S}_1 \cup \mathcal{S}_M \cup \mathcal{L}) \geq k + \eta k/4$ *and* $\overline{\deg}(Y, \mathcal{L}) \geq \tilde{r}k + \eta r'k/4$,

C) $\tilde{r}a_1 \leq (1 - \tilde{r})a_2$, $\overline{\deg}(X, \mathcal{S}_1 \cup \mathcal{S}_M \cup \mathcal{L}) \geq k + \eta k/4$ *and* $\overline{\deg}(Y, \mathcal{L}) \geq b_1 + \eta r'k/4$,

D) $\tilde{r}a_1 \geq (1 - \tilde{r})a_2$, $b_1 \leq \tilde{r}^2 k/(1 - \tilde{r})$, $\overline{\deg}(X, \mathcal{S}_M \cup \mathcal{L}) \geq k + \eta k/4$ *and* $\overline{\deg}(Y, \mathcal{L}) \geq b_1 + \eta k/4$, *and moreover, the neighbourhood of X does not contain both endpoints of any edge from $\mathbf{M}$.*

Proposition 4.11. *For each $\delta, q, d > 0$ and $\tilde{r}, r' \in \mathbb{Q}^+$ with $\tilde{r} \leq r' \leq 1/2$ there is $\varepsilon = \varepsilon(\delta, q, d, r') > 0$ such that for any $\tilde{N}_{max} \in \mathbb{N}$ there is a $\beta = \beta(\delta, q, r', \varepsilon, \tilde{N}_{max}) > 0$ and an $n_0 = n_0(\delta, q, \tilde{r}, \beta) \in \mathbb{N}$ such that for any $n \geq n_0$ and $k \geq qn$ the following holds.Let $\mathcal{D} = (W_A, W_B, \mathcal{D}_A, \mathcal{D}_B)$ be an βk-fine partition of a tree $T \in \mathcal{T}_k$ with colour classes T_1 and T_2 such that $|T_1| = \tilde{r}k$. Let H be an r'-skewed LKS-graph of order n,with parameters $(k, \delta, \varepsilon, d)$, let $\mathbf{H}$ be its corresponding cluster graph with $|V(\mathbf{H})| \leq \tilde{N}_{max}$ and $\mathcal{L}, \mathcal{S} \subseteq V(\mathbf{H})$ are sets of L-clusters and S-clusters, respectively.Let $\mathbf{M}$ be a matching in $\mathbf{H}$, let $\mathcal{S}_M = \mathcal{S} \cap V(\mathbf{M})$, $\mathcal{S}_1 := \{C \in \mathcal{S} \setminus V(\mathbf{M}) : \overline{\deg}(C) \geq (1 + \delta)\tilde{r}k\}$. Let A and B be two clusters of $\mathbf{H}$ such that $AB \in E(\mathbf{H})$ and one of the following holds.*

A) $\overline{\deg}(A, \mathcal{S}_1 \cup \mathcal{S}_M) \geq a_2\frac{1-\tilde{r}}{\tilde{r}} + \delta k$ *and* $\overline{\deg}(B, \mathcal{L}) \geq (\tilde{r} + \delta)k$

B) $\tilde{r}|V(\mathcal{D}_A) \cap V(T_2)| \geq (1 - \tilde{r})|V(\mathcal{D}_A) \cap V(T_1)|$,
$\overline{\deg}(A, \mathcal{S}_1 \cup \mathcal{S}_M \cup \mathcal{L}) \geq (1 + \delta)k$, *and* $\overline{\deg}(B, \mathcal{L}) \geq (\tilde{r} + \delta)k$,

C) $\tilde{r}|V(\mathcal{D}_A) \cap V(T_2)| \leq (1 - \tilde{r})|V(\mathcal{D}_A) \cap V(T_1)|$,
$\overline{\deg}(A, \mathcal{S}_1 \cup \mathcal{S}_M \cup \mathcal{L}) \geq (1 + \delta)k$, *and* $\overline{\deg}(B, \mathcal{L}) \geq |V(\mathcal{D}_B) \cap V(T_1)| + \delta k$,

D) $\tilde{r}|V(\mathcal{D}_A) \cap V(T_2)| \geq (1 - \tilde{r})|V(\mathcal{D}_A) \cap V(T_1)|$, $|V(\mathcal{D}_B) \cap V(T_1)| \leq \frac{\tilde{r}^2}{(1-\tilde{r})}k$
$\overline{\deg}(A, \mathcal{S}_M \cup \mathcal{L}) \geq (1 + \delta)k$, $\overline{\deg}(B, \mathcal{L}) \geq |V(\mathcal{D}_B) \cap V(T_1)| + \delta k$, *and moreover, the neighbourhood of* A *does not contain both endpoints of any edge from* $\mathbf{M}$.

Then $T \subseteq H$.

4.2.2 Proof of the theorem

Suppose r, q and η are fixed. If $r = 1/2$, then set $r' := r \in \mathbb{Q}$, $s := 1$, and $t := 2$. Otherwise, let $\rho := 1/2 - r > 0$ and $r' \in \mathbb{Q}$ be such that $r \leq r' \leq r(1 + \frac{\eta\rho q}{12})$ with $r' = s/t$, $s, t \in \mathbb{N}$ and $t \leq 12/(\eta\rho q r)$. Observe that $r' \leq 1/2$. Let $d := \frac{\eta^2 q^2 r'}{100}$. Let $\varepsilon = \min\{\frac{\eta d^2 q^2}{40}, \frac{1}{t}\varepsilon_{P4.11}(\frac{\eta r' q}{400}, q, d/2, r')\}$. Lemma 2.6 (Szemerédi regularity lemma) with input parameter $\varepsilon_{L2.6} := \varepsilon$ and $N_{min} := 1/\varepsilon$ outputs $n_R, N_{max} \in \mathbb{N}$. Set $\beta := \beta_{P4.11}(\frac{\eta r' q}{400}, q, r', t \cdot \varepsilon, tN_{max})$. Let $n_0 = \max\{2n_R, 2t \cdot N_{max}/\varepsilon, n_{0,P4.11}(\frac{\eta r' q}{400}, q, r', \beta)\}$ and let $n \geq n_0$. Suppose $k \geq qn$ is fixed. Let G be any graph on n vertices that has at least rn vertices of degree at least $(1 + \eta)k$.

We first find a subgraph H of G of size $n'' \geq (1 - \eta q/2)(1 - 2\varepsilon)n$ which is an r'-skew LKS-graph with parameters $(k, \frac{\eta q}{100}, t \cdot \varepsilon, \frac{d}{2})$ and construct the corresponding LKS-cluster graph $\mathbf{H}$.

Erase $\eta \cdot qn/2$ vertices from the set of vertices that have degree smaller than $(1+\eta)k$ and let G' be the resulting graph of order $n' = n(1 - \eta q/2)$. Observe that for all $v \in V(G')$, we have $\deg_{G'}(v) \geq \deg_G(v) - \eta k/2$ and hence at least $rn \geq r'n'(1 + \eta q/4)$ vertices of G' have degree at least $(1 + \eta/2)k$.

We apply Szemerédi regularity lemma (Lemma 2.6) on G' and obtain an ε-regular equitable partition $V(G') = V_0 \cup V_1 \cup \cdots \cup V_N$. Erase all edges within sets V_i, between irregular pairs, and between pairs of density lower than d. Hence, we erase at most $N \cdot \binom{n'/N}{2} \leq \varepsilon(n')^2/2$ edges within the sets V_i, at most $\varepsilon N^2 \cdot \left(\frac{n'}{N}\right)^2 = \varepsilon(n')^2$ edges in irregular pairs, and at most $\binom{N}{2} \cdot d \cdot \left(\frac{n'}{N}\right)^2 \leq \frac{d}{2} \cdot (n')^2$ edges in pairs of density less than d. In total we have thus erased less than $d \cdot (n')^2 = \frac{\eta^2 q^2 r'}{100} \cdot (n')^2$ edges.

Call a set V_i an L-set if the average degree of its vertices is at least $(1 + \eta q/4)k$ and otherwise an S-set. We have at least $(1 + \frac{\eta q}{20})r'N$ L-sets. Indeed, during the erasing process, less than $\eta r' q n'/6$ vertices dropped their degree by more than $\eta k/8$. Therefore, now there are at least $(1 + \frac{\eta q}{12})r'n'$ vertices of degree at least $(1 + 3\eta/8)k$. By regularity, in each S-set V_i there are at most $\varepsilon|V_i|$ of those vertices, as otherwise they form a subset of V_i of substantial size and thus the S-set V_i would have average degree at least $(1 + 3\eta/8)k - \varepsilon n' > (1 + \eta/4)k$. So we can have at most $\varepsilon n'$ vertices of degree at least $(1 + 3\eta/8)k$ distributed among all S-sets and at most $\varepsilon n'$ of them contained in V_0. Hence, at least $(1 + \frac{\eta q}{20})r'n'$ vertices of degree at least $(1 + 3\eta/8)k$ must be contained in L-sets, producing thus at least $(1 + \frac{\eta q}{20})r'N$ L-sets.

We subdivide any L-set into $t - s$ sets of the same size, which we call *L-clusters*, adding at most $t - s - 1$ leftover vertices to the garbage set V_0. Similarly, we subdivide any S-set into s sets, which we call *S-clusters*. In this way we have $(1-r')|C| = r'|D|$ for any L-cluster C, and any S-cluster D. By Lemma 2.5, if (V_i, V_j) is ε-regular and $C \subseteq V_i$ and $D \subseteq V_j$ are L or S clusters, then (C, D), is a ε'-regular pair for $\varepsilon' = t\varepsilon$ with density at least $d' := d - \varepsilon$. Observe that by the choice of n_0, we added in total less than $t \cdot N \leq \varepsilon n'$ vertices to the garbage set V_0. We delete at most $2\varepsilon n'$ vertices of the enlarged set V_0. Any L-cluster is a relatively large subset of the L-set it comes from, and thus basically inherits the average degree of the set it comes from. Together with the deletion of the enlarged garbage set, we obtain that each L-cluster has now average degree at least $(1 + \eta q/4)k - 3\varepsilon n' \geq (1 + \eta q/5)k$.

Denote by m_L the number of L-clusters and by m_S the number of S-clusters. We have $m_L \geq (1 + \frac{\eta q}{20})r'N \cdot (t - s)$, as each L-set divided in $t - s$ L-clusters. Similarly, we obtain $m_S < (1 - r')sN$. Therefore,

$$\begin{aligned}
m_L &\geq (1 + \eta q/100)m_L/2 + (1 - \eta q/100)(1 + \eta q/20) \cdot r'N \cdot (t - s)/2 \\
&> (1 + \eta q/100)m_L/2 + (1 + \eta q/100) \cdot \frac{s}{t} \cdot \frac{m_S}{s(1 - s/t)} \cdot (t - s)/2 \\
&= (1 + \eta q/100)m_L/2 + (1 + \eta q/100) \cdot \frac{m_S}{t - s} \cdot (t - s)/2 \\
&= (1 + \eta q/100)(m_L + m_S)/2 \ .
\end{aligned}$$

Finally, we delete all edges between S-clusters. We denote by L the set of vertices contained in L-clusters and by S the set of vertices contained in S-clusters.

Let H be the resulting graph. By construction, it is an r'-skew LKS-graph of order n'', where $(1 - 2\varepsilon)n' \leq n'' \leq n'$, with parameters $(k, \frac{\eta q}{100}, \varepsilon', d/2)$. The vertex set of the corresponding cluster graph $\mathbf{H}$ consists of the L- and S-clusters defined above, with edges corresponding to ε'-regular pairs of density at least $d/2$ in H. Observe that $|V(\mathbf{H})| \leq t \cdot N_{max}$.

After having processed the host graph, we turn our attention to the tree. Let T be any tree of order k with colour classes T_1 and T_2 and $|T_1| \leq rk \leq r'k$. Pick any vertex $R \in V(T)$ to be the root of T. Applying Lemma 2.9 on T with parameter $\ell_{P2.9} := \beta k$, we obtain its βk-fine partition $\mathcal{D} = (W_A, W_B, \mathcal{D}_A, \mathcal{D}_B)$. Without loss of generality, assume that $W_A \subseteq V(T_2)$. Let $\tilde{r} := |V(T_1) \setminus W_B|/k$. We then apply Proposition 4.10 with $\eta_{P4.10} := \eta q/100$, $r'_{P4.10} := r'$, $k_{P4.10} := k$, $n_{P4.10} := n''$, $H_{P4.10} := \mathbf{H}$, for $vu \in \mathbf{H}$, $a_1 := |V(\mathcal{D}_A) \cap V(T_2)|$, $a_2 := |V(\mathcal{D}_A) \cap V(T_1)|$, $b_1 := |V(\mathcal{D}_B) \cap V(T_1)|$, $b_2 := |V(\mathcal{D}_B) \cap V(T_2)|$, $\tilde{r}_{P4.10} := \tilde{r}$. We obtain a matching $\mathbf{M} \subseteq E(\mathbf{H})$ and two adjacent clusters $A, B \in V(\mathbf{H})$ satisfying one of four configurations.

For any of these four possible configurations, Proposition 4.11 with input $\delta_{P4.11} := \frac{\eta r' q}{400}$, $q_{P4.11} := q$, $d_{P4.11} := d/2$, $\varepsilon_{P4.11} := \varepsilon'$, $\tilde{N}_{max,P4.11} := tN_{max}$, $H_{P4.11} := H$, $\mathbf{H}_{P4.11} := \mathbf{H}$, and further input as in Proposition 4.10, gives an embedding of T in $H \subseteq G$, proving Theorem 4.4.

4.2.3 Proof of Proposition 4.10

We will prove Proposition 4.10 in several steps. We start by defining the desired matching $\mathbf{M}$ as well as several other subsets of $\mathbf{H}$.

Let $\mathbf{M} \subseteq \mathbf{H}[\mathcal{L}, \mathcal{S}]$ be a matching minimising the number of vertices in the set $\mathcal{S}_0 := \{X \in \mathcal{S} : \overline{\deg}(X) < (\tilde{r} + r'\eta/2)k\}$. It follows that $\mathcal{S}_1 = \mathcal{S} \setminus (\mathcal{S}_M \cup \mathcal{S}_0)$.

We define $\mathcal{B} \subseteq V(\mathbf{M})$ as the set of those clusters X, for which there is an alternating path $P = X_1X_2 \ldots X_k$, such that $X_1 \in \mathcal{S}_0$, $X_k = X$, $X_{2i} \in \mathcal{L}$, $X_{2i+1} \in \mathcal{S}_M$,

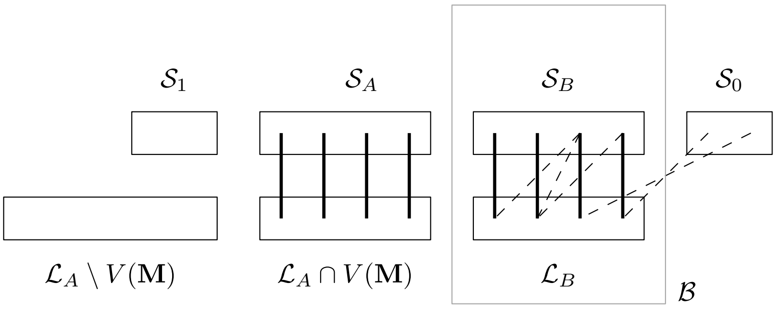

Figure 4.3: Various subsets of **H** used in the proof of Proposition 4.10.

$\{X_{2i}, X_{2i+1}\} \in \mathbf{M}$. Also let $\mathcal{L}_B = \mathcal{L} \cap \mathcal{B}$ and $\mathcal{S}_B = \mathcal{S}_M \cap \mathcal{B}$. Then we define $\mathcal{A} = V(\mathbf{M}) \setminus \mathcal{B}$, $\mathcal{L}_A = \mathcal{L} \setminus \mathcal{L}_B$, $\mathcal{S}_A = \mathcal{S}_M \setminus \mathcal{S}_B$.

Claim 4.12. *For all $X \in \mathcal{S}_B$ we have $\overline{\deg}(X) < (\tilde{r} + r'\eta/2)k$. Also, there are no edges between clusters from $\mathcal{L}_A$ and $\mathcal{S}_0 \cup \mathcal{S}_B$.*

Proof. If the first statement was not true, the symmetric difference of **M** and an alternating path between X and a vertex in $\mathcal{S}_0$ would yield a matching contradicting the choice of **M** as a matching minimising the size of $\mathcal{S}_0$.

If the second statement was not true, we would have an alternating path ending at X which is a contradiction with the definition of $\mathcal{L}_A$. □

Now we are going to define yet another subsets of $\mathcal{L}$ based on the average degrees of the clusters.

$$\mathcal{L}^* := \{X \in \mathcal{L} : \overline{\deg}(X, \mathcal{L}) \geq (\tilde{r} + r'\eta/2)k\},$$

$$\mathcal{L}^+ := \{X \in \mathcal{L} \setminus \mathcal{L}^* : \overline{\deg}(X, \mathcal{S}_M \cup \mathcal{S}_1) \geq (1 - \tilde{r} + \eta/2)k\}.$$

Next, we define $\mathcal{L}_A^* := \mathcal{L}^* \cap \mathcal{L}_A$ and $\mathcal{L}_A^+ := \mathcal{L}^+ \cap \mathcal{L}_A$. We have $\mathcal{L}_A^* = \mathcal{L}_A \setminus \mathcal{L}_A^+$ by Claim 4.12.We define $\mathcal{L}_B^+$ and $\mathcal{L}_B^*$ in a similar way. Finally, let

$$\mathcal{N} = N(\mathcal{L}_A^*) \cap \mathcal{L}.$$

Now suppose that none of the four configurations from statement of the theorem occurs in the cluster graph **H**. We are going to gradually constrain the structure of **H** until we find a contradiction.

Claim 4.13. *Let X and Y be two clusters such that $X \in \mathcal{L}$ and $\overline{\deg}(X, \mathcal{S}_0) = 0$ and $\overline{\deg}(Y, \mathcal{L}) \geq (\tilde{r} + r'\eta/2)k$. Then X and Y are not connected by an edge.*

Proof. If there is $X \in \mathcal{L}$ such that $\overline{\deg}(X, \mathcal{S}_0) = 0$, then we have $\overline{\deg}(X, \mathcal{L} \cup \mathcal{S}_1 \cup \mathcal{S}_M) \geq (1 + \eta)k$. Now suppose that there is an edge between such a cluster X and a cluster Y with $\overline{\deg}(Y, \mathcal{L}) \geq (\tilde{r} + r'\eta/2)k$. If $\tilde{r}a_1 > (1 - \tilde{r})a_2$, we have found Configuration B. If, on the other hand, $\tilde{r}a_1 \leq (1 - \tilde{r})a_2$, recall that $b_1 \leq a_2 + b_1 = \tilde{r}k$, meaning that we have found Configuration C. □

Corollary 4.14. *We have:*

1. *$e(\mathcal{L}_A, \mathcal{L}^* \cup \mathcal{S}_1) = 0$, thus $\mathcal{N}$ is a subset of $\mathcal{L}_B$,*

2. *$\forall X \in \mathcal{N} : \overline{\deg}(X, \mathcal{L}) < (\tilde{r} + r'\eta/2)k$,*

3. $\forall X \in \mathcal{S}_A : \overline{\deg}(X) = \overline{\deg}(X, \mathcal{L}) < (\tilde{r} + r'\eta/2)k$.

Proof.

1. Suppose that there is an edge between $X \in \mathcal{L}_A$ and $Y \in \mathcal{L}^* \cup \mathcal{S}_1$. From Claim 4.12 we get that $\overline{\deg}(X, \mathcal{S}_0) = 0$. From the definition of $\mathcal{L}^*$ and $\mathcal{S}_1$ we have $\overline{\deg}(Y, \mathcal{L}) \geq (\tilde{r} + r'\eta/2)k$. Thus we can apply Claim 4.13 for X and Y.

2. Each vertex $Y \in \mathcal{N}$ has a neighbour $X \in \mathcal{L}_A^*$. If $\overline{\deg}(Y, \mathcal{L}) \geq (\tilde{r} + r'\eta/2)k$ we are in the situation of the first part of this claim.

3. Each vertex $Y \in \mathcal{S}_A$ is matched to a vertex $X \in \mathcal{L}_A$. If $\overline{\deg}(Y, \mathcal{L}) \geq (\tilde{r} + r'\eta/2)k$, we are, yet again, in the situation of the first part of the claim.

□

Claim 4.15. *Every cluster in $\mathcal{N}$ has average degree at least $(\tilde{r} + \eta/2)k$ in $\mathcal{S}_0$.*

Proof. Suppose that it is not so. Then we have a cluster $Y \in \mathcal{N}$ such that

$$\begin{aligned} \overline{\deg}(Y, \mathcal{S}_1 \cup \mathcal{S}_M \cup \mathcal{L}) &\geq (1+\eta)k - (\tilde{r} + \eta/2)k \\ &\geq (1 - \tilde{r} + \eta/2)k \ . \end{aligned}$$

Now we consider separately three cases:

1. Suppose that $\tilde{r}a_1 \leq (1-\tilde{r})a_2$. Then either

$$\overline{\deg}(Y, \mathcal{L}) \geq b_1 + \eta k/4,$$

which leads to the Configuration C (consider Y and its neighbour in $\mathcal{L}_A^*$), or we have

$$\begin{aligned} \overline{\deg}(Y, \mathcal{S}_1 \cup \mathcal{S}_M) &= \overline{\deg}(Y, \mathcal{S}_1 \cup \mathcal{S}_M \cup \mathcal{L}) - \overline{\deg}(Y, \mathcal{L}) \\ &\geq (1-\tilde{r})k + \eta k/2 - (b_1 + \eta k/4) \\ &= \frac{1-\tilde{r}}{\tilde{r}}\tilde{r}k - b_1 + \eta k/4 \\ &= \frac{1-\tilde{r}}{\tilde{r}}(b_1 + a_2) - b_1 + \eta k/4 \\ &= \frac{1-2\tilde{r}}{\tilde{r}}b_1 + \frac{1-\tilde{r}}{\tilde{r}}a_2 + \eta k/4 \\ &\geq \frac{1-\tilde{r}}{\tilde{r}}a_2 + \eta k/4, \end{aligned}$$

where we used the bound on the average degree of Y and then the facts that $b_1 + a_2 = \tilde{r}k$ and $\tilde{r} \leq r' \leq 1/2$. This, on the other hand, leads to the Configuration A (again, consider Y and its neighbour in $\mathcal{L}_A^*$).

2. Suppose that $\tilde{r}a_1 > (1-\tilde{r})a_2$ and $b_1 \leq \frac{\tilde{r}^2}{1-\tilde{r}}k$. Following the same considerations as in the previous case we get that either $\overline{\deg}(Y, \mathcal{L}) \geq b_1 + \eta k/4$ or $\overline{\deg}(Y, \mathcal{S}_1 \cup \mathcal{S}_M) \geq \frac{1-\tilde{r}}{\tilde{r}}a_2 + \eta k/4$. The second case leads, again, to the Configuration A. We now proceed with the first case.

Let X be a neighbour of Y in $\mathcal{L}_A^*$. From Claim 4.12 we have $\overline{\deg}(X, \mathcal{S}_0) = 0$ and from Corollary 4.14.1 we have $\overline{\deg}(X, \mathcal{S}_1) = 0$, thus

$$\overline{\deg}(X, \mathcal{L} \cup \mathcal{S}_M) = \overline{\deg}(X) \geq (1+\eta)k > k + \eta k/4.$$

Moreover, all the matching edges containing clusters from $\mathcal{S} \cap N(X)$ must have both ends in the set $\mathcal{A}$ because there are no edges between vertices from $\mathcal{L} \cap \mathcal{A}$ and $\mathcal{S} \cap \mathcal{B}$ (Claim 4.12). On the other hand, all neighbours of X in $\mathcal{L}$ have to be in $\mathcal{B}$ (Corollary 4.14 (1)), so all matching edges containing vertices from $\mathcal{L} \cap N(X)$ are in $\mathcal{B}$. Thus, all of the assumptions of Configuration D for X and Y are satisfied.

3. Finally we are left with the case $\tilde{r}a_1 > (1-\tilde{r})a_2$ and $b_1 > \frac{\tilde{r}^2}{1-\tilde{r}}k$. Note that then we have

$$a_2 = \tilde{r}k - b_1 < \tilde{r}k - \frac{\tilde{r}}{1-\tilde{r}}\tilde{r}k = (1 - \frac{\tilde{r}}{1-\tilde{r}})\tilde{r}k = (1-2\tilde{r})\,\frac{\tilde{r}}{1-\tilde{r}}k.$$

Now either

$$\overline{\deg}(Y, \mathcal{L}) \geq \tilde{r}k + r'\eta k/4,$$

or

$$\begin{aligned}
\overline{\deg}(Y, \mathcal{S}_1 \cup \mathcal{S}_M) &= \overline{\deg}(Y, \mathcal{S}_1 \cup \mathcal{S}_M \cup \mathcal{L}) - \overline{\deg}(Y, \mathcal{L}) \\
&\geq (1-\tilde{r})k + \eta k/2 - (\tilde{r}k + r'\eta k/4) \\
&\geq (1-2\tilde{r})k + \eta k/4 \\
&= \frac{1-\tilde{r}}{\tilde{r}}\,(1-2\tilde{r})\,\frac{\tilde{r}}{1-\tilde{r}}k + \eta k/4 \\
&\geq \frac{1-\tilde{r}}{\tilde{r}}a_2 + \eta k/4.
\end{aligned}$$

The first option leads to Configuration B while the second one leads to Configuration A.

□

After restricting the structure of $\mathbf{H}$ we are ready to derive a contradiction by combining several properties of $\mathbf{H}$ together. At first we estimate the size of the set $\mathcal{L}_A$. Recall that we have $|\mathcal{L}| \geq (1+\eta)|\mathcal{S}|$, thus $|\mathcal{L}| > |\mathcal{S}|$. We also know that $|\mathcal{L}_B| = |\mathcal{S}_B|$, because the two sets are matched in $\mathbf{M}$. This means that

$$|\mathcal{L}_A| = |\mathcal{L}| - |\mathcal{L}_B| > |\mathcal{S}| - |\mathcal{S}_B| = |\mathcal{S}_A| + |\mathcal{S}_0| + |\mathcal{S}_1|. \tag{4.6}$$

Now we proceed by bounding the size of the set $\mathcal{N}$.

Lemma 4.16. *Suppose that the set $\mathcal{L}_A^*$ (and thus also $\mathcal{N}$) is nonempty. Then the following inequality holds:*

$$|\mathcal{N}|(\tilde{r} + r'\eta/2) > |\mathcal{S}_0|(1 - \tilde{r} + \eta/2). \tag{4.7}$$

Proof. We estimate the number of edges between L_A^+ and S_A. For $\mathcal{Y}, \mathcal{Z} \subseteq V(\mathbf{H})$, we set $\vec{w}(\mathcal{Z}, \mathcal{Y}) := \sum_{Z \in \mathcal{Z}} \overline{\deg}(Z, \mathcal{Y})$. On one hand we have

$$\vec{w}(\mathcal{L}_A^+, \mathcal{S}_A) = \sum_{Z \in \mathcal{L}_A^+} \overline{\deg}(Z, \mathcal{S}) \geq |\mathcal{L}_A^+|(1 - \tilde{r} + \eta/2)k,$$

because $\vec{w}(\mathcal{L}_A^+, \mathcal{S}_B \cup \mathcal{S}_0) = 0$ (Claim 4.12) and $\vec{w}(\mathcal{L}_A^+, \mathcal{S}_1) = 0$ (Corollary 4.14). On the other hand we have

$$\begin{aligned}
\vec{w}(\mathcal{L}_A^+, \mathcal{S}_A) &= \sum_{Z \in \mathcal{L}_A^+, W \in \mathcal{S}_A} \overline{\deg}(Z, W) \\
&= \sum_{Z \in \mathcal{L}_A^+, W \in \mathcal{S}_A} \frac{1 - r'}{r'} \overline{\deg}(W, Z) \\
&= \frac{1 - r'}{r'} \vec{w}(\mathcal{S}_A, \mathcal{L}_A^+) \\
&\leq \frac{1 - r'}{r'} (|\mathcal{S}_A|(\tilde{r} + r'\eta/2)k - \vec{w}(\mathcal{S}_A, \mathcal{L}_A^*)) \\
&\leq (1 - r')|\mathcal{S}_A|(1 + \eta/2)k - \vec{w}(\mathcal{S}_A, \mathcal{L}_A^*) \\
&\leq |\mathcal{S}_A|(1 + \eta/2)(1 - \tilde{r})k - \vec{w}(\mathcal{L}_A^*, \mathcal{S}_A) \ ,
\end{aligned}$$

because all the clusters from $\mathcal{S}_A$ (if there are any) have their average degree bounded by $(\tilde{r} + r'\eta/2)k$ (Corollary 4.14), and $\tilde{r} \leq r' \leq 1/2$. After combining the inequalities we get

$$|\mathcal{L}_A^+|(1 - \tilde{r} + \eta/2)k \leq |\mathcal{S}_A|(1 - \tilde{r} + \eta/2)k - \vec{w}(\mathcal{L}_A^*, \mathcal{S}_A). \tag{4.8}$$

We continue by estimating the number of edges between $\mathcal{L}_A^*$ and $\mathcal{N}$. On one hand we have

$$\vec{w}(\mathcal{L}_A^*, \mathcal{N}) = \vec{w}(\mathcal{N}, \mathcal{L}_A^*) \leq |\mathcal{N}|(\tilde{r} + r'\eta/2)k$$

due to Corollary 4.14 (2). On the other hand we have

$$\begin{aligned}
\vec{w}(\mathcal{L}_A^*, \mathcal{N}) &= \vec{w}(\mathcal{L}_A^*, V(\mathbf{H})) - \vec{w}(\mathcal{L}_A^*, \mathcal{S}_A) - \vec{w}(\mathcal{L}_A^*, \mathcal{S}_1 \cup \mathcal{S}_B \cup \mathcal{S}_0) \\
&= \vec{w}(\mathcal{L}_A^*, V(\mathbf{H})) - \vec{w}(\mathcal{L}_A^*, \mathcal{S}_A) \\
&\geq |\mathcal{L}_A^*|(1 + \eta)k - \vec{w}(\mathcal{L}_A^*, \mathcal{S}_A) \ ,
\end{aligned}$$

because there are neither edges between $\mathcal{L}_A^*$ and $\mathcal{S}_1$ (Corollary 4.14 (1)), nor edges between $\mathcal{L}_A^*$ and $\mathcal{S}_B \cup \mathcal{S}_0$ (Claim 4.12) and clusters in $\mathcal{L}$ have large degree.

By combining the two inequalities we get

$$|\mathcal{L}_A^*|(1 + \eta)k - \vec{w}(\mathcal{L}_A^*, \mathcal{S}_A) \leq |\mathcal{N}|(\tilde{r} + r'\eta/2)k. \tag{4.9}$$

Combining Inequalities (4.8), (4.9) and (4.6) in this order we get:

$$\begin{aligned}
|\mathcal{N}|(\tilde{r} + r'\eta/2)k &\geq |\mathcal{L}_A^*|(1 + \eta)k - \vec{w}(\mathcal{L}_A^*, \mathcal{S}_A) \\
&\geq |\mathcal{L}_A^*|(1 + \eta/2)k + |\mathcal{L}_A^+|(1 - \tilde{r} + \eta/2)k - |\mathcal{S}_A|(1 - \tilde{r} + \eta/2)k \\
&= |\mathcal{L}_A|(1 - \tilde{r} + \eta/2)k + |\mathcal{L}_A^*|\tilde{r}k - |\mathcal{S}_A|(1 - \tilde{r} + \eta/2)k \\
&= |\mathcal{L}_A^*|\tilde{r}k + (|\mathcal{L}_A| - |\mathcal{S}_A|)(1 - \tilde{r} + \eta/2)k \\
&> |\mathcal{L}_A^*|\tilde{r}k + (|\mathcal{S}_0| + |\mathcal{S}_1|)(1 - \tilde{r} + \eta/2)k \\
&\geq |\mathcal{S}_0|(1 - \tilde{r} + \eta/2)k
\end{aligned}$$

which concludes the proof. □

Corollary 4.17. *The set $\mathcal{L}_A$ is empty.*

Proof. Suppose that $\mathcal{N}$ (and thus also $\mathcal{L}_A^*$) is nonempty. Then on one hand we have

$$\vec{w}(\mathcal{N}, \mathcal{S}_0) = \frac{1-r'}{r'}\vec{w}(\mathcal{S}_0, \mathcal{N}) \leq \frac{1-r'}{r'}|\mathcal{S}_0|(\tilde{r} + r'\eta/2)k \leq |\mathcal{S}_0|(1+\eta/2)(1-r')k \ , \quad (4.10)$$

due to the definition of $\mathcal{S}_0$ and the fact $\tilde{r} \leq r'$. On the other hand we have

$$\vec{w}(\mathcal{N}, \mathcal{S}_0) \geq |\mathcal{N}|(\tilde{r} + \eta/2)k \quad (4.11)$$

due to Claim 4.15. After combining the inequalities we get that

$$|\mathcal{N}|(\tilde{r} + \eta/2) \leq |\mathcal{S}_0|(1-r')(1+\eta/2) \leq |\mathcal{S}_0|(1-\tilde{r})(1+\eta/2). \quad (4.12)$$

Combining with Lemma 4.16 we get

$$|\mathcal{S}_0|(1-\tilde{r}+\eta/2) < |\mathcal{N}|(\tilde{r} + r'\eta/2) < |\mathcal{N}|(\tilde{r} + \eta/2) \leq |\mathcal{S}_0|(1-\tilde{r})(1+\eta/2).$$

which gives a contradiction, because

$$1-\tilde{r}+\eta/2 > 1-\tilde{r}+\eta/2-\tilde{r}\eta/2 = (1-\tilde{r})(1+\eta/2).$$

Thus $\mathcal{L}_A^*$ and $\mathcal{N}$ are empty.

Now suppose that $\mathcal{L}_A^+ = \mathcal{L}_A$ is nonempty. Then on one hand we have

$$\begin{aligned}\vec{w}(\mathcal{L}_A^+, \mathcal{S}_A) &= \frac{1-r'}{r'}\vec{w}(\mathcal{S}_A, \mathcal{L}_A^+) \\ &< \frac{1-r'}{r'}|\mathcal{S}_A|(\tilde{r} + r'\eta/2)k \\ &\leq |\mathcal{S}_A|(1+\eta/2)(1-r')k \ ,\end{aligned}$$

because of Corollary 4.14 (3) and on the other hand we have

$$\begin{aligned}\vec{w}(\mathcal{L}_A^+, \mathcal{S}_A) &= \vec{w}(\mathcal{L}_A^+, \mathcal{S}_M \cup \mathcal{S}_1) \\ &\geq |\mathcal{L}_A^+|(1-\tilde{r}+\eta/2)k \\ &= |\mathcal{L}_A|(1-\tilde{r}+\eta/2)k \\ &\geq |\mathcal{S}_A|(1-r'+\eta/2)k \\ &> |\mathcal{S}_A|(1+\eta/2)(1-r')k \ ,\end{aligned}$$

where we used the definition of $\mathcal{L}_A^+$, Corollary 4.14 (1), Claim 4.12, Inequality (4.6), and the fact that $\tilde{r} \leq r'$.

Combining the inequalities gives a contradiction. Thus, the set $\mathcal{L}_A$ has to be empty. □

From Corollary 4.17 it follows that all L-clusters are in $\mathcal{L}_B$ and thus are matched to $\mathcal{S}_M$, i.e., $|\mathcal{L}| = |\mathcal{S}_M| \leq |\mathcal{S}|$, which contradicts our assumption that $|\mathcal{L}| > |\mathcal{S}|$.

4.2.4 Embedding

We call a pair (F, R) an *anchored τ forest* if F is a forest (possibly consisting of a single tree), $R \subseteq V(F_1)$, where F_1 is one of the colour classes of F, $F - R$ decomposes into components of size at least two and at most τ, each component K in $F - R$ is adjacent in F to at least one and at most two vertices from R and each two vertices in R are of distance at least 4. We shall use the notation $K \in F - R$ to denote that the tree K is one of the components of $F - R$.

First we state a proposition that will allow us to use matching edges in our r-skewed LKS-cluster graph to embed part of our tree T.

Specifically, in Proposition 4.18 we are given an anchored forest (F, R), an r-skewed LKS-graph which contains a cluster A with some nice average degree to some $L - S$-matching, and an injective mapping of R on ultratypical vertices of A. We want to extend it to an embedding of F.

Proposition 4.18. *For all $\eta, d > 0$ and $r \in \mathbb{Q}^+$, $0 < r \leq 1/2$, there is an $\varepsilon = \varepsilon(\eta, d, r) > 0$ such that for any $\tilde{N}_{max} \in \mathbb{N}$ there is a $\beta = \beta(\eta, r, \varepsilon, \tilde{N}_{max}) > 0$ such that for all $n \in \mathbb{N}$ the following holds.Let (F, R) be an anchored βn-forestwith colour classes F_1 and F_2 such that $R \subseteq F_2$ and for each component $K \in F - R$, we have $|F_1 \cap K| \leq |F_2 \cap K|$. Let H be an r-skewed LKS-graph of order n with parameters $(\cdot, \cdot, \varepsilon, d)$ with a corresponding cluster graph $\mathbf{H}$ of order at most $\tilde{N}_{max}$. Let $U \subseteq V(H)$ and let $\mathbf{M} \subseteq E(\mathbf{H})$ be a matching in $\mathbf{H}$ between L-clusters and S-clusters.*

If for $A \in V(\mathbf{H})$ we have

$$\overline{\deg}(A, \mathcal{S} \cap V(\mathbf{M})) \geq \frac{1-r}{r}|F_2| + \sum_{C \subseteq S \,:\, CD \in \mathbf{M}} \max\{|U \cap C|, \frac{1-r}{r}|U \cap D|\} + \eta n \;,$$

then for any injective mapping of R on ultratypical vertices of A, there is an embedding φ of F avoiding U and extending this mapping such that $\varphi(V(F_1)) \subseteq S \cap V(\mathbf{M})$, $\varphi(V(F_2) \setminus R) \subseteq L \cap \bigcup V(\mathbf{M})$, and $V(F_2)$ are mapped on ultratypical vertices. Moreover, for any cluster $C \in V(\mathbf{H})$ where we embedded vertices from $F - R$ it holds that $|C \setminus (U \cup \varphi(F))| \geq r\eta/8|C|$.

Next, we state a proposition allowing us to use high average degree of some clusters to embed further part of our tree T.

Specifically, in Proposition 4.19 we are given an anchored forest (F, R), an r-skewed LKS-graph which contains a cluster A with big enough average degree to a set of clusters with high average degree, and an injective mapping of R on ultratypical vertices of A. We want to extend it to an embedding of F.

When using the proposition, we always set $\mathcal{B}$ to be the set of L-clusters in (1) and the set of S_1-clusters in (2).

Proposition 4.19. *For all $\eta, d > 0$ and $0 < r \leq 1/2$, there is an $\varepsilon = \varepsilon(\eta, d, r) > 0$ such that for any $\tilde{N}_{max} \in \mathbb{N}$ there is a $\beta = \beta(\eta, r, \varepsilon, \tilde{N}_{max}) > 0$ such that for all $n \in \mathbb{N}$ the following holds. Let (F, R) be an anchored βn-forest with colour classes F_1 and F_2 such that $R \subseteq F_2$. Let H be an r-skewed LKS-cluster graph with parameters $(\cdot, \cdot, \varepsilon, d)$ of order n with an associated cluster graph $\mathbf{H}$ of order at most $\tilde{N}_{max}$. Let $U \subseteq V(H)$ and let $\mathcal{B} \subseteq V(\mathbf{H})$ be a set of clusters. Let $\varphi : R \to A$ with $A \in V(\mathbf{H})$ be an injective mapping on ultratypical vertices.*

1. *If* $\overline{\deg}(A,\mathcal{B}) \geq |F_1| + |\bigcup\mathcal{B} \cap U| + \eta n$, *then we can extend* φ *to* $N(R)$ *so that* $\varphi(N(R))$ *are ultratypical vertices in* $\bigcup\mathcal{B} \setminus U$ *and find a set* $W = W_1 \dot\cup W_2 \dot\cup \ldots \subseteq \bigcup\mathcal{B} \setminus (U \cup \varphi(R \cup N(R)))$ *of reserved vertices such that* $|W_i| = |(F_1 \cap K_i) \setminus N(R)|$, *with* $K_i \in F - R$ *and such that* W_i *lies in the same cluster as* $\varphi(K_i \cap N(R))$ *and for each cluster* $C \in \mathcal{B}$ *with* $C \cap \varphi(N(R)) \neq \emptyset$ *we have* $|C \setminus (U \cup W \cup \varphi(N(R)))| \geq r\eta/8 \cdot |C|$.

   Moreover, for any set $\tilde{U} \subseteq V(G) \setminus (U \cup W \cup \varphi(R \cup N(R)))$, *for which* $\overline{\deg}(B) \geq |F_1| + |F_2| + |U \cup \tilde{U}| + \eta n$ *for each* $B \in \mathcal{B}$ *and such that for any* $C \in V(\mathbf{H})$ *with* $C \cap \tilde{U} \neq \emptyset$ *we have* $|C \setminus (U \cup W \cup \tilde{U} \cup \varphi(N(R)))| \geq r\eta/8 \cdot |C|$, *we can further extend* φ *to the whole* F *avoiding* $U \cup \tilde{U}$ *such that* $\varphi(F_1) \subseteq \bigcup\mathcal{B}$. *Moreover, the extension* φ *is such that for any cluster* $C \in V(\mathbf{H})$ *with* $C \cap \varphi(F - (R \cup N(R)) \neq \emptyset$, *we have* $|C \setminus (\tilde{U} \cup U)| \geq r\eta/8 \cdot |C|$.

2. *If* $\overline{\deg}(A,\mathcal{B}) \geq |F_1| + |\bigcup\mathcal{B} \cap U| + \eta n$ *and* $\overline{\deg}(B, V(\mathbf{H}) \setminus \mathcal{B}) \geq |F_2| + |U| + \eta n$ *for each* $B \in \mathcal{B}$, *then we can extend* φ *to* F *in* $V(G)$ *avoiding* U *and such that* $\varphi(V(F_1)) \subseteq \bigcup\mathcal{B}$, $\varphi(V(F_2)) \subseteq \bigcup N_{\mathbf{H}}(\mathcal{B}) \setminus \mathcal{B}$, *and* $V(F_2)$ *are mapped on ultratypical vertices. Moreover, the embedding* φ *is such that for any cluster* $C \in V(\mathbf{H})$ *with* $C \cap \varphi(F - R) \neq \emptyset$, *we have* $|C \setminus (\varphi(F) \cup U)| \geq r\eta/8 \cdot |C|$.

We at first prove Proposition 4.18.

Proof of Proposition 4.18. Given $\eta, d > 0$ and $r \in \mathbb{Q}$ set $\varepsilon = \min\{(\frac{\eta r}{12})^2, \frac{dr\eta}{100}\}$.For any $\tilde{N}_{max} \in \mathbb{N}$ set $\beta = \frac{\varepsilon r\eta}{4\tilde{N}_{max}}$.

We shall define a set $\tilde{U}$ of vertices used for the embedding process. At the beginning $\tilde{U} = \varphi(R)$. At any time of the embedding process, let φ be the partial embedding of F. We shall embed one by one each component $K \in F - R$. The embedding φ will be defined in such a way that $\varphi(K \cap F_1) \subseteq S$ and $\varphi(K \cap F_2 \setminus R) \subseteq L$. During the whole embedding process, we shall ensure that the following holds

$$\overline{\deg}(A, \mathcal{S} \cap V(\mathbf{M})) \geq$$
$$\frac{1-r}{r}(|F_2| - |\varphi(F_2)|) + \sum_{C \subseteq \mathcal{S} : CD \in \mathbf{M}} \max\{|(U \cup \tilde{U}) \cap C|, \frac{1-r}{r}|(U \cup \tilde{U}) \cap D|\} + \eta n \,.$$

This holds at the beginning when $\tilde{U} = R$.

For each next $K \in F - R$ to be embedded, let R_K be the vertices in R adjacent to K (at least one, at most two). Let $\mathcal{S}' \subseteq \mathcal{S} \cap V(\mathbf{M})$ be such that both $\varphi(R_K)$ are typical to each cluster $C \in \mathcal{S}'$. By Lemma 2.7 we have that $|\mathcal{S} \cap \mathbf{M}| - |\mathcal{S}'| \leq 2\sqrt{\varepsilon}|V(\mathbf{H})|$ and thus similarly as in the proof of Proposition 4.9 we can calculate for $x_i \in R_K$, $i = 1, 2$ that $\deg(\varphi(x_i), \bigcup\mathcal{S}') \geq \overline{\deg}(A, \mathcal{S} \cap V(\mathbf{M})) - 3\sqrt{\varepsilon}n/r$ and thus

$$\begin{aligned}
&\deg(\varphi(x_i), \bigcup\mathcal{S}') \\
&\geq \frac{1-r}{r}(|F_2| - |\varphi(F_2)|) + \sum_{C \subseteq \mathcal{S} : CD \in \mathbf{M}} \frac{1-r}{r}|\tilde{U} \cap D| \\
&+ \sum_{C \subseteq \mathcal{S} : CD \in \mathbf{M}} \max\{|U \cap C|, \frac{1-r}{r}|U \cap D|\} + \eta n - 3\sqrt{\varepsilon}n/r \\
&\geq \frac{1-r}{r}(|F_2| - |\varphi(F_2)|) + \sum_{C \subseteq \mathcal{S} : CD \in \mathbf{M}} \max\{|(U \cup \tilde{U}) \cap C|, \frac{1-r}{r}|(U \cup \tilde{U}) \cap D|\} + \frac{3\eta n}{4} \\
&\geq \sum_{C \subseteq \mathcal{S}' : CD \in \mathbf{M}} \left(\max\{|(U \cup \tilde{U}) \cap C|, \frac{1-r}{r}|(U \cup \tilde{U}) \cap D|\} + 3\eta n/(4|\mathcal{S}'|)\right) .
\end{aligned}$$

Then there is a $C \in \mathcal{S}'$ with $CD \in \mathbf{M}$ such that

$$\deg(\varphi(x_i), C) \geq \max\{|(U \cup \tilde{U}) \cap C|, \frac{1-r}{r}|(U \cup \tilde{U}) \cap D|\} + 3\eta n/(4|\mathcal{S}'|) \ .$$

Thus,

$$\begin{aligned}
&|C| - \max\{|(\tilde{U} \cup U) \cap C|, \frac{1-r}{r}|(\tilde{U} \cup U) \cap D|\} \\
&\geq \deg(\varphi(x_i, C)) - \max\{|(\tilde{U} \cup U) \cap C|, \frac{1-r}{r}|(\tilde{U} \cup U) \cap D|\} \qquad (4.13) \\
&\geq 3\eta n/(4|\mathcal{S}'|) \\
&\geq \frac{1-r}{r}\beta n + \eta n/(2|V(\mathbf{H})|) \\
&\geq |F_1 \cap K| + \eta r|C|/2 \ , \qquad (4.14)
\end{aligned}$$

where the third inequality follows from the definition of β and the last inequality follows from Proposition 4.9 (1). Similarly we have

$$\begin{aligned}
|D \setminus (\tilde{U} \cup U)| &\geq \frac{r}{1-r}(|C| - \max\{|(\tilde{U} \cup U) \cap C|, \frac{1-r}{r}|(\tilde{U} \cup U) \cap D|\}) \\
&\geq \beta n + \frac{r}{1-r}\eta n/(2|V(\mathbf{H})|) \\
&\geq |F_2 \cap K| + \eta r|D|/2 \ , \qquad (4.15)
\end{aligned}$$

where we again use the definition of β and Proposition 4.9 (1).

In particular, in the neighbourhood of each vertex $u_i \in \varphi(R_K)$, $i = 1, 2$, there are at least $|F_1 \cap K|$ unused vertices of $C \setminus U$ that are typical w.r.t. $D \setminus (\tilde{U} \cup U)$. Let $\varphi(N(x_i) \cap K) = v_i$, $i = 1, 2$, be such vertices. Hence,

$$\deg(v_i, D \setminus (\tilde{U} \cup U)) \geq (d - \varepsilon)|D \setminus (\tilde{U} \cup U)| \geq (d - \varepsilon)r\eta|D|/8 > 3\varepsilon|D|$$

for $i = 1, 2$. Observe that $|K| \leq \beta n < \frac{\varepsilon r n}{|V(\mathbf{H})|} \leq \varepsilon \min\{|C|, |D|\}$. We can thus use Lemma 2.14 with $T_{L2.14} := K$, $X'_{L2.14} := C \setminus (\tilde{U} \cup U)$, $Y'_{L2.14} := D \setminus (\tilde{U} \cup U)$, $R_{L2.14} := \{N(x_i),\ i = 1, 2\}$ $\varepsilon_{L2.14} := \varepsilon$, $\alpha_{L2.14} := \frac{16\varepsilon}{\eta r}$, and $d_{L2.14} := d$ to embed K in $C \cup D$ with $\varphi(F_1 \cap K) \subseteq C \setminus (\tilde{U} \cup U) \subseteq \mathcal{S}$ and $\varphi(F_2 \cap K \setminus R) \subseteq D \setminus (\tilde{U} \cup U) \subseteq \mathcal{L}$. Add $\varphi(K)$ to $\tilde{U}$. From (4.14) and (4.15), we now have that $|C \setminus (\tilde{U} \cup U)| \geq r\eta/8|C|$, and $|D \setminus (\tilde{U} \cup U)| \geq r\eta/8|D|$. Observe also that for the partial embedding φ we have

$$\begin{aligned}
&\overline{\deg}(A, \mathcal{S} \cap V(\mathbf{M})) \\
&\geq \frac{1-r}{r}|F_2| + \sum_{C \subseteq \mathcal{S}\,:\,CD \in \mathbf{M}} \max\{|U \cap C|, \frac{1-r}{r}|U \cap D|\} + \eta n \\
&\geq \frac{1-r}{r}((|F_2| - |\varphi(F_2)|) + |\tilde{U} \cap L|) + \sum_{C \subseteq \mathcal{S}\,:\,CD \in \mathbf{M}} \max\{|U \cap C|, \frac{1-r}{r}|U \cap D|\} + \eta n \\
&\geq \frac{1-r}{r}(|F_2| - |\varphi(F_2)|) + \sum_{C \subseteq \mathcal{S}\,:\,CD \in \mathbf{M}} \max\{|(U \cup \tilde{U}) \cap C|, \frac{1-r}{r}|(U \cup \tilde{U}) \cap D|\} + \eta n \ ,
\end{aligned}$$

where the last inequality comes from the fact that $|F_1 \cap K| \leq |F_2 \cap K|$ for all $K \in F - R$, and that the embedding φ was defined in such a way that $\varphi(F_1) \subseteq \mathcal{S}$ and $\varphi(F_2 \setminus R) \subseteq \mathcal{L}$.

Proceeding in the same way for every $K \in F - R$, we extend $\varphi(R)$ to the whole anchored forest F in such a way that $\varphi(F_1) \subseteq \mathcal{S} \cap V(\mathbf{M})$, $\varphi(F_2 \setminus R) \subseteq \mathcal{L} \cap V(\mathbf{M})$, and for each cluster $C \in V(\mathbf{H})$ with $C \cap \varphi(F - R) \neq \emptyset$ we have $|C \setminus (\tilde{U} \cup U)| \geq r\eta/8|C|$.

□

We conclude this section by proving Proposition 4.19.

Proof of Proposition 4.19. Given $\eta, d > 0$ and $r \in \mathbb{Q}^+$, let $\varepsilon := \min\{\left(\frac{\eta r}{12}\right)^2, \frac{dr\eta}{100}\}$. Then for any $\tilde{N}_{max} \in \mathbb{N}$, set $\beta = \frac{r\eta\varepsilon}{4\tilde{N}_{max}}$.

We shall prove only the more difficult Case (1). Case (2) can be proven either analogously, or can be much simplified as F_2 will be mapped outside of $\mathcal{B}$ and thus does not need any reservation or cause any difficulties in embedding F_1.

We define a set $W = W_1 \cup W_2 \cup \dots$ of reserved vertices by setting $W = \emptyset$ at the beginning and progressively adding vertices to it. Also we shall define the set $\tilde{W}$ as the set of vertices used by the partial embedding of $F - R$. Hence at the beginning we have $\tilde{W} = \emptyset$. Suppose that for some s, we have already embedded $K_j \in F - R$, for $j \leq s$. Suppose that $W = W_1 \cup \dots \cup W_s$ is the corresponding set of reserved vertices, i.e., $|W \cup \tilde{W}| = \sum_{j=1}^{s} |K_j|$. For the next component $K_{s+1} \in F - R$ to be embedded, let R_{s+1} be the set of vertices in R adjacent to K_{s+1} (at least one, at most two). Let $\mathcal{B}' \subseteq \mathcal{B}$ be such that $\varphi(R_{s+1})$ are typical to each cluster $C \in \mathcal{B}'$. By Lemma 2.7 we have that $|\mathcal{B} \setminus \mathcal{B}'| \leq 2\sqrt{\varepsilon}|V(\mathbf{H})|$ and thus similarly as in Proposition 4.9 we get for $x \in R_{s+1}$ that

$$\begin{aligned}
\deg(\varphi(x), \bigcup \mathcal{B}') &\geq \overline{\deg}(A, \mathcal{B}) - 3\sqrt{\varepsilon}n/r \\
&\geq \sum_{j=1}^{s+1} |K_j \cap F_1| + |\bigcup \mathcal{B} \cap U| + \eta n - 3\sqrt{\varepsilon}n/r \\
&\geq |\bigcup_{j=1}^{s} W_j| + |\tilde{W}| + |K_{s+1} \cap F_1| + |\bigcup \mathcal{B} \cap U| + 3\eta n/4 \\
&\geq |K_{s+1} \cap F_1| + |W| + |\tilde{W}| + |\bigcup \mathcal{B} \cap U| + 3\eta n/4,
\end{aligned}$$

Hence there is a cluster $B \in \mathcal{B}'$ (not depending on the choice of vertex x in R_{s+1}) such that

$$\deg(\varphi(x), B \setminus (U \cup W \cup \tilde{W})) \geq 3\eta n/(4|\mathcal{B}'|) \geq \frac{\eta n}{4\tilde{N}_{max}} + \frac{\eta n}{2|V(\mathbf{H})|} > \beta n + \eta r/2 \cdot |B|.$$

In particular, in the neighbourhood of each vertex of R_{s+1} there are at least $|K_{s+1}|$ unused and unreserved ultratypical vertices in $B \setminus U$. For each $x \in R_{s+1}$, map its neighbor in K_{s+1} to one of these vertices and add the image to $\tilde{W}$. Choose a set of vertices of size $|K_{s+1} \cap F_1 \setminus N(R)|$ in $B \setminus (U \cup W \cup \tilde{W})$ and add it to W_{s+1} (i.e., also to W). Observe that $|B \setminus (U \cup W \cup \tilde{W})| \geq \eta r/8 \cdot |B|$. We proceed in the same way for every $K \in F - R$.

When we have embedded $N(R) \cap K$ of the last component $K \in F - R$, we have obtained an embedding of $N(R)$ and a reservation set $W = W_1 \cup W_2 \cup \dots$ for $F_1 \setminus N(R)$ such that W_j lies in the same cluster as $\varphi(N(R) \cap K_j)$ does, and in such a way, that for any cluster B where we embedded vertices from $N(R)$ (and possibly reserved space), we still have at least some unused and unreserved vertices, i.e., $|B \setminus (U \cup W \cup \tilde{W})| \geq \eta r/8 \cdot |B|$.

Now we shall proceed with the 'moreover' part, i.e., the embedding the left-over of the trees $K_j \in F - R$. Let u, v in cluster B be the images of $K_j \cap N(R)$ (alternatively there is only one such image). Set $W_j = \emptyset$ (and thus remove from W a set of vertices of the size $|K_j \cap F_1 \setminus N(R)|$). Similarly as above, we find $D \in V(\mathbf{H})$ such that

$$\deg(u, D \setminus (U \cup W \cup \tilde{W} \cup \tilde{U}) \geq |K_j \cap F_2| + \eta r/8 \cdot |D|,$$

and similarly

$$\deg(v, D \setminus (U \cup W \cup \tilde{W} \cup \tilde{U}) \geq |K_j \cap F_2| + \eta r/8 \cdot |D|.$$

As we have $|B\setminus(U\cup W\cup\tilde{W}\cup\tilde{U})| \geq |K_j\cap F_1\setminus N(R)|+r\eta/8\cdot|B|$ and $r\eta/8 > 3\varepsilon$, we may use Lemma 2.14 with $T_{L2.14} := K_j$, $R_{L2.14} := K_j\cap N(R)$, $X'_{L2.14} := B\setminus(U\cup W\cup\tilde{W}\cup\tilde{U})$, $Y'_{L2.14} := D\setminus(U\cup W\cup\tilde{U}\cup\tilde{W})$, $\alpha_{L2.14} := \frac{32\varepsilon}{r\eta}$, $\varepsilon_{L2.14} := \varepsilon$, and $d_{L2.14} := d$ to extend φ to the whole K_j with $F_1\cap K_j\subseteq B$ and $F_2\cap K_j\subseteq D$ and add the used vertices to $\tilde{W}$. Observe that after the embedding of K_j, we still have in each cluster B and D at least $r\eta/8\cdot|B|$ and $r\eta/8\cdot|D|$ vertices, respectively, outside U, W, $\tilde{U}$, and $\tilde{W}$. We continue until every $K\in F-R$ is embedded. □

4.2.5 Proof of Proposition 4.11

Given $\delta, q, d > 0$ and $\tilde{r}\leq r'\leq 1/2$ set

$$\begin{aligned}
\varepsilon &:= \min\left\{\varepsilon_{P4.18}(\frac{q\delta}{20}, d, r'), \varepsilon_{P4.19}(\frac{q\delta}{20}, d, r'), \left(\frac{\delta q}{3}\right)^2, d/17\right\},\\
\beta &:= \min\left\{\beta_{P4.18}(\frac{q\delta}{20}, r', \varepsilon, \tilde{N}_{max}), \beta_{P4.19}(\frac{q\delta}{20}, r', \varepsilon, \tilde{N}_{max}), \delta r'/8\right\},\\
n_0 &:= \frac{200}{\delta q r'\beta}.
\end{aligned}$$

We gradually construct an injective homomorphism φ of T into H. To this end we consider the four introduced cases.

In each case, we start by embedding the vertices of W_A and W_B to ultratypical vertices of A and B, respectively. This can be done by applying Lemma 2.14 with $X'_{L2.14}$ and $Y'_{L2.14}$ being the sets of ultratypical vertices of A and B, respectively, $T_{L2.14}$ being any tree with colour classes W_A and W_B such that $T[W_A\cup W_B]$ is a subgraph of $T_{L2.14}$, $\alpha_{L2.14} = 5\varepsilon$ and $R_{L2.14} = \emptyset$. Note that the assumptions of Lemma 2.14 are satisfied, since the pair (A,B) has density at least $d-\varepsilon > 15\varepsilon$, by Lemma 2.7 at least $1-\sqrt{\varepsilon} > 4/5$ of vertices of A or B, respectively, are ultratypical, and moreover $|W_A| < \varepsilon|A|$, $|W_B| < \varepsilon|B|$ by definition of fine partition.

We embed the rest of the tree T using different strategy for each case. In what follows, we use indexes 1 and 2 to denote that the structure is a substructure of T_1 or T_2, respectively.

When using Propositions 4.18 and 4.19, we shall always use (here we use the index $_P$ to indicate the parameter of the propositions) $d_P := d$, $r_P := r'$, $\tilde{N}_{max,P} := \tilde{N}_{max}$, $n_P := n$, $H_P := H$, $\mathbf{H}_P := \mathbf{H}$, and R_P will be either W_A or W_B depending whether we embed part of $\mathcal{D}_A$, or $\mathcal{D}_B$, respectively. In some cases, we shall use Proposition 4.19 several times. To avoid confusion, we shall use upper indices in parenthesis, e.g., $U^{(1)}_{P4.19}$, to indicate to which application of the proposition we refer. We will write $\mathcal{D}_{B1}$ as a shortcut for $\mathcal{D}_B\cap V(T_1)$ and $\mathcal{D}_{B2} := \mathcal{D}_B\cap V(T_2)$ and $\mathcal{D}_{A1}$ as a shortcut for $\mathcal{D}_A\cap V(T_2)$ (sic) and $\mathcal{D}_{A2} := \mathcal{D}_A\cap V(T_1)$. Thus, neighbours of W_A or W_B are in $\mathcal{D}_{A1}$ or $\mathcal{D}_{A2}$, respectively.

Case A

In this case we assume that there are two adjacent clusters A and B in H such that $\overline{\deg}(A, \mathcal{S}_1\cup\mathcal{S}_M) \geq \frac{1-\tilde{r}}{\tilde{r}}|\mathcal{D}_{A2}| + \delta k$ and $\overline{\deg}(B,\mathcal{L}) \geq (\tilde{r}+\delta)k$.

We start by embedding the vertices of W_A and W_B to ultratypical vertices of clusters A and B, respectively. We then further partition the rest of T and embed it in the

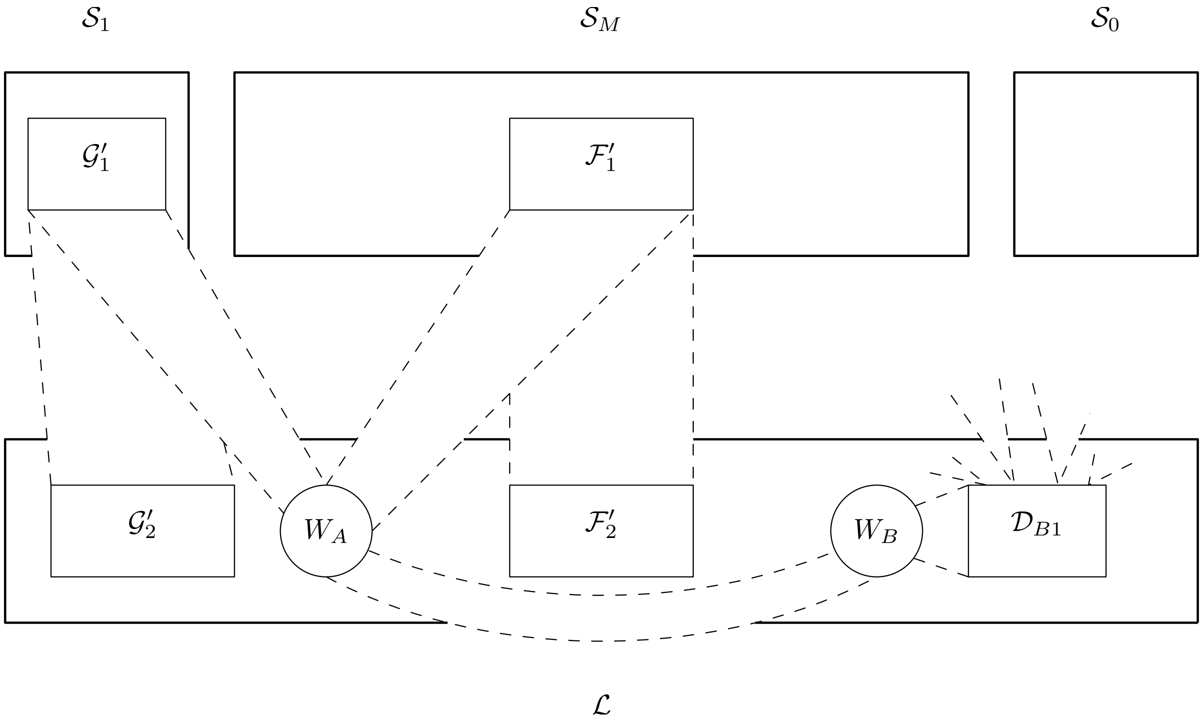

Figure 4.4: The embedding configuration in the case A. After inserting the vertices of W_A, W_B in the ultratypical vertices of clusters A and B we use Proposition 4.18 to embed $\mathcal{F}'$ in the matching connecting $\mathcal{S}_M := \mathcal{S} \cap M$ and $\mathcal{L} \cap M$. Then we invoke Proposition 4.19 to embed $\mathcal{G}'$ using the vertices in $\mathcal{S}_1$. Finally, we again invoke Proposition 4.19 to embed $\mathcal{D}_B$. Note that in this case, as well as in all of the subsequent cases, it may be the case that $B \in \mathcal{S}_1$.

following three steps which we describe in detail later. We partition the trees from $\mathcal{D}_A$ in two sets – $\mathcal{F}$ and $\mathcal{G}$ and define $\mathcal{F}'$ and $\mathcal{G}'$ as sets of subtrees of $\mathcal{F}$ and $\mathcal{G}$, respectively, with leaves in $\mathcal{D}_{A1}$ removed. We denote $\mathcal{F} \cap \mathcal{D}_{Ai}$ and $\mathcal{G} \cap \mathcal{D}_{Ai}$ by $\mathcal{F}_i$ and $\mathcal{G}_i$ respectively, for $i = 1, 2$. Analogously, we define $\mathcal{F}_i'$ and $\mathcal{G}_i'$ for $i = 1, 2$.

In the first step, we embed $\mathcal{F}'$ into the edges of the matching $\mathbf{M}$ using Proposition 4.18 and we embed $\mathcal{G}'$ through $\mathcal{S}_1$ vertices using Proposition 4.19 (i.e., $\varphi(\mathcal{G}_1') \subseteq \mathcal{S}_1$ and $\varphi(\mathcal{G}_2') \subseteq \mathcal{L}$).

In the second step, we embed the trees from $\mathcal{D}_{B1}$ using again Proposition 4.19. To this end we again use the bound on the degree of the cluster B – specifically, as $\overline{\deg}(B, \mathcal{L}) \geq \tilde{r}k + \delta k = |\mathcal{D}_{A2} \cup \mathcal{D}_{B1}| + \delta k$, the cluster B has enough neighbours for embedding $\mathcal{D}_{B1}$, even though $\mathcal{D}_{A2}$ is already embedded.

In the third step, we embed $\mathcal{F} \setminus \mathcal{F}'$ and $\mathcal{G} \setminus \mathcal{G}'$ greedily. The structure of the embedded tree is sketched in Figure 4.4.

1. In this step we embed the trees from the anchored forest $\mathcal{D}_A$ except of several leaves, ensuring that the neighbours of those left-out leaves are mapped to ultra-typical vertices in $\mathcal{L}$-clusters. We split the anchored βk-forest $\mathcal{D}_A$ into two disjoint forests $\mathcal{F}$ and $\mathcal{G}$ in the following way. Let $\mathcal{F}$ be a maximal subset of trees of $\mathcal{D}_A$ such that

$$|\mathcal{F}_2| \leq \frac{r'}{1-r'}\overline{\deg}(A, \mathcal{S}_M) - \frac{r'}{1-r'}\delta k/2, \tag{4.16}$$

and we choose it as an empty set if the size of the expression is less than zero. This means that if $\mathcal{G}$ is non-empty then

$$|\mathcal{F}_2| \geq \frac{r'}{1-r'}\overline{\deg}(A,\mathcal{S}_M) - \frac{r'}{1-r'}\delta k/2 - \beta k, \tag{4.17}$$

otherwise we could move a suitable tree from $\mathcal{G}$ to $\mathcal{F}$ while retaining the condition imposed on $\mathcal{F}$. By deleting the leaves of trees in $\mathcal{D}_A$ that are contained in $\mathcal{F}_1 \cup \mathcal{G}_1$ we get forests $\mathcal{F}'$ and $\mathcal{G}'$. For each tree $K \in \mathcal{F}' \cup \mathcal{G}'$ we have $|K \cap (\mathcal{F}'_1 \cup \mathcal{G}'_1)| \leq |K \cap (\mathcal{F}'_2 \cup \mathcal{G}'_2)|$, because each vertex from $K \cap (\mathcal{F}'_1 \cup \mathcal{G}'_1)$ has at least one child in $(\mathcal{F}'_2 \cup \mathcal{G}'_2)$. Specifically, $|\mathcal{G}'_1| \leq |\mathcal{G}'_2|$.

Now we apply Proposition 4.18 to our anchored forest $F_{P4.18} := \mathcal{F}'$ if it is non-empty. Set $U_{P4.18} := \varphi(W_A \cup W_B)$, $\eta_{P4.18} := q\delta/4$, $\mathbf{M}_{P4.18} := \mathbf{M}$, and $A_{P4.18} := A$. From Definition 2.8 we know that $|U_{P4.18}| = |W_A \cup W_B| \leq 12k/(\beta k) = 12/\beta$.

To apply the proposition it suffices to verify that the degree of A in $\mathcal{S}_M$ is sufficiently large, as by definition of $\mathcal{F}'$ we know that for each $K \in \mathcal{D}_A$ we have $|K \cap \mathcal{F}'_1| \leq |K \cap \mathcal{F}'_2|$. We have

$$\begin{aligned}
\overline{\deg}(A,\mathcal{S}_M) &\geq \frac{1-r'}{r'}|\mathcal{F}_2| + \delta k/2 \\
&\geq \frac{1-r'}{r'}|\mathcal{F}_2| + \frac{1-r'}{r'}|U_{P4.18}| + \delta k/4 \\
&\geq \frac{1-r'}{r'}|\mathcal{F}'_2| + \sum_{C\subseteq\mathcal{S}\,:\,CD\in\mathbf{M}} \max\{|U_{P4.18}\cap C|, \frac{1-r'}{r'}|U_{P4.18}\cap D|\} + \eta_{P4.18}n\ ,
\end{aligned}$$

where the first inequality is due to the definition of $\mathcal{F}$ (bound 4.16) and the second one and third one are due to the facts that $\delta k/4 \geq \frac{1-r'}{r'}|U_{P4.18}|$ (from the choice of n_0) and $\delta k/4 \geq \eta_{P4.18}n$ (from the choice of $\eta_{P4.18}$).

If $\mathcal{G}$ is non-empty and, thus, the bound 4.17 holds, we proceed by embedding $\mathcal{G}'$.

We apply Proposition 4.19 (Configuration 2) to the anchored forest $\mathcal{G}'$ and $\mathcal{B}^{(1)}_{P4.19} := \mathcal{S}_1$. As we know that $N_{\mathbf{H}}(\mathcal{S}_1)$ is disjoint from $\mathcal{S}_M$, there is no need to include $\varphi(\mathcal{F}_1) \subseteq \mathcal{S}_M$ in the forbidden set U that ensures the injectiveness of φ. Thus, we set $U^{(1)}_{P4.19} := \varphi(\mathcal{F}_2 \cup W_A \cup W_B)$.

Also note that $\bigcup \mathcal{B}^{(1)}_{P4.19} \cap U^{(1)}_{P4.19} \subseteq \varphi(W_A \cup W_B)$, because $\varphi(\mathcal{F}_2) \in \mathcal{L}$ (we could actually replace $W_A \cup W_B$ by W_B). Let $\eta^{(1)}_{P4.19} := \delta q/4$, and $A^{(1)}_{P4.19} := A$. Now we verify the first condition from Proposition 4.19. For the degree of the cluster A in $\mathcal{S}_1$ we have

$$\begin{aligned}
\overline{\deg}(A,\mathcal{S}_1) &= \overline{\deg}(A,\mathcal{S}_1\cup\mathcal{S}_M) - \overline{\deg}(A,\mathcal{S}_M) \\
\boxed{\text{assumption of this configuration}}\quad &\geq \frac{1-\tilde{r}}{\tilde{r}}|\mathcal{D}_{A2}| + \delta k - \overline{\deg}(A,\mathcal{S}_M) \\
\boxed{\tilde{r}\leq r',\ \text{bound (4.17)}}\quad &\geq \frac{1-r'}{r'}|\mathcal{F}_2\cup\mathcal{G}_2| + \delta k - \frac{1-r'}{r'}|\mathcal{F}_2| - \delta k/2 - \frac{1-r'}{r'}\beta k \\
\boxed{\text{bounding error terms}}\quad &\geq \frac{1-r'}{r'}|\mathcal{G}_2| + 3\delta k/8 \\
\boxed{|\mathcal{G}_2|\geq|\mathcal{G}'_2|\geq|\mathcal{G}'_1|}\quad &\geq |\mathcal{G}'_2| + 3\delta k/8 \\
&\geq |\mathcal{G}'_1| + |\bigcup \mathcal{B}^{(1)}_{P4.19} \cap U^{(1)}_{P4.19}| + \eta^{(1)}_{P4.19}n\ ,
\end{aligned}$$

where we at first used the lower bound on the degree of A in $\mathcal{S}_1 \cup \mathcal{S}_M$, then the lower bound on the size of $\mathcal{F}_2$, after bounding the error terms we used the fact that $|\mathcal{G}_2'| \geq |\mathcal{G}_1'|$ and then we again bounded the errors terms by using the facts that $|\bigcup \mathcal{B}^{(1)}_{P4.19} \cap U^{(1)}_{P4.19}| \leq |W_A \cup W_B| \leq \delta k/8$ and $\eta^{(1)}_{P4.19} n \leq \delta k/4$.

Further we verify that for each cluster $C \in \mathcal{S}_1$ we have

$$\begin{aligned}\overline{\deg}(C, V(\mathbf{H}) \setminus \mathcal{B}^{(1)}_{P4.19}) = \overline{\deg}(C, \mathcal{L}) &\geq \tilde{r}k + \delta k \\ \boxed{\text{bound on the skew of } T} \quad &\geq |\mathcal{D}_{A2}| + |\varphi(W_A \cup W_B)| + \eta^{(1)}_{P4.19} n \\ &= |\mathcal{F}_2| + |\mathcal{G}_2| + |\varphi(W_A \cup W_B)| + \eta^{(1)}_{P4.19} n \\ &\geq |\mathcal{G}_2'| + |U^{(1)}_{P4.19}| + \eta^{(1)}_{P4.19} n \ ,\end{aligned}$$

Thus we can extend φ to $\mathcal{G}$. Note that $\varphi(\mathcal{G}_2) \subseteq \mathcal{L}$.

2. In this step we embed the trees from $\mathcal{D}_B$ using Configuration 1 from Proposition 4.19. The appropriate set $U^{(2)}_{P4.19}$ guaranteeing the injectiveness of φ consists of $\varphi(\mathcal{F}_1' \cup \mathcal{F}_2' \cup \mathcal{G}_1' \cup \mathcal{G}_2' \cup W_A \cup W_B)$. We set $\mathcal{B}^{(2)}_{P4.19} := \mathcal{L}$, $\eta^{(2)}_{P4.19} := \delta q/2$ and $A^{(2)}_{P4.19} := B$. First we verify the first condition of the proposition. We have

$$\begin{aligned}\overline{\deg}(B, \mathcal{L}) &\geq \tilde{r}k + \delta k \\ \boxed{\text{bound on the skew of } T} \quad &= |\mathcal{F}_2 \cup \mathcal{G}_2 \cup \mathcal{D}_{B1}| + \delta k \\ \boxed{\text{bounding error terms}} \quad &\geq |\varphi(\mathcal{F}_2 \cup \mathcal{G}_2 \cup W_A \cup W_B)| + |\mathcal{D}_{B1}| + \delta k/2 \\ &\geq |\bigcup \mathcal{B}^{(2)}_{P4.19} \cap U^{(2)}_{P4.19}| + |\mathcal{D}_{B1}| + \eta^{(2)}_{P4.19} n \ ,\end{aligned}$$

We immediately use the 'moreover' part of the proposition with $\tilde{U}^{(2)}_{P4.19} = \emptyset$ and verify that for each L-cluster C we have

$$\begin{aligned}\overline{\deg}(C) &\geq k + \delta k \\ &\geq |\mathcal{D}_{B1}| + |\mathcal{D}_{B2}| + |\mathcal{D}_A| + |W_A \cup W_B| + \delta k \\ &\geq |\mathcal{D}_{B1}| + |\mathcal{D}_{B2}| + |U^{(2)}_{P4.19}| + \eta^{(2)}_{P4.19} n \ ,\end{aligned}$$

where we use mainly the fact that $|\mathcal{D}_A \cup \mathcal{D}_B \cup W_A \cup W_B| = k$.

3. We have defined an injective homomorphism φ of the whole tree T except of its leaves from $\mathcal{F}_1 \setminus \mathcal{F}_1'$ and $\mathcal{G}_1 \setminus \mathcal{G}_1'$. We know that their neighbours are embedded in ultratypical vertices of L-clusters. By Proposition 4.9, such vertices have degree at least $k + \delta k - 2\sqrt{\varepsilon}n/r' \geq k$ as $\delta q > 2\sqrt{\varepsilon}/r'$. Thus we can greedily extend φ to the whole tree T.

Case B

In this case we assume that $\tilde{r}|\mathcal{D}_{A1}| \geq (1-\tilde{r})|\mathcal{D}_{A2}|$ and that there are two adjacent clusters A, B such that $\overline{\deg}(A, \mathcal{S}_1 \cup \mathcal{S}_M \cup \mathcal{L}) \geq (1+\delta)k$ and $\overline{\deg}(B, \mathcal{L}) \geq (\tilde{r}+\delta)k$. The embedding procedure is roughly similar to the one from Case A. However, for embedding $\mathcal{D}_A$ we now also use $\mathcal{L}$.

We start by embedding certain part of the anchored forest $\mathcal{D}_A$ using the matching $\mathbf{M}$ and the set $\mathcal{S}_1$ similarly to the Case A. Then, we proceed by reserving $|\mathcal{D}_{B1}|$ vertices

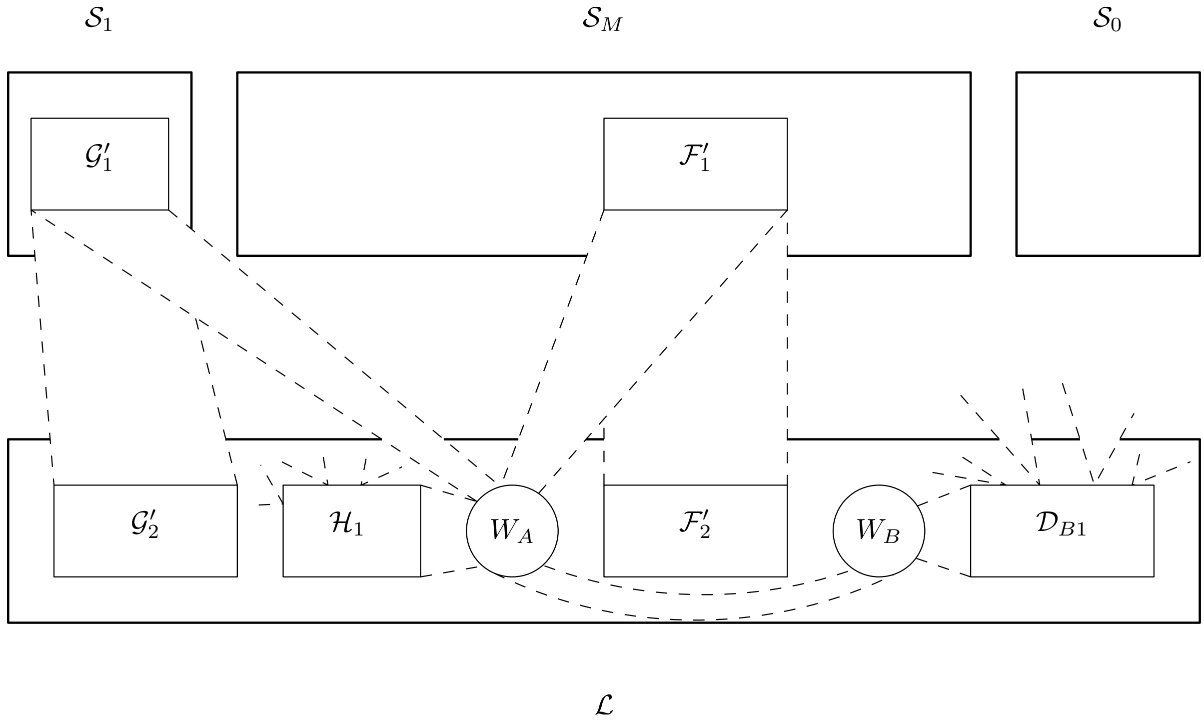

Figure 4.5: The embedding configuration in the case B. After inserting the vertices of W_A, W_B in the ultratypical vertices of clusters A and B we use Proposition 4.18 to embed $\mathcal{F}'$ in the matching $\mathbf{M}$. Then we invoke Proposition 4.19 to embed $\mathcal{G}'$ using the vertices in $\mathcal{S}_1$. Then we reserve suitable vertices in the neighbourhood of the cluster B in $\mathcal{L}$ that will later serve for embedding of $\mathcal{D}_{B1}$ using Proposition 4.19. Then we embed $\mathcal{H}_1$ using the same proposition and finally we embed $\mathcal{D}_B$ through the reserved vertices.

that will later help us to embed the anchored trees from $\mathcal{D}_B$. In the third part we embed the rest of the forest $\mathcal{D}_A$ using the high degree vertices in $\mathcal{L}$, and then proceed by embedding $\mathcal{D}_B$ using the reserved vertices. Finally, we argue that we can embed several leftover leaves of the tree as in the previous case.

1. Analogously to the preceding case we split the anchored forest $\mathcal{D}_A$ into three disjoint sets $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2$, $\mathcal{G} = \mathcal{G}_1 \cup \mathcal{G}_2$, and $\mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2$ in the following way.

   Let $K_1, K_2, \ldots$ be the trees of $\mathcal{D}_A$ sorted according to their skew, i.e., according to the ratio $|K_i \cap V(T_2)|/|K_i \cap V(T_1)|$ in descending order. We define $\mathcal{F}$ as the union $K_1 \cup \cdots \cup K_j$, where j is taken to be maximal such that

$$|\mathcal{F}_2| = \sum_{i=1}^{j} |K_i \cap V(T_1)| \leq \frac{r'}{1-r'}\overline{\deg}(A, \mathcal{S}_M) - \frac{r'}{1-r'}\delta k/3. \tag{4.18}$$

   If the right hand side is less than zero, define $\mathcal{F}$ as the empty set. Then we similarly define $\mathcal{G}$ as the union of trees $K_{j+1}, \ldots, K_{j'}$ where j' is maximal such that

$$|\mathcal{G}_2| = \sum_{i=j+1}^{j'} |K_i \cap V(T_1)| \leq \frac{r'}{1-r'}\overline{\deg}(A, \mathcal{S}_1) - \frac{r'}{1-r'}\delta k/3. \tag{4.19}$$

   Finally we set $\mathcal{H} = \mathcal{D}_A \setminus (\mathcal{F} \cup \mathcal{G})$.

As before, we have

$$|\mathcal{F}_2| \geq \frac{r'}{1-r'}\overline{\deg}(A,\mathcal{S}_M) - \frac{r'}{1-r'}\delta k/3 - \beta k \ , \tag{4.20}$$

if $\mathcal{F} \neq \mathcal{D}_A$ and

$$|\mathcal{G}_2| \geq \frac{r'}{1-r'}\overline{\deg}(A,\mathcal{S}_1) - \frac{r'}{1-r'}\delta k/3 - \beta k \ , \tag{4.21}$$

if $\mathcal{F} \cup \mathcal{G} \neq \mathcal{D}_A$. Additionally, we also have

$$\tilde{r}|\mathcal{F}_1 \cup \mathcal{G}_1| \geq (1-\tilde{r})|\mathcal{F}_2 \cup \mathcal{G}_2| \ , \tag{4.22}$$

because of the assumption $\tilde{r}|\mathcal{D}_{A1}| \geq (1-\tilde{r})|\mathcal{D}_{A2}|$ and the fact that in $\mathcal{F} \cup \mathcal{G}$ there are the anchored trees with biggest skew.

We define $\mathcal{F}'$ and $\mathcal{G}'$ as in the previous case. We have $|K \cap \mathcal{F}'_1| \leq |K \cap \mathcal{F}'_2|$ for each $K \in \mathcal{F}'$ and $|\mathcal{G}'_1| \leq |\mathcal{G}'_2|$.

If $\mathcal{F}'$ is non-empty we apply Proposition 4.18 to embed the anchored forest $F_{P4.18} := \mathcal{F}'$ in the same way as in the previous case. Set $U_{P4.18} = \varphi(W_A \cup W_B)$, $\eta_{P4.18} = \delta q/4$, $r_{P4.18} := r'$, $\mathbf{M}_{P4.18} := \mathbf{M}$, and $A_{P4.18} := A$. Similarly to the previous case we verify that

$$\begin{aligned}\overline{\deg}(A,\mathcal{S}_M) &\geq \frac{1-r'}{r'}|\mathcal{F}_2| + \delta k/3 \\ &\geq \frac{1-r'}{r'}|\mathcal{F}_2| + \frac{1-r'}{r'}|U_{P4.18}| + \delta k/4 \\ &\geq \frac{1-r'}{r'}|\mathcal{F}'_2| + \sum_{C\subseteq\mathcal{S}\,:\,CD\in\mathbf{M}} \max\{|U_{P4.18}\cap C|, \frac{1-r'}{r'}|U_{P4.18}\cap D|\} + \eta_{P4.18} n \ .\end{aligned}$$

If $\mathcal{G}$ is non-empty we proceed by embedding $\mathcal{G}'$. This is also done in an analogous way to the preceding case.

We apply Proposition 4.19 (Configuration 2) to the anchored forest $F^{(1)}_{P4.19} := \mathcal{G}'$ and set $\mathcal{B}^{(1)}_{P4.19} := \mathcal{S}_1$. By the properties of a skew LKS graph, the set $N_{\mathbf{H}}(\mathcal{S}_1) \cup \mathcal{S}_1$ is disjoint from $\mathcal{S}_M \supseteq \varphi(\mathcal{F}'_1)$, thus for ensuring injectiveness of φ it suffices to set $U^{(1)}_{P4.19} := \varphi(\mathcal{F}'_2 \cup W_A \cup W_B)$ and then we also have $\bigcup \mathcal{B}^{(1)}_{P4.19} \cap U^{(1)}_{P4.19} \subseteq W_A \cup W_B$. Set $\eta^{(1)}_{P4.19} := \delta q/4$, and $A^{(1)}_{P4.19} := A$.

Now we verify the first condition from the proposition. For the degree of the cluster A in $\mathcal{S}_1$ we have

$$\begin{aligned}\overline{\deg}(A,\mathcal{S}_1) &\geq \frac{1-r'}{r'}|\mathcal{G}_2| + \delta k/3 \\ &\geq |\mathcal{G}'_1| + |\bigcup \mathcal{B}^{(1)}_{P4.19} \cap U^{(1)}_{P4.19}| + \eta^{(1)}_{P4.19} n \ ,\end{aligned}$$

where we use the definition of $\mathcal{G}$, the fact that $|\mathcal{G}_2| \geq |\mathcal{G}'_1|$ and the fact that $|\bigcup \mathcal{B}^{(1)}_{P4.19} \cap U^{(1)}_{P4.19}| \leq \delta q/12$.

Further, we verify that for each cluster $C \in \mathcal{S}_1$ we have

$$\begin{aligned}\overline{\deg}(C, V(\mathbf{H}) \setminus \mathcal{B}^{(1)}_{P4.19}) = \overline{\deg}(C,\mathcal{L}) &\geq \tilde{r}k + \delta k \\ &\geq |\mathcal{D}_{A2}| + |\varphi(W_A \cup W_B)| + \delta k/2 \\ &\geq |\mathcal{F}_2| + |\mathcal{G}_2| + |\varphi(W_A \cup W_B)| + \delta k/2 \\ &\geq |\mathcal{G}'_2| + |U^{(1)}_{P4.19}| + \eta^{(1)}_{P4.19} n \ ,\end{aligned}$$

where we used the facts that $|\mathcal{D}_{A2}| \leq \tilde{r}k$ and bounded the error terms in the usual manner.

2. In this step we reserve suitable vertices for embedding $\mathcal{D}_B$ and use Proposition 4.19, Configuration 1, to this end.

   We apply the proposition the anchored forest $F^{(2)}_{P4.19} := \mathcal{D}_B$, $A^{(2)}_{P4.19} := B$, the set $U^{(2)}_{P4.19} := \varphi(W_A \cup W_B \cup \mathcal{F}_2 \cup \mathcal{G}_2)$, and $\mathcal{B}^{(2)}_{P4.19} := \mathcal{L}$. Take $\eta^{(2)}_{P4.19} := q\delta/20$. We start by verifying the first condition:

$$\begin{aligned}\overline{\deg}(B,\mathcal{L}) &\geq \tilde{r}k + \delta k\\ &\geq |\mathcal{F}_2 \cup \mathcal{G}_2 \cup \mathcal{H}_2 \cup \mathcal{D}_{B1}| + \delta k\\ &\geq |\mathcal{F}_2 \cup \mathcal{G}_2 \cup \mathcal{D}_{B1}| + \delta k\\ &\geq |\mathcal{D}_{B1}| + |\varphi(W_A \cup W_B \cup \mathcal{F}_2 \cup \mathcal{G}_2)| + \delta k/2\\ &\geq |\mathcal{D}_{B1}| + |U^{(2)}_{P4.19}| + \eta^{(2)}_{P4.19} n\ ,\end{aligned}$$

   where we use the upper bound on the smaller colour class of T and then we bound the error terms as usual. This gives us an embedding of $N(W_B) \cap \mathcal{D}_B$ as well as the reservation set W that will help us later for embedding $\mathcal{D}_{B1}$.

   Before finishing the embedding of $\mathcal{D}_B$ by invoking the 'moreover' part of Proposition 4.19, Configuration 1, we shall embed the anchored forest $\mathcal{H}$, which will define the set $\tilde{U}^{(2)}_{P4.19} := \varphi(\mathcal{H})$.

3. We proceed with embedding of $F^{(3)}_{P4.19} := \mathcal{H}$, using a third time Proposition 4.19, Configuration 1. Let $U' = \varphi(N(W_B) \cap \mathcal{D}_B) \cup W$, $|U'| = |\mathcal{D}_{B1}|$ and set $U^{(3)}_{P4.19} := \varphi(W_A \cup W_B \cup \mathcal{F} \cup \mathcal{G}) \cup U'$. Thus $U^{(3)}_{P4.19} \cap \mathcal{L} \subseteq \varphi(W_A \cup W_B \cup \mathcal{F}'_2 \cup \mathcal{G}'_2) \cup U'$. Further set $\mathcal{B}^{(3)}_{P4.19} := \mathcal{L}$, $\eta^{(3)}_{P4.19} := \delta q/4 \geq \eta^{(2)}_{P4.19}$, and $A^{(3)}_{P4.19} := A$. We verify the first condition of the proposition:

$$\begin{aligned}&\overline{\deg}(A,\mathcal{L})\\ &\geq k + \delta k - \overline{\deg}(A,\mathcal{S}_M) - \overline{\deg}(A,\mathcal{S}_1)\\ \text{bounds (4.20) and (4.21)}\quad &\geq k + \delta k - \Big(\frac{1-r'}{r'}|\mathcal{F}_2| + \delta k/3 + \frac{1-r'}{r'}\beta k\Big) -\\ &\quad \Big(\frac{1-r'}{r'}|\mathcal{G}_2| + \delta k/3 + \frac{1-r'}{r'}\beta k\Big)\\ \text{bounding error terms}\quad &\geq k - \frac{1-r'}{r'}(|\mathcal{F}_2| + |\mathcal{G}_2|) + \delta k/4\\ \tilde{r} \leq r'\quad &\geq k - \frac{1-\tilde{r}}{\tilde{r}}(|\mathcal{F}_2| + |\mathcal{G}_2|) + \delta k/4\\ \text{bound (4.22)}\quad &\geq k - (|\mathcal{F}_1| + |\mathcal{G}_1|) + \delta k/4\\ T \text{ is of size } k\quad &\geq |\mathcal{F}_2| + |\mathcal{G}_2| + |\mathcal{H}| + |\mathcal{D}_B| + |W_A \cup W_B| + \delta k/4\\ &\geq |\mathcal{H}_1| + |\varphi(W_A \cup W_B \cup \mathcal{F}'_2 \cup \mathcal{G}'_2)| + |\mathcal{D}_{B1}| + \delta k/4\\ &\geq |\mathcal{H}_1| + |U^{(3)}_{P4.19} \cap \bigcup \mathcal{B}^{(3)}_{P4.19}| + \eta^{(3)}_{P4.19} n\ ,\end{aligned}$$

   where we at first used our bounds on $|\mathcal{F}_2|$ and $|\mathcal{G}_2|$. Then we used the inequality $\tilde{r}|\mathcal{F}_1 \cup \mathcal{G}_1| \geq (1-\tilde{r})|\mathcal{F}_2 \cup \mathcal{G}_2|$, we followed by interpreting k as the size of T and used trivial bounds on error term throughout the computation.

We immediately use the second part of the proposition with $\tilde{U}^{(3)}_{P4.19} = \emptyset$. We verify that for each $C \in \mathcal{L}$ we have

$$\begin{aligned}\overline{\deg}(C) &\geq k + \delta k \\ &= |\mathcal{F} \cup \mathcal{G} \cup \mathcal{H} \cup \mathcal{D}_B \cup W_A \cup W_B| + \delta k \\ &\geq |\mathcal{H}_1| + |\mathcal{H}_2| + |U^{(3)}_{P4.19} \cup \tilde{U}^{(3)}_{P4.19}| + \eta^{(3)}_{P4.19} n \,.\end{aligned}$$

Thus, we can extend φ to $\mathcal{H}$. Note that $|C \setminus (U \cup U' \cup \tilde{U})| \geq r'\eta^{(3)}_{P4.19}|C|/8$ for each cluster C with $C \cap \varphi(\mathcal{H})$.

4. Now, we finish up the embedding of $\mathcal{D}_B$, using the 'moreover' part of the second application of Proposition 4.19. The first condition of the proposition is satisfied, as for each $C \in \mathcal{L}$ we have

$$\begin{aligned}\overline{\deg}(C) &\geq k + \delta k \\ &= |\mathcal{D}_A \cup \mathcal{D}_B \cup W_A \cup W_B| + \delta k \\ &\geq |\mathcal{D}_{B1}| + |\mathcal{D}_{B2}| + |\varphi(W_A \cup W_B \cup \mathcal{F}' \cup \mathcal{G}' \cup \mathcal{H})| + \eta^{(2)}_{P4.19} n \\ &\geq |\mathcal{D}_{B1}| + |\mathcal{D}_{B2}| + |U^{(2)}_{P4.19} \cup \tilde{U}^{(2)}_{P4.19}| + \eta^{(2)}_{P4.19} n \,.\end{aligned}$$

The second condition is that for each cluster C with $C \cap \varphi(\mathcal{H})$ we have $|C \setminus (U \cup U' \cup \tilde{U})| \geq r'\eta^{(2)}_{P4.19}|C|/8$. This is satisfied as $\eta^{(3)}_{P4.19} \geq \eta^{(2)}_{P4.19}$ and by the property of the embedding of $\mathcal{H}$, guaranteed by the third application of Proposition 4.19.

5. We have defined an injective homomorphism φ on the whole tree T except of its leaves from $\mathcal{F}_1 \setminus \mathcal{F}'_1$ and $\mathcal{G}_1 \setminus \mathcal{G}'_1$. As we know that their neighbours are embedded in ultratypical vertices of L-clusters, we can greedily extend the embedding to the whole tree T, as in Case A.

Case C

In this case we assume that $\tilde{r}|\mathcal{D}_{A1}| \leq (1-\tilde{r})|\mathcal{D}_{A2}|$ and that there are adjacent clusters A and B such that $\overline{\deg}(A, \mathcal{S}_1 \cup \mathcal{S}_M \cup \mathcal{L}) \geq (1+\delta)k$ and $\overline{\deg}(B, \mathcal{L}) \geq |\mathcal{D}_{B2}| + \delta k = |\mathcal{D}_{B1}| + \delta k$. The embedding procedure is very similar to the one from the preceding case, the difference being in the order in which we embed the parts of T in the host graph.

We start by reserving vertices for the embedding of the anchored forest $\mathcal{D}_B$ using Proposition 4.19. Then we embed parts of $\mathcal{D}_A$ using the matching $\mathbf{M}$ and $\mathcal{S}_1$ as in the previous cases. We have to be more careful, though, as the vertices reserved for $\mathcal{D}_B$ can cover substantial part of $\mathbf{M}$. We finish by embedding the rest of $\mathcal{D}_A$ through high degree L-clusters using Proposition 4.19.

1. We start by reserving vertices for embedding the anchored forest $\mathcal{D}_B$ such that $\mathcal{D}_{B1} := \mathcal{D}_B \cap V(T_1)$ will be embedded in the neighbourhood of the cluster B. Set $\mathcal{B}^{(1)}_{P4.19} := \mathcal{L}$ and $U^{(1)}_{P4.19} := \varphi(W_A \cup W_B)$. Set $\eta^{(1)}_{P4.19} := q\delta/20$, and $A^{(1)}_{P4.19} := B$. We apply Proposition 4.19, Configuration 1, to reserve vertices in $\mathcal{L}$ that will later serve for embedding of $\mathcal{D}_B$. We verify that the first condition of the proposition is satisfied. Indeed:

$$\begin{aligned}\overline{\deg}(B, \mathcal{L}) &\geq |\mathcal{D}_{B1}| + \delta k \\ &\geq |\mathcal{D}_{B1}| + |\varphi(W_A \cup W_B)| + \eta^{(1)}_{P4.19} n \,,\end{aligned}$$

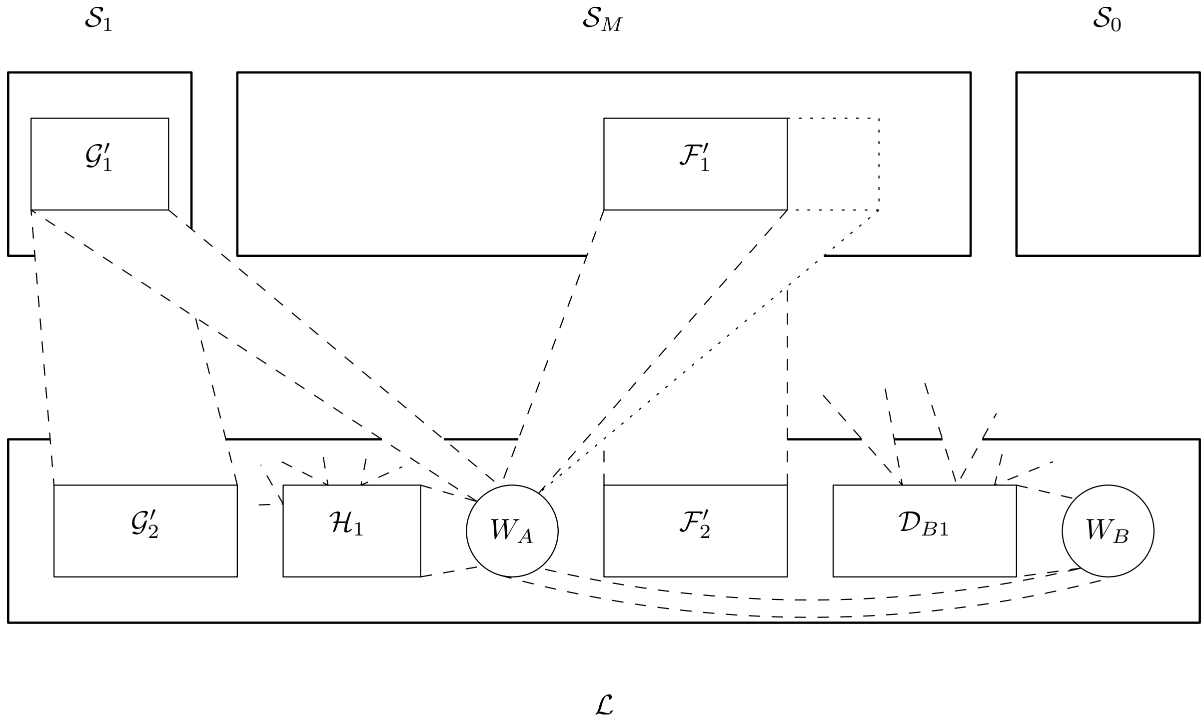

Figure 4.6: The embedding configuration in the case C. The configuration is very similar to the preceding one from case B. However, in this case we start by embedding $\mathcal{D}_B$ in the neighbourhood of the cluster B. The figure suggests that because of the vertices reserved for $\mathcal{D}_{B1}$ we must be more careful in the application of Proposition 4.18 and add those vertices in the forbidden set $U_{P4.18}$.

where we used the standard error estimation.

This gives us embedding of $N(W_B) \cap \mathcal{D}_B$ as well as a reserved set W. We set $U' = \varphi(N(W_B) \cap \mathcal{D}_B) \cup W$, $|U'| = |\mathcal{D}_{B1}|$. After embedding the whole T except of several of its leaf neighbours, we will invoke the second part of the proposition with $\tilde{U}^{(1)} = \varphi(\mathcal{F}' \cup \mathcal{G}' \cup \mathcal{H})$ where $\mathcal{F}' \cup \mathcal{G}' \cup \mathcal{H} \subseteq \mathcal{D}_A$. Note that if we set $\tilde{U}^{(1)}$ to such value, we will satisfy the first condition needed for the actual embedding of $\mathcal{D}_B$, because for any cluster $C \in \mathcal{L}$ we have

$$\begin{aligned}\overline{\deg}(C) &\geq k + \delta k \\ &= |\mathcal{D}_A \cup \mathcal{D}_B \cup W_A \cup W_B| + \delta k \\ &\geq |\mathcal{D}_{B1}| + |\mathcal{D}_{B2}| + |U^{(1)}_{P4.19} \cup \tilde{U}^{(1)}| + \eta^{(1)}_{P4.19} n \ .\end{aligned}$$

To satisfy the second condition we will ensure that for all subsequent applications of Propositions 4.18 and 4.19 we choose the value η being greater than $\eta^{(1)}_{P4.19}$.

2. We now proceed by embedding the anchored forest $\mathcal{D}_A$ analogously to the previous case. We split the forest $\mathcal{D}_A$ into three (possibly empty) forests $\mathcal{F}, \mathcal{G}, \mathcal{H}$ in such a way that $\mathcal{F}$ is maximal with

$$|\mathcal{F}_2| \leq \frac{r'}{1-r'}\overline{\deg}(A, \mathcal{S}_M) - |U'| - \frac{r'}{1-r'}\delta k/3 \ , \tag{4.23}$$

or $\mathcal{F}$ is empty if the value of right hand side is smaller then zero. Moreover, if

$\mathcal{F} \neq \mathcal{D}_A$, we have

$$|\mathcal{F}_2| \geq \frac{r'}{1-r'}\overline{\deg}(A, \mathcal{S}_M) - |U'| - \frac{r'}{1-r'}\delta k/3 - \beta k \ . \tag{4.24}$$

Then we similarly define $\mathcal{G}$ to be maximal such that

$$|\mathcal{G}_2| \leq \frac{r'}{1-r'}\overline{\deg}(A, \mathcal{S}_1) - \frac{r'}{1-r'}\delta k/3 \ , \tag{4.25}$$

or $\mathcal{G}$ is empty if the value of right hand side is smaller then zero. Moreover, if $\mathcal{F} \cup \mathcal{G} \neq \mathcal{D}_A$, we have

$$|\mathcal{G}_2| \geq \frac{r'}{1-r'}\overline{\deg}(A, \mathcal{S}_1) - \frac{r'}{1-r'}\delta k/3 - \beta k \ . \tag{4.26}$$

We have $\mathcal{H} := \mathcal{D}_A \setminus (\mathcal{F} \cup \mathcal{G})$ and, as in the previous case, $\mathcal{F} \cup \mathcal{G}$ consist of the trees with big skew, so if $\mathcal{D}_{A2}$ is non-empty we have:

$$\frac{1-\tilde{r}}{\tilde{r}} \geq \frac{|\mathcal{D}_{A1}|}{|\mathcal{D}_{A2}|} \geq \frac{|\mathcal{H}_1|}{|\mathcal{H}_2|} \ . \tag{4.27}$$

We define $\mathcal{F}'$ and $\mathcal{G}'$ as usual. We use Proposition 4.18 to embed the forest $F_{P4.18} := \mathcal{F}'$ as in the previous cases. Set $U_{P4.18} := \varphi(W_A \cup W_B) \cup U'$, $\eta_{P4.18} := \delta q/4$, $\mathbf{M}_{P4.18} := \mathbf{M}$, and $A_{P4.18} := A$. We verify that

$$\begin{aligned}
\overline{\deg}(A, \mathcal{S}_M) &\geq \frac{1-r'}{r'}|\mathcal{F}_2| + \frac{1-r'}{r'}|U'| + \delta k/3 \\
&\geq \frac{1-r'}{r'}|\mathcal{F}_2| + \frac{1-r'}{r'}|U_{P4.18}| + \delta k/4 \\
&\geq \frac{1-r'}{r'}|\mathcal{F}_2'| + \sum_{C \subseteq \mathcal{S} \,:\, CD \in \mathbf{M}} \max\{|U_{P4.18} \cap C|, \frac{1-r'}{r'}|U_{P4.18} \cap D|\} + \eta_{P4.18} n \ ,
\end{aligned}$$

where we used the fact that $|U_{P4.18}| = |\varphi(W_A \cup W_B)| + |U'| \leq |U'| + \delta k/12$.

If $\mathcal{G}$ is non-empty, we proceed by embedding $\mathcal{G}'$. As in the preceding cases, we apply Proposition 4.19, Configuration 2, to $F^{(2)}_{P4.19} := \mathcal{G}'$ and set $\mathcal{B}^{(2)}_{P4.19} := \mathcal{S}_1$. As we know that $N_{\mathbf{H}}(\mathcal{S}_1) \cup \mathcal{S}_1$ is disjoint from $\bigcup \mathcal{S}_M \supseteq \varphi(\mathcal{F}_1')$, for ensuring the injectiveness of φ it suffices to set $U^{(2)}_{P4.19} := \varphi(\mathcal{F}_2' \cup W_A \cup W_B) \cup U'$. Because $\varphi(\mathcal{F}_2') \cup U' \subseteq \mathcal{L}$, we have $\bigcup \mathcal{B}^{(2)}_{P4.19} \cap U^{(2)}_{P4.19} \subseteq \varphi(W_A \cup W_B)$. Set $\eta^{(2)}_{P4.19} := \delta q/4$, and $A^{(2)}_{P4.19} := A$. We start by verifying the first condition from the proposition. We have

$$\begin{aligned}
\overline{\deg}(A, \mathcal{S}_1) &\geq \frac{1-r'}{r'}|\mathcal{G}_2| + \delta k/3 \\
&\geq |\mathcal{G}_1'| + |\bigcup \mathcal{B}^{(2)}_{P4.19} \cap U^{(2)}_{P4.19}| + \eta^{(2)}_{P4.19} n \ ,
\end{aligned}$$

where we use the definition of $\mathcal{G}$, the fact that $|\mathcal{G}_2| \geq |\mathcal{G}_1'|$ and the fact that $|\bigcup \mathcal{B}^{(2)}_{P4.19} \cap U^{(2)}_{P4.19}| \leq 12/\beta$.

Further we verify that for each cluster $C \in \mathcal{S}_1$ we have

$$\begin{aligned}
\overline{\deg}(C, V(\mathbf{H}) \setminus \bigcup \mathcal{B}^{(2)}_{P4.19}) = \overline{\deg}(C, \mathcal{L}) &\geq \tilde{r}k + \delta k \\
\boxed{\text{bound on skew of } T} \quad &\geq |\mathcal{D}_{A2}| + |\mathcal{D}_{B2}| + \delta k \\
&\geq (|\mathcal{F}_2| + |\mathcal{G}_2|) + |\mathcal{D}_{B1}| + |\varphi(W_A \cup W_B)| + \delta k/2 \\
&\geq |\mathcal{G}_2'| + |\varphi(\mathcal{F}_2' \cup W_A \cup W_B)| + |U'| + \delta k/2 \\
&\geq |\mathcal{G}_2'| + |U^{(2)}_{P4.19}| + \eta n \ ,
\end{aligned}$$

where we started by using the bound on the skew of T, i.e., $|\mathcal{D}_{A2}| + |\mathcal{D}_{B2}| \leq \tilde{r}k$, then bounded the error terms and rearranged suitable terms.

3. Now we apply Proposition 4.19, the first part, to embed the forest $F^{(3)}_{P4.19} := \mathcal{H}$. Set $\mathcal{B}^{(3)}_{P4.19} := \mathcal{L}$ and $U^{(3)}_{P4.19} := \varphi(W_A \cup W_B \cup \mathcal{F}' \cup \mathcal{G}') \cup U'$, thus $U^{(3)}_{P4.19} \cap \mathcal{L} \subseteq \varphi(W_A \cup W_B \cup \mathcal{F}'_2 \cup \mathcal{G}'_2) \cup U'$. Set $\eta^{(3)}_{P4.19} := \delta q/8$, and $A^{(3)}_{P4.19} := A$. We start by verifying the first condition:

$$\begin{aligned}
\overline{\deg}(A, \mathcal{L}) &\geq k + \delta k - \overline{\deg}(A, \mathcal{S}_M) - \overline{\deg}(A, \mathcal{S}_1) \\
\boxed{\text{bounds (4.24) and (4.26)}} \quad &\geq k + \delta k - \left(\frac{1-r'}{r'}|\mathcal{F}_2| + \delta k/3 + \frac{1-r'}{r'}|\mathcal{D}_{B1}| + \frac{1-r'}{r'}\beta k\right) \\
&\quad - \left(\frac{1-r'}{r'}|\mathcal{G}_2| + \delta k/3 + \frac{1-r'}{r'}\beta k\right) \\
\boxed{\text{bounding error terms}} \quad &\geq k - \frac{1-r'}{r'}(|\mathcal{F}_2 \cup \mathcal{G}_2 \cup \mathcal{D}_{B1}|) + \delta k/4 \\
\boxed{\tilde{r} \leq r'} \quad &\geq k - \frac{1-\tilde{r}}{\tilde{r}}(|\mathcal{F}_2 \cup \mathcal{G}_2 \cup \mathcal{D}_{B1}|) + \delta k/4 \\
&= k - \frac{1}{\tilde{r}}(|\mathcal{F}_2 \cup \mathcal{G}_2 \cup \mathcal{D}_{B1}|) + (|\mathcal{F}_2 \cup \mathcal{G}_2 \cup \mathcal{D}_{B1}|) + \delta k/4 \\
\boxed{\text{bound on skew of } T} \quad &\geq k - \frac{1}{\tilde{r}}(\tilde{r}k - |\mathcal{H}_2|) + (|\mathcal{F}_2 \cup \mathcal{G}_2 \cup \mathcal{D}_{B1}|) + \delta k/4 \\
&= \frac{1}{\tilde{r}}|\mathcal{H}_2| + (|\mathcal{F}_2 \cup \mathcal{G}_2 \cup \mathcal{D}_{B1}|) + \delta k/4 \\
&\geq \frac{1-\tilde{r}}{\tilde{r}}|\mathcal{H}_2| + (|\mathcal{F}_2 \cup \mathcal{G}_2 \cup \mathcal{D}_{B1}|) + \delta k/4 \\
\boxed{\text{bound (4.27)}} \quad &\geq |\mathcal{H}_1| + |\varphi(W_A \cup W_B \cup \mathcal{F}'_2 \cup \mathcal{G}'_2)| + |\mathcal{D}_{B1}| + \eta^{(3)}_{P4.19} n \\
&\geq |\mathcal{H}_1| + |U^{(3)}_{P4.19} \cap \bigcup \mathcal{B}^{(3)}_{P4.19}| + \eta^{(3)}_{P4.19} n \ ,
\end{aligned}$$

We set $\tilde{U}^{(3)}_{P4.19} = \emptyset$ and immediately invoke the second part of proposition. We verify that for each $C \in \mathcal{L}$ we have

$$\begin{aligned}
\overline{\deg}(C) &\geq k + \delta k \\
&= |\mathcal{D}_A \cup \mathcal{D}_B \cup W_A \cup W_B| + \delta k \\
&\geq |\mathcal{H}_1| + |\mathcal{H}_2| + |\varphi(W_A \cup W_B \cup \mathcal{F}' \cup \mathcal{G}') \cup U'| + \eta^{(3)}_{P4.19} n \\
&\geq |\mathcal{H}_1| + |\mathcal{H}_2| + |U^{(3)}_{P4.19} \cup \tilde{U}^{(3)}_{P4.19}| + \eta^{(3)}_{P4.19} n \ .
\end{aligned}$$

Thus we can extend φ to $\mathcal{H}$. Moreover, note that after each application of Propositions 4.18 and 4.19 it was true that φ avoided at least $r'\eta^{(1)}_{P4.19}|C|/8$ vertices of each cluster C. Thus, we can extend φ to $\mathcal{D}_B$ as we promised in the first part of the analysis of this case.

4. We have defined φ on the whole tree T except for $\mathcal{F}_1 \setminus \mathcal{F}'_1$ and $\mathcal{G}_1 \setminus \mathcal{G}'_1$. We can again extend φ to the whole T in the usual greedy manner.

Case D

In this case we assume the existence of two adjacent clusters A, B such that $\overline{\deg}(A, \mathcal{S}_M \cup \mathcal{L}) \geq k + \delta k$ and $\overline{\deg}(B, \mathcal{L}) \geq |\mathcal{D}_{B2}| + \delta k$. Moreover, we assume that $\tilde{r}|\mathcal{D}_{A1}| \geq (1-\tilde{r})|\mathcal{D}_{A2}|$

and $|\mathcal{D}_{B2}| \leq \frac{\tilde{r}}{1-\tilde{r}} rk$ and for each edge $(C, D) \subseteq \mathbf{M}$ either $\overline{\deg}(A, C) = 0$ or $\overline{\deg}(A, D) = 0$.

We proceed in the same way as in the previous case, although the analysis is different.

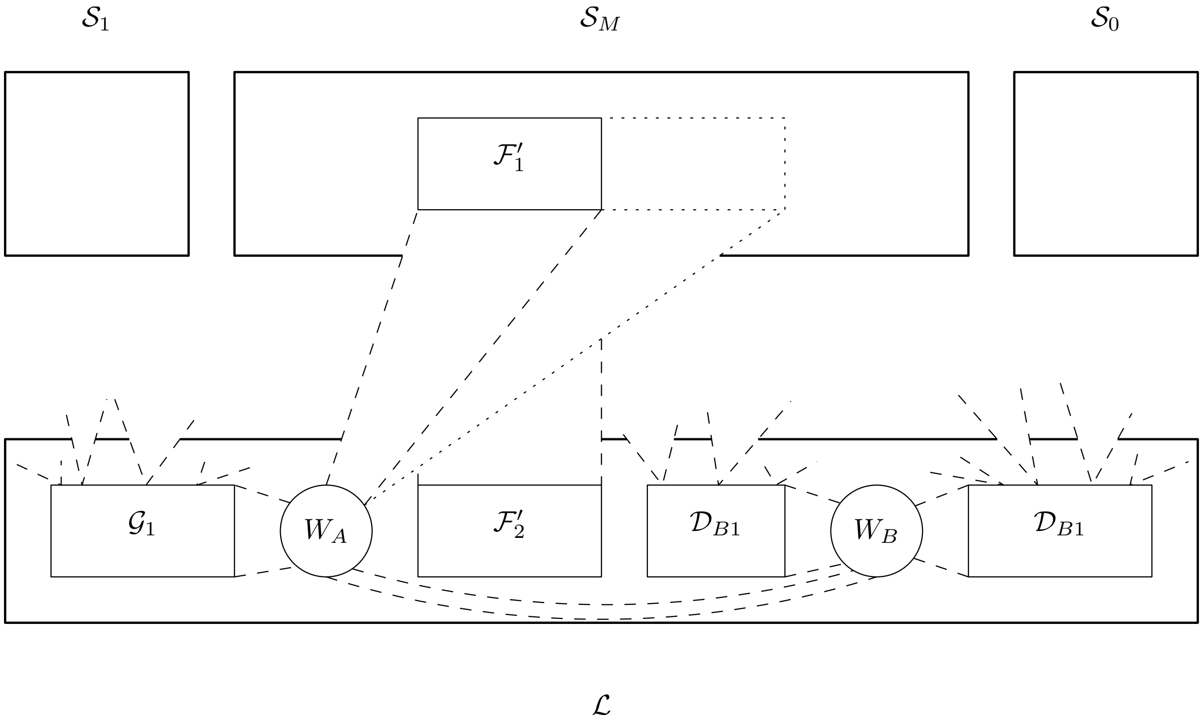

Figure 4.7: The embedding configuration in the case D. The order of operations is the same as in the preceding case, but the analysis is different. The figure suggests that as in the previous case we have to me more careful in the application of Proposition 4.18. The special condition on the neighbourhood of the cluster A plays the following role: we split the reserved vertices for $\mathcal{D}_{B1}$ into two parts – the vertices in the neighbourhood of A (the right rectangle on the figure) and those that are not neighbours of A (the left rectangle). Now the condition implies that the first type of vertices does not play a role in the embedding of $\mathcal{F}'$ using the matching, whilst the second type of vertices does not have to be considered in the embedding of $\mathcal{G}$ through the $\mathcal{L}$-neighbourhood of A.

1. We start by reserving vertices for embedding the anchored forest $F^{(1)}_{P4.19} := \mathcal{D}_B = \mathcal{D}_{B1} \cup \mathcal{D}_{B2}$ such that $\mathcal{D}_{B1}$ will be embedded in the L-neighbourhood of the cluster B. This is done using Proposition 4.19 in the exactly same way as in the previous case. We get an embedding of $N(W_B) \cap \mathcal{D}_B$ and a set of reserved vertices W. We set $U' = \varphi(N(W_B) \cap \mathcal{D}_B) \cup W$, $|U'| = |\mathcal{D}_{B1}|$. We will also invoke the 'moreover' part Proposition 4.19 after embedding the rest of T and then we set $\tilde{U}^{(1)}_{P4.19} = \varphi(\mathcal{F}' \cup \mathcal{G})$ for $\mathcal{F}' \cup \mathcal{G} \subseteq \mathcal{D}_A$. We have to ensure that for subsequent applications of Propositions 4.18 and 4.19 we have $\eta \geq \eta^{(1)}_{P4.19} = q\delta/20$.

   Moreover, we split the set $U' \subseteq \mathcal{L}$ in two sets U'_1 and U'_2 such that U'_1 contains the vertices from U' contained in clusters C such that $C \in N_{\mathbf{H}}(A)$ (we define $N_{\mathbf{H}}(A)$ as the set of clusters C with $\overline{\deg}(A, C) > 0$) and $U'_2 := U' \setminus U'_1$. Note that our assumption on the neighbourhood of cluster A states that if we have $(C, D) \subseteq \mathbf{M}$ with $D \cap U'_1 \neq \emptyset$, we have then $\overline{\deg}(A, C) = 0$.

2. We continue by embedding the anchored forest $\mathcal{D}_A$ analogously to previous cases. Partition $\mathcal{D}_A = \mathcal{F} \cup \mathcal{G}$, ordering the components by decreasing order f their skew, in such a way that $\mathcal{F}$ is maximal with

$$|\mathcal{F}_2| \leq \frac{r'}{1-r'}\overline{\deg}(A, \mathcal{S}_M) - |U'_2| - \frac{r'}{1-r'}\delta k/2 \;, \tag{4.28}$$

or $\mathcal{F}$ is empty if the right hand side is smaller than zero. We define $\mathcal{F}'$ as usual. If $\mathcal{F} \neq \mathcal{D}_A$, we have

$$|\mathcal{F}_2| \geq \frac{r'}{1-r'}\overline{\deg}(A, \mathcal{S}_M) - |U'_2| - \frac{r'}{1-r'}\delta k/2 - \beta k \;. \tag{4.29}$$

Moreover, $\mathcal{F}$ is chosen so that it contains the trees with maximal skew, thus if it is non-empty we have

$$\frac{|\mathcal{F}_1|}{|\mathcal{F}_2|} \geq \frac{|\mathcal{F}_1 \cup \mathcal{G}_1|}{|\mathcal{F}_2 \cup \mathcal{G}_2|} \geq \frac{1-\tilde{r}}{\tilde{r}} \;. \tag{4.30}$$

Now we use Proposition 4.18 to embed $F_{P4.18} := \mathcal{F}'$. Set $U_{P4.18} := \varphi(W_A \cup W_B) \cup U'_2$ and $\mathbf{M}_{P4.18}$ be only those matching pairs $(C, D), C \subseteq \mathcal{S}$ such that $\overline{\deg}(A, C) > 0$. Observe that U'_1 is disjoint from $\bigcup V(\mathbf{M}_{P4.18})$. Set $\eta_{P4.18} := \delta q/3$, and $A_{P4.18} := A$. As in the previous cases we easily verify that

$$\begin{aligned}\overline{\deg}(A, \mathcal{S}_M) &\geq \frac{1-r'}{r'}|\mathcal{F}_2| + \frac{1-r'}{r'}|U'_2| + \delta k/2 \\ &\geq \frac{1-r'}{r'}|\mathcal{F}_2| + \frac{1-r'}{r'}|U_{P4.18}| + \delta k/3 \\ &\geq \frac{1-r'}{r'}|\mathcal{F}'_2| + \sum_{C\subseteq\mathcal{S}\,:\,CD\in\mathbf{M}} \max\{|U_{P4.18} \cap C|, \frac{1-r'}{r'}|U_{P4.18} \cap D|\} \; + \eta n.\end{aligned}$$

Thus we can extend φ to $\mathcal{F}'$. Note that $\mathcal{F}'_2$ is embedded in L-clusters that are not in the neighbourhood of A. Indeed, from our assumption on the cluster A we have $\overline{\deg}(A, D) = 0$ for any edge $CD \in \mathbf{M}_{P4.18}$, with $C \subseteq \mathcal{S}$.

3. We now apply Proposition 4.19, first part, to embed $F^{(2)}_{P4.19} := \mathcal{G}$ if it is non-empty. Set $\mathcal{B}^{(2)}_{P4.19} := \mathcal{L} \cap N_{\mathbf{H}}(A)$ and $U^{(2)}_{P4.19} := \varphi(W_A \cup W_B \cup \mathcal{F}) \cup U'$. Note that $U^{(2)}_{P4.19} \cap \bigcup \mathcal{B}^{(2)}_{P4.19} \subseteq \varphi(W_A \cup W_B) \cup U'_1$, as we know that neither U'_2, nor $\varphi(\mathcal{F}'_2)$ is in $N_{\mathbf{H}}(A)$ and $\varphi(\mathcal{F}'_1) \cap \mathcal{L} = \emptyset$. Set $\eta^{(2)}_{P4.19} := q\delta/4$, and $A^{(2)}_{P4.19} := A$.

   We verify the first condition of the proposition:

$$
\begin{aligned}
\overline{\deg}(A,\mathcal{L}) &\geq k+\delta k-\overline{\deg}(A,\mathcal{S}_M)\\
\boxed{\text{bound (4.29)}}\quad &\geq k+\delta k-\left(\frac{1-r'}{r'}|\mathcal{F}_2|+\frac{1-r'}{r'}|U_2'|+\delta k/2+\frac{1-r'}{r'}\beta k\right)\\
\boxed{\text{definition of } U'}\quad &\geq k+\delta k-\frac{1-r'}{r'}|\mathcal{F}_2|-\frac{1-r'}{r'}(|\mathcal{D}_{B1}|-|U_1'|)-\delta k/2-\frac{1-r'}{r'}\beta k\\
\boxed{\tilde{r}\leq r'}\quad &\geq k-\frac{1-\tilde{r}}{\tilde{r}}|\mathcal{F}_2|-\frac{1-\tilde{r}}{\tilde{r}}|\mathcal{D}_{B1}|+\frac{1-r'}{r'}|U_1'|+\delta k/3\\
\boxed{\text{bound (4.30)}}\quad &\geq k-|\mathcal{F}_1|-\frac{1-\tilde{r}}{\tilde{r}}|\mathcal{D}_{B1}|+|U_1'|+\delta k/3\\
\boxed{\text{assumed bound on } \mathcal{D}_{B1}}\quad &\geq k-|\mathcal{F}_1|-\frac{1-\tilde{r}}{\tilde{r}}\frac{\tilde{r}}{1-\tilde{r}}\tilde{r}k+|U_1'|+\delta k/3\\
&=(1-\tilde{r})k-|\mathcal{F}_1|+|U_1'|+\delta k/3\\
\boxed{\text{bound on the skew of } T}\quad &\geq |\mathcal{D}_{A1}|-|\mathcal{F}_1|+|U_1'|+\delta k/3\\
&\geq |\mathcal{G}_1|+|U_1'|+\delta k/3\\
&\geq |\mathcal{G}_1|+|U^{(2)}_{P4.19}\cap\bigcup\mathcal{B}^{(2)}_{P4.19}|+\eta^{(2)}_{P4.19}n\,.
\end{aligned}
$$

We set $\tilde{U}^{(2)}_{P4.19}:=\emptyset$ and immediately apply the second part of the proposition. We verify that for each $C\in\mathcal{L}$ we have

$$
\begin{aligned}
\overline{\deg}(C) &\geq k+\delta k\\
&=|\mathcal{D}_A\cup\mathcal{D}_B\cup W_A\cup W_B|+\delta k\\
&\geq |\mathcal{G}_1|+|\mathcal{G}_2|+|\varphi(W_A\cup W_B\cup\mathcal{F}')|+|\mathcal{D}_{B1}|+\eta^{(2)}_{P4.19}n\\
&\geq |\mathcal{G}_1|+|\mathcal{G}_2|+|U^{(2)}_{P4.19}\cup\tilde{U}^{(2)}_{P4.19}|+\eta^{(2)}_{P4.19}n\,.
\end{aligned}
$$

Thus, we can extend φ to $\mathcal{G}$. Moreover, after each operation it was true that φ avoided at least $r'\eta^{(1)}_{P4.19}|C|/8$ vertices of each cluster C. Thus, we can extend φ to $\mathcal{D}_B$ using the 'moreover' part of Proposition 4.19.

4. We again extend the embedding of T greedily to $\mathcal{F}_1\setminus\mathcal{F}_1'$ as usual.

Chapter 5

The skew Loebl-Komlós-Sós conjecture for paths

In this chapter we verify that the skew Loebl-Komlós-Sós conjecture holds for paths. Since paths have very simple structure, we will be able to verify the conjecture for them in full generality.

The path on k vertices is arguably the most important tree from $\mathcal{T}_k$, thus it make sense to verify that our embedding conjectures hold for them. That the Erdős-Sós conjecture holds for paths for observed already in 1959 by Erdős and Gallai. In Section 5.2 we interpret their proof as a simple algorithm for finding the path. This will serve as a motivation for a later algorithm for finding paths in LKS graphs.

Theorem 5.1. *Every (r, k)-LKS graph G contains a path on $\lfloor 2rk \rfloor + 1$ vertices.*

The proof that we present in the next section is basically the one from [BLW00]. Although it appears there only for the special case $r = 1/2$, the generalisation for all $r \leq 1/2$ is straightforward and does not affect the combinatorial part of the proof. We still give the proof here, since it is not clear that all of the computations really work out as desired for $r < 1/2$.

Note that after seeing the proof of Theorem 5.1, one can observe that there is an appropriate polynomial time algorithm for constructing long paths in (r, k)-LKS graphs. In the last section we actually provide an algorithm with linear time complexity for fixed $r < 1/2$ (we however state the theorem with the dependence of time complexity on r).

Theorem 5.2. *There is an algorithm for an (r, k)-LKS graph G finds a path on $\lfloor 2rk \rfloor + 1$ vertices in time $\mathcal{O}(\min(m/(1 - 2r), mk))$.*

The author implemented the algorithm as his software project.

5.1 Proof of Theorem 5.1

For fixed r suppose that k is the smallest integer for which the theorem fails, i.e., there is an (r, k)-LKS graph G such that $P_{\lfloor 2rk \rfloor + 1} \notin G$. Since k is minimal, G contains a path on $\lfloor 2r(k - 1) \rfloor + 1 \geq \lfloor 2rk - 1 \rfloor + 1 = \lfloor 2rk \rfloor$ vertices. Now we define $r' < r$ to be such that $2r'k = \lfloor 2rk \rfloor$. Hence we have an (r', k)-LKS graph that does not contain a path on $2r'k + 1$ vertices, but does contain a path on $2r'k$ vertices. We may, moreover, assume that G is an inclusion-wise minimal counterexample to r' and k, hence it is connected and does not contain edges between its S-vertices. It can be checked that the theorem holds for any r, k if $rk \leq 1$, hence we suppose that $k > \frac{1}{r'} \geq 2$.

We now derive a contradiction by finding a path on $2r'k + 1$ vertices in G. For convenience and since we do not use the value of r anymore, we use the letter r to denote the value of r'.

Before proving the theorem we gradually show that the following configurations cannot occur in G:

Proposition 5.3. *The following cannot occur in G:*

1. *G contains a path on $2rk$ vertices with one of its endpoints in L.*
2. *G contains a path on $2rk - 1$ vertices with both of its endpoints in L.*
3. *G contains a cycle C on at least $2rk$ vertices.*
4. *G contains a cycle C on $2rk - 1$ vertices.*
5. *G contains a cycle C on $2rk - 2$ vertices.*

Proof. 1–2 We use the greedy argument to prolong the path and get a path on $2rk + 1$ vertices.

3 G is connected and has at least $k + 1 \geq 2rk + 1$ vertices. Thus there are $u \in C$ and $v \in G \setminus C$ connected by an edge. Then $vu\overrightarrow{C}u^-$ is a path with at least $2rk + 1$ vertices.

4 We will consider two cases.

(a) There is an L-vertex u outside of C.

Thus we have a path $\overrightarrow{P}$ from u to some $v \in C$. Now the path $u\overrightarrow{P}v\overrightarrow{C}v^-$ has at least $2rk$ vertices and one of its endpoint is in L. This contradicts the first statement of this proposition.

(b) All L-vertices are in C.

If there are two consecutive vertices in C that are both in L, the rest of the cycle forms a path with both of its endpoints in L, which contradicts the second statement of this proposition. Otherwise the size of C is at least $2|L| \geq 2rn \geq 2r(k + 1) \geq 2rk$, which contradicts the third statement from this lemma.

5 Let $\overrightarrow{C}$ be an oriented cycle of length $2rk - 2$ in L such that the number of L vertices on C is the largest possible. We will again consider two cases.

(a) There is an L-vertex u outside of C.

Thus, there is a path $\overrightarrow{P}$ from u to some $v \in C$. The path has to consist of a single edge, otherwise we would have a path on at least $2rk$ vertices with one endpoint in L. We can also assume that both v^- and v^+ are in S, otherwise the path $uv\overrightarrow{C}v^-$ or $uv\overleftarrow{C}v^+$ would have $2rk - 1$ vertices and both endpoints in L. This means that v has to be in L.

If $v, w \in C$ and there are edges uv and uw, it cannot happen that $w = v^+$ or $w = v^{++}$, because then the cycle $uw\overrightarrow{C}vu$ would have either more vertices or more L-vertices than C. It also cannot be the case that $w = v^{+++}$, as this would mean that v^+ and $v^{++} = w^-$ are two neighbouring S-vertices.

So there are at most $(2rk-2)/4$ edges from u to C. The rest of the neighbours of u has to be leaves, because otherwise we would again have a path contradicting the first part of this lemma. Number of those neighbours is at least

$$k - \frac{2rk-2}{4} = \left(1 - \frac{r}{2}\right)k + \frac{1}{2} > \left(1 - \frac{r}{2}\right)\frac{1}{r} + \frac{1}{2} = \frac{1}{r}.$$

But this contradicts Lemma 2.3 for X defined as the leaf neighbours of u.

(b) All L-vertices are on C.

Thus $2rk-2 = |C| \geq |L| \geq rn$, implying that $k \geq \frac{n}{2} + \frac{1}{r} > \frac{n}{2}$. Let $S_C = S \cap C$ and $S_R = S \setminus C$. Additionally, we denote $A \subseteq C \cap L$ the set of L vertices lying on C such that their successors on C are also in L. We have $|A| = |L| - |S_C|$.

We may suppose that two different vertices u, v from A have different neighbourhoods in S_R (i.e. $N(u) \cap N(v) \cap S_R = \emptyset$) because if there were two edges uw and vw (for some $w \in S_R$) the path $u^+ \overrightarrow{C} vwu \overleftarrow{C} v^+$ of length $2rk-1$ would have both of its endpoints in L, contradicting the second statement of this proposition.

Each vertex in L has at most $2rk-3$ neighbours in C and therefore at least $k-(2rk-3)$ neighbours in S_R. From this we can estimate size of S_R, which gives us a bound on the size of S_C. Specifically, we have

$$\begin{aligned}(1-r)n - |S_C| &\geq |S| - |S_C| = |S_R| \geq |A|(k-2rk+3) \\ &= (|L| - |S_C|)(k-2rk+3) \geq (rn - |S_C|)(k-2rk+3).\end{aligned}$$

By rearranging the terms we get

$$|S_C|(k-2rk+2) \geq (k-2rk+3)rn - (1-r)n,$$

hence

$$\begin{aligned}|S_C| &\geq \frac{k-2rk+2+1}{k-2rk+2}rn - \frac{1-r}{k-2rk+2}n \\ &= rn - \frac{1-2r}{k-2rk+2}n > rn - \frac{1-2r+2/k}{k-2rk+2}n = rn - \frac{n}{k}.\end{aligned}$$

But we know that $k > \frac{n}{2}$, therefore $|S_C| > rn-2$. From this we finally deduce a contradiction with the size of C.

$$|C| = |L| + |S_C| \geq rn + rn - 1 = 2rn - 1 \geq 2rk - 1.$$

□

We are now ready to prove Theorem 5.1.

Proof. We know that G contains a path on $2rk$ vertices. By Proposition 5.3 (1) we know that both endpoints of the path are in S, thus after trimming the endpoints we get a path $\overrightarrow{P} = u\overrightarrow{P}v$ on $2rk-2$ vertices such that $u, v \in L$.

The neighbours of u and v outside P cannot lie in L and they cannot have L-neighbours outside P, since otherwise we would have a path on at least $2rk-1$ vertices with both endpoints in L, contradicting Proposition 5.3 (2).

Define $W_1 = N(u) \setminus P$ and $W_2 = N(v) \setminus P$. It must hold that $W_1 \cap W_2 = \emptyset$, because if there were a vertex $w \in W_1 \cap W_2$, the cycle $u\overrightarrow{P}vw$ would contradict Proposition 5.3 (4).

We now bound the number of edges between the set $\{u, v\}$ and P. There cannot be a vertex $w \in P$ such that both edges uw and vw^- are present in G, because then the cycle $uw\overrightarrow{P}vw^-\overleftarrow{P}u$ would contradict Proposition 5.3 (5). Using a simple pairing argument we infer that there are at most $|P| - 1 \leq 2rk - 3$ edges between $\{u, v\}$ and P. Thus,

$$|W_1| + |W_2| \geq 2k - (2rk - 3) = 2k - 2rk + 3.$$

Now we can use Lemma 2.3 to bound the size of $N(W_1 \cup W_2)$. We have

$$|N(W_1 \cup W_2)| > r|W_1 \cup W_2| = r(|W_1| + |W_2|) \geq r(2k - 2rk + 3). \tag{5.1}$$

Now let $D = (N(W_1) \setminus u)^- \cup (N(W_2) \setminus v)^+$. At first note that $D \subseteq S$. Indeed, if there were a vertex $w \in N(u)$ and its neighbour t such that $t^- \in L$, the path $t^-\overleftarrow{P}uwt\overrightarrow{P}v$ of length $2rk - 1$ would contradict Proposition 5.3 (1). Thus $D \cap N(W_1 \cup W_2) = \emptyset$.

Finally we observe that $N(W_1)^- \cap N(W_2)^+ = \emptyset$. Indeed, if there were $w \in P$ such that $w^+ \in N(W_1)$ and $w^- \in N(W_2)$, we could find a cycle $usw^+\overrightarrow{P}vtw^-\overleftarrow{P}u$ of length $2rk - 1$ contradicting 5.3 (4). Consider the mapping from $N(W_1 \cup W_2)$ to D that assigns each vertex from $N(W_1) \setminus u$ its predecessor and each vertex from $N(W_2) \setminus v$ its successor. This mapping is injective and its image is disjoint from its domain. Hence we can bound the number of vertices of P as follows:

$$\begin{aligned}
|P| &\geq |N(W_1 \cup W_2) \cup (N(W_1) \setminus u)^- \cup (N(W_2) \setminus v)^+| \\
&\geq |N(W_1 \cup W_2)| + |N(W_1 \cup W_2)| - 2 \\
&= 2|N(W_1 \cup W_2)| - 2 \\
&\geq 2r(2k - 2rk + 3) - 2 = 4rk - 4r^2k + 6r - 2 \\
&= 2rk - 2 + 2rk(1 - 2r) + 6r > 2rk - 2,
\end{aligned}$$

so P is a path on at least $2rk - 1$ vertices, a contradiction. □

5.2 Proof of Theorem 5.2

To give an intuition behind the algorithm, we start by stating the well-known algorithm of Erdős and Gallai that can be used to verify that the Erdős-Sós conjecture holds for paths via Lemma 2.2.

Theorem 5.4. *There is an algorithm that for a graph G finds its path of length $\min(2\delta(G) + 1, |V(G)|)$ in time $\mathcal{O}(m)$.*

Proof. We provide such an algorithm. In the ith step of the algorithm we have found a subgraph $T_i \subseteq G$[1] that is either a path or a cycle. We do the following:

1. Suppose that $T_i = u\overrightarrow{T}_iv$ is a path on at most $2\delta(G)$ vertices. If there exists a vertex $w \notin T_i$ adjacent to either u or v, we prolong T_i by adding to it this adjacent vertex. If there is $w \in T_i$ such that uw and w^-v are edges in G we define T_{i+1} as the cycle $u\overrightarrow{T}_iw^-v\overrightarrow{T}_iwu$. If neither of these cases occurs we get by a simple pairing argument that $\deg(u) + \deg(v) \leq |T_i| - 1 \leq 2\delta(G) - 1$, a contradiction.

[1]Note that for notational convenience we use the same symbol both for a path and a cycle which is a difference compared to the notation from Chapter 5.

2. Suppose that T_i is a cycle not spanning the whole G. Then we find a vertex $a \notin T_i$ adjacent to a vertex $w \in T_i$. The desired path T_{i+1} is then $aw\overrightarrow{T}_i w^-$.

The algorithm starts with an arbitrary edge uv and stops when the path/cycle T_i has length $\min(2\delta(G), |V(G)|)$. The number of steps is at most twice the number of vertices of the final path.

If we use suitable data structures, the time complexity of each step is either $\mathcal{O}(\delta(G))$ or $\mathcal{O}(\delta(G)) + \deg(v)$ if we add a vertex v to T_i. Thus, the total time complexity can be then estimated as

$$\mathcal{O}(\delta(G)^2 + \sum_{v \in G} \deg(v)) \subseteq \mathcal{O}(n\delta(G) + m) \subseteq \mathcal{O}(m).$$

□

We are now ready to prove Theorem 5.2. Prepare for the worst. At first we give its high-level overview. Then we introduce invariants that will be preserved during the run-time of the algorithm and sketch the necessary data structures. We continue by describing the algorithm as a sequence of *steps*. In the end we at first bound the number of steps executed by the algorithm, thus proving its correctness, and finally deduce that its time complexity is, indeed, the one advertised above. Because specifying all the details of the algorithm would be quite tedious and it is far from the aim of this thesis we knowingly omit several implementation details and encourage the interested reader to fill in the gaps.

Overview The algorithm we are going to sketch follows the proof of Theorem 5.1 and we will several times refer to specific places in its proof. The algorithmic idea is the same as the one behind the algorithm from Theorem 5.4. More specifically, we work with a subgraph $T_i \subseteq G$ that is either a path or a cycle and we gradually try to make it longer in a greedy manner. However, there are several additional difficulties:

- Although we still without loss of generality assume that $2rk \in \mathbb{N}$ we can no longer assume that G is a minimal counterexample. Thus, in each step we rather work with a graph $G_i \subseteq G$. Sometimes we get G_{i+1} by erasing several edges or vertices of G_i while maintaining that each such G_i is an (r, k)-LKS graph.

- We will have to consider more cases than in Theorem 5.4. Sometimes it can happen that we only replace a vertex from our path/cycle by another one and it may even happen that the size of T_i is reduced. In one specific case we forget the whole T_i and start anew from the beginning. Thus, to bound the number of steps of the algorithm, we introduce several invariants and use amortisation arguments.

- Because we aim for linear time complexity, we cannot afford to start the search for the desired path every time we delete some vertices or edges from G_i. Although we will never erase a vertex from T_i it may well happen that a vertex inside T_i becomes an S-vertex and thus T_i contains an edge between two S-vertices. This forces the following technical difficulty: it will no longer hold that there are no edges between S-vertices of G_i but there can be such edges only between consecutive vertices of T_i. Fortunately, this technical issue does no affect the analysis of the algorithm significantly.

The algorithm works in *steps* and for the sake of the analysis of its time complexity is divided into two *epochs*. Each step consists of finding desired Configurations from Theorem 5.1 that we, for the sake of clarity, group into short procedures that we call *Configurations*. Each Configuration is *successful* if we manage to find T_{i+1} during the course of the procedure. The overall structure of the algorithm can be found in Figure 5.2. If we succeed in finding one of the desired Configurations we update our path/cycle as well as other corresponding data structures. If we do not find any desired improving Configuration, we delete several vertices or edges from G_i or we delete several edges between S-vertices contained in T_i.

We will show that the time complexity of most of the steps scales with k plus the sum of degrees of L-vertices from $T_{i+1} \setminus T_i$ plus the number of deleted vertices and edges from G_i. The exceptions are the steps in which we run Configuration 3 and those will be treated separately.

Data: An (r, k)-LKS graph G
Result: a path P embedded in G on $2rk + 1$ vertices
Epoch 1: **while** $|T_i| \leq 2rk + 1$ **do**
  if *T_i is a path* **then**
    run Configuration 1, **if** *successful* **then continue**;
    run Configuration 2, **if** *successful* **then continue**;
    run Configuration 3, **if** *successful* **then continue**;
    erase all vertices from U_i;
  else
    if *$N(P_i) \setminus U_i$ is not empty* **then**
      run Configuration 4, **if** *successful* **then continue**;
      erase several leaf-neighbours of a vertex not in T_i;
    else
      compute the ratio r' of number of L-vertices in the graph $T_i \cup U_i$;
      if $r' \geq r$ **then** goto Epoch 2;
      delete all vertices from $T_i \cup U_i$ and set $T_{i+1} = \emptyset$;
    end
  end
end
Epoch 2: **while** $|T_i| \leq 2rk + 1$ **do**
  if *T_i is a path* **then**
    run Configuration 1 `/* always successful */`
  else
    run Configuration 5 `/* always successful */`
  end
end

Algorithm 1: The algorithm from Theorem 5.2.

Notation, invariants, data structures We use the following notation. The set of L-vertices of G_i is called L_i, its set of S-vertices is S_i and $n_i := |V(G_i)|$. We denote by U_i the set of neighbours of T_i such that all their L-neighbours are in T_i. Note that all vertices of U_i are S-vertices. Following invariants will hold during the course of the algorithm.

1. T_i contains at least one L-vertex. If T_i is a path then both of its endpoints are L-vertices.

2. G_i is an (r, k)-LKS graph. We do not require G_i to be connected and there can be edges between its S-vertices but such are always between consecutive vertices of T_i.

3. $(T_i \setminus T_{i+1}) \cap G_{i+1} \subseteq S_{i+1}$ and $(U_i \setminus U_{i+1}) \cap G_{i+1} = \emptyset$. In other words, vertices erased from T_i are always S-vertices in the new host graph and a vertex can get off the set U_i only by being erased from G_i.

4. At least one of any two consecutive vertices of T_i was an L-vertex in the step $j < i$ in which it was added to T_j.

5. $|U_i| \leq 2k$.

6. In Epoch 1 if T_i is a non-empty cycle then the set $N(T_i) \setminus (U_i \cup T_i)$ is non-empty.

7. In Epoch 2 all L-vertices of G_i are also in T_i.

We will not argue separately for each invariant that it holds during the whole course of the algorithm. We encourage the reader to do so after she reads the description of the algorithm.

We continue by sketching main data structures that we are going to use. The main structures needed will be several lists – each $v \in V(T_i)$ stores the following lists: *Neighbours* containing all neighbours of v, *LNeighbours* containing all of its neighbours in L , *TNeighbours* containing all neighbours in T_i, and finally *LL^+Neighbours1* and *LL^-Neighbours*. Each $v \in U_i$ stores the list of its L-neighbours such that their predecessor on T_i or their successor on T_i, respectively, is also an L-vertex. There is also the list *Tvertices* containing the vertices of T_i sorted according to their order in T_i and several others. For each list we also store its inverse counterparts. This means that every vertex has lists with information, for what other vertices it is in their list Neighbour, TNeighbour, ... One can observe that we can add or delete vertex u from T_i or erase several incident edges and change its state to an S-vertex or erase the vertex entirely from G_i, and updating the corresponding information in all lists can be done in time $\mathcal{O}(\deg(u))$.

If we change the state of a vertex from L to S we change appropriate data structures, erase edges connecting the vertex to its S-neighbours (unless they are its neighbours in T_i) and we also recursively check all of its S-neighbours and potentially do the same. Observe that the amortised cost of all of these operations is bounded by $\mathcal{O}(\sum_{v \in L} \deg(v)) = \mathcal{O}(m)$.

Epoch 1

In the next two sections we finally give a description of the algorithm. At several places we refer to the proof of Theorem 5.1. Delete all edges between S-vertices of G and continue by repeating the following steps until the size of T_i is sufficient or until we proceed to Epoch 2. If T_i is empty, we define T_{i+1} as an edge between two L-vertices (the case, when there is none, is easy). Suppose that T_i is a path. Set $\overrightarrow{T}_i = u\overrightarrow{T}_i v$.

Configuration 1 If there is an L-neighbour w of u or v, respectively, we set $T_{i+1} = wu\overrightarrow{T}_i v$ or $T_{i+1} = u\overrightarrow{T}_i vw$, respectively. If this does not happen, but there is an S-neighbour w of u or v, respectively, such that w has an L-neighbour t outside of T_i, we set $T_{i+1} = twu\overrightarrow{T}_i v$ or $T_{i+1} = u\overrightarrow{T}_i vwt$, respectively. If there is $w \in T_i$ such that there are two edges uw and vw^- in G_i, we set T_{i+1} as the cycle $uw\overrightarrow{T}_i vw^-\overleftarrow{T}_i u$. With some care the whole process can be done in time $\mathcal{O}(k)$.

Configuration 2 If there are $w \in N(u) \cap U_i$ and $t \in T_i$ such that wt is an edge in G_i and, moreover, t^- is an L-vertex, we define T_{i+1} as $t^-\overleftarrow{T}_i uwt\overrightarrow{T}_i v$ and similarly for the other case. The existence of such Configuration can be checked in time $\mathcal{O}(k)$ if we use the lists LL^+Neighbours1, or LL^-Neighbours, respectively.

Configuration 3 In this Configuration we search for three vertices s, t, and w such that $s \in N(U) \cap U_i, t \in N(v) \cap U_i, w \in T_i$ and, further, sw^+ and tw^- are edges of G_i. In this case, we get a cycle T_{i+1} as $u\overleftarrow{T}_i w^- tv\overrightarrow{T}_i w^+ su$.

The existence of such Configuration can be checked in time $\mathcal{O}(k + |e(U_i, T_i)|) = \mathcal{O}(rk^2)$ by at first enumerating the L-neighbours of $N(u) \cap U_i$ and labeling those L-vertices and then enumerating L-neighbours of $N(v)$ and checking whether their double successor is already labeled.

If we were not successful in any of the preceding Configurations, we observe that we can follow the arguments from Theorem 5.1 and get that $r|U_i| > |T_i|$. If this inequality is, indeed, satisfied, we delete the whole U_i while keeping the LKS property of G_{i+1} (Lemma 2.3).

We continue with the case when T_i is a cycle. If the set $N(T_i) \setminus (U_i \cup T_i)$ is empty, we check in time $\mathcal{O}(k)$, whether the proportion of L-vertices to all vertices in the component of G_i induced by $T_i \cup U_i$ at least r. If it is so, we delete all other vertices of G_i and proceed to Epoch 2. Otherwise we delete all vertices from $T_i \cup U_i$ and start again with an empty T_{i+1}. Invariant 4 ensures that the number of L-vertices of G that we have deleted is proportional to the size of T_i. The graph G_{i+1} is still an (r, k)-LKS graph. Now we assume that there is a vertex $u \notin T_i$ with a neighbour in $v \in T_i$ but $N(u) \not\subseteq T_i$. In the following Configuration we create a path T_{i+1} containing u, or erase all neighbours of u that are not in T_i.

Configuration 4

Suppose that u has an L-neighbour $w \notin T_i$. Now consider path $wuv\overrightarrow{T}_i v^-$ of length $|T_i| + 2$. We trim all the S-vertices from the end of this path and possibly delete some edges between S-vertices that were contained in T_i. We set T_{i+1} to be this resulting path. Note that either we have trimmed at most one vertex and T_{i+1} is longer than T_i or the number of trimmed vertices is proportional to the number of deleted edges between two S-neighbours in T_i. Every S-vertex from $N(T_i) \setminus (U_i \cup T_i)$ has an L-neighbour outside of T_i. Therefore, we further assume that u is an L-vertex with no L-neighbours outside of T_i. If u has an S-neighbour $w \notin T_i$, it is either a leaf or it has another L-neighbour $t \notin T_i$ (neighbours in T_i would give the previous Configuration). If it has such neighbour t, we solve this Configuration analogously to the first Configuration of this Configuration using the path $twuv\overrightarrow{T}_i v^-$. If u has a neighbour w in T_i such that w^+ or w^-, respectively, is an L-vertex, we get T_{i+1} as a path $uw\overrightarrow{T}_i u^+$ or $uw\overleftarrow{T}_i u^-$, respectively. If both v^+ and v^{++} are S-vertices, we define T_{i+1} as the path $uv\overleftarrow{T}_i v^+$ without the consecutive sequence of S-vertices at the end of it. We delete edges between S-vertices. Such path is shorter than T_i but the number of trimmed vertices is again linear in the number of deleted edges between S-vertices from T_i. We proceed in an analogous manner if both v^- and v^{--} are S-vertices. If we do not find any of these Configurations, there are no two neighbours of u in T_i that are at distance one or three apart. We finally deal with the case when two neighbours of u in T_i are at distance two apart. If we have $v \in T_i$ such that u is incident both with v and v^{++}, we define T_{i+1} as the cycle $vuv^{++}\overrightarrow{T}_i v$ that is of same length as the cycle T_i but possess more L-vertices. Note that both v and v^{++} are L-vertices, otherwise we would get the preceding Configuration. Further we may assume that any two neighbours of u in P^i are at least at distance four apart. Now we use the arguments from the first part of the proof of the statement 5 of the Proposition 5.3 that there are at least $1/r$ leaf neighbours of u. We delete this set of leaves and keep G_{i+1} an (r,k)-LKS graph according to Lemma 2.3.

All this can be done in time $\mathcal{O}(k)$ plus the degree of potentially added L-vertices.

Epoch 2

The Epoch 2 is conceptually similar to the first one. We know that all L-vertices of G_i lie in T_i, which means that G_i has to be small. Specifically, $2rk > |T_i| \geq |L_i| \geq rn_i$ implies that $n_i < 2k$. Furthermore, we get that $|S_i| \leq (1-r)n_i < 2(1-r)k$. If T_i is a path we proceed in the same manner as in Epoch 1 but we always succeed in Configuration 1, because otherwise we would have $|S_i| \geq |(N(u) \cup N(v)) \cap U_i| \geq 2(1-r)k+3$ contradicting the preceding inequality.

If T_i is a cycle and contains at least one consecutive pair of S-vertices, we delete the whole interval of S-vertices in T_i containing this pair, making T_{i+1} a path. Note that the number of vertices trimmed from T_i is then linear in the number of deleted edges between S-vertices from T_i. Otherwise we proceed with the last Configuration.

Configuration 5 If there are two vertices $u, v \in T_i$ such that $u, u^+, v, v^+ \in L_i$ and moreover there is $w \notin T_i$ incident with both u and v we define T_{i+1} as the path $u^+\overrightarrow{T}_i vwu\overleftarrow{T}_i v^+$.

If this does not happen we can follow the proof of Proposition 5.3 parts 4b and 5b and infer that we either get a longer path T_{i+1} (part 4b) or get a contradiction with the bound on size of T_i. All this can be done in time $\mathcal{O}(k)$.

Time complexity analysis

We will at first bound the number of steps done by the algorithm. Note that whenever we deleted vertices from T_i, except of Configuration 3, we argued that the number is proportional to the number of deleted edges between two S-neighbours in T_i. Because of invariant 4 we, thus, get that the overall amount of vertices deleted in all of these steps is $\mathcal{O}(|L|)$ and the number of these steps is also bounded by $\mathcal{O}(|L|)$.

This means that the number of steps with $|T_{i+1}| > |T_i|$ is also $\mathcal{O}(|L|)$. Finally, in some steps we have $|T_{i+1}| = |T_i|$ and we have not trimmed any vertices from T_i. Then we either switched an S-vertex with an L-vertex (Configuration 4) or created a cycle from a path. Number of steps of the first kind is at most $|L|$ because of invariant 3. Finally, there cannot be two consecutive steps of the algorithm of the second kind, so we finally get that the overall number of steps done is $\mathcal{O}(|L|)$. This also concludes the proof of correctness of the algorithm.

Now we analyse the complexity of steps of the algorithm. Note that, as we mentioned earlier, the complexity of all deletion operations can be amortised and its time complexity sums up to $\mathcal{O}(m)$. Now let us consider all steps of the algorithm in which we do not execute Configuration 3. All other Configurations and updating of used data structures spend time proportional to k plus the sum of degrees of L-vertices added to T_{i+1}. By invoking invariant 3 we conclude that the overall time complexity of all such steps is $\mathcal{O}(|L|k) \subseteq \mathcal{O}(m)$.

It remains to deal with the steps in which we executed Configuration 3. At first note that this means that we are analysing only the first Epoch. Further, when we, after the unsuccessful execution of Configuration 3, decided to erase S-neighbours of u and v, the time spent on the procedure is proportional to the number of erased edges and can be, thus, amortised in a straightforward manner. From now on let us consider only such steps that ended by successful execution of Configuration 3, thus setting T_{i+1} as the cycle created from the path T_i. We will call these steps *bad steps*.

Each bad step runs in time $\mathcal{O}(rk^2)$. We claim that between every two consecutive bad steps we have added at least $(1-2r)k$ vertices to U_i. For this we invoke the following observation: if T_j is a cycle and in steps $j, j+1, \ldots, j'$ we do not execute Configuration 2 (thus, we do not execute Configuration 3 neither) then $T_i, j \leq i \leq j'$ is either a cycle or a path with at least one endpoint outside T_j. This can be seen by checking the procedures Configuration 1 and Configuration 4. Note that this observation fails for Configuration 5 which is the reason for dividing the algorithm in two Epochs. When we are executing Configuration 2 we know that both endpoints of T_i have at least $(1-2r)k$ neighbours in U_i. But we know that the neighbours of the endpoint that was not in T_j could not be in U_j, thus $|U_i \setminus U_j| \geq (1-2r)k$.

Therefore the number of bad steps is at most $\frac{n}{(1-2r)k}+1$. Thus, the overall time spent on all these steps is $\mathcal{O}(\frac{1}{1-2r}\frac{n}{k} \cdot rk^2) = \mathcal{O}(\frac{1}{1-2r}nrk) = \mathcal{O}(\frac{1}{1-2r}m)$. For $r = 1/2$ we still get complexity $\mathcal{O}(m + |L|rk^2) = \mathcal{O}(mk)$. This finishes the analysis of time complexity. Finally note that the space complexity is linear even for $r = 1/2$.

Chapter 6

Conclusion

The thesis may be seen as an illustration of the remarkable power of the celebrated regularity lemma. Indeed, this deep result allows us not only to understand the structure of dense graphs, but also to apply this knowledge to solve various problems that seem otherwise hopelessly unapproachable. Embedding trees is certainly one of such problems.

The sparse decomposition of Ajtai, Komlós, and Szemerédi [HPS+15] shows, how an extremal graph theoretical question may lead to a general structural result. Hopefully, the results of the thesis will also eventually lead to better understanding of the structure of graphs.

Bibliography

[AKSSa] M. Ajtai, J. Komlós, M. Simonovits, and E. Szemerédi. On the approximative solution of the Erdős-Sós conjecture on trees. In preparation.

[AKSSb] M. Ajtai, J. Komlós, M. Simonovits, and E. Szemerédi. The solution of the Erdős-Sós conjecture for large trees. In preparation.

[AKSSc] M. Ajtai, J. Komlós, M. Simonovits, and E. Szemerédi. Some elementary lemmas on the Erdős-Sós conjecture on trees. In preparation.

[BD96] Stephan Brandt and Edward Dobson. The Erdős-Sós conjecture for graphs of girth 5. *Discrete Mathematics*, 150(1):411 – 414, 1996.

[BD07] Suman Balasubramanian and Edward Dobson. Constructing trees in graphs with no k2,s. *Journal of Graph Theory*, 56(4):301–310, 2007.

[BLW00] C. Bazgan, H. Li, and M. Woźniak. On the Loebl-Komlós-Sós conjecture. *J. Graph Theory*, 34(4):269–276, 2000.

[Coo09] O. Cooley. Proof of the Loebl-Komlós-Sós conjecture for large, dense graphs. *Discrete Math.*, 309(21):6190–6228, 2009.

[Die97] Reinhard Diestel. *Graph Theory*. Number 173 in Graduate Texts in Mathematics. Springer, 1997.

[Dob02] Edward Dobson. Constructing trees in graphs whose complement has no K 2,s. *Combinatorics, Probability and Computing*, 11(4):343–347, 2002.

[Dvo] Zdeněk Dvořák. Regularity lemma. https://iuuk.mff.cuni.cz/~rakdver/index.php?which=uceni&subject=kg3.

[EG59] P. Erdos and T. Gallai. On maximal paths and circuits of graphs. *Acta Mathematica Hungarica*, 10(3-4):337–356, 9 1959.

[ES46] P. Erdős and A. H. Stone. On the structure of linear graphs. *Bulletin of the American Mathematical Society*, 52:1087–1091, 1946.

[Fan13] Genghua Fan. The Erdős–Sós conjecture for spiders of large size. *Discrete Mathematics*, 313(22):2513 – 2517, 2013.

[FS07] Genghua Fan and Lingli Sun. The Erdős–Sós conjecture for spiders. *Discrete Mathematics*, 307(23):3055 – 3062, 2007.

[FS10] P Flajolet and R Sedgewick. *Analytic combinatorics*. Citeseer, 2010.

[GZ16] Agnieszka Görlich and Andrzej Zak. On Erdős-Sós conjecture for trees of large size. *Electr. J. Comb.*, 23(1):P1.52, 2016.

[HKP+17a] J. Hladký, J. Komlós, D. Piguet, M. Simonovits, M. Stein, and E. Szemerédi. The approximate Loebl–Komlós–Sós Conjecture I: The sparse decomposition. *SIAM J. Discrete Math.*, 31(2):945–982, 2017.

[HKP+17b] J. Hladký, J. Komlós, D. Piguet, M. Simonovits, M. Stein, and E. Szemerédi. The approximate Loebl–Komlós–Sós Conjecture II: The rough structure of LKS graphs. *SIAM J. Discrete Math.*, 31(2):983–1016, 2017.

[HKP+17c] J. Hladký, J. Komlós, D. Piguet, M. Simonovits, M. Stein, and E. Szemerédi. The approximate Loebl–Komlós–Sós Conjecture III: The finer structure of LKS graphs. *SIAM J. Discrete Math.*, 31(2):1017–1071, 2017.

[HKP+17d] J. Hladký, J. Komlós, D. Piguet, M. Simonovits, M. Stein, and E. Szemerédi. The approximate Loebl–Komlós–Sós Conjecture IV: Embedding techniques and the proof of the main result. *SIAM J. Discrete Math.*, 31(2):1072–1148, 2017.

[HLT02] P. E. Haxell, T. Luczak, and P. W. Tingley. Ramsey numbers for trees of small maximum degree. *Combinatorica*, 22(2):287–320, 2002. Special issue: Paul Erdős and his mathematics.

[HP16] J. Hladký and D. Piguet. Loebl-Komlós-Sós Conjecture: dense case. *J. Combin. Theory Ser. B*, 116:123–190, 2016.

[HPS+15] J. Hladký, D. Piguet, M. Simonovits, M. Stein, and E. Szemerédi. The approximate Loebl–Komlós–Sós conjecture and embedding trees in sparse graphs. *Electron. Res. Ann. Math. Sci.*, 22:1–11, 2015.

[HRSW16] F. Havet, B. Reed, M. Stein, and D. R. Wood. A Variant of the Erdős-Sós Conjecture. *ArXiv e-prints*, June 2016.

[KPR17] Tereza Klimosova, Diana Piguet, and Václav Rozhon. A skew version of the Loebl-Komlós-Sós conjecture. *Electronic Notes in Discrete Mathematics*, 61:743–749, 2017.

[KPR18] T. Klimošová, D. Piguet, and V. Rozhoň. A version of the Loebl-Komlós-Sóos conjecture for skewed trees. *ArXiv e-prints*, February 2018.

[McL05] Andrew McLennan. The Erdős-Sós conjecture for trees of diameter four. *Journal of Graph Theory*, 49(4):291–301, 2005.

[PS08] D. Piguet and M. J. Stein. Loebl-Komlós-Sós conjecture for trees of diameter 5. *Electron. J. Combin.*, 15(1):Research Paper 106, 11 pp. (electronic), 2008.

[PS12] D. Piguet and M. J. Stein. An approximate version of the Loebl-Komlós-Sós conjecture. *J. Combin. Theory Ser. B*, 102(1):102–125, 2012.

[Sof00] Sara Nadiv Soffer. The Komlós-Sós conjecture for graphs of girth 7. *Discrete Mathematics*, 214(1-3):279–283, 2000.

[SW97] Jean-François Saclé and Mariusz Woźniak. The Erdős-Sós conjecture for graphs without C4. *J. Comb. Theory Ser. B*, 70(2):367–372, July 1997.

[Tin10] Gary Tiner. On the Erdos-Sos conjecture for graphs on n = k+3 vertices. *Ars Comb.*, 95, 2010.

[WLL00] Min Wang, Guo-jun Li, and Ai-de Lui. A result of Erdős-Sós conjecture. *Ars Combinatoria*, 55, 01 2000.

[Woz96] Mariusz Wozniak. On the Erdős-Sós conjecture. *Journal of Graph Theory*, 21(2):229–234, 1996.

[Zha11] Y. Zhao. Proof of the $(n/2 - n/2 - n/2)$ conjecture for large n. *Electron. J. Combin.*, 18(1):Paper 27, 61, 2011.