

## THE APPROXIMATE LOEBL–KOMLÓS–SÓS CONJECTURE AND EMBEDDING TREES IN SPARSE GRAPHS

JAN HLADKÝ, DIANA PIGUET, MIKLÓS SIMONOVITS, MAYA STEIN,  
AND ENDRE SZEMERÉDI

(Communicated by Timothy Gowers)

ABSTRACT. Loeb, Komlós and Sós conjectured that every  $n$ -vertex graph  $G$  with at least  $n/2$  vertices of degree at least  $k$  contains each tree  $T$  of order  $k + 1$  as a subgraph. We give a sketch of a proof of the approximate version of this conjecture for large values of  $k$ .

For our proof, we use a structural decomposition which can be seen as an analogue of Szemerédi’s regularity lemma for possibly very sparse graphs. With this tool, each graph can be decomposed into four parts: a set of vertices of huge degree, regular pairs (in the sense of the regularity lemma), and two other objects each exhibiting certain expansion properties. We then exploit the properties of each of the parts of  $G$  to embed a given tree  $T$ .

The purpose of this note is to highlight the key steps of our proof. Details can be found in [arXiv:1211.3050].

### 1. INTRODUCTION

Szemerédi’s Regularity Lemma from 1975 allows us to decompose each dense graph into a bounded collection of random-like subgraphs. The lemma and its variants have found numerous applications in graph theory, number theory, and theoretical computer science. In particular, in the last two decades, it is crucial to some developments on extremal problems concerning dense graphs. On the other hand, extremal problems concerning sparse graphs have been lacking a general framework. We present a new tool which generalizes the Regularity Lemma and which applies to all graphs. This tool seems particularly suitable for embedding

---

Received by the editors April 14, 2014 and, in revised form, March 27, 2015.

1991 *Mathematics Subject Classification*. 05C35 (primary), 05C05 (secondary).

*Key words and phrases*. Extremal graph theory, tree-containment problems, Loeb–Komlós–Sós conjecture, regularity lemma, sparse graphs.

The work on this project lasted from the beginning of 2008 and we are very grateful to the funding bodies for their support. JH was funded by a BAYHOST fellowship, a DAAD fellowship, Charles University grant GAUK 202-10/258009, EPSRC award EP/D063191/1, and by an EPSRC Postdoctoral Fellowship hosted by the Mathematics Institute at the University of Warwick. JK and ESz acknowledge the support of NSF grant DMS-0902241. DP acknowledges the support of the Marie Curie fellowship FIST, DFG grant TA 309/2-1, a DAAD fellowship, Czech Ministry of Education project 1M0545, EPSRC award EP/D063191/1, grant PIEF-GA-2009-253925 of the European Union’s FP7/2007-2013, and EPSRC Additional Sponsorship EP/J501414/1. MS was supported by a FAPESP fellowship, and by FAPESP travel grant PQ-EX 2008/50338-0, also CMM-Basal, and FONDECYT grants 11090141 and 1140766.

trees, even in very sparse graphs. As an application, we prove an approximate version of the Loeb–Komlós–Sós Conjecture.

The Loeb–Komlós–Sós Conjecture is a typical example of a problem in extremal graph theory. These often are of the following type: “does a certain density condition imposed on a graph of order  $n$  guarantee a given subgraph?” Statements of this spirit include Mantel’s theorem (which states that an average degree of more than  $n/2$  ensures a triangle as a subgraph), the more general Turán theorem, and Dirac’s theorem, which states that a minimum degree of at least  $n/2$  forces the appearance of a Hamilton cycle. Other prominent results include the Erdős–Stone–Simonovits theorem, which determines asymptotically the average degree threshold for appearance of any fixed graph, or the solution [22] of the Pósa–Seymour conjecture about containment of powers of Hamilton cycles.

Some of the progress in the area in the last two decades has been enabled by the developments around the regularity lemma. The regularity lemma allows us to approximate an original graph by a so-called cluster graph. The point in doing so is that an original combinatorial problem translates to an easier one on the cluster graph. Let us illustrate this fundamental feature with the examples mentioned above: the modern, regularity lemma based approach reduces the Erdős–Stone–Simonovits theorem to Turán’s theorem on the cluster graph. In a similar spirit, the proof of the Pósa–Seymour conjecture about the appearance of the  $k$ -th power of a Hamilton cycle can be reduced to an easier question of tiling with copies of  $K_{k+1}$  on the cluster graph level, an answer to which is given by the Hajnal–Szemerédi theorem. These and other applications of the regularity lemma in extremal graph theory are surveyed in [24, 21].

Containment of trees is a particularly important case to study, as trees constitute a relatively simple graph class. An easy greedy embedding argument shows that each graph with a minimum degree of at least  $k$  contains each tree with  $k$  edges. A graph formed by a union of cliques of order  $k$  shows that this result is optimal.

Two important conjectures have been made as to how the minimum degree condition can be relaxed. The first of these is the famous Erdős–Sós conjecture from 1963:

**Conjecture 1.** *Every graph of average degree greater than  $k - 1$  contains all trees with  $k$  edges as subgraphs.*

Conjecture 1 trivially holds for the containment of a star with  $k$  leaves, and it is a classical result of Erdős and Gallai [10] that it also holds for paths (for more history, see [11]). Further partial results include [5, 13, 29, 31]. A proof of the conjecture for large graphs has been announced by Ajtai, Komlós, Simonovits and Szemerédi [2].

Loeb, Komlós, and Sós (see [9]) conjectured the same assertion holds when replacing the average degree condition with the median degree.

**Conjecture 2.** *Every graph of median degree at least  $k$  contains all trees with  $k$  edges as subgraphs.*

Previous work on Conjecture 2 includes solutions which use additional restrictions on the host graph [28, 8], or on the trees [6, 26]. Most notably, Conjecture 2 has been solved for large dense graphs, i.e., for  $k$  linear in  $n$ , in [20, 7], building on an approximate version given in [27]. For the exact value  $k = n/2$ , this had been achieved earlier in [1, 32].

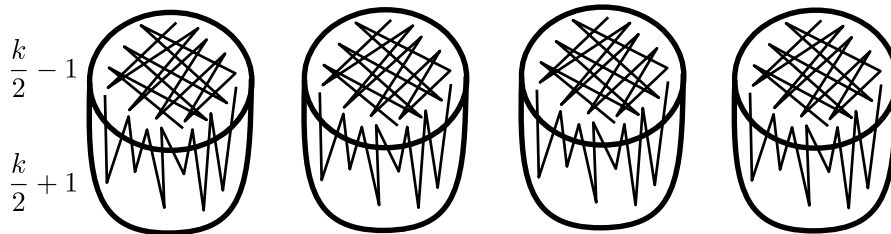


FIGURE 1. An extremal graph for the Loebel-Komlós-Sós Conjecture.

Note that because of the stars, the median degree in Conjecture 2 has to be at least  $k$ . Further note that Conjecture 2 is almost the best possible in the sense that we cannot decrease much the number of vertices, namely  $n/2$ , that are required to have degree at least  $k$ . For this, first assume that  $n$  is even, and that  $n = k + 1$ . Let  $G^*$  be obtained from the complete graph on  $n$  vertices by deleting all edges inside a set of  $\frac{n}{2} + 1$  vertices. Then  $G^*$  has  $\frac{n}{2} - 1$  vertices of degree  $k$ . It is easy to check that  $G^*$  does not contain the path with  $k$  edges (or any other tree with  $k$  edges and independence number less than  $\frac{n}{2} + 1$ ). Now, taking the union of several disjoint copies of  $G^*$  we obtain examples for other values of  $n$ . (And adding a small complete component we can get to any value of  $n$ .) See Figure 1 for an illustration.

It is not difficult to see that each of the two conjectures implies that the Ramsey number of two trees  $T_k$  and  $T_m$  with  $k$  and  $m$  edges, respectively, is at most  $k + m$ . This is the best possible for stars of even order, but not for all trees [12, 19].

Our main result is an approximate version of Conjecture 2, which reads as follows:

**Theorem 3.** *For every  $\varepsilon > 0$  there exists  $k_0$  such that for every  $k > k_0$ , every  $n$ -vertex graph  $G$  with at least  $(1 + \varepsilon)n/2$  vertices of degree at least  $(1 + \varepsilon)k$  contains each tree  $T$  with  $k$  edges.*

Previous results [1, 27, 32, 20, 7] on the dense case of Conjecture 2 relied on Szemerédi's regularity lemma. The strategy of these proofs is explained in the next section. The (original) regularity lemma is void when the host graph is sparse, i.e., when  $k = o(n)$ . To circumvent this shortcoming, we present an extension of the regularity lemma which is tailored to tree-embedding problems, and which applies even to sparse graphs. We then show how this decomposition, which we call *sparse decomposition*, can be used to embed the tree  $T$  given by Theorem 3.

In this paper we show some of the key ideas behind the proof. The actual implementation of these ideas is technical and can be found in [14], split into four parts [15, 16, 17, 18] for publication purposes.

## 2. THE DENSE CASE

In this section, we recall the solution of the dense approximate version of Conjecture 2 due to Piguet and Stein. Their proof provides several key ingredients which are common to the proof of Theorem 3.

**Theorem 4** ([27]). *For every  $C, \varepsilon > 0$  there exists  $k_0$  such that for every  $k > k_0$  we have that every graph  $G$  of order  $n < Ck$  with at least  $n/2$  vertices of degree at least  $(1 + \varepsilon)k$  contains each tree  $T$  with  $k$  edges.*

The proof follows a strategy typical for graph theory results which employ the regularity lemma. It has three main steps: partitioning  $T$ , finding a suitable matching structure in the cluster graph  $\mathbf{G}$  of the graph  $G$ , and embedding  $T$  into  $G$  using this matching structure. The right way to partition  $T$  is given by the following lemma. We say that a subtree  $F \subset T$  is *adjacent* to a vertex  $v \in V(T) \setminus V(F)$  if there is an edge from  $v$  to a vertex in  $F$ .

**Lemma 5.** *For each  $\tau > 0$ ,  $k \in \mathbb{N}$ , for any tree  $T$  with  $k$  edges there is a set  $W = W_A \cup W_B \subset V(T)$ , and a set  $\mathcal{T} = \mathcal{T}_A \cup \mathcal{T}_B$  of disjoint subtrees covering all of  $V(T) - W$  such that*

- (a) *the trees in  $\mathcal{T}$  have order less than  $\tau k$ ,*
- (b) *the trees in  $\mathcal{T}$  are not adjacent to each other,*
- (c)  *$|W| < 100/\tau$ ,*
- (d) *each side of the bipartition of  $V(T)$  contains one of the sets  $W_A, W_B$ ,*
- (e) *each tree in  $\mathcal{T}_B$  is adjacent to only one vertex of  $W$ , and that vertex lies in  $W_B$ ,*
- (f) *each tree in  $\mathcal{T}_A$  is adjacent to at most two vertices of  $W$ , and these lie in  $W_A$ ,*
- (g)  *$|V(\bigcup \mathcal{T}_B)| < k/2$ .*

We call a tree in  $\mathcal{T}$  *internal* if it is adjacent to two vertices of  $W$ , and call it an *end tree* otherwise. Note that  $\mathcal{T}_B$  only contains end trees by property (e).

In order to obtain the partition from Lemma 5, we traverse  $T$  from the leaves to a fixed root, sequentially chopping off ‘branches’ of  $T$  that have reached the critical size of  $\tau k$ . Similar strategies of dividing larger objects into smaller pieces have been used in other proofs employing embedding with the regularity method. As we shall see, the partition of  $T$  from Lemma 5 will be useful for the sparse case as well.

The bulk of the work is on finding a suitable structure in the graph  $G$ . To this end we use the regularity lemma [30]. Let us first introduce the key notion of regular pairs. Given  $\eta > 0$ , a pair  $(A, B)$  of disjoint sets is  $\eta$ -*regular* if  $|d(A, B) - d(U, W)| < \eta$  for each  $U \subseteq A, W \subseteq B$  with  $|U| > \eta|A|, |W| > \eta|B|$ . The regularity lemma then reads as follows.

**Lemma 6** (Regularity lemma). *For each  $\eta > 0, m \in \mathbb{N}$  there are  $n_0, M$  such that every graph on  $n > n_0$  vertices allows for a partition of all but at most  $\eta n$  of its vertices into  $m < k < M$  sets (the ‘clusters’) such that all but at most  $\eta k^2$  pairs of clusters form  $\eta$ -regular pairs.*

With the help of Lemma 6, we regularize the graph  $G$  and obtain a cluster graph  $\mathbf{G}$  with clusters of size  $\nu k$ . Let  $\mathbf{L}$  be the set of those clusters of  $\mathbf{G}$  whose typical vertices have degree more than  $k$ . Piguet and Stein show that  $\mathbf{G}$  contains a matching  $M$  plus two adjacent vertices  $A, B \in \mathbf{L}$  such that the vertices in  $A$  have degree more than  $k$  into  $M$ , and those in  $B$  have degree larger than  $k/2$  into  $V(M) \cup \mathbf{L}$ .

The tree  $T$  can be embedded into  $G$  by suitably mapping the vertices of  $W_A$  into  $A$ , the vertices of  $W_B$  into  $B$ , and packing subtrees  $\mathcal{T}$  (viewed as bipartite graphs) into the edges of  $M$  and the edges  $E_{\mathbf{L}}$  emanating from  $\mathbf{L}$  using basic properties of regular pairs. Here, it is crucial we choose the parameter  $\tau$  for Lemma 5 such that  $\tau \ll \nu$ . That is, individual subtrees of  $\mathcal{T}$  are much smaller than the clusters.

The large degrees of  $A$  and  $B$  into  $V(M) \cup \mathbf{L}$  guarantee that there is enough space for embedding all trees from  $\mathcal{T}$ . More precisely, each time we wish to embed

a tree  $T' \in \mathcal{T}$ , there are two things we have to ensure. The first is that we have enough free space in the neighbourhood of either  $A$  or  $B$  to map the root of  $T'$ . The second is that we have sufficient free space in some regular pair meeting this neighbourhood, to map the rest of  $T'$ . For this, a degree of about  $k$  for the vertices in  $A$  and a degree of about  $k/2$  for the vertices in  $B$  is sufficient.

### 3. THE SPARSE DECOMPOSITION

In this section we introduce the basis of our proof of Theorem 3, the *sparse decomposition*. This tool has been conceived by Ajtai, Komlós, Simonovits and Szemerédi during their work on the Erdős-Sós conjecture. It allows us to decompose any given graph, after a removal of a small fraction of the edges, into four sets: a set  $\Psi$  of vertices of high degree, a graph  $\mathbf{G}_{\text{reg}}$  whose edges span regular pairs, an expanding graph  $G_{\text{exp}}$ , and a set  $\mathfrak{A}$  of vertices which has a yet different expansion property.

Throughout this section, let us fix a graph  $G$  on  $n$  vertices, and let  $k$  have the same order of magnitude as the average degree of  $G$ .<sup>1</sup> We use Greek majuscules and minuscules to denote sufficiently large and sufficiently small constants, respectively. Some of these constants may depend on  $G$  and  $k$  but are absolutely bounded from above and from below. We make the subtle relations between these constants explicit only when it adds to the clarity of this rough sketch.

The first step of the sparse decomposition is to separate the vertices of very high degree from those of comparatively low degree. By deleting only a few well-chosen edges, we arrive at a subgraph  $G'$  of  $G$  that has a gap in its degree sequence. More precisely, there are numbers  $\Omega^*$  and  $\Omega^{**}$  with  $\Omega^* \ll \Omega^{**}$  such that no vertex of  $G'$  has degree between  $\Omega^*k$  and  $\Omega^{**}k$  in  $G'$ . Let us indicate how to create the gap. We fix constants  $1 \ll \Omega_1 \ll \Omega_2 \ll \dots \ll \Omega_{\lceil \frac{1}{\varepsilon} \rceil + 1} := \infty$ . There is an index  $i \in [\lceil \frac{1}{\varepsilon} \rceil + 1]$  such the total degree of the vertices  $v$  with  $\deg(v) \in [\Omega_i k, \Omega_{i+1} k]$  is at most  $\varepsilon kn$ . Deleting the edges incident with these vertices almost yields the gap with  $\Omega^* = \Omega_i$ , and  $\Omega^{**} = \Omega_{i+1}$ . The problem is that the edge deletion may cause degrees of other vertices fall into the forbidden region  $[\Omega_i k, \Omega_{i+1} k]$ . This can be resolved using an additional argument, which we omit here.

Let  $\Psi$  denote the set of all vertices of degree at least  $\Omega^{**}k$ . The mere structural information about the vertices  $\Psi$  is that they have huge degrees. On the other hand, this property turns out to be so powerful for tree embeddings that it compensates the lack of any finer description.

Before proceeding with the decomposition, we need a few concepts. The *density* of a bipartite graph  $D = (U, W; F)$  is  $d(D) := \frac{|F|}{|U||W|}$ , where  $F = E(D)$  are the edges of  $D$ . An  $(m, \gamma)$ -dense spot in a graph is a non-empty bipartite subgraph  $D$  with density  $d(D) > \gamma$  and minimum degree  $\delta(D) > m$ . A graph  $H$  is  $(m, \gamma)$ -nowhere-dense if it does not contain any  $(m, \gamma)$ -dense spot.

Let  $\mathcal{D}$  be a maximal set of edge-disjoint  $(\gamma k, \gamma)$ -dense spots in  $G' - \Psi$ . Let  $G_{\text{exp}}$  be the  $(\gamma k, \gamma)$ -nowhere dense graph obtained from  $G' - \Psi$  by removing the edges of  $\mathcal{D}$ . (We chose the name  $G_{\text{exp}}$  for this graph in order to emphasize its expansion property given by the fact it is nowhere dense.) We now sequentially remove from

<sup>1</sup>This setting is compatible with the one of Theorem 3. Indeed, a straightforward calculation gives that in that case the average degree  $d$  of  $G$  satisfies  $d > k/2$ . If, on the other hand,  $d \geq 2k$  then there is no need to use the sparse decomposition as we can pass to a subgraph with minimum degree at least  $d/2 \geq k$  and embed  $T$  greedily.

$G_{\text{exp}}$  any vertex of degree less than  $\rho k$ , where  $\rho$  is a certain constant much smaller than  $\varepsilon$ , but greater than  $\gamma \ll \rho \ll \varepsilon$ . Note that in the cleaning procedure we lose less than  $\rho kn$  edges, and the obtained graph, which we still call  $G_{\text{exp}}$ , has minimum degree at least  $\rho k$ .

The next step consists of regularizing the dense spots in  $\mathcal{D}$ . By this we mean we wish to find a graph spanning almost all of  $\bigcup \mathcal{D}$  and consisting of clusters that pairwise mostly form highly regular pairs, in the sense above. For each single one of these spots this is possible by Lemma 6 above, but such a naive regularization of each dense spot separately is useless. Indeed, for embedding  $T$  we may need to traverse many different spots  $\mathcal{D}$ . Thus the cluster structure of different dense spots must agree on their intersection.

Consider all the Venn cells  $\mathcal{V}$  with respect to the system  $\{U, W : (U, W; F) \in \mathcal{D}\}$ . We shall not attempt to regularize those Venn cells  $\mathcal{A} \subset \mathcal{V}$  which are of size less than  $\alpha k$  (for  $\alpha \ll \eta$ ), as those cells themselves may be smaller than the anticipated cluster sizes. We now construct an auxiliary graph  $\mathcal{G} = (\mathcal{V} \setminus \mathcal{A}, \mathcal{E})$  on the larger Venn cells, joining two Venn cells  $X, Y$  with an edge if there is a dense spot  $(U, W; F) \in \mathcal{D}$  with  $X \subset U, Y \subset W$ .

Let us sketch how to regularize simultaneously all the dense pairs corresponding to the edges of  $\mathcal{G}$  in this setting. It can be shown that the maximum degree of  $\mathcal{G}$  is bounded from above by a constant  $\Delta$  that is independent of  $k$ . By Vizing's theorem, we can cover  $\mathcal{G}$  with  $\Delta + 1$  matchings  $M_1, \dots, M_{\Delta+1}$ . We follow the idea of Szemerédi's proof of the regularity lemma, pumping up the mean square energy when refining an irregular partition. The key difference is that we track  $\Delta + 1$  mean-square energies, one for each matching  $M_i$ , rather than just a single one. This is similar to the proof of the multi-coloured version of the regularity lemma, which tracks a mean-square energy for each colour separately. We thus obtain a system of clusters, of size  $\nu k$ , say, refining  $\mathcal{V} \setminus \mathcal{A}$  and regular pairs between some of these clusters. Let  $\mathbf{G}_{\text{reg}}$  be the graph spanned by the regular pairs of positive density.

It remains to make use of vertices in  $\mathfrak{A} := \bigcup \mathcal{A}$ , i.e., those in small Venn cells. An elementary double-counting argument gives the following expansion property of  $\mathfrak{A}$ , which we call the  $(\Lambda, \beta, \gamma)$ -avoiding property: For every  $X \subseteq V(G)$  with  $|X| \leq \Lambda k$  for all but at most  $\beta k$  vertices  $v \in \mathfrak{A}$  there is a dense spot  $D \in \mathcal{D}$  which contains  $v$  and which satisfies  $|X \cap V(D)| \leq \gamma^2 k$ . The constant  $\beta$  will be chosen much smaller than  $\gamma$ , but still larger than  $\tau \ll \beta \ll \gamma$ .

Putting all of the above together, we obtained a sparse decomposition  $(\Psi, \mathbf{G}_{\text{reg}}, G_{\text{exp}}, \mathfrak{A})$  which captures all but at most  $o(kn)$  edges of  $G$ .

#### 4. EMBEDDING THE TREE $T$

The proof of Theorem 4 as outlined in Section 2 is a combination of two elements: a global embedding strategy based on the matching structure given by the clusters  $A, B$ , and the matching  $M$ , and a local strategy applied sequentially for embedding the individual subtrees from  $\mathcal{T}$ . The local strategy there is the standard technique of filling up regular pairs.

Also in the proof of Theorem 3 we shall find a suitable global structure, now in the sparse decomposition instead of in the cluster graph. We will discuss this structure later on. First, we indicate local strategies for embedding subtrees  $\mathcal{T}$  in each of the ingredients  $\Psi, \mathbf{G}_{\text{reg}}, G_{\text{exp}}, \mathfrak{A}$  of the sparse decomposition. The starting point of

using either  $\Psi$ ,  $\mathbf{G}_{\text{reg}}$ ,  $G_{\text{exp}}$ , or  $\mathfrak{A}$  is that in our sequential embedding procedure we wish to extend the partial embedding from a vertex with a substantial degree into the respective object. For this, suppose that  $U \subset V(G)$  is the set of vertices used in earlier steps of the embedding.

To work with  $\mathfrak{A}$ , we use the avoiding property. Say we have to embed a tree  $T' \in \mathcal{T}$ . For simplicity, let us assume that  $T' \in \mathcal{T}_B$ ; this guarantees that  $T'$  is an end tree. Suppose its parent in  $W_B$  has been embedded already in a vertex  $v$  of degree more than  $\beta k$  into  $\mathfrak{A} \setminus U$ . Then, by the definition of the avoiding property, there is a neighbour of  $v$  in  $\mathfrak{A} \setminus U$  that is contained in a dense spot  $D$  which does not meet  $U$  much. We can place the first vertex of  $T'$  appropriately into  $D$ . We then use the minimum degree  $\gamma k$  of  $D$  to embed the rest of  $T'$  greedily. Here, we use that  $\tau + \gamma^2 \leq \gamma$ .

Next, we show how to use  $G_{\text{exp}}$ . Again, suppose we are in the process of embedding a subtree  $T' \in \mathcal{T}_B$ . Say  $x$  is the last vertex of  $T'$  that has been embedded already, and we now wish to embed the children of  $x$ . Assume the image  $v$  of  $x$  has neighbourhood  $N_v \subset V(G_{\text{exp}})$  of size at least  $\rho k/2$  in  $G_{\text{exp}} - U$ . (Below it will become clear why we may assume this.) Since  $G_{\text{exp}}$  does not contain any  $(\gamma k, \gamma)$ -dense spots, in particular not between  $U$  and  $N_v$ , we know that most vertices in  $N_v$  have less than  $\rho k/3$  neighbours in  $U$ . (Here, we used that  $\gamma \ll \rho$ .) Thus, it is possible to embed the children of  $x$  in vertices that have degree at least  $\rho k/2$  in  $G_{\text{exp}} - (U \cup x)$ , placing them in equally good positions as  $x$  earlier. Following this strategy successively, we manage to embed all of  $T'$ .

Regular pairs in  $\mathbf{G}_{\text{reg}}$  are used in the usual way for embedding trees of  $\mathcal{T}$ . That is, we view these trees as bipartite graphs, and embed them in regular pairs using the regularity property.

It only remains to explain the role of  $\Psi$ . This set is used very rarely for embedding; in particular we always have  $|U \cap \Psi| < \lambda k/2$ . The idea is that after mapping a vertex  $x \in T$  to a vertex  $v \in \Psi$ , we have an affluence of choices to extend the embedding. However, the huge degree of  $v$  alone is not enough. For example, we cannot map non-leaf vertices of  $T$  to leaves of  $G$ , and these may potentially comprise the entire neighbourhood of  $v$ . To circumvent this issue, we employ a rather delicate cleaning procedure prior to starting the embedding. That is, we find a set  $\Psi' \subset \Psi$  of vertices which have degree at least  $\Omega' k$  (for suitable  $\Omega^* \ll \Omega' \ll \Omega^{**}$ ) into a ‘useful part of  $G$ ’ in such a way that we do not lose many edges during the cleaning. Having done so, we wish to map the neighbours of  $x$  to neighbours of  $v$  that send not more than a few edges to  $U$ . This will guarantee that we can extend the embedding avoiding  $U$  in subsequent steps. To this end, consider the set  $\tilde{U} = \{u \in V(G) : \deg(u, U) \geq \lambda k\}$ . We have  $\tilde{U} \subset \{u \in V : \deg(u, U \setminus \Psi) > \lambda k/2\}$ , and double-counting the edges between  $\tilde{U}$  and  $U \setminus \Psi$  gives

$$|\tilde{U}| < \frac{\Omega^* k |U|}{\lambda k/2} \ll \Omega' k.$$

In particular, the neighbours of  $x$  can be embedded outside of  $\tilde{U}$ .

In Lemma 7, we describe the structural counterpart of the matching structure  $(A, B, M)$  from the dense case (again, the structure is much simplified for presentation reasons). Note that this global structure must combine properties of all the objects  $\Psi$ ,  $\mathbf{G}_{\text{reg}}$ ,  $G_{\text{exp}}$ ,  $\mathfrak{A}$ , as it could happen that none of them alone suffices for embedding  $T$ .

We write  $\mathbb{L}$  for the set of those vertices of  $G$  that have degree at least  $(1 + \varepsilon)k$  in  $G$ . We call a collection  $\mathcal{M}$  of  $\eta$ -regular pairs of positive density with clusters of sizes  $\mu k$  an  $(\eta, \mu)$ -regular matching if all the regular pairs are disjoint.

**Lemma 7.** *The graph  $G$  contains two disjoint sets  $\mathbb{X}, \mathbb{Y} \subseteq \mathbb{L}$ , and an  $(\eta, \mu)$ -regular matching  $\mathcal{M}$  with the following properties.*

- (i) *The bipartite graph  $G[\mathbb{X}, \mathbb{Y}]$  has minimum degree at least  $100/\tau$ .*
- (ii) *The vertices in  $\mathbb{X}$  each have degree at least  $(1 + \varepsilon/2)k$  into*

$$Q := (\Psi \cup V(G_{\text{exp}}) \cup \mathfrak{A} \cup \mathbb{L} \cup V(\mathcal{M})) \setminus (\mathbb{X} \cup \mathbb{Y}).$$

- (iii) *The vertices in  $\mathbb{Y}$  have degrees at least  $(1 + \varepsilon/2)k/2$  into  $Q$ .*
- (iv)  *$\mathcal{M}$  and  $\mathbb{X} \cup \mathbb{Y}$  are disjoint.*

The sets  $\mathbb{X}$  and  $\mathbb{Y}$  will host  $W_A$  and  $W_B$ , that is, we can think of  $\mathbb{X}$  and  $\mathbb{Y}$  as counterparts to the sets  $A$  and  $B$  from the dense case. Property (i) then guarantees that the edges between  $W_A$  and  $W_B$  can be embedded greedily (cf. Lemma 5(c)). Properties (ii) and (iii) guarantee that subtrees  $\mathcal{T}_A$  and  $\mathcal{T}_B$  can be embedded. Instead of embedding  $\mathcal{T}_A$  and  $\mathcal{T}_B$  using the matching  $\mathcal{M}$  and edges  $E_{\mathbb{L}}$  as in the proof of Theorem 4, we have to make use of embedding techniques developed above.

There are two major issues with the indicated approach. The first difficulty is encountered when embedding an internal tree  $T' \in \mathcal{T}_A$ . As such a tree may be adjacent to two vertices of  $W_A$ , and we plan to embed  $W_A$  in  $\mathbb{X}$ , we have to return to  $A$  after embedding  $T'$ .

To understand this difficulty better, it is instructive to first see how an internal tree  $T'$  with head  $x \in W_A$  and tail  $y \in W_A$  is embedded in the dense case (head and tail are the two vertices from Lemma 5(e)). Say  $x$  has been embedded into vertex  $v \in A$ . We choose an edge  $XY \in M \cup E_{\mathbb{L}}$  that will host  $T'$ , such that  $X$  is an edge in the cluster graph. The regularity of  $(A, X)$  guarantees that the embedding of  $T'$  can be extended from  $x$ , but also, that after embedding  $T'$  we can embed  $y$  back in  $A$ .

In the sparse case, we do not have any similar property for the set  $Q$ . To resolve this issue, we introduce certain cleaning procedures which guarantee that the last vertex before a tail of an internal tree is always embedded in a vertex of  $Q$  which has degree at least  $100/\tau$  into  $\mathbb{X}$ .

The second difficulty arises when constructing the regular matching  $\mathcal{M}$ . In analogy to the dense case, it would be a natural guess that  $\mathcal{M}$  is a matching in  $\mathbf{G}_{\text{reg}}$ . In Figure 1 we give an example showing that it is not always possible to choose  $\mathcal{M}$  like this.

Given the example of Figure 1, one might wonder why we bother to construct the cluster graph  $\mathbf{G}_{\text{reg}}$  at all. The answer is that for constructing the regular matching  $\mathcal{M}$ , the graph  $\mathbf{G}_{\text{reg}}$  is helpful either directly or via the information it gives by lacking a suitable matching.

## 5. CONCLUDING REMARKS

Let us conclude with several comments.

- Our proof builds on techniques developed by Ajtai, Komlós, Simonovits and Szemerédi for their work on the Erdős–Sós conjecture. However, there is a substantial difference between the proofs already on the level of the sparse decomposition. In their proof, a suitable matching structure can



always be found in the cluster graph  $\mathbf{G}_{\text{reg}}$ . That means examples like the one in Figure 2 do not enter the picture in the Erdős–Sós conjecture.

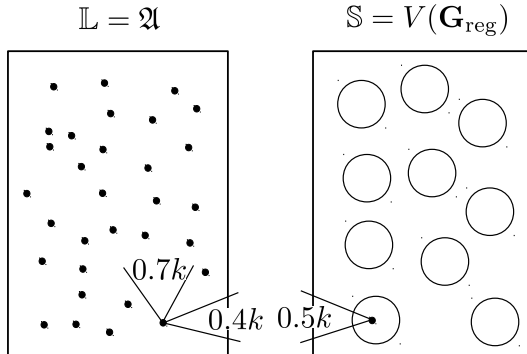


FIGURE 2. The graph  $G$  has a set  $\mathbb{L}$  of  $\frac{5n}{9}$  vertices of degree  $1.1k$ , and a set  $\mathbb{S}$  of  $\frac{4n}{9}$  vertices of degree  $0.5k$ . For each  $v \in \mathbb{L}$ , we have  $\deg(v, \mathbb{L}) = 0.7k$  and  $\deg(v, \mathbb{S}) = 0.4k$ . For each  $v \in \mathbb{S}$ , we have  $\deg(v, \mathbb{L}) = 0.5k$  and  $\deg(v, \mathbb{S}) = 0$ . Further, there is a sparse decomposition of  $G$  such that the union of the dense spots  $\mathcal{D}$  covers all the edges of  $G$ . The dense spots  $\mathcal{D}$  intersect in such a way that the small Venn cells  $\mathcal{A}$  cover  $\mathbb{L}$ . The cluster graph  $\mathbf{G}_{\text{reg}}$  is edgeless, and all vertices have degree less than  $k$  into  $\Psi \cup V(G_{\text{exp}}) \cup \mathbb{Q} \cup \mathbb{L}$ . Thus, for our set  $\mathbb{X}$ , we need to find  $\mathcal{M}$  elsewhere. In this particular case, one can show that there is a regular matching between  $\mathbb{L}$  and  $\mathbb{S}$  covering almost all of  $\mathbb{S}$ , and that such a matching is a good choice for  $\mathcal{M}$ .

- Similarly to the Erdős–Sós setting, it seems that our approach can be combined with the stability approach of Simonovits. We hope to resolve the Loeb–Komlós–Sós conjecture exactly, for  $k$  sufficiently large (this is work in progress).
- The sparse decomposition of a graph is not uniquely determined and can actually vary vastly. This is caused by the arbitrariness in the choice of the dense spots from which we obtain the regularized graph  $\mathbf{G}_{\text{reg}}$ . This situation is in acute contrast with the situation of decomposition of dense graphs (given by the regularity lemma). Indeed, in the dense setting the structure of the cluster graph is essentially unique, cf. [3].<sup>2</sup>
- Another important question is whether there is an alternative approach to proving Conjecture 2 that avoids the notion of sparse decomposition and even the notion of regular pairs. Such a programme has been developed in the dense setting by Szemerédi and his collaborators, see [25] for a particular instance of “deregularizing” a result originally resolved [23] using the regularity method. However, this programme has not given a general alternative view, as of yet.

<sup>2</sup>In order to have uniqueness, the setting needs to be somewhat strengthened; see Theorem 1 and Theorem 2 in [3]. The uniqueness phenomenon can be nicely expressed in the language of graph limits [4].

## REFERENCES

- [1] M. Ajtai, J. Komlós and E. Szemerédi, On a conjecture of Loeb1, in *Graph Theory, Combinatorics, and Algorithms, Vol. 1, 2* (Kalamazoo, MI, 1992), Wiley-Intersci. Publ., Wiley, New York, 1995, 1135–1146. [MR 1405889](#)
- [2] M. Ajtai, J. Komlós, M. Simonovits and E. Szemerédi, Proof of the Erdős-T. Sós conjecture for large trees, in preparation.
- [3] N. Alon, A. Shapira and U. Stav, Can a graph have distinct regular partitions?, *SIAM J. Discrete Math.*, **23** (2008/09), 278–287. [MR 2476828](#)
- [4] C. Borgs, J. Chayes and L. Lovász, Moments of two-variable functions and the uniqueness of graph limits, *Geom. Func. Anal.*, **19** (2010), 1597–1619. [MR 2594615](#)
- [5] S. Brandt and E. Dobson, The Erdős–Sós conjecture for graphs of girth 5, *Discr. Math.*, **150** (1996), 411–414. [MR 1392751](#)
- [6] C. Bazgan, H. Li and M. Woźniak, On the Loeb1-Komlós-Sós conjecture, *J. Graph Theory*, **34** (2000), 269–276. [MR 1771461](#)
- [7] O. Cooley, Proof of the Loeb1-Komlós-Sós conjecture for large, dense graphs, *Discrete Math.*, **309** (2009), 6190–6228. [MR 2551974](#)
- [8] E. Dobson, Constructing trees in graphs whose complement has no  $K_{2,s}$ , *Combin. Probab. Comput.*, **11** (2002), 343–347. [MR 1918720](#)
- [9] P. Erdős, Z. Füredi, M. Loeb1 and V. T. Sós, Discrepancy of trees, *Studia Sci. Math. Hungar.*, **30** (1995), 47–57. [MR 1341566](#)
- [10] P. Erdős and T. Gallai, On maximal paths and circuits of graphs, *Acta Math. Acad. Sci. Hungar.*, **10** (1959), 337–356 (unbound insert). [MR 0114772](#)
- [11] Z. Füredi and M. Simonovits, The history of degenerate (bipartite) extremal graph problems, in *Erdős Centennial*, Bolyai Soc. Math. Stud., 25, János Bolyai Math. Soc., Budapest, 2013, 169–264. [MR 3203598](#)
- [12] L. Gerencsér and A. Gyárfás, On Ramsey-type problems, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.*, **10** (1967), 167–170. [MR 0239997](#)
- [13] P. E. Haxell, Tree embeddings, *J. Graph Theory*, **36** (2001), 121–130. [MR 1814529](#)
- [14] J. Hladký, J. Komlós, D. Piguet, M. Simonovits, M. Stein and E. Szemerédi, The approximate Loeb1-Komlós-Sós conjecture, [arXiv:1211.3050](#).
- [15] J. Hladký, J. Komlós, D. Piguet, M. Simonovits, M. Stein and E. Szemerédi, The approximate Loeb1-Komlós-Sós conjecture I: The sparse decomposition, [arXiv:1408.3858](#).
- [16] J. Hladký, J. Komlós, D. Piguet, M. Simonovits, M. Stein and E. Szemerédi, The approximate Loeb1-Komlós-Sós conjecture II: The rough structure of LKS graphs, [arXiv:1408.3871](#).
- [17] J. Hladký, J. Komlós, D. Piguet, M. Simonovits, M. Stein and E. Szemerédi, The approximate Loeb1-Komlós-Sós conjecture III: The finer structure of LKS graphs, [arXiv:1408.3866](#).
- [18] J. Hladký, J. Komlós, D. Piguet, M. Simonovits, M. Stein and E. Szemerédi, The approximate Loeb1-Komlós-Sós conjecture IV: Embedding techniques and the proof of the main result, [arXiv:1408.3870](#).
- [19] P. E. Haxell, T. Luczak and P. W. Tingley, Ramsey numbers for trees of small maximum degree, *Combinatorica*, **22** (2002), 287–320. [MR 1909088](#)
- [20] J. Hladký and D. Piguet, Loeb1-Komlós-Sós Conjecture: Dense case, [arXiv:0805.4834](#).
- [21] D. Kühn and D. Osthus, Embedding large subgraphs into dense graphs, in *Surveys in Combinatorics 2009*, London Math. Soc. Lecture Note Ser., 365, Cambridge Univ. Press, Cambridge, 2009, 137–167. [MR 2588541](#)
- [22] J. Komlós, G. Sárközy and E. Szemerédi, Proof of the Seymour conjecture for large graphs, *Ann. Comb.*, **2** (1998), 43–60. [MR 1682919](#)
- [23] J. Komlós, G. N. Sárközy and E. Szemerédi, Proof of the Seymour conjecture for large graphs, *Ann. Comb.*, **2** (1998), 43–60.
- [24] J. Komlós, A. Shokoufandeh, M. Simonovits and E. Szemerédi, The regularity lemma and its applications in graph theory, in *Theoretical Aspects of Computer Science* (Tehran, 2000), Lecture Notes in Comput. Sci., 2292, Springer, Berlin, 2002, 84–112. [MR 1966181](#)
- [25] I. Levitt, G. N. Sárközy and E. Szemerédi, How to avoid using the regularity lemma: Pósa’s conjecture revisited, *Discrete Math.*, **310** (2010), 630–641. [MR 2564819](#)

- [26] D. Piguet and M. J. Stein, The Loebel-Komlós-Sós conjecture for trees of diameter 5 and for certain caterpillars, *Electron. J. Combin.*, **15** (2008), Research Paper 106, 11 pp. [MR 2438578](#)
- [27] D. Piguet and M. J. Stein, An approximate version of the Loebel-Komlós-Sós conjecture, *J. Combin. Theory Ser. B*, **102** (2012), 102–125. [MR 2871771](#)
- [28] S. N. Soffer, The Komlós-Sós conjecture for graphs of girth 7, *Discrete Math.*, **214** (2000), 279–283. [MR 1743645](#)
- [29] J.-F. Saclé and M. Woźniak, The Erdős-Sós conjecture for graphs without  $C_4$ , *J. Combin. Theory (Series B)*, **70** (1997), 367–372. [MR 1459879](#)
- [30] E. Szemerédi, Regular partitions of graphs, in *Problèmes Combinatoires et Théorie des Graphes* (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976), Colloq. Internat. CNRS, 260, CNRS, Paris, 1978, 399–401. [MR 540024](#)
- [31] M. Woźniak, On the Erdős-Sós conjecture, *J. Graph Theory*, **21** (1996), 229–234. [MR 1368749](#)
- [32] Y. Zhao, Proof of the  $(n/2 - n/2 - n/2)$  conjecture for large  $n$ , *Electron. J. Combin.*, **18** (2011), Paper 27, 61 pp. [MR 2776803](#)

INSTITUTE OF MATHEMATICS, CZECH ACADEMY OF SCIENCE. ŽITNÁ 25, 110 00, PRAHA, CZECH REPUBLIC. WITH INSTITUTIONAL SUPPORT RVO:67985840.

*E-mail address:* [honzahladky@gmail.com](mailto:honzahladky@gmail.com)

INSTITUTE OF COMPUTER SCIENCE, CZECH ACADEMY OF SCIENCES, POD VODÁRENSKOU VĚŽÍ 2, 182 07 PRAGUE, CZECH REPUBLIC. WITH INSTITUTIONAL SUPPORT RVO:67985807

*E-mail address:* [piguet@cs.cas.cz](mailto:piguet@cs.cas.cz)

RÉNYI INSTITUTE, BUDAPEST, HUNGARY

*E-mail address:* [miki@renyi.hu](mailto:miki@renyi.hu)

CENTRO DE MODELAMIENTO MATEMÁTICO, UNIVERSIDAD DE CHILE, BEAUCHEF 851, SANTIAGO CENTRO, RM, CHILE

*E-mail address:* [mstein@dim.uchile.cl](mailto:mstein@dim.uchile.cl)

DEPARTMENT OF MATHEMATICS, RUTGERS UNIVERSITY, 110 FRELINGHUYSEN RD., PISCATAWAY, NJ 08854-8019, USA

*E-mail address:* [szemered@nyuszik.rutgers.edu](mailto:szemered@nyuszik.rutgers.edu)