Induced trees in triangle-free graphs

Jiří Matoušek Robert Šámal*

Department of Applied Mathematics and
Institute of Theoretical Computer Science (ITI)
Charles University
Malostranské nám. 25, 118 00 Praha 1
Czech Republic

Abstract

We prove that every connected triangle-free graph on n vertices contains an induced tree on $\exp(c\sqrt{\log n}\,)$ vertices, where c is a positive constant. The best known upper bound is $(2+o(1))\sqrt{n}$. This partially answers questions of Erdős, Saks, and Sós and of Pultr.

1 Introduction

For a graph G, let $t(G)$ denote the maximum number of vertices of an induced subgraph of G that is a tree (i.e., connected and acyclic). There are arbitrary large graphs G with $t(G) \le 2$, namely graphs in which every connected component is a clique. To rule out these trivial examples, we need to put some restrictions on G.

Motivated by study of forbidden configurations in Priestley spaces [1], Pultr (private communication, 2002) asked how big $t(G)$ can be if G is connected and bipartite. Formally, he was interested about asymptotic properties of the function

$$f_B(n) = \min\{t(G) : |V(G)| = n, G \text{ connected and bipartite}\}.$$

Pultr's question was the starting point of our work. However, the function $t(G)$ was studied earlier and in a more general context by Erdős, Saks, and Sós [2]. They describe the influence of the number of edges of G on $t(G)$ and, more to our point, they study how small $t(G)$ can be if $\omega(G)$ is given. They observe that $t(G) \le 2\alpha(G)$, and this allows them to use estimates for Ramsey numbers. This way, they show that for any fixed $k > 3$ there are constants c_1, c_2 such that

$$c_1 \frac{\log n}{\log\log n} \le \min\{t(G) : |V(G)| = n, G \not\supseteq K_k\} \le c_2 \log n\,.$$

For $k = 3$ the lower bound still applies, but the upper bound obtained using Ramsey numbers is only $O(\sqrt{n}\log n)$. We concentrate on this case $k = 3$, that is we put

$$f_T(n) = \min\{t(G) : |V(G)| = n, G \text{ connected and triangle-free}\}.$$

Instead of applying Ramsey theory, we approach the problem directly.

It is easy to show that $f_T(n) \le f_B(n) = O(\sqrt{n}\,)$. The best construction we are aware of yields $f_B(n) \le (2+o(1))\sqrt{n}$; see Section 2. A simple "blow-up" construction, also presented in Section 2, shows that if $f_T(n_0) < \sqrt{n_0}$ for some n_0,

*Currently on leave from Institute for Theoretical Computer Science (ITI). The paper was finished while the second author was a PIMS postdoctoral fellow at Department of Mathematics, Simon Fraser University, Burnaby, B.C. V5A 1S6, Canada.

then $f_T(n) = O(n^{1/2-\varepsilon})$ for a positive constant $\varepsilon > 0$, and similarly for f_B. Hence, $f_T(n)$ either is of order exactly $\sqrt{n}$, or it is bounded above by some power strictly smaller than $1/2$. We conjecture that the second possibility holds, and that another power of n is a lower bound. We can prove the following lower bound.

Theorem 1.1 *There is a constant $c > 0$ such that for all n*

$$f_T(n) \geq e^{c\sqrt{\log n}}.$$

Conjecture 1.2 *There are constants $0 < \alpha < \beta < 1/2$ such that for all n*

$$n^{\alpha} \leq f_B(n) \leq n^{\beta}.$$

We finish the introduction by mentioning further results concerning $t(G)$. It is interesting to consider the problem of finding induced trees in (sparse) random graphs. Vega [3] shows that $t(G_{n,c/n}) = \Omega(n)$ a.s.; Palka and Ruciński [6] prove that $t(G_{n,c\log n/n}) = \Theta(n \log n / \log\log n)$ a.s.

Another way of looking at the problem is to consider it as a Ramsey-type problem: we look for minimal n such that each graph on n vertices contains an induced tree on t vertices or a K_3. A result in induced Ramsey theory [4] implies that there is a graph F on $n = t^c$ vertices (c an absolute constant) such that every subgraph of F either contains a tree on t vertices as an induced subgraph (we can even specify the tree in advance) or it contains a K_3. This perhaps supports our Conjecture 1.2.

Krishnan and Ochem [5] search for values of $f_T(n)$ (for small n) using a computer; they succeed to find $f_T(n)$ for $n \leq 15$. They also extend results of [2] about the decision problem: "given a connected graph G and an integer t, does G have an induced tree with t vertices?". Not only this is NP-complete for general graphs (which is proved in [2]), but it remains NP-complete even if we restrict to bipartite graphs, or to triangle-free graphs of maximum degree 4.

2 Initial observations

Observation 2.1 *$f_B(n) \leq (2 + o(1))\sqrt{n}$.*

Proof: We only prove here $f_B(n) \leq (3 + o(1))\sqrt{n}$. Put $s = \lceil \sqrt{n}\, \rceil$, take a path on s vertices and replace each edge by a complete bipartite graph $K_{s,s}$. The resulting graph G satisfies $t(G) \leq 3s$. □

Lemma 2.2 (Blow-up construction) *Let G be a connected triangle-free graph and let $W \subseteq V(G)$ be a subset of m vertices such that any induced tree in G contains at most t vertices of W. Then we have $f_T(n) = O(n^{\ln(t-1)/\ln(m-1)})$. The same result holds with "triangle-free" replaced by "bipartite" and with f_T replaced by f_B.*

Proof: (Omitted in the extended abstract.) □

Corollary 2.3 *If $f_T(n_0) < \sqrt{n_0}$ for some n_0, then $f_T(n) = O(n^{1/2-\varepsilon})$ for a positive constant $\varepsilon > 0$. (The same is true for f_B.)*

Proof: Let G be the graph on n_0 vertices for which $t(G) = t < \sqrt{n_0}$. We let $W = V(G)$ and $m = n_0$ and apply Lemma 2.2. □

As mentioned in the introduction, Krishnan and Ochem [5] search for values of $f_T(n)$ using a computer. This was motivated by hope that Corollary 2.3 would apply. It turns out, however, that for small n Observation 2.1 gives a precise estimate even for $f_T(n)$ (e.g., $f_T(15) = 7$); therefore Corollary 2.3 does not apply.

Remark. If we consider the construction from Lemma 2.2 for $G = K_3$, $W = V(G)$, $m = 3$, and $t = 2$ we recover a result of [2] that there is a graph G containing triangles (but no K_4) such that $t(G) = O(\log n)$.

3 A weaker lower bound for bipartite graphs

In this extended abstract we prove only a statement weaker than Theorem 1.1—we give a bound on $f_B(n)$ instead of $f_T(n)$. The proof is simpler than that of Theorem 1.1 and it serves as an introduction to it.

We begin with a lemma about selecting induced forests of a particular kind in a bipartite graph. We introduce some terminology. Let H be a bipartite graph with color classes A and B. We will think of A as the "top" class and B as the "bottom" class (in a drawing of G in the plane, say). We write $a = |A|$ and $b = |B|$. For a subgraph F of H we write $A(F) = V(F) \cap A$, we set $a(F) = |A(F)|$, and we define $B(F)$ and $b(F)$ similarly.

Whenever we say *forest* we actually mean an induced subgraph of H that is a forest. An *up-forest* F is a forest such that every vertex in $A(F)$ has degree (in F) precisely 1 and every vertex in $B(F)$ has degree (in F) at least 1. A *matching* is a forest F in which all vertices have degrees (in F) exactly 1.

Lemma 3.1 *Let H be a bipartite graph with color classes A and B as above, let Δ be the maximum degree of H, and let $\eta \in (0, 1)$ be a real parameter. Let us suppose that every vertex in A is connected to at least one vertex in B. Then at least one of the following cases occurs:*

(M) *There is a matching with at least $(1 - \eta)a$ edges.*

(B) *There is an up-forest F with*

$$b(F) \geq \frac{\eta}{\Delta^3} \cdot a$$

that is 2-branching, meaning that every vertex in $B(F)$ has degree at least 2 in F.

Proof: (Omitted in the extended abstract.) □

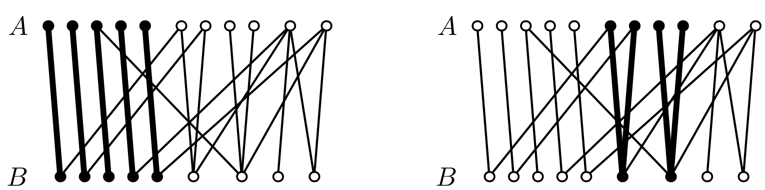

Figure 1: An illustration of Lemma 3.1

Now we prove the lower bound

$$f_B(n) \geq e^{c\sqrt{\log n}}$$

for a constant $c > 0$.

Let G be a given connected bipartite graph. We assume that $n = |V(G)|$ is sufficiently large whenever convenient. We let t be the "target size" of an induced tree in G we are looking for; namely, $t = \lceil \exp(c\sqrt{\log n}\,) \rceil$. If G has a vertex of degree at least $t-1$, then we can take its star for the induced tree and we are done, so we may assume that the maximum degree satisfies $\Delta \le t-2$.

Let us fix an arbitrary vertex of G as a root, and let L_i be the set of vertices of G at distance precisely i from the root. All edges of G go between L_{i-1} and L_i for some i, since an edge within some L_i would close an odd cycle.

We may assume that $L_t = \emptyset$, for otherwise G contains an induced path of length t. Hence there is a k with $|L_k| \ge n/t$.

Let us fix such a k. We are going to construct sets $M_i \subseteq L_i$, $i = k, k-1, \ldots$, inductively, until we first reach an i with $|M_i| = 1$ (this happens for $i = 0$ at the latest since $|L_0| = 1$). We let ℓ be this last i.

Suppose that nonempty sets $M_k, M_{k-1}, \ldots, M_i$ have already been constructed, in such a way that the subgraph of G induced by $M_k \cup \cdots \cup M_i$ is a forest, each of whose components intersects M_i in at most one vertex. We are going to construct M_{i-1}.

Let us put $A = M_i$, $B = L_{i-1}$, and let us consider the bipartite graph H with color classes A and B and with $E(H) = \{\{u,v\} \in E(G) : u \in A, v \in B\}$. Every vertex of A is connected to at least one vertex in B. We set $\eta = \frac{1}{t}$ and apply Lemma 3.1. This yields an up-forest F in H as in the lemma. We define $M_{i-1} = B(F)$.

If F is a matching, i.e., case (M) occurred in the lemma, we call the step from M_i to M_{i-1} a *matching step*. In this case, we have $|M_{i-1}| \ge (1-\frac{1}{t})|M_i|$. Otherwise, F is a 2-branching forest; then we call the step a *branching step* and we have $|M_{i-1}| \ge |M_i|/(t\Delta^3) \ge |M_i|/t^4$.

Suppose that the sets $M_k, \ldots, M_\ell$ have been constructed, $|M_\ell| = 1$. We claim that the number b of branching steps in the construction is at least $c_1\sqrt{\log n}$ for a suitable constant $c_1 > 0$. Indeed, there are no more than t matching steps, and so $1 = |M_l| \ge |M_k|(1-1/t)^t t^{-4b} \ge (n/t)e^{-1}/2 \cdot t^{-4b} = \Omega(nt^{-4b-1})$. Thus $b = \Omega(\log n/\log t) = \Omega(\sqrt{\log n}\,)$, since $t = \lceil \exp(c\sqrt{\log n}\,)\rceil$.

It is easy to see that $M_k \cup M_{k-1} \cup \cdots \cup M_l$ induces a forest in G. We let T be the component of this forest containing the single vertex of M_l. Since every vertex of M_{i-1}, $l < i \le k$, has at least one neighbor in M_i, and if the step from M_i to M_{i-1} was a branching step then each vertex of M_{i-1} has at least two neighbors in M_i, it follows that T has at least $2^b = \exp(\Omega(\sqrt{\log n}\,))$ vertices. This finishes the proof of the lower bound $f_B(n) \ge \exp(c\sqrt{\log n}\,)$. $\square$

Remark. The above proof may seem wasteful in many respects. However, the result is tight up to the value of the constant in the exponent if we insist on selecting an induced tree "growing up" from the root (i.e., made of up-forests). Indeed, any such induced tree in the graph G_q in Figure 2 may contain at most two of the q vertices at the topmost level of the graph. Let us put $q = \exp(c\sqrt{\log n}\,)$ and glue copies of G_q according to the pattern of a complete q-ary tree (as in the proof of Lemma 2.2), so that the resulting graph has approximately n vertices (that is, the depth is $l = \Theta(\log n)$. We obtain a graph with all up-growing induced trees having size at most $2^l = \exp(O(\sqrt{\log n}\,))$.

Acknowledgment

We would like to thank participants of a research seminar where initial steps of this work have been made, for a stimulating environment, and in particular, we are indebted to Robert Babilon, Martin Bálek, Helena Nyklová, Ondra Pangrác, and Pavel Valtr, for useful discussions and observations.

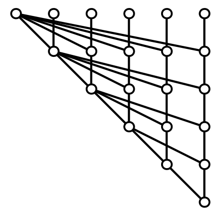

Figure 2: Graph G_6 in which all "up-growing trees" contain at most two vertices of the uppermost level.

References

[1] Richard N. Ball and Aleš Pultr, *Forbidden forests in Priestley spaces*, Cah. Topol. Géom. Differ. Catég. **45** (2004), no. 1, 2–22.

[2] Paul Erdős, Michael Saks, and Vera T. Sós, *Maximum induced trees in graphs*, J. Combin. Theory Ser. B **41** (1986), no. 1, 61–79.

[3] Wenceslas Fernandez de la Vega, *Induced trees in sparse random graphs*, Graphs Combin. **2** (1986), no. 3, 227–231.

[4] Yoshiharu Kohayakawa, Hans-Jürgen Prömel, and Vojtěch Rödl, *Induced Ramsey numbers*, Combinatorica **18** (1998), no. 3, 373–404.

[5] Sivarama Krishnan and Pascal Ochem, *Searching for induced trees*, DocCourse Prague 2005, Project Report, Charles University, Prague 2005.

[6] Zbigniew Palka and Andrzej Ruciński, *On the order of the largest induced tree in a random graph*, Discrete Appl. Math. **15** (1986), no. 1, 75–83.