An approximate version of the Loebl-Komlós-Sós conjecture

Diana Piguet [1,2]

Katedra Aplikované Matematiky
Univerzita Karlova v Praze
Prague

Maya Stein [1,3]

Instituto de Matemática e Estatística
Universidade de São Paulo
São Paulo

Abstract

The Loebl–Komlós–Sós conjecture states that, given a graph G and a natural number k, if at least half the vertices of G have degree at least k, then any tree with at most k edges is a subgraph of G.

We prove an approximate version of this conjecture for large graphs and k linear in $|V(G)|$.

Keywords: Loebl–Komlós–Sós conjecture, Szemerédi's regularity lemma.

[1] The first author was supported by the ITI grant 1M00216220808, the second author by the FAPESP grant 05/54051-9.

[2] Email: `diana@kam.mff.cuni.cz`

[3] Email: `maya@ime.usp.br`

1 Introduction

Loebl conjectured (see [3]) that given a graph G of order n, if at least $n/2$ vertices of G have degree at least $n/2$, then G contains any tree with at most $n/2$ edges as a subgraph [4]. Recently, Zhao [6] proved this conjecture for large n. An interesting consequence of this result is that large trees T have a Ramsey number [5] $r(T, 2)$ of at most $2|E(T)|$.

Komlós and Sós (see [3]) generalised Loebl's conjecture to the following.

Conjecture 1.1 (Loebl–Komlós–Sós conjecture) *Let G be a graph so that at least half of its vertices have degree at least some $k \in \mathbb{N}$. Then G contains every tree with at most k edges as a subgraph.*

Ajtai, Komlós and Szemerédi [1] have proved an approximative version of the Loebl conjecture. Their proof uses Szemerédi's regularity lemma and the Gallai-Edmonds matching theorem (for details on these two well-known results, see for example [2]). Inspired by this result, we prove in [5] the corresponding approximate version of the Loebl–Komlós–Sós conjecture. However, we have to restrict k to be linear in n.

Theorem 1.2 [5] *For every $\varepsilon > 0$ and $q > 0$ 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 $(1+\varepsilon)n/2$ vertices of G have degree at least $(1+\varepsilon)k$, then every tree T with at most k edges is a subgraph of G.

In the case when $k \geq \frac{n}{2}$, our proof follows the method developed in [1]. The case when $k < \frac{n}{2}$ turns out to be more complicated and requires some additional ideas. For a sketch of our proof, see Section 2.

We can even strengthen our result to the following.

Theorem 1.3 [5] *For every $\varepsilon > 0$ and $q > 0$, 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 for each $k \geq qn$ the following is true.*

If at least $(1+\varepsilon)n/2$ vertices of G have degree at least $(1+\varepsilon)k$, then any bipartite graph of order $k+1$ with $k+c$ edges and at most c components is a subgraph of G.

---

[4] The subgraphs we consider need not be induced.

[5] The Ramsey number $r(H, \ell)$ of a graph H is defined as the minimal natural number r such that any edge-coloring of a complete graph of order $n \geq r$ with ℓ colours yields a monochromatic subgraph of K_n which is isomorphic to H.

2 Sketch of the proof of Theorem 1.2

In this section, we wish to present the main steps of the proof of Theorem 1.2, point out the problems that arise when $k < \frac{n}{2}$, and highlight the ideas we employ to resolve these problems. The details of the proof of Theorem 1.2 as well as a proof of Theorem 1.3 can be found in [5].

Let $\eta > 0$ and $q > 0$ be given. Applied to parameters depending on η and q Szemerédi's regularity lemma provides us with an $n_0 \in \mathbb{N}$. Now, let $k \in \mathbb{N}$, let G be a graph of order at least n_0 that satisfies the condition of Theorem 1.2, and let T be a tree with k edges.

In order to embed T in G, let us now define the *weighted cluster graph* H on the partition classes, or *clusters*, of G that are given by the regularity lemma. For each regular pair (V_i, V_j) of density $d_{ij} \geq p$, this graph H has an edge e_{ij} of weight d_{ij}, where $p = p(\varepsilon, q)$ is a certain threshold density. In what follows, the degree $\deg_Y(X)$ of a cluster X into a subset $Y \subseteq V(H)$ is the sum of the weights of the corresponding edges, multiplied by s, where s is the size of the clusters.

Denote by $L \subseteq V(H)$ the set of those clusters of degree at least $(1+\varepsilon')k$ in H, where $\varepsilon' = \varepsilon'(\varepsilon, q) > 0$. Using the Gallai-Edmonds matching theorem, we show that H contains one of two useful configurations.

Namely, we show that there is an edge $AB \in E(H)$ and a matching M of $H - \{A, B\}$ such that

(i) both clusters A and B have degree at least $(1+\varepsilon')k$ into $V(M)$, or

(ii) only cluster A has degree at least $(1+\varepsilon')k$ into $V(M)$, and cluster B has degree at least $(1+\varepsilon')\frac{k}{2}$ into $M \cup L$. Furthermore, each edge of M meets $N(A)$ in at most one cluster.

For $k \geq n/2$, one can always guarantee for the first alternative. In this case, the rest of our proof follows the lines of the proof in [1].

So suppose we are in the situation of Case (i) above. First, we decompose the tree T into small subtrees (of order much below εk) and a small set $\mathcal{R}$ of vertices (of constant order in n), so that between any two of our subtrees lies a vertex of $\mathcal{R}$. In fact, $\mathcal{R}$ is the disjoint union of two sets $\mathcal{R}_A$ and $\mathcal{R}_B$, and each component of $T - \mathcal{R}$ is adjacent to only one of these two sets. Let us denote the set of trees adjacent to $\mathcal{R}_A$ by $\mathcal{T}_A$, and the set of trees adjacent to $\mathcal{R}_B$ by $\mathcal{T}_B$.

Next, we partition the matching M into M_A and M_B in a way that $\deg_{M_A}(A)$ and $\deg_{M_B}(B)$ are such that $\bigcup \mathcal{T}_A$ fits in $V(M_A)$, and $\bigcup \mathcal{T}_B$ fits in $V(M_B)$, where $V(M_A)$ and $V(M_B)$ are viewed as the respective subsets of $V(G)$.

Finally, we embed $\mathcal{R}_A$ in A and $\mathcal{R}_B$ in B and use the regularity of the pairs of clusters that define the matching edges to embed the small trees of $\mathcal{T}_A \cup \mathcal{T}_B$, one after the other, levelwise, into $V(M_A \cup M_B)$. 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 the tree, and the fact that we embed the trees from $\mathcal{T}_A \cup \mathcal{T}_B$ in the matching edges, ensures that the predecessor of any vertex $r \in \mathcal{R}_A \cup \mathcal{R}_B$ is embedded in a cluster that is adjacent to A, respectively to B (in which we wish to embed r). This enables us to embed all of $\mathcal{R}$ in $A \cup B$, as planned.

An important detail of this 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 has about the same size. 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'. 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 $\mathcal{T}_A \cup \mathcal{T}_B$, we shall choose the side carefully.

In this manner, we can ensure that all of T fits into M, or more precisely into the corresponding subgraph of G, and thus finish the proof for Case (i).

In Case (ii), it is not possible to partition the matching M into M_A and M_B so that $\bigcup \mathcal{T}_A$ fits in $V(M_A)$ and $\bigcup \mathcal{T}_B$ fits in $V(M_B)$, as in Case (i). 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 $V(M_A)$, 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 $V(M_B) \cup L' \cup N(L')$, 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. But, we shall exclude in the first phase only trees that are (each) adjacent to only one vertex from $\mathcal{R}$. This has the advantage that the part of the tree embedded in the first phase is connected (otherwise, in the second phase we would see ourselves confronted with the rather difficult problem of having to connect components).

This idea suggests a natural modification of our sets $\mathcal{T}_A \cup \mathcal{T}_B$, in the following way. Denote by $\mathcal{T}'_A$ the set of those trees from $\mathcal{T}_A$ that are adjacent to only one vertex from $\mathcal{R}_A$, and in the same way define $\mathcal{T}'_B$. Assume that

$$|V(\bigcup \mathcal{T}'_A)| \geq |V(\bigcup \mathcal{T}'_B)|,$$

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

Our plan now is to first embed the trees from $\mathcal{T}' \cup \mathcal{T}_B$ and to postpone the embedding of $\bigcup \mathcal{T}'_A$ to the second phase. Observe that by what we explained above, even leaving out $\mathcal{T}'_A$, we might be unable to find a partition of M so that we can embed the trees from $\mathcal{T}_A \setminus \mathcal{T}'_A$ in one side, and $\bigcup \mathcal{T}_B$ in the other. In fact, we cannot hope to fit more than $\bigcup \mathcal{T}'_B$ into $V(M_B) \cup L' \cup N(L')$.

So, let us partition M into M_A and M_B so that $\deg_{M_A}(A)$ allows for the embedding of $\bigcup \mathcal{T}'$, and $\deg_{M_B \cup L}(B)$ allows for the embedding of $\bigcup \mathcal{T}'_B$. This actually means that the place we reserved for the embedding of $\bigcup(\mathcal{T}_B \setminus \mathcal{T}'_B)$ lies in M_A.

Therefore, we shall 'switch' this forest to $\mathcal{T}_A$. Let us explain what we mean by *switching*. For every tree $t \in \mathcal{T}_B \setminus \mathcal{T}'_B$, delete all vertices from t that are adjacent to $\mathcal{R}_B$ in T and add them to $\mathcal{R}_A$. Put the components of what remains of t into $\mathcal{T}_A$.

After switching all trees $t \in \mathcal{T}_B \setminus \mathcal{T}'_B$, denote by $\mathcal{T}_F$ the (enlarged) set $\mathcal{T}_A \setminus \mathcal{T}'_A$. That is, $\mathcal{T}_F$ consists of all trees from the original $\mathcal{T}_A \setminus \mathcal{T}'_A$, together with all trees we generated by switching. Note that we have thus at most tripled the size of $\mathcal{R}$. Also, each tree from $\mathcal{T}_F$ and $\mathcal{T}'_A$ is adjacent only to (the enlarged) $\mathcal{R}_A$ and each tree from $\mathcal{T}'_B$ is still adjacent only to $\mathcal{R}_B$.

We now embed the vertices from $\mathcal{R}_A \cup \mathcal{R}_B$ in $A \cup B$, embed $\bigcup \mathcal{T}_F$ in $V(M_A)$, and embed $\bigcup \mathcal{T}'_B$ in $V(M_B)$ and $L' \cup N(L')$, in the same way as in Case (i). Note that we have to use not only edges from M as before, but also edges of H 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 (ii).

In the second phase we wish to embed $\bigcup \mathcal{T}'_A$. We shall now use all of M, forgetting about the partition. The vertices of $\mathcal{R}_A$ to which the trees of $\mathcal{T}'_A$ are adjacent have already been embedded in the first phase. Having chosen their images carefully then ensures that now they have still large enough degree into the unused parts of the clusters from $V(M)$. Hence there is enough place in M for $\bigcup \mathcal{T}'_A$.

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 $\mathcal{T}'_B$. This could be a problem, if we now calculate with 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 irrevelant whether the embedding is balanced or not in these edges.

The embedding itself of $\bigcup \mathcal{T}'_A$ is done as before. This finishes Case (ii), and thus the proof of Theorem 1.2.

3 Open problems

A question that naturally arises is whether an extension of Theorem 1.2 for arbitrary $k \in \mathbb{N}$, i. e. to sparse graphs, is possible. Instead of the original regularity lemma, which is of little use in sparse graphs, one might use the recently developed regularity lemma for sparse graphs (see [4]).

One of the problems that arise then is the following. When embedding the tree T, 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 $O(n)$.

References

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

[2] R. Diestel, "Graph Theory," 3rd Ed., Springer–Verlag, 2005.

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

[4] Y. Kohayakawa, V. Rödl, *Regular pairs in sparse random graphs I,* Random Structures Algorithms 22, no. 4, (2003), 359-434.

[5] D. Piguet, M. Stein, *An approximative version of the Loebl-Komlós-Sós conjecture*, in preparation.

[6] Y. Zhao, *Proof of the* $(n/2, n/2n/2)$ *Conjecture for large* n, preprint.