An approximate version of the Loebl-Komlós-Sós conjecture

Diana Piguet* and Maya Jakobine Stein†

Abstract

Loebl, Komlós, and Sós conjecture that if at least half of the vertices of a graph G have degree at least some $k \in \mathbb{N}$, then every tree with at most k edges is a subgraph of G. Our main result is an approximate version of this conjecture for large enough $n = |V(G)|$, and k linear in n.

We extend our result to a slightly larger class of subgraphs. Namely, we show that G contains as subgraphs all bipartite connected graphs of order $k+1$ with at most $k+c$ edges, where c is some constant in n.

Also, we derive from our result an asymptotic bound for the Ramsey number of trees. We prove that $r(T_k, T_m) \leq k + m + o(k+m)$, provided that $\liminf(k/m), \liminf(m/k) > 0$.

1 Introduction

We explore how certain global assumptions on a graph G ensure the existence of specific subgraphs. More precisely, we are interested in finding trees as (not necessarily induced) subgraphs. The central conjecture in our investigations makes, to this end, assumptions on the median degree of G.

Conjecture 1 (Loebl, Komlós, Sós [5]). *Every graph on $n \in \mathbb{N}$ vertices of which at least $n/2$ have degree at least some $k \in \mathbb{N}$, contains as subgraphs all trees with at most k edges.*

*Supported by ITI grant no. 1M00216220808.
†Supported by FAPESP grant no. 05/54051-9.

The original version for $k = n/2$ was formulated by Loebl, the generalisation to arbitrary k is due to Komlós and Sós (see [5]).
A generalisation of an example due to Zhao [13] shows that the bound for the number of vertices with high degree in Conjecture 1 is close to best possible. It cannot be replaced by $n/2 - n/\sqrt{k} - n/k$ (if k divides n).

Our main result is an approximate version of Conjecture 1 for $k \in \Theta(n)$.

Theorem 2. *For every $\eta, q > 0$ there is an $n_0 \in \mathbb{N}$ such that for each graph G on $n \geq n_0$ vertices and each $k \geq qn$ the following is true.*
If at least $(1+\eta)n/2$ vertices of G have degree at least $(1+\eta)k$, then G contains as subgraphs all trees with at most k edges.

For arbitrary k, this has been conjectured by Ajtai, Komlós and Szemerédi in [1]. There, also a proof for the special case $k = n/2$ is given.

The exact version, Conjecture 1, is trivial for stars, and for trees that consist of two stars with adjacent centres. Bazgan, Li, and Woźniak [2] prove the conjecture for paths, and for other special cases. In a forthcoming article [9], the authors of this paper prove the Loebl–Komlós–Sós conjecture for trees of diameter at most 5 and second special case.
In Loebl's version with $k = n/2$, the conjecture has recently been proved by Zhao [13] for large enough graphs.

Zhao's result implies that the Ramsey number[1] $r(T_{k+1})$ of a tree T_{k+1} with k edges is at most $2k$, for large k. Bounds for Ramsey numbers of trees have been studied so far only for special cases (e.g. see [6]).
In the same way as the bound for $r(T_{k+1})$ follows from the Loebl conjecture, one can deduce from Conjecture 1, if true, a bound for the Ramsey number of trees T_{k+1}, T_{m+1} of different order $k+1$ and $m+1$. Namely, if the Loebl–Komlós–Sós conjecture holds, then $r(T_{k+1}, T_{m+1}) \leq k + m$. This upper bound has been conjectured in [5].
Using Theorem 2, we prove this to be asymptotically true.

Proposition 3. *Let T_{k+1} and T_{m+1} be trees of order $k+1$, resp. $m+1$. Then $r(T_{k+1}, T_{m+1}) \leq k + m + o(k+m)$, provided that $\liminf k/m > 0$ and $\liminf m/k > 0$.*

[1] The Ramsey number $r(H, H')$ of two graphs H, H' (or short $r(H)$, if $H = H'$) is defined as the minimal integer n such for every graph G or order at least n either H is a subgraph of G, or H' is a subgraph of the complement $\bar{G}$ of G.

It is not difficult to see that the exact bound of $r(T_{k+1}, T_{m+1}) \leq k+m$ also follows from a positive answer to the Erdős–Sós conjecture. This well-known conjecture states that each graph with average degree greater than $k-1$ contains all trees with at most k edges as subgraphs.
For partial results on the Erdős–Sós conjecture, see e.g. [3, 10, 12].

Our proof of Theorem 2 is inspired by Ajtai, Komlós and Szemerédi's proof of their approximate version of the Loebl conjecture [1]. We use the regularity lemma followed by a Gallai-Edmonds decomposition of the reduced cluster graph. This enables us to find a certain substructure in the cluster graph, which contains a large matching, and captures the degree condition on G. The tree T is then embedded mainly into the matching edges.
We shall see that in the case that $k \geq n/2$, it is not difficult to obtain the same structure as in [1]. Our proof then follows [1], providing all details.
In the case that $k < n/2$, however, the situation is more complex. We will have to content ourselves with a less favorable structure in the cluster graph, which complicates the embedding of T. For a brief outline of the crucial ideas we then employ, see Section 3.1. The full proof is given in the remainder of Section 3.

Using similar ideas of proof, we extend Theorem 2 in a new direction. We pursue the question which other subgraphs are contained in our graph G from Theorem 2.
Our third result affirms that we can replace the trees with bipartite graphs that have few more edges than trees.

Theorem 4. *For every $\eta, q > 0$ and for every $c \in \mathbb{N}$ there is an $n_0 \in \mathbb{N}$ so that for each graph G on $n \geq n_0$ vertices and each $k \geq qn$ the following is true.*
If at least $(1+\eta)n/2$ vertices of G have degree at least $(1+\eta)k$, then each connected bipartite graph Q on $k+1$ vertices with at most $k+c$ edges is a subgraph of G.

In particular, our graph G contains all even cycles of length at most $k+1$ (see Corollary 11).

An interesting observation due to A. Pór [personal communication] is the following. If Conjecture 1 holds for every k constant in n, then it is true for all k and n. Moreover, if we could prove the conjecture for some specific function k, tending to infinity, and all multiples ck with $0 < c \leq 1$, then it would also hold for all $k' \geq k$ (see Proposition 17).

The idea is to start with an assumed counterexample for the non-constant case (resp. for k'), and then take a large number of disjoint copies of it, so that k becomes constant in n. We thus reach a contradiction. For more details, see Section 4.3.
The same argument applies to the Erdős–Sós conjecture, and to the approximate version of the Loebl-Komlós-Sós conjecture. In the latter, the argument reflects that the case $k < n/2$ of Theorem 2 is more difficult to prove that the case $k \geq n/2$, and why the sparse case of Theorem 2 appears to be even harder.
In fact, by Pór's argument, a version for constant k would imply our Theorem 2, and moreover, we would no longer need the bound $n_0 \in \mathbb{N}$. Even more is true: if the version of Theorem 2 for constant k is true, then we get arbitrarily close to the exact version. More precisely, in that case a variant of Conjecture 1 holds, where the bound k on the degrees is replaced with the bound $k + 1$, and the bound $n/2$ on the number of vertices of large degree is replaced with $(n + 1)/2$. See Proposition 16.

Our paper is organised as follows. In Section 2.1, we introduce the regularity lemma and discuss some basic properties of regularity. Our tool for finding the desired structure of the cluster graph, Lemma 7, will be proved in Section 2.2. All of Section 3 is dedicated to the proof of our main result, Theorem 2. A detailed overview can be found in the first subsection.
In Section 4, we explore applications and generalisations of Theorem 2. Our asymptotic bound for Ramsey numbers of trees will be derived in Section 4.1. In Section 4.2, we extend Theorem 2, replacing trees with bipartite graphs that have few cycles. We investigate the potential of a possible sparse version of our result in Section 4.3.

2 Preliminaries

The purpose of this section is to introduce the two main tools used in the proofs of Theorem 2 and Theorem 4. The first of these tools is the well-known regularity lemma. The second is Lemma 7, which will give structural information on our graph G from Theorem 2 (and Theorem 4). We derive it from the Gallai-Edmonds matching theorem.

2.1 Regularity

In this subsection, we introduce the notion of regularity, state Szemerédi's regularity lemma, and review a few useful properties of regularity. All of this is well-known, so the advanced reader is invited to skip this section. For an instructive survey on the regularity lemma and its applications, consult [8].

Let us first go through some necessary notation. For a graph $G = (V, E)$, with $W \subseteq E$ and $S \subseteq V$, we will write $G - W$ for the subgraph $(V, E \setminus W)$ of G, and $G - S$ the subgraph of G which is obtained by deleting all vertices of S and all incident edges. If X and Y are disjoint subsets of the vertex set $V(G)$, then let $e(X, Y)$ denote the number of edges between X and Y. Write $N_Y(X)$ for the set of all neighbours in Y of vertices from X. For X, Y with $X \cap Y \neq \emptyset$, define $N_Y(X) := N_{Y \setminus X}(X)$.

A bipartite graph G with partition classes V_1 and V_2 is called (α, ε)-*regular* if for all subsets $V_1' \subseteq V_1$, $V_2' \subseteq V_2$ with $|V_1'| \geq \alpha|V_1|$ and $|V_2'| \geq \alpha|V_2|$, it is true that $|d(V_1, V_2) - d(V_1', V_2')| < \varepsilon$.
A partition $V_0 \cup V_1 \cup \cdots \cup V_N$ of $V(G)$ is called $(\alpha, \varepsilon; N)$-*regular*, if

- $|V_0| \leq \varepsilon n$ and $|V_i| = |V_j|$ for $i, j = 1, \ldots, N$,
- all but at most εN^2 pairs (V_i, V_j) with $i \neq j$ are (α, ε)-regular.

We are now able to state Szemerédi's regularity lemma.

Theorem 5 (Regularity lemma, Szemerédi [11]). *For every $\varepsilon, \alpha > 0$ and $m_0 \in \mathbb{N}$, there exist $M_0, N_0 \in \mathbb{N}$ so that every graph G of order $n \geq N_0$ admits an $(\alpha, \varepsilon; N)$-regular partition of its vertex set $V(G)$ with $m_0 \leq N \leq M_0$.*

Call the partition classes V_i of G *clusters*. Now, for each graph G, for each $(\alpha, \varepsilon; N)$-regular partition of $V(G)$, and for any density p define the *cluster graph*, or *reduced graph*, in the following standard way.
First, we construct an auxiliary graph G_p obtained from G by deleting all edges inside the clusters V_i, all edges that are incident with V_0, all edges between irregular pairs, and all edges between regular pairs (V_i, V_j) of density

$$\frac{e(V_i, V_j)}{|V_i||V_j|} < p.$$

Observe that

$$|E(G-G_p)| \leq N\frac{s^2}{2} + \varepsilon n^2 + \varepsilon N^2 s^2 + \frac{N^2}{2}ps^2 \leq (\frac{1}{2m} + 2\varepsilon + \frac{p}{2})n^2, \quad (1)$$

where $s := |V_i|$.
Now, the *cluster graph* $H = H_p$ on the vertex set $\{V_i\}_{1\leq i\leq N}$ has an edge V_iV_j for each pair (V_i, V_j) of clusters that has positive density in G_p. We shall prefer to work with the *weighted cluster graph* $\bar{H} = \bar{H}_p$ which we obtained from H by assigning weights

$$w(V_iV_j) := \frac{e(V_i, V_j)}{s}$$

to the edges $V_iV_j \in E(H)$.
In the setting of weighted graphs, the (weighted) *degree* of vertex v is defined as

$$\bar{\deg}(v) := \sum_{w\in N(v)} \omega(vw),$$

and the degree into a subset $W \subseteq V(\bar{H})$, where we only count the weights of v–W edges, is denoted by $\bar{\deg}_W(v)$. We shall adopt this notation for our weighted cluster graph $\bar{H}$. Similarly, for a subset $X \subseteq V_j$, we write

$$\bar{\deg}_X(V_i) := \frac{e(X, V_i)}{s}.$$

For a set $\mathcal{Y}$ of subsets of distinct clusters from $G_p - V_i$, we shall write $\bar{\deg}_{\mathcal{Y}}(V_i)$ for $\sum_{Y\in\mathcal{Y}} \bar{\deg}_Y(V_i)$.

We shall often use edges of $\bar{H}$ to represent the respective subgraph of G_p, or its vertex set. For example, an edge $e = CD \in E(\bar{H})$, might refer to the subgraph of G_p induced by $C \cup D$, or to $C \cup D$ itself, i.e. in $\bar{\deg}_e(V_i)$. And for a set $U \subseteq C \cup D$, we sometimes use the shorthand $e \cap U$ for $(C \cup D) \cap U$.

Let us review some basic properties of G_p and $\bar{H}$. A simple calculation shows that for $X \in V(\bar{H})$ and for $\mathcal{Y} \subseteq V(\bar{H})$ with $X \notin \mathcal{Y}$ we have that

$$\deg_{\mathcal{Y}}(v) > \bar{\deg}_{\mathcal{Y}}(X) - \varepsilon|\mathcal{Y}|s \text{ for all but at most } \alpha s \text{ vertices } v \text{ of } X, \quad (2)$$

and if $\mathcal{Y}'$ is a set of subsets of the clusters in $\mathcal{Y}$, each of size at least αs, then

$$\deg_{\mathcal{Y}'}(v) > \bar{\deg}_{\mathcal{Y}'}(X) - 2\varepsilon|\mathcal{Y}|s \text{ for all but at most } \alpha s \text{ vertices } v \text{ of } X. \quad (3)$$

These vertices v will be called *typical with respect to* $\mathcal{Y}$, or w.r.t. $\mathcal{Y}'$. If no confusion is possible, we call v simply *typical*.

2.2 The matching

The main interest in this subsection is Lemma 7, which will give us important structural information on the cluster graph H that corresponds to the graph G from Theorem 2 (or Theorem 4). A weaker variant of this lemma, Lemma 8 below, appeared in [1].

For the proof of Lemma 7, we need a simplified version of the Gallai-Edmonds matching theorem, a proof of which can be found for example in [4].
A 1-*factor*, or *perfect matching*, of a graph G is a 1-regular spanning subgraph of G. We call G 1-*factor-critical*, if for each $v \in V(G)$, there exists a perfect matching of $G - v$.

Theorem 6 (Gallai, Edmonds). *Every graph G contains a set $S \subseteq V(G)$ so that each component of $G - S$ is 1-factor-critical, and so that there is a matching in G that matches the vertices of S to vertices of different components of $G - S$.*

We are now ready for one of the key tools in the proof of Theorem 2.

Lemma 7. *Let $\bar{H}$ be a weighted graph on N vertices, and let $K \in \mathbb{Q}$. Let L be the set of those vertices $v \in V(\bar{H})$ with $\bar{\deg}(v) \geq K$. If $|L| > N/2$, then there are two adjacent vertices $A, B \in L$, and a matching M in $\bar{H}$ such that one of the following holds.*

(a) *M covers $N(A \cup B)$,*

(b) *M covers $N(A)$, and $\bar{\deg}_{L \cup M}(B) \geq K/2$. Moreover, each edge in M has at most one endvertex in $N(A)$.*

Proof. Observe that we may assume that $Y := V(\bar{H}) - L$ is independent.
Now, Theorem 6 applied to the unweighted version of $\bar{H}$ yields a set $S \subseteq V(\bar{H})$ and a matching M'. Fix S and choose M' so that it contains a maximal number of vertices of Y. Let M consist of the union of M' and some maximal matching of $\bar{H} - V(M')$.
Set $L' := L \setminus S$. Clearly, if there is an edge AB with endvertices $A, B \in L'$, then (a) holds. Therefore, we may assume that L' is independent.

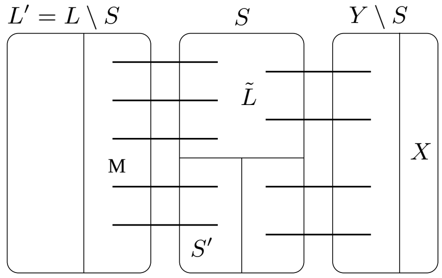


Figure 1: The graph $\bar{H}$ with the matching M, and sets L, S and Y.

Then, each edge of $\bar{H}$ that is not incident with S has one endvertex in L', and one in Y. Now, consider any component C of $\bar{H}-S$. Since C is 1-factor-critical, we have that $|(C-x)\cap Y| = |(C-x)\cap L'|$, for every $x \in V(C)$. Hence, C consists of only one vertex, and so must every component of $\bar{H}-S$. Denote by X the subset of Y that is not covered by M. Set $\tilde{L} := N(L')\cap L \subseteq S$ (see Figure 1). Now, if there is a vertex $B \in \tilde{L}$ whose weighted degree into $\bar{H}-X$ is at least $K/2$, then B, together with any of its neighbours A in L', satisfies (b). So, we may assume that for each $B \in \tilde{L}$,

$$\bar{\deg}_{\bar{H}-X}(B) < K/2, \tag{4}$$

and hence $\bar{\deg}_X(B) \geq K/2$.
On the other hand, $\bar{\deg}_{\tilde{L}}(x) < K$ for each $x \in X$. Thus, by double (weighted) edge-counting, it follows that

$$|X| \geq \frac{|\tilde{L}|}{2}. \tag{5}$$

Set $S' := S \cap Y$. By (4), the weighted degree of $\tilde{L}\cup S'$ into L' is less than $|\tilde{L}|K/2 + |S'|K$, while each vertex of L' has weighted degree at least K into $\tilde{L}\cup S'$. Thus, again by double edge-counting, and by (5),

$$|X| + |S'| \geq \frac{|\tilde{L}|}{2} + |S'| > |L'|. \tag{6}$$

Furthermore, since Y is independent, M matches $S' \subseteq Y$ to L'. Thus $|L'| \geq |S'| + |L \setminus M|$, and so, by (6),

$$|X| > |L \setminus M|.$$

Since $|L| > \frac{N}{2}$, this implies that M contains a edge AB with both $A, B \in L$. We may assume that $A \in L'$ and $B \in \tilde{L}$. By (4), B has a neighbour D in X. Hence, the matching $M' \cup BD \setminus AB$ covers more vertices of Y than M' does, a contradiction to the choice of M'. □

Note that in the case $K \geq N/2$ the situation in Lemma 7 is less complicated. Then, it suffices to observe that since clearly $|S| \leq |V(\bar{H} - S)|$, not all components of $\bar{H} - S$ can be of order 1. (Indeed, otherwise there would be a vertex $v \in L$ with $N(v) \subseteq S$, a contradiction.) Thus, as each component of $\bar{H} - S$ is 1-factor-critical, there exists an L'–L' edge, and conclusion (a) of Lemma 7 holds.
This proves the following lemma, which appeared in [1].

Lemma 8. *If $K \geq N/2$, then Lemma 7 always yields case (a).*

In the case $k \geq n/2$, this observation simplifies our proof of Theorem 2 considerably. The reader only interested in this special case may skip Subsections 3.4, the last third of Subsection 3.5 and Subsection 3.7.
We shall not make use of Lemma 8 in our proof of Theorem 2.

3 Proof of Theorem 2

The organisation of this section is as follows. The first subsection is devoted to an outline of our proof, highlighting the main ideas, leaving out all details. In Subsection 3.2, we shall apply the regularity lemma to our graph G and use Lemma 7 to find a substructure of the corresponding weighted cluster graph $\bar{H}$, that will facilitate the embedding of the tree T^*.
We shall prepare T^* for this by cutting it into small pieces in Subsections 3.3 and 3.4. Then, in Subsection 3.5, we partition the matching given by Lemma 7, according to the decomposition of the tree T^*. What remains is the actual embedding procedure, which we divide into the two cases given by Lemma 7, and treat separately in Subsections 3.6 and 3.7.

3.1 Overview

In this subsection, we shall give an outline of our proof of Theorem 2. All details can be found in the subsequent subsections. The overview is meant to help in understanding the general ideas we shall employ, but may also be skipped, as it contains no necessary information needed to follow the proof of Theorem 2.

In order to facilitate reading, let us restate our result before we sketch its proof.

Theorem 3. *For every $\eta, q > 0$ there is an $n_0 = n_0(\eta, q) \in \mathbb{N}$ such that for each graph G on $n \geq n_0$ vertices and each $k \geq qn$ the following is true.*
If at least $(1+\eta)n/2$ vertices of G have degree at least $(1+\eta)k$, then G contains as subgraphs all trees with at most k edges.

So, assume that we are given $\eta > 0$ and $q > 0$. The regularity lemma applied to parameters depending on η and q yields an $n_0 \in \mathbb{N}$. Now, let $n \geq n_0$, let $k \geq qn$, let G be a graph of order n that satisfies the condition of Theorem 2, and let T^* be a tree with k edges. We wish to find a subgraph of G that is isomorphic to T^*, i.e. we would like to *embed* T^* in G.
In order to do so, consider the weighted cluster graph $\bar{H}$ corresponding to G that is given by the regularity lemma. Denote by $L \subseteq V(\bar{H})$ the set of those clusters that have degree at least $(1+\pi')k$ in $\bar{H}$, where $\pi' = \pi'(\eta, q) > 0$. Apply Lemma 7 to $\bar{H}$ and $K := (1+\pi')k$. The rest of our proof will be divided into two cases, corresponding to the two possible conclusions (a) and (b) of Lemma 7.
As the technical details for these two cases overlap, we chose not to separate the two cases completely from each other later on. In this outline, however, we think it is more instructive to present first the easier proof for case (a), and then turn our attention to case (b).

In case (a) of Lemma 7, we shall decompose T^* into small subtrees (of order much below ηk) and a small set SD of vertices (of constant order in n), so that between any two of our subtrees lies a vertex from SD (the name SD stands for 'seeds'). In fact, SD is the disjoint union of two sets SD^A and SD^B, and each tree of $T^* - SD$ is adjacent to only one of these two sets. Denote the set of trees adjacent to SD^A by T_A, and the set of trees adjacent to SD^B by T_B. The formal definition of SD, T_A and T_B can be found in Section 3.3.
Next, in Section 3.5, we partition the matching M from Lemma 7 into M_A and M_B. This is done in a way so that $\deg_{M_A}(A)$ is large enough so that $\bigcup T_A$ fits into M_A, and $\deg_{M_B}(B)$ is large enough so that $\bigcup T_B$ fits into M_B.
Finally, in Section 3.6, we embed SD^A in A and SD^B in B and use the regularity of the edges in H to embed the small trees of $T_A \cup T_B$, one after the other, levelwise, into $M_A \cup M_B$. The order of this embedding procedure will be such that the already embedded part of T^* is always connected.

Moreover, the structure of our decomposition of T^*, and the fact that we embed the trees from $T_A \cup T_B$ in the matching edges, ensures that the predecessor of any vertex $r \in SD^A \cup SD^B$ is embedded in a cluster that is adjacent to A, respectively to B (in which we wish embed r). This enables us to embed all of SD in $A \cup B$, as planned.
An important detail of our embedding technique is that we shall always try to *balance* the embedding in the matching edges, in the sense that the used part of either side should have about the same size (cf. conditions (iv) and (v) of Section 3.6). We only allow for an unbalanced embedding if the degree of A resp. B into one of the endclusters of the concerned edge is already 'exhausted' (cf. condition (a) in Section 3.6). In practice, this means that whenever we have the choice into which endcluster of an edge $e \in M$ we embed the root of some tree of $T_A \cup T_B$, we shall choose the side carefully.
In this manner, we can ensure that all of T^* will fit into M (or more precisely into the corresponding subgraph of G). This finishes the embedding of T^* in case (a) of Lemma 7.

In case (b) of Lemma 7, it is not possible to partition the matching M into M_A and M_B so that $\bigcup T_A$ fits into M_A and $\bigcup T_B$ fits into M_B, as in case (a). More precisely, for any partition of M into M_A and M_B, if $\deg_{M_A}(A)$ allows for the embedding of a forest with vertex set of size t_A in M_A, then $\deg_{M_B \cup L}(B)$ only guarantees for the embedding of a forest with vertex set of size at most $(k - t_A)/2$ in the subgraph of G_p induced by M_B and the edges incident with L', where $L' := L \setminus M$.
Hence, we will embed only part of T^* in a first phase, and deal with the rest of T^* in a second phase. In the first phase, we shall exclude from the embedding only trees that are (each) adjacent to only one vertex from SD. This has the advantage that the part of the tree embedded in the first phase is connected (in this way, we avoid the difficulty of having to connect already embedded parts of T^* in the second phase).
This idea suggests a natural modification of our sets $T_A \cup T_B$, in the following way. Denote by $\bar{T}_A$ the set of those trees from T_A that are adjacent to only one vertex from SD^A, and similarly define $\bar{T}_B$. Assume that

$$|V(\bigcup \bar{T}_A)| \geq |V(\bigcup \bar{T}_B)|,$$

and set $T' := (T_A \cup T_B) \setminus (\bar{T}_A \cup \bar{T}_B)$.

Our plan now is to first embed the trees from $T' \cup T_B$ and to postpone the embedding of $\bigcup \bar{T}_A$ to later. Observe that as we explained above, even leaving out $\bar{T}_A$, we might be unable to find a partition of M so that we can embed the trees from $T_A \setminus \bar{T}_A$ in one side, and $\bigcup T_B$ in the other. In fact, if we wish to embed all of $T^* - \bar{T}_A$ in M, we cannot hope to fit more than $\bigcup \bar{T}_B$ into M_B and edges incident with L'.

So, we shall partition M into M_F and $\bar{M}_B$ so that $\deg_{M_F}(A)$ allows for the embedding of $\bigcup T'$, and $\deg_{\bar{M}_B \cup L}(B)$ allows for the embedding of $\bigcup \bar{T}_B$. This actually means that the place we reserved for the embedding of $\bigcup(T_B \setminus \bar{T}_B)$ lies in M_F. Therefore, we shall 'switch' this forest to T_A.

Let us explain what we mean by *switching*. For each tree $t \in T_B \setminus \bar{T}_B$, delete all vertices from t that are adjacent to SD^B in T^* and add them to SD^A. Put the components of what remains of t into T_A.

After switching all trees $t \in T_B \setminus \bar{T}_B$, denote by T_F the (enlarged) set $T_A \setminus \bar{T}_A$. That is, T_F consists of all trees from the original $T_A \setminus \bar{T}_A$, together with all trees we generated by switching. Note that we have thus at most tripled the size of SD. Also, each tree from T_F and $\bar{T}_A$ is adjacent only to the enlarged SD^A, which we denote by $\bar{SD}^A$, and each tree from $\bar{T}_B$ is still adjacent only to SD^B. For details on the switching procedure, consult Section 3.4.

It remains to embed T^* in G, which is done in Section 3.7. We first embed the vertices from $\bar{SD}^A \cup SD^B$ in $A \cup B$, embed $\bigcup T_F$ in M_F, and embed $\bigcup \bar{T}_B$ in $\bar{M}_B$ and edges incident with L', in the same way as in case (a). Note that we have to use not only edges from M as before, but also edges of H that are incident with L'. But this is not a problem: for each tree, we are able to find a suitable edge because of the high degree of the clusters from L'. This completes the first phase of our embedding in case (b), which for better readability we shall split into two subphases later on.

In the remaining third phase we wish to embed $\bigcup \bar{T}_A$. We shall now use all of M, forgetting about the partition. The neighbours of the trees from $\bar{T}_A$ in $\bar{SD}^A$ have already been embedded in the first phase. Having chosen their images carefully then, ensures that now they have still large enough degree into what is not yet used of M. This ensures that there is enough place for $\bigcup \bar{T}_A$ in M.

Also, it is essential here that each edge of M meets $N(A)$ in at most one cluster. The reason is that parts of these clusters might have been used in the first phase of the embedding. So, some of the edges involved might be unbalanced, in the sense above, because e. g. the degree of B was such that we were not able to choose the endcluster in which we embedded the roots of the trees from T'_B. This could be a problem, if we now count on using the neighbours of A in both endclusters for our embedding. However, if each

edge of M has at most one endcluster in $N(A)$, then it is irrelevant whether the embedding is balanced or not in these edges.
The embedding itself of $\bigcup \bar{T}_A$ is done as before. This finishes the sketch of our proof in case (b).

3.2 Preparations

We shall now prove Theorem 2. First of all, we fix a few constants depending on η and q. Set

$$\pi := \min\{\eta, q\}, \quad \varepsilon := \frac{\pi^4 q}{5 \cdot 10^5}, \quad \alpha := \frac{\pi^5 q}{25 \cdot 10^7} \quad \text{and} \quad m_0 := \frac{500}{q\pi^2}.$$

The regularity lemma (Theorem 5) applied to ε, α and m_0 yields natural numbers M_0 and N_0.
Fix

$$\beta := \frac{\varepsilon}{M_0}, \quad p := \frac{\pi^2 q}{250} \quad \text{and} \quad n_0 := \max\{N_0, \frac{9\varepsilon}{p/2 - 7\alpha}\}.$$

Thus our constants satisfy the following relations

$$\frac{1}{n_0} \ll \beta \ll \alpha \ll \varepsilon \ll \frac{1}{m_0} < p \ll \pi \leq q,$$

where $a \ll b$ stands for the fact that $a < \frac{\pi}{100} b$.
In particular, p satisfies

$$4\varepsilon + \frac{1}{m_0} < p < \frac{\pi^2}{250}. \tag{7}$$

Let $n \geq n_0$, let $k \geq qn$, and let G be a graph of order n which has at least $(1+\eta)\frac{n}{2}$ vertices of degree at least $(1+\eta)k$. Suppose T^* is a tree of order $k+1$. Our aim is to find an embedding $\varphi : V(T^*) \to V(G)$ that preserves adjacency.
Now, by Theorem 5 there exists an $(\alpha, \varepsilon; N)$-regular partition of $V(G)$, with $m_0 \leq N \leq M_0$. As in Section 2.1, let G_p be the subgraph of G that preserves exactly the edges between regular pairs of density at least p.
By (1) and by (7),

$$|E(G - G_p)| < pn^2 < \frac{\pi^2}{8} kn.$$

Thus, for all but at most $\frac{\pi}{4}n$ vertices v, we have that $\deg_{G_p}(v) \geq \deg_G(v) - \frac{\pi}{2}k$. Hence,

$$G_p \textit{ has at least } (1+\frac{\pi}{2})\frac{n}{2} \textit{ vertices of degree at least } (1+\frac{\pi}{2})k.$$

Let $\bar{H} = \bar{H}_p$ be the weighted cluster graph corresponding to G_p. A simple calculation shows that there are more than $(1+\frac{\pi}{10})\frac{N}{2}$ clusters that (each) contain more than αs vertices of degree at least $(1+\frac{\pi}{2})k$ in G_p.
By (2), each such cluster X must itself have degree

$$\bar{\deg}(X) > (1+\frac{\pi}{2})k - \varepsilon n > (1+\frac{\pi}{5})k.$$

Then Lemma 7 applied to $\bar{H}$ and $K := (1+\frac{\pi}{5})k$ yields an edge $AB \in E(\bar{H})$ with $A, B \in L$, together with a matching M' of $\bar{H}$, which satisfy (a) or (b) of Lemma 7. Obtain M from M' by deleting all edges that are incident with A or with B. In case (a) of Lemma 7, we calculate that

$$\bar{\deg}_M(A), \bar{\deg}_M(B) \geq (1+\frac{\pi}{5})k - \frac{3n}{N} \geq (1+\frac{\pi}{5} - \frac{3}{qm_0})k \geq (1+\frac{\pi}{10})k,$$

and similarly in case (b), it follows that

$$\bar{\deg}_M(A) \geq (1+\frac{\pi}{10})k \quad \text{and} \quad \bar{\deg}_{M\cup L}(B) \geq (1+\frac{\pi}{10})\frac{k}{2}.$$

Thus, for the remainder of our proof of Theorem 2 we shall work with the assumption that there is a matching M of $\bar{H}$ and vertices $A, B \notin V(M)$ so that

1. $\bar{\deg}_M(A), \bar{\deg}_M(B) \geq (1+\frac{\pi}{10})k$, or
2. $\bar{\deg}_M(A) \geq (1+\frac{\pi}{10})k$, $\bar{\deg}_{M\cup L}(B) \geq (1+\frac{\pi}{10})\frac{k}{2}$, and each cluster in $N(A)$ meets a different edge of M.

3.3 Partitioning the tree

In this section, we shall cut our tree into small pieces. More precisely, we shall define a set $SD \subseteq V(T^*)$, and sets T_A and T_B of disjoint small subtrees of T^* which are connected through the vertices from SD. Moreover, SD together with the union of all trees from $T_A \cup T_B$ will span T^*.

Fix a root R of T^*. For a vertex x of a subtree $T \subseteq T^*$, denote by $T(x)$ the subtree of T induced by x and all vertices v greater than x in the tree-order of T^*, i. e. all vertices v such that the path between the root R and v contains the vertex x. If $R \notin V(T)$, then define the *seed* $sd(T)$ of T as the maximal vertex of T^* which is smaller than every vertex of T.
Our sets $SD = SD^A \cup SD^B$, T_A and T_B will satisfy:

(I) $SD^A \cap SD^B = \emptyset$,

(II) $R \in SD^A$, and any other vertex $x \in SD$ lies at even distance to R if and only if $x \in SD^A$,

(III) $T_A \cup T_B$ consists of the components of $T^* - SD$,

(IV) $|V(T)| \leq \beta k$, and $sd(T) \in SD$, for each $T \in T_A \cup T_B$,

(V) $\max\{|SD^A|, |SD^B|\} < \frac{2}{\beta}$, and

(VI) $e(V(\bigcup T_A), SD^B) = \emptyset$, and $e(V(\bigcup T_B), SD^A) = \emptyset$.

Let us first define SD. To this end, we shall inductively find vertices x_i, and define auxiliary trees $T^i \subseteq T^*$. Set $T^0 := T^*$.
In step $i \geq 1$, let $x_i \in V(T^*)$ be maximal in the tree-order of $V(T^{i-1})$ with

$$|V(T^{i-1}(x_i))| > \beta k, \tag{8}$$

and define

$$T^i := T^{i-1} - (T^{i-1}(x_i) - x_i),$$

as illustrated in Figure 2(a). If there is no vertex satisfying (8), then set $x_i := R$, and stop the definition process. Say our process stops in some step j. Let A' be the set of all x_i, $i \leq j$, with even distance to the root R, and let B' be the set of all other x_i.
Since for each $x \in A' \cup B'$ we have deleted at least $\beta k - 1$ vertices from T^*, it follows that

$$|A' \cup B'| \leq \frac{k}{\beta k - 1} < \frac{2}{\beta}.$$

For the sake of condition (VI), we shall add a few more vertices to our sets A' and B', which will result in the desired SD.
Let $\mathcal{C}$ be the set of the components of $T^* - (A' \cup B')$. For each $T \in \mathcal{C}$ with $sd(T) \in A'$, denote by $A(T)$ the set of vertices adjacent to B'. Similarly,

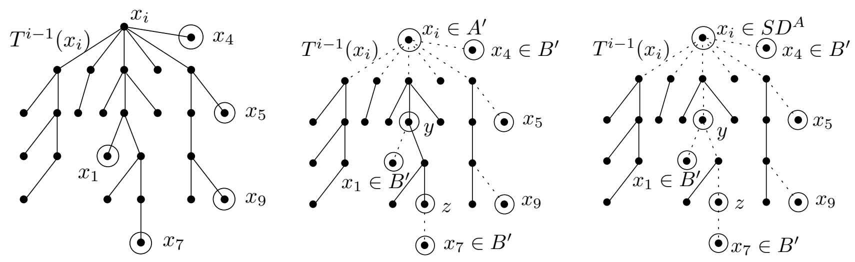
if $sd(T) \in B'$, then denote by $B(T)$ the set of vertices adjacent to A' (cf. Figure 2(b)). Set

$$SD^A := A' \cup \bigcup_{T \in \mathcal{C}} A(T), \quad \text{and} \quad SD^B := B' \cup \bigcup_{T \in \mathcal{C}} B(T)$$

and set $SD := SD^A \cup SD^B$.
Finally, we shall define T_A and T_B. Let $\mathcal{C}'$ be the set of the components of $T^* - SD$. Set

$$T_A := \{T \in \mathcal{C}' : \; sd(T) \in SD^A\} \qquad \text{and} \qquad T_B := \{T \in \mathcal{C}' : \; sd(T) \in SD^B\},$$

as shown in Figure 2(c), and set $V_A := \bigcup_{T \in T_A} V(T)$ and $V_B := \bigcup_{T \in T_A} V(T)$.



(a) In step i, we find vertex x_i.

(b) Say $x_i \in A'$. We add y and z to $A(T)$.

(c) $T^{i-1}(x_i) - SD \subseteq T_A$.

Figure 2: Phases of the partition of T^*.

Observe that our six conditions are clearly met. This finishes our manipulation of the tree T^* in Case 1.

3.4 The switching

In Case 2 from Section 3.2, we shall not only cut our tree to small pieces (cf. Section 3.3), but also switch some of our small subtrees from one of the two sets T_A, T_B to the other. We achieve this by adding some more vertices to SD, thus naturally refining our partition of T^*.

Set

$$\begin{aligned}\bar{T}_A :=& \{T \in T_A : e(V(T), SD - sd(T)) = \emptyset\}, \quad \text{and}\\ \bar{T}_B :=& \{T \in T_B : e(V(T), SD - sd(T)) = \emptyset\}.\end{aligned}$$

We may assume that

$$|V(\bigcup_{T \in \bar{T}_A} T)\ | \geq |V(\bigcup_{T \in \bar{T}_B} T)\ |. \tag{9}$$

Now, consider a tree $T \in T_B \setminus \bar{T}_B$ as in Figure 3(a). Denote by $S(T)$ the set of all vertices in $V(T)$ that in T^* are adjacent to some vertex of SD^B. For illustration see Figure 3(b).
Set

$$\bar{SD}^A := SD^A \cup \bigcup_{T \in T_B \setminus \bar{T}_B} S(T) \quad \text{and} \quad \bar{SD} := \bar{SD}^A \cup SD^B.$$

Finally, define

$$T'_A := \bigcup_{T \in T_B \setminus \bar{T}_B} \{C : C \text{ is a component of } T - S(T)\}, \quad T_F := (T_A \setminus \bar{T}_A) \cup T'_A,$$

and set

$$V_F := \bigcup_{T \in T_F} V(T) \quad \text{and } \bar{V}_B := \bigcup_{T \in \bar{T}^B} V(T).$$

Observe that our sets $\bar{SD}$, T_F, and $\bar{T}_B$ still satisfy conditions (I)-(IV) and (VI) from Section 3.3. Instead of (V), we now have the similar

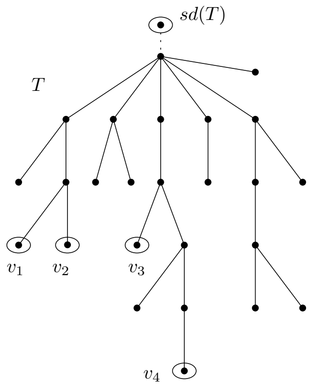
(V)' $|\bar{SD}| \leq \frac{8}{\beta}$,

since by the definition of $\bar{SD}^A$ we know that $|\bar{SD}^A| \leq |SD^A| + 2|SD^B| \leq \frac{6}{\beta}$.

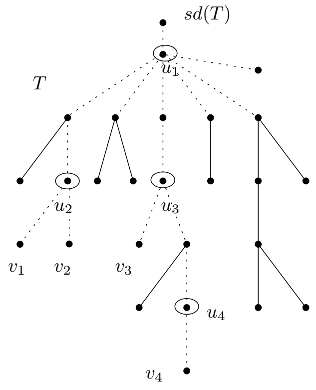
3.5 Partitioning the matching

In this subsection, we shall divide the matching M into two, into which we will later embed the two parts $\bigcup T_A$, $\bigcup T_B$, respectively $\bigcup T_F$, $\bigcup \bar{T}_B$, of T^* that we defined in Subsection 3.3, resp. in Subsection 3.4.

For this, we will need the following number-theoretic lemma, which appeared also in [1]. We give a short proof.



(a) A tree $T \in T_B \setminus \bar{T}_B$, with $sd(T), v_1, v_2, v_3, v_4 \in SD^B$.



(b) The set $S(T) = \{u_1, \dots, u_4\}$, and the subtrees of T generated by the switching.

Figure 3: The switching procedure.

Lemma 9. *Let I be a finite set, and let $a, b, \Delta > 0$. For $i \in I$, let $\alpha_i, \beta_i \in (0, \Delta]$. Suppose that*

$$\frac{a}{\sum_{i\in I}\alpha_i} + \frac{b}{\sum_{i\in I}\beta_i} \leq 1. \tag{10}$$

Then there is a partition of I into I_a and I_b such that $\sum_{i\in I_a}\alpha_i > a - \Delta$ and $\sum_{i\in I_b}\beta_i \geq b$.

Proof. Define a total order $\preceq$ on I in a way that $i \preceq j$ implies $\frac{\alpha_i}{\beta_i} \leq \frac{\alpha_j}{\beta_j}$ for all $i, j \in I$. Let $\ell \in I$ be minimal in this order with $a \geq \sum_{i \succ \ell}\alpha_i$.
Set $I_a := \{i \in I : \; i \succ \ell\}$ and set $I_b := I \setminus I_a$. It is clear that $\sum_{i\in I_a}\alpha_i > a - \Delta$, by the definition of ℓ and as $\alpha_\ell \leq \Delta$. So, all we have to show is that $\sum_{i\in I_b}\beta_i \geq b$.
Indeed, suppose otherwise. Then by (10), and by the definition of ℓ, we have that

$$\frac{\sum_{i\in I_b}\beta_i}{\sum_{i\in I}\beta_i} < \frac{b}{\sum_{i\in I}\beta_i} \leq \frac{a - \sum_{i\in I_a}\alpha_i}{\sum_{i\in I}\alpha_i} + \frac{b}{\sum_{i\in I}\beta_i} \leq 1 - \frac{\sum_{i\in I_a}\alpha_i}{\sum_{i\in I}\alpha_i} = \frac{\sum_{i\in I_b}\alpha_i}{\sum_{i\in I}\alpha_i}.$$

Multiply the two sides of this inequality with $\sum_{i\in I}\alpha_i\cdot\sum_{i\in I}\beta_i$, subtract $\sum_{i\in I_b}\alpha_i\cdot\sum_{i\in I_b}\beta_i$, and divide by $\sum_{i\in I_a}\beta_i\sum_{i\in I_b}\beta_i$ to obtain

$$\frac{\alpha_\ell}{\beta_\ell}\leq\frac{\sum_{i\in I_a}\alpha_i}{\sum_{i\in I_a}\beta_i}<\frac{\sum_{i\in I_b}\alpha_i}{\sum_{i\in I_b}\beta_i}\leq\frac{\alpha_\ell}{\beta_\ell},$$

(where the first and last inequality follow from the definition of $\preceq$). This yields the desired contradiction. $\square$

We shall now apply Lemma 9 to partition our matching $M=\{e_i\}_{i\leq|M|}$. We do this separately for the two cases from Section 3.2.
In Case 1, we set

$$a:=|V_A|+\frac{\pi k}{20},\quad b:=|V_B|+\frac{\pi k}{20},\quad\text{and}\quad\Delta:=2s.$$

For $i\leq|M|$, set $\alpha_i:=\bar{\deg}_{e_i}(A)\leq\Delta$, and $\beta_i:=\bar{\deg}_{e_i}(B)\leq\Delta$. Observe that

$$\frac{a}{\sum_{i=1}^{|M|}\alpha_i}+\frac{b}{\sum_{i=1}^{|M|}\beta_i}\leq\frac{|V_A|+|V_B|+\frac{\pi k}{10}}{(1+\frac{\pi}{10})k}\leq 1.$$

Hence, Lemma 9 yields a partition of M into M_A and M_B such that

$$\bar{\deg}_{M_A}(A)>|V_A|+\frac{\pi k}{40}\quad\textit{and}\quad\bar{\deg}_{M_B}(B)>|V_B|+\frac{\pi k}{40}.\tag{11}$$

In Case 2, set

$$a:=|V_F|+\frac{\pi k}{20},\quad b:=|\bar{V}_B|+\frac{\pi k}{40},\quad\text{and}\quad\Delta:=2s.$$

For $i=1,\dots|M|$, again set $\alpha_i:=\bar{\deg}_{e_i}(A)$, and $\beta_i:=\bar{\deg}_{e_i}(B)$. For $i=|M|+1,\dots,|M|+|L'|$, set $\alpha_i:=0$, and set $\beta_i:=\bar{\deg}_{C_i}(B)$, where C_i is the ith cluster in L', where $L':=L\setminus M$. Observe that by (9),

$$|\bar{V}_B|\leq\frac{(1-|V_F|)}{2}k.$$

Now, it is easy to verify that the condition of Lemma 9 holds. We thus obtain a partition of M into M_F and $\bar{M}_B$ such that

$$\bar{\deg}_{M_F}(A)>|V_F|+\frac{\pi k}{40}\quad\text{and}\quad\bar{\deg}_{\bar{M}_B\cup L'}(B)\geq|\bar{V}_B|+\frac{\pi k}{40}.\tag{12}$$

3.6 The embedding in Case 1

In this subsection, we shall complete the proof of Theorem 2 under the assumption that Case 1 of Section 3.2 holds. So, we assume that there are an edge $AB \in E(\bar{H})$ and disjoint matchings $M_A = \{e_1, e_2, \dots, e_{m_A}\}$ and $M_B = \{e_{m_A+1}, e_{m_A+2}, \dots, e_{|M|}\}$ in $\bar{H} - \{A, B\}$ as in Section 3.5. These matchings together with the sets SD, T_A and T_B from Section 3.3 satisfy (11).

We wish to embed T^* in G_p. Our embedding φ will be defined in $|SD|$ steps.
In each step $i \geq 1$, we choose a suitable vertex $r_i \in SD$ and embed it together with all trees from

$$T_i := \{T \in T_A \cup T_B : \ sd(T) = r_i\}.$$

Set $V_0 := \emptyset$ and for $i \geq 1$ let

$$V_i := V_{i-1} \cup \{r_j\} \cup \bigcup_{T \in T_j} V(T).$$

We start with $r_1 := R$, and in each step $i > 1$, we shall choose a vertex $r_i \in SD$ that is adjacent to some V_j with $j < i$. The seed r_i will be embedded in a vertex $v_i \in A \cup B$, while T_i will be mapped to an edge from M (or more precisely, to the corresponding subgraph of G_p).
Set $U_0 := \emptyset$, and once φ is defined on V_i, set $U_i := \varphi(V_i)$. Furthermore, we shall define indices $a_i \geq a_{i-1}$ and $b_i \geq b_{i-1}$, starting with $a_0 := 1$ and $b_0 := m_A + 1$. These indices will mark the edges we are currently embedding in.
For each $i \geq 0$, the following conditions will hold.

(i) $|(A \cup B) \cap U_i| \leq i$,

(ii) if $v \in V_i \cap N(SD^A)$, resp. $v \in V_i \cap N(SD^B)$, then $\varphi(v)$ has at least $\frac{p}{2}s$ neighbours in A, resp. in B,

(iii) $U_i \cap e_j = \emptyset$ for each $j \in (a_i, m_A] \cup (b_i, m]$,

(iv) for $CD = e_{a_i}$, if $||C \cap U_i| - |D \cap U_i|| > \beta k$, then $\min\{\bar{\deg}_C(A), \bar{\deg}_D(A)\} \leq \min\{|C \cap U_i|, |D \cap U_i|\} + (2\alpha + \varepsilon)s + \beta k$,

(v) for $CD = e_{b_i}$, if $||C \cap U_i| - |D \cap U_i|| > \beta k$, then $\min\{\bar{\deg}_C(B), \bar{\deg}_D(B)\} \leq \min\{|C \cap U_i|, |D \cap U_i|\} + (2\alpha + \varepsilon)s + \beta k$,

(vi) $\Sigma_i^{m_A}(A) \geq |V_A \setminus V_i|$ and $\Sigma_i^{m}(B) \geq |V_B \setminus V_i|$,

where for $\ell \in \mathbb{N}$ we define

$$\Sigma_i^{\ell}(A) := \sum_{j=a_i}^{\ell} \left(\bar{\deg}_{e_j}(A) - |e_j \cap U_i| - \mu_A\right), \qquad \textit{with} \quad \mu_A := \frac{\pi k}{40|M_A|},$$

and

$$\Sigma_i^{\ell}(B) := \sum_{j=b_i}^{\ell} \left(\bar{\deg}_{e_j}(B) - |e_j \cap U_i| - \mu_B\right), \qquad \textit{with} \quad \mu_B := \frac{\pi k}{40|M_B|}.$$

Observe that properties (i)–(v) trivially hold for $i = 0$. Property (vi) holds for $i = 0$ because of (11).

So, suppose now that we are in some step $i \geq 1$ of our embedding process. Choose $r_i \in SD$ as detailed above.
Let us assume that $r_i \in SD^A$, the case when $r_i \in SD^B$ is analogous. Set $b_i := b_{i-1}$, and let $a_i \in \{a_{i-1}, \dots, m_A\}$ be minimal with

$$\Sigma_{i-1}^{a_i}(A) \geq |V_i \setminus V_{i-1}|. \tag{13}$$

We embed r_i in a vertex $\varphi(r_i) \in A$ that is typical with respect to B, typical w. r. t. $\bigcup_{a_{i-1}<j<a_i} e_j$, and typical w. r. t. each endcluster of $e_{a_{i-1}}$ and of e_{a_i}. properties (i) and (ii) for $i-1$ ensure that if v is the predecessor of r_i in T^*, then $\varphi(v)$ has at least $\frac{ps}{2} - i$ neighbours in $A \setminus U_{i-1}$. By (2), at most $6\alpha s$ of these vertices do not have the required properties. Hence, there are at least $(\frac{p}{2} - 6\alpha)s - i \geq 1$ suitable vertices we may choose $v_i := \varphi(r_i)$ from.
Next, we embed the trees from T_i. This is done inductively, in $|T_i|$ substeps of step i. In each substep $j \geq 1$, we shall embed one tree $t^j \in T_i$. Denote by $V^{<j} \subseteq V(T_i)$ the set $\bigcup_{\ell<j} V(t^\ell)$ of vertices we have already embedded before substep j and set $U^{<j} := U_{i-1} \cup \varphi(V^{<j})$. Note that in particular, $U_{i-1}^{<1} = U_{i-1}$.
For each $j \geq 0$, and for every edge $CD \in \{e_{a_{i-1}}, \dots, e_{a_i}\}$ with

$$||C \cap U_{i-1}| - |D \cap U_{i-1}|| \leq \beta k$$

the following will hold.

(a) If $||C \cap U^{<j+1}| - |D \cap U^{<j+1}|| > \beta k$, then
$\min\{\deg_C(v_i), \deg_D(v_i)\} \leq \min\{|C \cap U^{<j+1}|, |D \cap U^{<j+1}|\} + 2\alpha s + \beta k.$

For $j = 0$, assertion (a) is void, and there is nothing to show.
So, assume now that we are in substep $j \geq 1$. That is, $\varphi(v)$ has been defined for all $v \in V^{<j}$, and we are about to embed t^j. We claim that there is an edge $e \in \{e_{a_{i-1}}, \ldots, e_{a_i}\}$ which satisfies

$$\deg_e(v_i) - |e \cap U^{<j}| \geq \frac{8}{p}(\alpha s + \beta k). \tag{14}$$

Indeed, suppose there is no such edge. Then, by (2), by (13), and by the choice of v_i, we have that

$$\begin{aligned}
&\frac{8}{p}(\alpha s + \beta k)(a_i - a_{i-1} + 1) > \\
&> \sum_{\ell = a_{i-1}}^{a_i} (\bar{\deg}_{e_\ell}(A) - 2\varepsilon s - |e_\ell \cap U_{i-1}|) - |U^{<j} \setminus U_{i-1}| \\
&\geq \Sigma_{i-1}^{a_i}(A) + (a_i - a_{i-1} + 1)(\mu_A - 2\varepsilon s) - |V^{<j}| \\
&\geq (a_i - a_{i-1} + 1)(\mu_A - 2\varepsilon s),
\end{aligned}$$

implying that $\frac{8}{p}(\alpha + \varepsilon q) + 2\varepsilon > \frac{\pi q}{20}$, a contradiction. This proves the existence of an edge e that satisfies (14).
Moreover, a calculation similar to the one above shows that,

$$\text{as long as } |V^{<j}| < \Sigma_{i-1}^{a_i - 1}(A), \text{ we can choose } e \neq e_{a_i}. \tag{15}$$

We shall do so whenever we can.

So, assume now that we have chosen an edge e for which (14), and if possible, different from e_{a_i}. Clearly, we can write $e = CD$ such that

$$\deg_C(v_i) - |C \cap U^{<j}| \geq \frac{4}{p}(\alpha s + \beta k). \tag{16}$$

We claim that furthermore

$$|D \setminus U^{<j}| \geq \frac{2}{p}(2\alpha s + \beta k). \tag{17}$$

Indeed, suppose for contradiction that (17) does not hold. Then (16) implies that

$$|C \cap U^{<j}| \leq s - \frac{4}{p}(\alpha s + \beta k) < |D \cap U^{<j}| - \frac{2}{p}\beta k$$

(recall that $s = |C| = |D|$). Hence, by (a) for $j-1$, and by (iv) for $i-1$,

$$\min\{\deg_C(v_i), \deg_D(v_i)\} \leq |C \cap U^{<j}| + 2\alpha s + \beta k.$$

Thus, by (14),

$$\begin{aligned} \frac{8}{p}(\alpha s + \beta k) \leq \; & \deg_C(v_i) + \deg_D(v_i) - |e \cap U^{<j}| \\ & < s + 2\alpha s + \beta k - |D \cap U^{<j}| \\ & < |D \setminus U^{<j}| + 2(\alpha s + \beta k). \end{aligned}$$

So, $|D \setminus U^{<j}| > \frac{6}{p}(\alpha s + \beta k)$, a contradiction to our assumption that (17) does not hold. We have thus shown (17).

Finally, we shall embed t^j into the endclusters of e. Write $V(t^j) = r \cup L_1 \cup L_2 \cup \ldots$, where r is the root of t^j and L_ℓ is the ℓth level of t^j (i. e. the set of vertices at distance ℓ to r).
First, suppose that $\deg_{D \setminus U^{<j}}(v_i) \leq 2\alpha s$. In this case, choose $\varphi(r) \in N_{C \setminus U^{<j}}(v_i)$, avoiding the (by (3)) at most $2\alpha s$ vertices that are not typical w. r. t. one of the sets A and $D \setminus U^{<j}$. This is possible because of (16).
Embed the rest of $V(t^j)$ levelwise, choosing for $\varphi(L_\ell)$ unused vertices of $D \setminus U^{<j}$ that are typical with respect to $C \setminus U^{<j}$, if ℓ is odd; and choosing vertices of $C \setminus U^{<j}$ that are typical with respect to A and w. r. t. $D \setminus U^{<j}$, if ℓ is even. Observe that hence the image of any vertex $v \in V(t^j) \cap N(SD^A)$ has at least $\frac{p}{2}s$ neighbours in A, as required for (ii).
Now, suppose that $\deg_{D \setminus U^{<j}}(v_i) > 2\alpha s$. In this case, we may alternatively wish to embed r in D. We do so in either of the following cases

1. $|\bigcup_{\ell \in \mathbb{N}} L_{2\ell-1}| > |\bigcup_{\ell \in \mathbb{N}} L_{2\ell}|$ and $|C \setminus U^{<j}| \geq |D \setminus U^{<j}|$, or
2. $|\bigcup_{\ell \in \mathbb{N}} L_{2\ell-1}| < |\bigcup_{\ell \in \mathbb{N}} L_{2\ell}|$ and $|C \setminus U^{<j}| \leq |D \setminus U^{<j}|$,

and otherwise embed r in C, as before. After having thus chosen a place for the root r, the rest of t^j is embedded analogously as above (possibly swapping the roles of C and D). This completes the embedding of t^j.

Let us prove property (a) for j. To this end, assume that there is an edge $CD \in \{e_{a_{i-1}}, \ldots e_{a_i}\}$ such that $||C \cap U_{i-1}| - |D \cap U_{i-1}|| \leq \beta k$ and $||C \cap U^j_{i-1}| - |D \cap U^j_{i-1}|| > \beta k$. Now, if $||C \cap U^{<j}| - |D \cap U^{<j}|| > \beta k$, then (a) for j follows from (a) for $j-1$.

So suppose otherwise, that is

$$||C \cap U^{<j}| - |D \cap U^{<j}|| \leq \beta k. \tag{18}$$

This is only possible if in step j, we could not choose into which of C and D we would embed the root of t^j. This implies that

$$\min\{\deg_{C \setminus U^{<j}}(v_i), \deg_{D \setminus U^{<j}}(v_i)\} \leq 2\alpha s.$$

Using (18), this gives

$$\begin{aligned}
\min\{\deg_C(v_i), \deg_D(v_i)\} \leq\ & \max\{|C \cap U^{<j}|, |D \cap U^{<j}|\} + 2\alpha s \\
\leq\ & \min\{|C \cap U^{<j}|, |D \cap U^{<j}|\} + 2\alpha s + \beta k \\
\leq\ & \min\{|C \cap U^{<j+1}|, |D \cap U^{<j+1}|\} + 2\alpha s + \beta k,
\end{aligned}$$

as desired for (a). This completes substep j.

Once all $t^j \in T_i$ are embedded, step i terminates. Conditions (i), (ii), (iii), (v), and the second part of (vi) hold for i, as they hold for $i-1$, and by our choice of $\varphi(V_i \setminus V_{i-1})$. Property (iv) for i follows either from (iv) for $i-1$, or from (a) for $j = |T_i|$.
In order to see the first part of (vi) for i, recall that we chose $\varphi(\bigcup T_i)$ in edges $e \neq e_{a_i}$ whenever possible. So, by (15), it follows that

$$|e_{a_i} \cap (U_i \setminus U_{i-1})| \leq |V_i \setminus V_{i-1}| - \Sigma_{i-1}^{a_i - 1}(A).$$

Hence, by (vi) for $i-1$, and by (iii) for i, we have that

$$\begin{aligned}
\Sigma_i^{m_A}(A) &= \Sigma_{i-1}^{m_A}(A) - \Sigma_{i-1}^{a_i-1}(A) - \sum_{j=a_i}^{m_A} |e_j \cap (U_i \setminus U_{i-1})| \\
&\geq |V_A \setminus V_{i-1}| + |e_{a_i} \cap (U_i \setminus U_{i-1})| - |V_i \setminus V_{i-1}| - |e_{a_i} \cap (U_i \setminus U_{i-1})| \\
&= |V_A \setminus V_i,
\end{aligned}$$

as desired.
This completes the embedding of the tree T^* in G_p for Case 1. It remains to prove Theorem 2 for Case 2, which we shall do in the next subsection.

3.7 The embedding in Case 2

We shall now complete the proof of Theorem 2 under the assumption that Case 2 of Section 3.2 holds. That is, there are an edge $AB \in E(\bar{H})$ and disjoint matchings $M_F := \{e_1, \dots, e_{m_F}\}$ and $\bar{M}_B := \{e_{m_F+1}, \dots, e_{|M|}\}$ in $\bar{H} - \{A, B\}$ together with sets $\bar{SD}$, T_F and $\bar{T}_B$ from Sections 3.3 and 3.4 satisfying (12) from Section 3.5. Set $m_B := |\bar{M}_B|$.

First, we determine which trees of $\bar{T}_B$ are going to be embedded in $\bar{M}_B$ and which in edges incident with L'. So, let $T_B^M \subseteq \bar{T}_B$ be maximal with

$$\bar{\deg}_{\bar{M}_B}(B) \geq |\bigcup_{T \in T_B^M} V(T)| + \frac{\pi k m_B}{40(m_B + |L'|)}. \tag{19}$$

Set $T_B^L := \bar{T}_B \setminus T_B^M$. Let $V_B^M := \bigcup_{T \in T_B^M} V(T)$ and let $V_B^L := \bar{V}_B \setminus V_B^M$. Observe that if $T_B^M \neq \bar{T}_B$, then the maximality of T_B^M ensures that

$$\bar{\deg}_{\bar{M}_B}(B) < |V_B^M| + \frac{\pi k m_B}{40(m_B + |L'|)} + \beta k.$$

Hence, by (12), either $T_B^L = \emptyset$, or

$$\bar{\deg}_{L'}(B) \geq |V_B^L| + \frac{\pi k |L'|}{80(m_B + |L'|)}. \tag{20}$$

Our embedding will be defined in three phases. In the first phase, we shall embed all vertices from $\bar{SD}$ in $A \cup B$, embed the trees from T_F in edges of M_F, and embed $\bigcup T_B^M$ in edges of $\bar{M}_B$. In the second phase, we shall embed $\bigcup T_B^L$ in edges incident with $L' \cap N(B)$, and in the third phase, we shall embed $\bigcup \bar{T}_A$ in the remaining space inside edges from both M_F and $\bar{M}_B$.

The first phase is defined in $|\bar{SD}|$ steps, and will be similar to the embedding from Section 3.6. In each step $i \geq 1$, we shall embed a vertex $r_i \in \bar{SD}$ together with all trees from

$$T_i := \{T \in T_F \cup T_B^M : sd(T) = r_i\}.$$

Set $V_0 := \emptyset$, and for $i \geq 1$, set

$$V_i := V_{i-1} \cup \{r_j\} \cup \bigcup_{T \in T_j} V(T).$$

Start with $r_1 := R$. In each subsequent step, we choose an $r_i \in \bar{SD}$ that is adjacent to some V_j with $j < i$.
Let $U_i := \varphi(V_i)$.Furthermore, for $i \geq 0$, set

$$V_i^F := V_i \cap V_F, \quad \text{and} \quad V_i^B := V_i \cap V_B^M.$$

We again define an index b_i, which plays the same role as b_i from Section 3.6, starting with $b_0 := m_F + 1$.

For $i \geq 0$, our embedding φ satisfies the following conditions.

(i) $|(A \cup B) \cap U_i| \leq i$,

(ii) if $v \in V_i \cap N(\bar{SD}^A)$, resp. $v \in V_i \cap SD^B$, then $\varphi(v)$ has at least $\frac{p}{2}s$ neighbours in A, resp. in B,

(iii) $e_\ell \cap U_i = \emptyset$ for each ℓ with $b_i < \ell \leq m$,

(iv) for $e_{b_i} = CD$, if $||C \cap U_i| - |D \cap U_i|| > \beta k$, then $\min\{\bar{\deg}_C(B), \bar{\deg}_D(B)\} \leq \min\{|C \cap U_i|, |D \cap U_i|\} + (2\alpha + \varepsilon)s + \beta k$,

(v) $\Sigma_i^M(B) \geq |V_B^M \setminus V_i^B|$,

where

$$\Sigma_i^\ell(B) := \sum_{j=b_i}^{\ell} \left(\bar{\deg}_{e_j}(B) - |e_j \cap U_i| - \mu_B,\right) \quad \textit{with} \quad \mu_B := \frac{\pi k}{40(|M_B| + |L'|)}.$$

These properties reflect those from Case 1. Observe that we no longer need to define an index a_i, nor do we require properties (iii)–(v) for the A-side. This is due to the fact that each edge from M has only one endvertex in $N(A)$, which will simplify our embedding technique.
Observe that for $i = 0$, conditions (i)–(iv) trivially hold, and (v) holds because of (19). In addition to conditions (i)-(v), we shall require for $i \geq 1$ and for $v_i := \varphi(r_i)$ that

(vi) $\deg_M(v_i) \geq (1 + \frac{\pi}{20})k$, if $r_i \in \bar{SD}^A$,

(vii) $\deg_{L'}(v_i) \geq |V_B^L| + |L'|\frac{\pi k}{100N}$ or $V_B^L = \emptyset$, if $r_i \in SD^B$.

These latter two conditions will be needed in the second and the third phase to embed the trees from $\bigcup T_B^L$ and $\bigcup \bar{T}_A$.

Suppose now that we are in some step $i \geq 1$, and wish to embed a seed r_i together with the trees from T_i.
First, assume that $r_i \in \bar{SD}^B$. Then the embedding process is very similar to the one in Case 1. The only difference is that we shall have a few more conditions on the typicality of $\varphi(v_i)$. As before, choose $b_i \geq b_{i-1}$ minimal with

$$\Sigma_{i-1}^{b_i}(B) \geq |V_i \setminus V_{i-1}|. \tag{21}$$

Then, we choose $\varphi(r_i) = v_i \in B$ not only typical with respect to A, w. r. t. $\bigcup_{\ell=b_{i-1}+1}^{b_i-1} e_\ell$, and w. r. t. each endcluster of $e_{b_{i-1}}$ and of e_{b_i}, but also typical with respect to L'. We can thus ensure that v_i satisfies (vii). Indeed, by (2) and by (20), we have that

$$\deg_{L'}(v_i) \geq \bar{\deg}_{L'}(B) - |L'|\varepsilon s \geq |V_B^L| + |L'|\frac{\pi k}{100N},$$

unless $V_B^L = \emptyset$.
The rest of the embedding is analogous to Case 1. Conditions (i)–(v) are shown in the same way, employing an analogue of property (a) from Case 1. Property (vi) holds trivially.

Now, suppose that $r_i \in \bar{SD}^A$. This case is even simpler, as each edge from M meets $N(A)$ in at most one cluster. We choose $\varphi(r_i) = v_i \in A$ typical w. r. t. B, w. r. t. $\bigcup_{e \in M_F} e$, and also typical with respect to $\bigcup_{C \in N(A)} C$, thus ensuring property (vi). With the help of (2) and (12) we find suitable edges into which we embed the trees from T_i. The verification of conditions (i)–(v) is identical. Property (vii) holds trivially.
This ends the first phase of our embedding process.

The second phase of our embedding process is again defined in $|\bar{SD}|$ steps. Assume that $V_B^L \neq \emptyset$ (otherwise we skip the second phase). In each step $i \geq 1$, we embed the trees from

$$T_i^B := \{T \in T_B^L : \ sd(T) = r_i\}$$

in edges incident with L'. Suppose that we are at substep j of this procedure, i. e. that we have already embedded the trees $t^1, \ldots, t^{j-1}$ of T_i^B. Denote by $U^{<j}$ the set of vertices used so far for the embedding. By (vii), we have that

$$\sum_{C\in L'} \deg_{C\setminus U^{<j}}(v_i) \geq |V_B^L| + |L'|\frac{\pi}{100}\frac{k}{N} - |\bigcup_{C\in L'} C\cap U^{<j}| > |L'|\frac{\pi}{100}\frac{k}{N}.$$

It is not difficult to calculate that hence there is a cluster $C \in L'$ with

$$\deg_{C\setminus U^{<j}}(v_i) \geq \frac{2}{p}(\alpha s + \beta k).$$

Embed the root r of t^j in a vertex $v \in C$ which is typical with respect to $\bigcup_{X\in V(\bar{H})} X$. Then, by (2), and by the definition of L, we have that

$$\deg_{V(G_p)\setminus U^{<j}}(v) \geq (1+\frac{\pi}{5})k - \varepsilon n - |U^{<j}| \geq \frac{\pi}{10}k. \tag{22}$$

Next, we wish to find a suitable neighbour D of C in $V(\bar{H})$ so that we can embed t^j into the edge CD. Using (22), one can show that there is a $D \in V(H)$ such that

$$deg_{D\setminus U^{<j}}(v) \geq \frac{2}{p}(\alpha s + \beta k).$$

We can now embed t^j levelwise in the edge CD, in the same way as before, when we embedded in edges from M.
This ends the definition of the embedding of $\bigcup \bar{T}_B$, and thus the second phase.

Also the third phase of our embedding process takes place in $|\bar{SD}|$ steps, where in each step $i \geq 1$, we embed the trees from

$$T_i^F := \{T \in \bar{T}_A : \ sd(T) = r_i\}.$$

The embedding process is identical to the one from the first phase when $r_i \in \bar{SD}^A$, with the difference that now we use edges from both M_F and M_B. That is, at substep j, when we have embedded the trees $t^1, \dots, t^{j-1}$ of T_i^F, property (vi) enables us to find an edge $CD \in M$ with

$$\min\{\deg_{C\setminus U^{<j}}(v_i), |D \setminus U^{<j}|\} \geq \frac{2}{p}(\alpha s + \beta k).$$

Then, embed t^j levelwise in the edges CD as before.
We thus embed all of $\bigcup \bar{T}_A$, which terminates the third phase of our embedding process. This completes the proof of Theorem 2.

4 Extensions and applications

In this last section, we explore applications and generalisations of Theorem 2. In Section 4.1 we show how our theorem implies an asymptotic upper bound on the Ramsey number of trees. We extend Theorem 2 so that it allows for embedding subgraphs other than trees in Section 4.2, and in Section 4.3, we turn our attention to the yet unresolved sparse case.

4.1 A bound on the Ramsey number of trees

Based on ideas from [5] and using Theorem 2, we prove the following.

Lemma 10. *For all $0 < \varepsilon < 1/4$, and for each $r \in (0,1]$, there exists an $n_0 \in \mathbb{N}$ such that the following holds for each $n \geq n_0$.*
If G is a graph on $\lceil(1+\varepsilon)n\rceil$ vertices, and T_{k+1} and T_{m+1} are trees of order $k+1, m+1 \in \mathbb{N}$, such that $k/m = r$ and $k+m = n$, then either T_{k+1} is a subgraph of G, or T_{m+1} is a subgraph of the complement $\bar{G}$ of G.

Recall that $r(T_{k+1}, T_{m+1})$ denotes the Ramsey number of the trees T_{k+1} and T_{m+1}. So, Lemma 10 implies an asymptotic upper bound on the Ramsey number of trees.

Proposition 3. *The Ramsey number $r(T_{k+1}, T_{m+1})$ of T_{k+1} and T_{m+1} is at most $k+m+o(k+m)$, provided that $\liminf(k/m), \liminf(m/k) > 0$.*

The sharp bound $k+m$ has been conjectured in [5].

Proof of Lemma 10. Apply Theorem 2 to $\varepsilon/8$ and $r/4$ to obtain an $n_0 \in \mathbb{N}$. Now, assume that k, m and n are as in Lemma 10. Let G be a graph on $\lceil(1+\varepsilon)n\rceil$ vertices.
For simplicity, let us assume that the order of G is $(1+\varepsilon)n$. Clearly, either at least half of the vertices of G have degree at least $k + \frac{\varepsilon}{2}n$ or in the complement $\bar{G}$ of G, at least half of the vertices have degree at least $m + \frac{\varepsilon}{2}n$.
First, suppose that the former of these assertions is true. We show that then $T_{k+1} \subseteq G$.
Choose a set L of size $\lceil\frac{1}{2}(1+\varepsilon)n\rceil$ which contains only vertices of degree at least $k + \frac{\varepsilon}{2}n$. Delete $\lfloor\frac{\varepsilon}{4}n\rfloor$ vertices from $V(G) \setminus L$. This yields a graph G' on $n' = \lceil(1+\frac{3}{4}\varepsilon)n\rceil \geq n_0$ vertices with at least $|L| \geq \frac{1}{2}(1+\frac{\varepsilon}{8})n'$ vertices of degree at least

$$k + \frac{\varepsilon}{4}n \geq (1 + \frac{\varepsilon}{4})k.$$

So, since $k \geq \frac{r}{4}n'$, and by Theorem 2, T_{k+1} is a subgraph of G'. Hence, $T_{k+1} \subseteq G$, as desired.

Now, assume that the second assertion above holds, that is, in the complement $\bar{G}$ of G, at least $\frac{1}{2}(1 + \varepsilon)n$ vertices have degree at least $m + \frac{\varepsilon}{2}n$. We then embed T_{m+1} in $\bar{G}$. This is done analogously. It suffices to observe that

$$m \geq rm \geq \frac{r}{2}n \geq \frac{r}{4}n'.$$

□

4.2 Graphs with few cycles

The question we pursue in this subsection is whether the condition of Theorem 2 allows for embedding other graphs on $k+1$ vertices, apart from trees. J. Foniok [personal communication] asked if we may add an edge to our tree T^* and still embed it in G. Inspired by this question, we show that, if we add only constantly many edges to a tree of order $k+1$, forming only even cycles, then the thus obtained graph embeds in G.

Theorem 4. *For every $\eta, q > 0$ and for every $c \in \mathbb{N}$, there exists an $n_0 \in \mathbb{N}$ such that for each graph G on $n \geq n_0$ vertices and each $k \geq qn$, the following holds.*
If G has at least $(1 + \eta)n/2$ vertices of degree at least $(1 + \eta)k$, then each connected bipartite graph Q on $k + 1$ vertices with at most $k + c$ edges is a subgraph of G.

In particular, the condition of Theorem 2 allows for embedding even cycles in G:

Corollary 11. *For every $\eta, q > 0$ there is an $n_0 \in \mathbb{N}$ so that for all graphs G on $n \geq n_0$ vertices and each $k \geq qn$ the following is true.*
If at least $(1 + \eta)n/2$ vertices of G have degree at least $(1 + \eta)k$, then G contains all even cycles of length at most $k + 1$.

Note that the graph Q from Theorem 4 need not even be connected (since we may simply add some edges to make it connected and then apply Theorem 4).

Corollary 12. *For every $\eta, q > 0$ and for every $c \in \mathbb{N}$, there exists an $n_0 \in \mathbb{N}$ such that for each graph G on $n \geq n_0$ vertices and each $k \geq qn$, the following holds.*
If G has at least $(1+\eta)n/2$ vertices of degree at least $(1+\eta)k$, then each bipartite graph Q on $k+1$ vertices, with $\ell \in \mathbb{N}$ components and at most $k+c-\ell$ edges, is a subgraph of G.

Observe that our argument for the bound on Ramsey number from Subsection 4.1 also applies here. With an identical proof, we obtain the following.

Corollary 13. *Let Q_1 and Q_2 be any two bipartite graphs of order $k+1$, and $m+1$ respectively, as in Theorem 4 or in Corollary 12. Then, $r(Q_1, Q_2) \leq k+m+o(k+m)$, as long as $\liminf k/m, \liminf m/k > 0$.*

Our proof of Theorem 4 follows closely the lines of the proof of Theorem 2. We embed a rooted spanning tree (T^*, R) of Q, and choosing φ carefully, we ensure the adjacencies for the edges from $E(Q) \setminus E(T^*)$.

Proof of Theorem 4. Set $\pi := \min\{\eta, q\}$ and set

$$\alpha' := \frac{1}{c}\alpha^{c+1}, \quad \varepsilon := \frac{\pi^4 q}{5 \cdot 10^5}, \quad \text{and} \quad m_0 := \frac{500}{\pi^2 q},$$

where α is the constant from the proof of Theorem 2. As in the proof of Theorem 2, the regularity lemma applied to α', ε, and m_0, yields natural numbers N_0 and M_0'. Set $M_0 := \max\{M_0', c\}$, and define β, n_0 and p accordingly.
Now, let G be a graph on $n \geq n_0$ vertices which satisfies the condition of Theorem 4, let $k \geq qn$, and let Q be a connected bipartite graph of order $k+1$ with at most $k+c$ edges, with a spanning tree T^*. Fix a root R in T^*. Denote by M^* the subgraph of Q induced by the edges in $E(Q) \setminus E(T^*)$ and let N^* be the set of predecessors of $V(M^*)$ in the tree order of T^*.
We decompose T^* as in Section 3.3, with the difference that we now add the vertices from $V(M^*) \cup N^*$ to the sets A' and B' (from the definition of SD), depending on the parity of their distance in T^* to R. In this way, and since Q is bipartite, we obtain, after the switching, two independent sets $\bar{SD}^A$ and SD^B so that

$$|\bar{SD}^A| + |SD^B| \leq \frac{8}{\beta} + 8c < \frac{9}{\beta},$$

which is constant in n.

The definition of our the embedding φ is similar as in the proof of Theorem 2, except for some extra precautions we take for vertices from $V(M^*) \cup N^*$. At step i, for each vertex $v \in \bar{SD}^A$, define

$$N_v^i := \bigcap_{\ell=1}^{j} N(\varphi(u_\ell)) \cap A,$$

where $u_1, \dots u_j$ are the already embedded neighbours of v in SD^B. If none of the neighbours of v in SD^B has been embedded before step i, then set $N_v^i := A$. Analogously define N_v^i for $v \in SD^B$.

In each step i of our embedding process, we shall ensure the following.

(i) If $v \in V(M^*)$ is not yet embedded, then $|N_v^i| \geq \left(\frac{p}{2}\right)^j s$,

where $j = j(v, i)$ is the number of neighbours of v in $\bar{SD}^A$ resp. SD^B that have already been embedded before step i.

Observe that in step $i = 0$, either $N_v^0 = A$ or $N_v^0 = B$, and therefore condition (i) is met.

Suppose that at step $i \geq 1$ of our embedding process, we are about to embed a vertex $v = r_i \in V(M^*) \cup N^*$. Assume that $v \in \bar{SD}^A$ (the case when $v \in SD^B$ is analogous). Denote by $w_1, \dots, w_\ell$ the neighbours of v in $V(M^*)$ that have not been embedded yet.

Now, embed v in a vertex from N_v^{i-1} that satisfies the at most six conditions of typicality from the proof of Theorem 2, except the typicality w. r. t. B, which we replace by typicality w. r. t. each $N_{w_j}^{i-1}$, for $1 \leq j \leq \ell$. This is possible, since our embedding scheme and the condition on the number of edges of Q ensure that v has at most $c+1$ neighbours in Q that are already embedded. Thus, it follows from (i) for $i-1$ and for v that

$$|N_v^{i-1}| \geq \left(\frac{p}{2}\right)^{c+1} s > \alpha' s.$$

So, there are at least

$$\left(\left(\frac{p}{2}\right)^{c+1} - (6+c)\alpha'\right) s - |\bar{SD}| \geq \left(\frac{\pi^2 q}{250}\right)^{c+1} s - \frac{9}{\beta} > 1$$

unused typical vertices we can choose $\varphi(v)$ from.

Finally, observe that since we chose v typical w. r. t. each $N_{w_j}^{i-1}$, we have ensured property (i) for i and for every $v' \in V(M^*)$ that is not yet embedded.

This completes the proof of Theorem 4. $\square$

4.3 The sparse case

In this final section we shall discuss the sparse version of our main result, Theorem 2. Recall that we had to restrict k to be linear in n in our result. So, a question naturally arising is whether Theorem 2 remains true for arbitrary k. Such a version would make at the same time the trees we look for tiny, and the host graph sparse.

Question 14. *Does Theorem 2 remain true for $k \in o(n)$?*

In other words, we would like to know whether for every $\eta > 0$ and every $k(n)$, possibly sublinear, there exists an n_0 such that the following is true. If at least $(1+\eta)n/2$ vertices of a graph G of order $n \geq n_0$ have degree at least $(1+\eta)k(n)$, then every tree with $k(n)$ edges is a subgraph of G.

A. Pór [personal communication] observed that we actually only need to resolve the problem for all cases when k is constant in n.

Indeed, suppose that for every $k \in \mathbb{N}$ and every $\eta > 0$ there is an $n_0(k, \eta)$ such that the assertion of Theorem 2 is true for k, η, and all $n \geq n_0(k, \eta)$. In order to reach a contradiction, suppose that the assertion is false for some specific choice of $k'(n)$, and η'. Then, there exist a graph G on of order say n', and a tree T with $k'(n')$ edges, so that at least $(1+\eta')n'/2$ vertices of G have degree at least $(1+\eta')k'(n')$, but $T \nsubseteq G$.

Now, consider the graph G' which consists of $n_0(k'(n'), \eta')$ disjoint copies of G. Still, at least $(1+\eta')|G'|$ vertices of G' have degree at least $(1+\eta')k'(n')$, and still, $T \nsubseteq G'$. Thus, the approximate version of Conjecture 1 is false for (the constant) $k = k'(n')$, for η', and for $|G'| \geq n_0(k'(n'), \eta')$, a contradiction to our assumption above.

In order to prove the approximate version of Conjecture 1 for all k, it would thus be enough to prove it for all k constant in n. Hence, a positive answer to the following question would imply a positive answer to Question 14.

Question 15. *Does Theorem 2 remain true for all $k \in O(1)$?*

Actually, a positive answer to Question 15 implies even more. As first observed by O. Pangrác [personal communication], the argument above in fact proves the approximate version of Conjecture 1 for all n (without any bound n_0).

So, if the answer to Question 15 is positive, then for every $\eta > 0$, and all $k, n \in \mathbb{N}$ it holds that each graph G of order n with at least $(1+\eta)n/2$ vertices of degree at least $(1+\eta)k$ contains as subgraphs all trees with at most k edges.

As we may choose $\eta = 1/n$, we can restate the conclusion of these observations more dramatically as follows.

Proposition 16. *If the answer to Question 15 is positive, then also the following is true for all $k, n \in \mathbb{N}$.*
Each graph G of order n with at least $(n+1)/2$ vertices of degree at least $k+1$ contains as subgraphs all trees with at most k edges. $\square$

A variation of Pór's argument shows that if there is a function $k = k(n)$, tending to infinity, so that the approximate version of Conjecture 1 holds for all $ck(n)$ with $0 < c \leq 1$, then it holds for for all choices of $k' = k'(n) \geq k(n)$.

Proposition 17. *Suppose there is a $k = k(n)$, with $\lim_{n\to\infty} k(n) = \infty$, so that for every $\eta > 0$ there exists an $n_0 = n_0(\eta)$ so that the conclusion of Theorem 2 holds for all graphs of order some $n \geq n_0$, for $ck(n)$ with $0 < c \leq 1$, and for $\eta > 0$.*
Then for all choices of $k'(n) > k(n)$ and $\eta' > 0$ there exists an $n'_0 = n'_0(k', \eta')$ so that the conclusion of Theorem 2 holds for all graphs of order $n \geq n'_0$, for $k'(n)$ and for η'.

Proof. Suppose otherwise. Then there exists a $k = k(n)$, so that for each $\eta > 0$ there is an $n_0 = n_0(\eta)$ as above. Furthermore, there are $k'(n) > k(n)$ and η' for which the assertion does not hold. This means that for every $n'_0 = n'_0(k', \eta')$ there exist a graph G on $n'_1 \geq n'_0$ vertices, and a tree T with $k'(n'_1)$ edges, so that at least $(1+\eta')n'_1/2$ vertices of G have degree at least $(1+\eta')k'(n'_1)$, but $T \not\subseteq G$. Fix such an $n'_0 \geq n_0(\eta')$, and an $n'_1 \geq n'_0$.
Let $m \in \mathbb{N}$ be such that $k'(n'_1) \leq k(mn'_1)$. Set

$$c := k'(n'_1)/k(mn'_1).$$

Observe that by our choice of n'_0, we can assume that

$$m \geq 1 \geq n_0(\eta')/n'_1.$$

Now, consider the graph G' that consists of m disjoint copies of G. Still, at least $(1+\eta')|G'|/2$ vertices of G' have degree $(1+\eta')k'(n'_1)$, and still, $T \not\subseteq G'$. Thus, Conjecture 1 is false for $k'(n'_1) = ck(|G'|)$, for η', and for $|G'| \geq n_0(\eta')$, a contradiction to our assumption. □

Observe that we can use the same arguments also for Conjecture 1 (in the exact version), or for the Erdős–Sós conjecture. In particular, if these conjectures can be solved for constant k, then they hold in general.

While at first glance Question 15 might seem easier to solve than Theorem 2, since we are only looking for tiny trees, one has to keep in mind that also the host graph is very sparse.
For this reason, Szemerédi's original regularity lemma is of little use in the sparse setting. However, one may circumvent this problem by using instead the regularity lemma for sparse graphs that was recently developed by Kohayakawa and Rödl (see [7]).
But, more obstacles arise in the sparse case, if we try to follow our approach from Theorem 2. One of these is the following. When embedding the tree, we have to avoid in each step the atypical vertices in the neighbourhood of

the already embedded vertices. But, in the sparse case, the expected size of the neighbourhood of a vertex is $o(n)$, while the size of the set of atypical vertices might be $\Theta(n)$.

Acknowledgment

The authors would like to thank Martin Loebl for helpful discussions.

References

[1] M. Ajtai, J. Komlós, and E. Szemerédi. On a conjecture of Loebl. In *Proc. of the 7th International Conference on Graph Theory, Combinatorics, and Algorithms*, pages 1135–1146, Wiley, New York, 1995.

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

[3] S. Brandt and E. Dobson. The Erdős–Sós conjecture for graphs of girth 5. *Discr. Math.*, 150:411–414, 1996.

[4] R. Diestel. *Graph Theory* (3rd ed.). Springer-Verlag, 2005.

[5] P. Erdős, Z. Füredi, M. Loebl, and V. T. Sós. Discrepency of trees. *Studia sci. Math. Hungar.*, 30(1-2):47–57, 1995.

[6] P. Haxell, T. Luczak, and P. Tingley. Ramsey numbers for trees of small maximum degree. *Combinatorica*, 22(2):287–320, 2002.

[7] Y. Kohayakawa. Szemerédi's regularity lemma for sparse graphs. In F. Cucker and M. Shub, editors, *Foundations of Computational Mathematics*, pages 295–352. Springer-Verlag, 1997.

[8] J. Komlós and M. Simonovits. Szemerédi's regularity lemma and its applications in graph theory. In D. Miklós, V. T. Sós, and T. Szönyi, editors, *Combinatorics: Paul Erdős is eighty*, volume 2 of *Bolyai Society Mathematical Studies*, pages 295–352. János Bolyai Mathematical Society, 1996.

[9] D. Piguet and M. Stein. The Loebl–Komlós–Sós conjecture for trees of diameter 5 and other special cases. In preparation.

[10] J.-F. Saclé and M. Woźniak. A note on the Erdős–Sós conjecture for graphs without C_4. *J. Combin. Theory (Series B)*, 70(2):229–234, 1997.

[11] E. Szemerédi. Regular partitions of graphs. *Colloques Internationaux C.N.R.S. 260 – Problèmes Combinatoires et Théorie des Graphes, Orsay*, pages 399–401, 1976.

[12] M. Woźniak. On the Erdős–Sós conjecture. *J. Graph Theory*, 21(2):229–234, 1996.

[13] Y. Zhao. Proof of the $(n/2-n/2-n/2)$ conjecture for large n. Preprint.