A CARO-WEI BOUND FOR INDUCED LINEAR FORESTS IN GRAPHS

GWENAËL JORET AND ROBIN PETIT

Abstract. A well-known result due to Caro (1979) and Wei (1981) states that every graph G has an independent set of size at least $\sum_{v\in V(G)} \frac{1}{d(v)+1}$, where $d(v)$ denotes the degree of vertex v. Alon, Kahn, and Seymour (1987) showed the following generalization: For every $k \geqslant 0$, every graph G has a k-degenerate induced subgraph with at least $\sum_{v\in V(G)} \min\{1, \frac{k+1}{d(v)+1}\}$ vertices. In particular, for $k = 1$, every graph G with no isolated vertices has an induced forest with at least $\sum_{v\in V(G)} \frac{2}{d(v)+1}$ vertices. Akbari, Amanihamedani, Mousavi, Nikpey, and Sheybani (2019) conjectured that, if G has minimum degree at least 2, then one can even find an induced *linear* forest of that order in G, that is, a forest where each component is a path.

In this paper, we prove this conjecture and show a number of related results. In particular, if there is no restriction on the minimum degree of G, we show that there are infinitely many "best possible" functions f such that $\sum_{v\in V(G)} f(d(v))$ is a lower bound on the maximum order of a linear forest in G, and we give a full characterization of all such functions f.

1. Introduction

Caro [3] and Wei [7] proved the following well-known lower bound on the independence number of a graph, where $d(v)$ denotes the degree of vertex v.

Theorem 1 (Caro [3] and Wei [7]). *Every graph G has an independent set of size at least $\sum_{v\in V(G)} \frac{1}{d(v)+1}$.*

Alon, Kahn and Seymour [2] generalized this result for k-degenerate induced subgraphs as follows. (Recall that a graph is *k-degenerate* if every subgraph has a vertex of degree at most k.)

Theorem 2 (Alon, Kahn, and Seymour [2]). *For every integer $k \geqslant 0$, every graph G has an induced subgraph with at least $\sum_{v\in V(G)} \min\left\{1, \frac{k+1}{d(v)+1}\right\}$ vertices that is k-degenerate.*

The bound in Theorem 2 is best possible, as shown by complete graphs. Theorem 2 with $k = 0$ corresponds to the Caro-Wei bound for independent sets. In this paper, we focus on the $k = 1$ case. For $k = 1$, Theorem 2 shows that every graph G with no isolated vertex has an induced forest with at least $\sum_{v\in V(G)} \frac{2}{d(v)+1}$ vertices.[1] It is natural to ask whether one can guarantee some additional properties for the trees in this induced forest, such as a bound on the maximum degree, or a specific structure. This is the topic that we explore in this paper.

Our starting point is a conjecture of Akbari, Amanihamedani, Mousavi, Nikpey, and Sheybani [1]: They conjectured that one can impose an upper bound of 2 on the maximum degree of the trees—or equivalently, that each tree is a path—*provided* that G has minimum degree at least 2.

Computer Science Department, Université libre de Bruxelles, Brussels, Belgium
E-mail address: gwenael.joret@ulb.be, robin.petit@ulb.be.
Date: March 27, 2024.
G. Joret is supported by a PDR grant from the Belgian National Fund for Scientific Research (FNRS).
[1]We remark that Punnim [6] also proved the result for $k = 1$, apparently unaware of [2], with a probabilistic proof that differs from the constructive one given in [2].

Conjecture 3 (Akbari *et al.* [1]). *Every graph G with minimum degree at least 2 has an induced linear forest with at least $\sum_{v \in V(G)} \frac{2}{d(v)+1}$ vertices.*

Here, a forest is *linear* if every component is a path. Note that it is necessary to rule out vertices of degree 1 in the above conjecture, as shown by the claw $K_{1,3}$. As supporting evidence for their conjecture, the authors of [1] proved it in the case where the graph G is regular.

Our first contribution is a proof of Conjecture 3, which follows from the following theorem.

Theorem 4. *Let $f : \mathbb{N} \to \mathbb{R}$ be defined as follows:*

$$f(d) = \begin{cases} 1 & \text{if } d = 0 \\ \frac{5}{6} & \text{if } d = 1 \\ \frac{2}{d+1} & \text{if } d \geqslant 2. \end{cases}$$

Then, every graph G has an induced linear forest with at least $\sum_{v \in V(G)} f(d(v))$ vertices.

Our proof of Theorem 4 is a short inductive argument. We remark that Theorem 4 is best possible in the sense that the value of $f(d)$ cannot be increased for any d. (This is clear for $d = 0$, for $d = 1$ this is witnessed by $K_{1,3}$, and for $d \geqslant 2$ by complete graphs.)

If we compare Theorem 2 for $k = 1$ with Theorem 4, we see that the lower bounds are the same except for the contribution of degree-1 vertices, which are respectively 1 and 5/6. As it turns out, one can keep a contribution of 1 for degree-1 vertices and get an induced forest that is *almost* linear, namely, every component is a caterpillar (a path plus some hanging leaves).

Corollary 5. *Every graph G with no isolated vertices has an induced forest with at least $\sum_{v \in V(G)} \frac{2}{d(v)+1}$ vertices, every component of which is a caterpillar.*

Corollary 5 follows easily from Theorem 4, the proof consists in first removing the degree-1 vertices and then applying Theorem 4.

It is interesting to rephrase the previous results in terms of treewidth and pathwidth: graphs of treewidth 0 correspond to edgeless graphs, and same for pathwidth 0. Graphs of treewidth 1 are exactly forests, while graphs of pathwidth 1 are exactly forests where each component is a caterpillar. Thus, the Caro-Wei bound gives the extremal bound as a function of the degrees for finding induced subgraphs with treewidth/pathwidth 0. Theorem 2 with $k = 1$ does the same for induced subgraphs with treewidth 1, and Corollary 5 shows that the extremal function is the same for pathwidth 1 as for treewidth 1.

Now, let us come back to Theorem 4. As already mentioned, the function f in the lower bound cannot be improved, thus the result is best possible in this sense. However, and perhaps surprisingly, there are other "best possible" functions f such that $\sum_{v \in V(G)} f(d(v))$ is a lower bound on the maximum order of an induced linear forest in G. A simple one is the function f with $f(0) = f(1) = 1$, $f(2) = 2/3$, and $f(d) = 0$ for all $d \geqslant 3$. (We skip the easy proof that it is a lower bound; it is best possible as shown by K_3 and $K_{1,t}$ for $t \geqslant 3$.) This naturally raises the following question:

> *Given the degree sequence of G, what is the best possible lower bound on the maximum order of an induced linear forest in G that we can write?*

We answer this question as follows: First, we give a full description of the set of functions f providing valid lower bounds and that are extremal (that is, that cannot be improved). Second, given the latter description and the degree sequence of a graph G, we explain how to choose the best lower bound for G.

Theorem 6. *Let $\varepsilon \in \mathbb{R}$ with $0 \leqslant \varepsilon \leqslant \frac{1}{6}$ and define*

$$f_\varepsilon : \mathbb{N} \to [0,1] : d \mapsto \begin{cases} 1 & \text{if } d = 0 \\ 1 - \varepsilon & \text{if } d = 1 \\ \frac{2}{3} & \text{if } d = 2 \\ \min\left\{3\varepsilon, \frac{2}{d+1}\right\} & \text{if } d \geqslant 3. \end{cases}$$

Then f_ε provides a lower bound for induced linear forests, that is, every graph G has an induced linear forest with at least $\sum_{v \in V(G)} f_\varepsilon(d(v))$ vertices. Moreover, this lower bound is extremal, and every lower bound for induced linear forests is dominated by some lower bound of this form.

The optimal choice for ε given the degree distribution of a graph G is given in Theorem 9 in Section 4.

Theorem 6 is the main result of this paper. Note that Theorem 4 follows from Theorem 6 by taking $\varepsilon = 1/6$. However, we were unable to find a short proof for Theorem 6, the inductive argument that we use in the proof of Theorem 4 only works for $\varepsilon = 1/6$. The heart of the proof of Theorem 6 is an auxiliary lemma, Lemma 11, which we call the *ABC Lemma*. Informally, given a tripartition of the vertex set of G into sets A, B, C, the lemma provides a large induced linear forest in G satisfying the constraint that vertices in B are leaves and vertices in C are isolated. How large is the forest depends on the tripartition, with vertices in B contributing less than those in A, and vertices in C less than those in B.

Using the ABC lemma, we were able to generalize Theorem 6 to the setting of induced subgraphs that are forests of caterpillars and have maximum degree at most k for some $k \geqslant 2$. For $k = 2$, this corresponds to induced linear forests. For $k = +\infty$, this corresponds to induced forests of caterpillars, as in Corollary 5. Thus, varying k gives a way of interpolating between these two extremes. Again, for fixed k, we describe all the corresponding extremal functions, see Theorem 8. As it turns out, the proof for general k is not more difficult than for $k = 2$, which is why we include this result.

We conclude this introduction by mentioning one last contribution. In order to motivate it, let us recall that treewidth is a lower bound on pathwidth, which in turn is a lower bound on treedepth minus 1. These three graph invariants are closely related to each other and play a central role in structural graph theory. Graphs of treewidth 1 are forests, graphs of pathwidth 1 are forests of caterpillars, and graphs of treedepth 2 are forests of stars. Since the extremal lower bounds are unique and the same in the first two cases, one may wonder if this remains true for the last case as well, that is, for finding induced forests of stars in a given graph. However, this is not the case. It turns out that there are infinitely many extremal lower bounds, as for induced linear forests. These are described in the following theorem.

Theorem 7. *For every $\varepsilon \in \mathbb{R}$ with $0 \leqslant \varepsilon \leqslant \frac{1}{6}$, the following function provides a lower bound for induced forests of stars:*

$$f_\varepsilon : \mathbb{N} \to [0,1] : d \mapsto \begin{cases} 1 & \text{if } d = 0 \\ 1 - \varepsilon & \text{if } d = 1 \\ \min\left\{\frac{3}{5}, \frac{1}{2} + \varepsilon\right\} & \text{if } d = 2 \\ \min\left\{\frac{2}{d+1}, \frac{1}{d} + \varepsilon\right\} & \text{if } d \geqslant 3. \end{cases}$$

That is, every graph G contains an induced forest of stars with at least $\sum_{v \in V(G)} f_\varepsilon(d(v))$ vertices. Moreover, this lower bound is extremal, and every lower bound for induced forests of stars is dominated by some extremal lower bound of this form.

The proof of Theorem 7 follows a similar high level strategy than that of Theorem 6. In particular, it introduces and uses a so-called *AB Lemma* (Lemma 12) specifically designed for forests of stars.

The paper is organized as follows. In Section 2 we introduce the necessary notations and preliminary results. In Section 3, we first give the short proof of Theorem 4 and then turn to the proof of Theorem 6, the main result of this paper. The latter proof relies on the ABC Lemma, which is proved in its own section, Section 5. Finally, Section 6 is devoted to the proof of Theorem 7.

2. Preliminaries

All graphs in this paper are finite, simple, and undirected. The maximum degree of a graph G is denoted $\Delta(G)$. Given a vertex v in a graph G, the degree of v is denoted $d_G(v)$ and the set of neighbors of v is denoted by $N_G(v)$. We use the notation $N_G[S]$ for the set of vertices at distance at most 1 from a given set S of vertices in G, and $N^i_G(S)$ for the set of vertices at distance exactly i from S. We drop the subscript G when the graph is clear from the context. In this paper, a *leaf* of a graph G is any vertex v with degree exactly 1. (Thus, two adjacent vertices of degree 1 are both considered to be leaves.)

Given a class of graphs $\mathcal{C}$ and a graph G, let $\alpha_{\mathcal{C}}(G)$ denote the maximum order of an induced subgraph H of G such that H belongs to $\mathcal{C}$.

Given a function $f : \mathbb{N} \to [0, 1]$ and a graph G, we define $f(G)$ as follows:

$$f(G) := \sum_{v \in V(G)} f(d(v)).$$

(This is a slight abuse of notation but it will be convenient in the paper.)

Given a graph class $\mathcal{C}$ and a function $f : \mathbb{N} \to [0, 1]$, the function f is called a *lower bound* for the graph invariant $\alpha_{\mathcal{C}}$ if $\alpha_{\mathcal{C}}(G) \geqslant f(G)$ holds for all graphs G. Such a lower bound f is said to be *extremal* if f cannot be augmented, that is, if there does not exist $g : \mathbb{N} \to [0, 1]$ with $g \not\equiv 0$ such that $f + g$ is also a lower bound for $\alpha_{\mathcal{C}}$. We remark that, for the graph parameters studied in this paper, it will be the case that every lower bound f is dominated by some extremal lower bound f', in the sense that $f(d) \leqslant f'(d)$ for every $d \in \mathbb{N}$.

A forest is *linear* if every component is a path. A *caterpillar* is tree obtained from a path by adding leaves to it. In other words, a tree T is a caterpillar if $T - V(P)$ has no edge for some path P of T. A *forest of caterpillars* is a forest, every component of which is a caterpillar. Forests of caterpillars can equivalently be characterized as the graphs with pathwidth at most 1. A special case of a forest of caterpillars is a *forest of stars*, a forest whose components are stars. We note that forests of star coincide with graphs of treedepth at most 2. (We skip the definitions of treewidth, pathwidth, and treedepth, as they will not be needed in this paper; we refer the interested reader to [5].)

We use the following notations: $\mathcal{L}$ denotes the set of all linear forests and $\mathcal{S}$ denotes the set of all forests of stars.

3. Linear forests

Let us start with the proof of Theorem 4, which we restate here for convenience.

Theorem 4. *Let $f : \mathbb{N} \to \mathbb{R}$ be defined as follows:*

$$f(d) = \begin{cases} 1 & \text{if } d = 0 \\ \frac{5}{6} & \text{if } d = 1 \\ \frac{2}{d+1} & \text{if } d \geqslant 2. \end{cases}$$

Then, every graph G has an induced linear forest with at least $\sum_{v \in V(G)} f(d(v))$ vertices.

Proof. We proceed by induction on $|V(G)|$. Let $\Delta := \Delta(G)$. First, suppose that $\Delta \leqslant 2$. Then every component of G is a path or a cycle. Each path component P contributes at most $|V(P)|$ to $f(G)$ and can be taken entirely. Each cycle component C contributes $2|V(C)|/3$ to $f(G)$, and we can take all vertices but one from C. Note that $|V(C)| - 1 \geqslant 2|V(C)|/3$ since $|V(C)| \geqslant 3$. Hence, this results in an induced linear forest of G with at least $f(G)$ vertices, as desired.

Next, assume that $\Delta \geqslant 3$. Observe that:

$$f(k-1) - f(k) \geqslant f(d-1) - f(d) \qquad \forall k, d \quad \text{with } 1 \leqslant k \leqslant d \tag{1}$$

and

$$d \cdot (f(d-1) - f(d)) = f(d) \qquad \forall d \qquad \text{with } d \geqslant 3. \tag{2}$$

(Indeed, the value of $f(1)$ was chosen so that $f(2) - f(3) = f(1) - f(2) = f(0) - f(1) = 1/6$ and (1) holds.)

Let v be a vertex of maximum degree in G, let $G' := G - v$ and, using induction on G', let F be an induced linear forest in G' with at least $f(G')$ vertices. Using (1) and (2), we obtain

$$f(G') - f(G) = \sum_{w \in N_G(v)} \big(f(d_G(w) - 1) - f(d_G(w))\big) - f(\Delta) \geqslant \Delta\,\big(f(\Delta - 1) - f(\Delta)\big) - f(\Delta) = 0$$

implying that $f(G') \geqslant f(G)$. Therefore, F has the desired size for G. $\square$

Corollary 5 follows from Theorem 4, as we now explain.

Proof of Corollary 5. Let L be the set of vertices of G with degree 1 and let $G' := G - L$. By Theorem 4, there exists an induced linear forest F' of G' with at least $f(G')$ vertices, where f is the function from Theorem 4. Observe that $d_G(v) \geqslant d_{G'}(v)$ for every $v \in V(G')$. In particular this implies $f(d_G(v)) \leqslant f(d_{G'}(v))$ for every $v \in V(G')$ since f is nonincreasing. Let $F := G[V(F') \cup L]$. Then F is a forest of caterpillars that is induced in G. Furthermore,

$$\begin{aligned} \big|V(F)\big| &= |L| + \big|V(F')\big| \\ &\geqslant |L| + \sum_{v \in V(G')} f(d_{G'}(v)) \\ &\geqslant \sum_{v \in L} 1 + \sum_{v \in V(G) - L} f(d_G(v)) \\ &= \sum_{v \in V(G)} \frac{2}{d_G(v) + 1}, \end{aligned}$$

where the last equality holds because every vertex in $V(G')$ has degree at least 2 in G. Therefore, F has the desired size. $\square$

As mentioned in the introduction, the lower bound f provided in Theorem 4 for induced linear forests is extremal but is not the only extremal lower bound. For instance, the following function f' is also a lower bound that is extremal, yet f and f' are not comparable.

$$f' : \mathbb{N} \to [0,1] : d \mapsto \begin{cases} 1 & \text{if } d = 0 \\ 1 & \text{if } d = 1 \\ \frac{2}{3} & \text{if } d = 2 \\ 0 & \text{if } d \geqslant 3. \end{cases}$$

As it turns out, every lower bound is dominated by some extremal lower bound, and there are infinitely many extremal lower bounds. A full characterization is given in Theorem 6; each extremal lower bound is uniquely determined by the weight it sets to degree-1 vertices: This should be at least 5/6 (as in Theorem 4) and at most 1 (as in the function f' above). Every intermediate value is feasible, provided the weights of all vertices with degree at least 3 are adapted in the right way.

Proving Theorem 6 is the main goal of this paper. However, our proof approach works in a slightly more general context (with no extra work), that of finding induced forest of caterpillars with maximum degree at most k, where $k \geqslant 2$ is a fixed constant. For $k = 2$, this corresponds to induced linear forests. We consider this more general setting in the next section and prove the full characterization of the lower bounds for fixed k, see Theorem 8 in the next section. Theorem 6 follows then by taking $k = 2$.

4. Caterpillars of bounded degree

Given an integer k with $k \geqslant 2$, let $\mathcal{C}_k$ denote the set of all forests of caterpillars of maximum degree at most k. In particular $\mathcal{L} = \mathcal{C}_2$. Here is the characterization of extremal lower bounds for the parameter $\alpha_{\mathcal{C}_k}$.

Theorem 8. *For every integer $k \geqslant 2$ and every $\varepsilon \in \mathbb{R}$ with $0 \leqslant \varepsilon \leqslant \frac{2}{(k+1)(k+2)}$, the following function is an extremal lower bound for $\alpha_{\mathcal{C}_k}$:*

$$f_{k,\varepsilon} : \mathbb{N} \to [0,1] : d \mapsto \begin{cases} 1 & \textit{if } d = 0 \\ 1 - \varepsilon & \textit{if } d = 1 \\ \frac{2}{d+1} & \textit{if } 2 \leqslant d \leqslant k \\ \min\left\{(k+1)\varepsilon, \frac{2}{d+1}\right\} & \textit{if } d \geqslant k+1. \end{cases}$$

Furthermore, these functions completely characterize the possible lower bounds: Every lower bound for $\alpha_{\mathcal{C}_k}$ is dominated by $f_{k,\varepsilon}$ for some ε.

Let us make a few remarks before proving Theorem 8. As mentioned earlier, Theorem 6 follows from the above theorem by taking $k = 2$. Note also that for $\varepsilon = \frac{2}{(k+1)(k+2)}$, one obtains the following extremal lower bound for $\alpha_{\mathcal{C}_k}$:

$$f_k : \mathbb{N} \to [0,1] : d \mapsto \begin{cases} 1 & \text{if } d = 0 \\ \frac{k(k+3)}{(k+1)(k+2)} & \text{if } d = 1 \\ \frac{2}{d+1} & \text{if } d \geqslant 2, \end{cases}$$

which generalizes Theorem 4 (for $k = 2$). By varying k from 2 to $+\infty$, one can think of the above lower bound f_k for $\alpha_{\mathcal{C}_k}$ as interpolating between the bounds in Theorem 4 and in Corollary 5.

Given the degree distribution of a graph G, it is not difficult to find a value of ε in Theorem 8 giving the best possible lower bound on $\alpha_{\mathcal{C}_k}$, that is, such that $f_{k,\varepsilon}(G)$ is maximized. This is the content of the following theorem.

Theorem 9. *Let G be a graph, and for $d \geqslant 0$ let n_d denote the number of vertices in G with degree d. Let $\varepsilon_k^* := \frac{2}{(k+1)(D^*+1)}$, where D^* is the smallest integer $D \geqslant k+1$ such that $(k+1)\sum_{d=k+1}^{D} n_d \geqslant n_1$ if there is such an integer, otherwise let $\varepsilon_k^* := 0$.*

Then, the function $t_{G,k} : [0, \frac{2}{(k+1)(k+2)}] \to \mathbb{R} : \varepsilon \mapsto \sum_{v \in V(G)} f_{k,\varepsilon}(d(v))$ is maximized at ε_k^*. In other words, f_{k,ε_k^*} provides the best lower bound that can be achieved on G in Theorem 8.

Proof. First observe that $t_{G,k}$ is piecewise linear. Indeed, for every integer $D \geqslant k+1$, on the interval $I_D = [\frac{2}{(k+1)(D+2)}, \frac{2}{(k+1)(D+1)}]$ we have:

$$t_{G,k}(\varepsilon) = \left(n_0 + n_1 + \sum_{d=2}^{k} \frac{2n_d}{d+1} + \sum_{d=D+1}^{\Delta(G)} \frac{2n_d}{d+1}\right) + \varepsilon\left((k+1)\sum_{d=k+1}^{D} n_d - n_1\right).$$

Thus, the linear components of $t_{G,k}$ consist of the interval $[0, \frac{2}{(k+1)(\Delta(G)+1)}]$ and the intervals I_D for $k+1 \leqslant D \leqslant \Delta(G)-1$, and $t_{G,k}$ is monotone on these intervals. Hence, it is enough to look at the values $t_{G,k}(0)$ and $t_{G,k}(\frac{2}{(k+1)(D+1)})$ for $D \geqslant k+1$.

If $(k+1)\sum_{d=k+1}^{\Delta(G)} n_d < n_1$, then for every $D \geqslant k+1$ and $\varepsilon = \frac{2}{(k+1)(D+1)}$:

$$t_{G,k}(0) - t_{G,k}(\varepsilon) = n_1\varepsilon - \sum_{d=k+1}^{\Delta(G)} n_d f_{k,\varepsilon}(d) \geqslant n_1\varepsilon - \sum_{d=k+1}^{\Delta(G)} n_d (k+1)\varepsilon > 0,$$

and so $t_{G,k}$ is maximized at $\varepsilon = \varepsilon_k^* = 0$.

If $(k+1)\sum_{d=k+1}^{\Delta(G)} n_d \geqslant n_1$, for $D \geqslant k+1$ define

$$T_{G,k}(D) := t_{G,k}\left(\frac{2}{(k+1)(D+1)}\right) - t_{G,k}\left(\frac{2}{(k+1)(D+2)}\right).$$

Now, since

$$\begin{aligned} T_{G,k}(D) &= \frac{2n_1}{k+1}\left(\frac{1}{D+2} - \frac{1}{D+1}\right) + \sum_{d=k+1}^{D} 2n_d\left(\frac{1}{D+1} - \frac{1}{D+2}\right) \\ &= \frac{2}{(k+1)(D+1)(D+2)}\left((k+1)\sum_{d=k+1}^{D} n_d - n_1\right), \end{aligned}$$

we see that $T_{G,k}(D) \geqslant 0$ for every $D \geqslant D^*$, and conversely, $T_{G,k}(D) < 0$ for every $D < D^*$. Therefore, $t_{G,k}(\varepsilon)$ is maximized at $\varepsilon = \varepsilon_k^* = \frac{2}{(k+1)(D^*+1)}$. $\square$

Theorem 8 provides an exact characterization of all the lower bounds for $\alpha_{\mathcal{C}_k}$ that only depend on the degree distribution of the graph G under consideration. That is, knowing only the degree distribution of G, it is not possible to write a better lower bound for $\alpha_{\mathcal{C}_k}$ than those described in the theorem. However, if we know some extra local information about the degree-1 vertices of G, namely the degrees of their neighbors, it is possible to state a more precise lower bound on $\alpha_{\mathcal{C}_k}$ that dominates all the bounds provided in Theorem 8. To state this lower bound, we need to introduce the following function $h_{k,G} : V(G) \to [0,1]$ where G is a graph and k is an integer with $k \geqslant 2$:

$$h_{k,G}(v) := \begin{cases} 1 & \text{if } d(v) = 0 \\ 1 & \text{if } d(v) = 1 \text{ and } d(w) \leqslant k \text{ where } N(v) = \{w\} \\ 1 - \frac{2}{(k+1)(d(w)+1)} & \text{if } d(v) = 1 \text{ and } d(w) \geqslant k+1 \text{ where } N(v) = \{w\} \\ \frac{2}{d(v)+1} & \text{if } d(v) \geqslant 2. \end{cases}$$

Here is the refined lower bound on $\alpha_{\mathcal{C}_k}$.

Theorem 10. *For every integer $k \geqslant 2$, every graph G has an induced forest of caterpillars of maximum degree at most k with at least $\sum_{v \in V(G)} h_{k,G}(v)$ vertices.*

Theorem 10 is the main technical result of the paper. Theorem 8 follows easily from Theorem 10, as we now explain. From a technical point of view, Theorem 10 can be thought of as a "local strengthening" of Theorem 8 to help the proof by induction go through.

Proof of Theorem 8 assuming Theorem 10. Given an integer $k \geqslant 2$ and a real number ε with $0 \leqslant \varepsilon \leqslant \frac{2}{(k+1)(k+2)}$, we need to show that $f_{k,\varepsilon}$ is an extremal lower bound for $\alpha_{\mathcal{C}_k}$. First, we show that it is a lower bound (using Theorem 10), and then we construct graphs showing that it is extremal.

To show that $f_{k,\varepsilon}$ is a lower bound for $\alpha_{\mathcal{C}_k}$, consider any graph G. We need to show that $\alpha_{\mathcal{C}_k}(G) \geqslant \sum_{v \in V(G)} f_{k,\varepsilon}(v)$. Clearly, it is enough to prove it in the case where G is connected. Furthermore, the inequality clearly holds if $|V(G)| \leqslant 2$, thus we may assume that $|V(G)| \geqslant 3$. In particular, $\Delta(G) \geqslant 2$.

Let L be the set of leaves of G. (Recall that a *leaf* is defined as a vertex of degree 1.) Given a vertex v of G with degree at least 2, let $\ell(v)$ denote the number of neighbors of v that are leaves, and let

$$D(v) \coloneqq \left(h_{k,G}(v) + \sum_{w \in N(v) \cap L} h_{k,G}(w) \right) - \left(f_{k,\varepsilon}(d(v)) + \sum_{w \in N(v) \cap L} f_{k,\varepsilon}(d(w)) \right).$$

Since every leaf is adjacent to a non-leaf vertex (since G is connected and $|V(G)| \geqslant 3$), it follows that

$$\sum_{v \in V(G)} (h_{k,G}(v) - f_{k,\varepsilon}(d(v))) = \sum_{v \in V(G)-L} D(v).$$

Consider some vertex v with $d(v) \geqslant 2$. If $2 \leqslant d(v) \leqslant k$, then

$$D(v) = \left(\frac{2}{d(v)+1} + \ell(v) \right) - \left(\frac{2}{d(v)+1} + \ell(v) \cdot (1-\varepsilon) \right) = \ell(v)\varepsilon \geqslant 0.$$

If on the other hand $d(v) \geqslant k+1$, then

$$f_{k,\varepsilon}(d(v)) + \sum_{w \in N(v) \cap L} f_{k,\varepsilon}(d(w)) = \ell(v) + \min\left\{ (k+1)\varepsilon, \frac{2}{d(v)+1} \right\} - \ell(v)\varepsilon.$$

Thus, either $(k+1)\varepsilon \leqslant \frac{2}{d(v)+1}$, in which case:

$$D(v) = \frac{2(k+1-\ell(v))}{(k+1)(d(v)+1)} - (k+1-\ell(v))\varepsilon \geqslant \frac{2(k+1-\ell(v))}{(k+1)(d(v)+1)} - \frac{2(k+1-\ell(v))}{(k+1)(d(v)+1)} = 0,$$

or $(k+1)\varepsilon > \frac{2}{d(v)+1}$, in which case:

$$D(v) = \sum_{w \in N(v) \cap L} \big(h_{k,G}(w) - f_{k,\varepsilon}(d(w))\big) = \ell(v) \left(\varepsilon - \frac{2}{(k+1)(d(v)+1)} \right) \geqslant 0.$$

Therefore, in all possible cases for v we have $D(v) \geqslant 0$.

It follows that

$$\sum_{v \in V(G)} h_{k,G}(v) = \sum_{v \in V(G)} f_{k,\varepsilon}(d(v)) + \sum_{v \in V(G)-L} D(v) \geqslant \sum_{v \in V(G)} f_{k,\varepsilon}(d(v)).$$

Since $\alpha_{\mathcal{C}_k}(G) \geqslant \sum_{v \in V(G)} h_{k,G}(v)$ by Theorem 10, this concludes the proof that $f_{k,\varepsilon}$ is a lower bound for $\alpha_{\mathcal{C}_k}$.

Now it remains to show that (i) every lower bound for $\alpha_{\mathcal{C}_k}$ is bounded by $f_{k,\varepsilon}$ for some ε satisfying $0 \leqslant \varepsilon \leqslant \frac{2}{(k+1)(k+2)}$, and (ii) that these bounds are all extremal. Let us first show (i). Let $\varphi : \mathbb{N} \to \mathbb{R}$ be a lower bound for $\alpha_{\mathcal{C}_k}$ and define $\varepsilon \coloneqq 1 - \varphi(1)$. Clearly,

$$\varphi(d) \leqslant 1 \qquad \forall d \geqslant 0.$$

Also,

$$\varphi(d) \leqslant \frac{2}{d+1} \qquad \forall d \geqslant 2,$$

since

$$(d+1)\varphi(d) = \varphi(K_{d+1}) \leqslant \alpha_{\mathcal{C}_k}(K_{d+1}) = 2.$$

If $\varepsilon \geqslant \frac{2}{(k+1)(k+2)}$, then $(k+1)\varepsilon \geqslant \frac{2}{k+2} \geqslant \frac{2}{d+1}$ for all $d \geqslant k+1$. We deduce that $\varphi \leqslant f_{k,\frac{2}{(k+1)(k+2)}}$, and we are done.

Now, assume that $\varepsilon < \frac{2}{(k+1)(k+2)}$. We claim that φ is dominated by $f_{k,\varepsilon}$. Note that

- $\varphi(0) \leqslant 1 = f_{k,\varepsilon}(0)$;
- $\varphi(1) = 1 - \varepsilon = f_{k,\varepsilon}(1)$, and
- $\varphi(d) \leqslant \frac{2}{d+1} = f_{k,\varepsilon}(d)$ for $2 \leqslant d \leqslant k$.

It remains to show that $\varphi(d) \leqslant f_{k,\varepsilon}(d)$ for all $d \geqslant k+1$. To do so, let $n \geqslant 1$, and define the graph $H_{n,k}$ as follows:

$$V(H_{n,k}) = \{(v,i) : 1 \leqslant v \leqslant n, 0 \leqslant i \leqslant k+1\},$$

and there is an edge between vertex (v,i) and vertex (w,j) in $H_{n,k}$ if and only if $v = w$ and $i = 0$, or $v \neq w$ and $i = j = 0$. Informally, $H_{n,k}$ is the graph obtained by starting with K_n (whose vertices are $\{(v,0) : 1 \leqslant v \leqslant n\}$) and adding exactly $k+1$ leaves to each vertex.

We claim that $\alpha_{\mathcal{C}_k}(H_{n,k}) = (k+1)n$. It is clear that for every $1 \leqslant v \leqslant n$, at most $k+1$ of the vertices $(v,0), \dots, (v,k+1)$ can be in any induced forest of maximum degree at most k in $H_{n,k}$, otherwise $(v,0)$ must be in it and must have degree at least $k+1$, hence the inequality $\alpha_{\mathcal{C}_k}(H_{n,k}) \leqslant (k+1)n$. This upper bound holds with equality, as witnessed by the forest induced by the set $\{(v,i) : 1 \leqslant v \leqslant n, 1 \leqslant i \leqslant k+1\}$.

We deduce that the function φ must satisfy

$$n\varphi(n+k) + (k+1)n\varphi(1) \leqslant \alpha_{\mathcal{C}_k}(H_{n,k}) = (k+1)n.$$

In particular, for $d = n+k \geqslant k+1$:

$$\varphi(d) \leqslant \frac{1}{n}\Big((k+1)n\left(1 - \varphi(1)\right)\Big) = (k+1)\varepsilon.$$

Therefore, $\varphi(d) \leqslant \min\{(k+1)\varepsilon, \frac{2}{d+1}\} = f_{k,\varepsilon}(d)$ holds for all $d \geqslant k+1$, and we conclude that $\varphi \leqslant f_{k,\varepsilon}$.

It remains to show property (ii), stating that $f_{k,\varepsilon}$ is extremal for every ε. Let ε_1 and ε_2 be such that $0 \leqslant \varepsilon_1 < \varepsilon_2 \leqslant \frac{2}{(k+1)(k+2)}$. Then

$$f_{k,\varepsilon_1}(1) = 1 - \varepsilon_1 > 1 - \varepsilon_2 = f_{k,\varepsilon_2}(1)$$

and

$$f_{k,\varepsilon_1}(3) = 3\varepsilon_1 < 3\varepsilon_2 = f_{k,\varepsilon_2}(3).$$

We deduce that neither of f_{k,ε_1} and f_{k,ε_2} dominates the other. It follows that $f_{k,\varepsilon}$ is extremal for every ε with $0 \leqslant \varepsilon \leqslant \frac{2}{(k+1)(k+2)}$. □

The key to the proof of Theorem 10 is the following technical lemma, which we call the ABC lemma.

Lemma 11 (ABC lemma)**.** *Let G be a graph and let (A,B,C) be a partition of its vertex set (with some parts possibly empty). Then G contains an induced linear forest F satisfying*

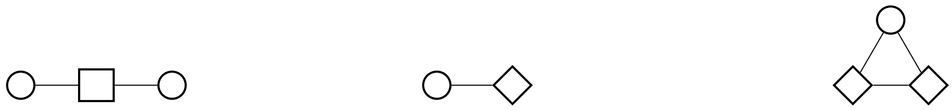
- $d_F(v) \leqslant 2$ *for all vertices* $v \in V(F) \cap A$,
- $d_F(v) \leqslant 1$ *for all vertices* $v \in V(F) \cap B$,
- $d_F(v) \leqslant 0$ *for all vertices* $v \in V(F) \cap C$, *and*
- $\big|V(F)\big| \geqslant f(G,A,B,C) := \sum_{v \in V(G)} f(v;G,A,B,C)$,

where

$$f(v;G,A,B,C)=\begin{cases} f_A(d_G(v)) & \text{if } v\in A\\ f_B(d_G(v)) & \text{if } v\in B\\ f_C(d_G(v)) & \text{if } v\in C\end{cases}\qquad f_A(d)=\begin{cases}1 & \text{if } d=0\\ \frac{5}{6} & \text{if } d=1\\ \frac{2}{d+1} & \text{if } d\geqslant 2\end{cases}$$

$$f_B(d)=\begin{cases}1 & \text{if } d=0\\ \frac{5}{6} & \text{if } d=1\\ \frac{1}{3} & \text{if } d=2\\ \frac{4}{3(d+1)} & \text{if } d\geqslant 3\end{cases}\qquad f_C(d)=\begin{cases}1 & \text{if } d=0\\ \frac{1}{6} & \text{if } d=1\\ \frac{1}{6} & \text{if } d=2\\ \frac{2}{3(d+1)} & \text{if } d\geqslant 3.\end{cases}$$

Before proving Theorem 10 using Lemma 11, let us point out that the values of $f_A(d)$ are best possible, as already discussed before. The values of $f_B(d)$ and $f_C(d)$ for $d=0,1,2$ are best possible too. This is clear for $f_B(0)$, $f_B(1)$, and $f_C(0)$. For $f_B(2), f_C(1), f_C(2)$, this is shown by the examples in Figure 1. (For $d\geqslant 3$, the values of $f_B(d)$ and $f_C(d)$ in Lemma 11 are most likely not best possible but they are good enough for our purposes.)



$f_B(2)\leqslant 2(1-f_A(1))=\frac{1}{3}$ $\qquad$ $f_C(1)\leqslant 1-f_A(1)=\frac{1}{6}$ $\qquad$ $f_C(2)\leqslant \frac{1}{2}(1-f_A(2))=\frac{1}{6}$

FIGURE 1. Examples showing that the values of $f_B(2)$, $f_C(1)$, and $f_C(2)$ in Lemma 11 are best possible. Vertices in A are shown as circles, vertices in B are shown as squares, and vertices in C as diamonds.

In order to lighten the notations somewhat, we will abbreviate $f(v;G,A,B,C)$ into $f(v;G)$ in the proofs when the partition (A,B,C) of $V(G)$ is clear from the context.

Now, let us show that Lemma 11 implies Theorem 10. The next section will be devoted to the proof of Lemma 11.

Proof of Theorem 10 assuming Lemma 11. Let $k\geqslant 2$ and let G be a graph. We need to show that G contains an induced forest of caterpillars of maximum degree at most k with at least $\sum_{v\in V(G)} h_{k,G}(v)$ vertices. The proof is by induction on $|V(G)|$, with the base case $|V(G)|\leqslant 2$ being easily seen to be true.

For the inductive part, suppose that $|V(G)|\geqslant 3$ and that the theorem holds for graphs with a smaller number of vertices. If G contains a vertex u adjacent to at least $k+1$ leaves, then $\sum_{v\in V(G-u)} h_{k,G-u}(v)\geqslant \sum_{v\in V(G)} h_{k,G}(v)$, and we are done by induction on $G-u$. Thus, we may assume that there is no such vertex in G. Similarly, if G is not connected, we are done by applying induction on the connected components of G, so we may assume that G is connected.

Let L be the set of leaves of G. Since G is connected and $|V(G)|\geqslant 3$, no leaf is adjacent to another leaf; that is, L is an independent set.

Define the following sets:

$$\begin{aligned} A' &:= \{v\in V(G)-L \text{ s.t. } \left|N(v)\cap L\right|\leqslant k-2\}\\ B' &:= \{v\in V(G)-L \text{ s.t. } \left|N(v)\cap L\right| = k-1\}\\ C' &:= \{v\in V(G)-L \text{ s.t. } \left|N(v)\cap L\right| = k\}.\end{aligned}$$

Applying Lemma 11 on $G' := G-L$ with the partition (A',B',C') of $V(G')$, we obtain an induced linear forest F' in G' satisfying the degree constraints of the lemma. Now define $F :=$

$G[V(F') \cup L]$. Clearly, F is an induced forest of caterpillars in G satisfying $\Delta(F) \leqslant k$. We will show that F has the desired number of vertices.

For a vertex v of G, let $\ell(v)$ denote the number of neighbors of v in G that are leaves. Note that every vertex $v \in V(G')$ satisfies $d_G(v) \geqslant 2$ and thus $h_{k,G}(v) = \frac{2}{d_G(v)+1}$. For such a vertex v, define

$$D(v) \coloneqq \big(\ell(v) + f(v; G')\big) - \left(h_{k,G}(v) + \sum_{w \in N(v) \cap L} h_{k,G}(w)\right).$$

We claim that $D(v) \geqslant 0$ holds for every vertex $v \in A' \cup B' \cup C'$. To show this, let us fix some vertex $v \in V(G')$.

Case 1: $d_G(v) \leqslant k$. Then $h_{k,G}(v) = \frac{2}{d_G(v)+1}$ and $h_{k,G}(w) = 1$ for every vertex $w \in N(v) \cap L$. Thus, $D(v) = f(v; G') - \frac{2}{d_G(v)+1}$.

If $v \in A'$, then $D(v) = f_A(d_{G'}(v)) - \frac{2}{d_G(v)+1} \geqslant 0$ since $d_{G'}(v) \leqslant d_G(v)$.

If $v \in B'$, then $\ell(v) = k - 1 \geqslant 1$, so $d_{G'}(v) \leqslant 1$ and $f_B(d_{G'}(v)) \geqslant \frac{5}{6}$, implying $D(v) \geqslant \frac{5}{6} - \frac{2}{d_G(v)+1} > 0$.

If $v \in C'$, then $d_G(v) = \ell(v) = k$, $N(v) \subseteq L$ and thus $D(v) = 1 - \frac{2}{d_G(v)+1} > 0$.

Case 2: $d_G(v) \geqslant k + 1$. Then $h_{k,G}(w) = 1 - \frac{2}{(k+1)(d_G(v)+1)}$ for every vertex $w \in N(v) \cap L$, and thus

$$D(v) = f(v; G') - \frac{2}{d_G(v) + 1} \frac{k + 1 - \ell(v)}{k + 1}.$$

If $v \in A'$, then $f(v; G') = f_A(d_{G'}(v)) \geqslant \frac{2}{d_G(v)+1}$, and thus $D(v) \geqslant 0$ since $\ell(v) \geqslant 0$.

If $v \in B'$, then $\ell(v) = k - 1$, implying

$$D(v) = f_B(d_{G'}(v)) - \frac{2}{d_G(v) + 1} \frac{2}{k + 1} \geqslant f_B(d_{G'}(v)) - \frac{4}{3(d_G(v) + 1)} \geqslant f_B(d_{G'}(v)) - \frac{1}{3}.$$

Observe that $d_{G'}(v) \geqslant 2$. If $d_{G'}(v) = 2$ then $f_{B'}(d_{G'}(v)) = \frac{1}{3}$ and thus $D(v) \geqslant 0$. If $d_{G'}(v) \geqslant 3$ then $f_{B'}(d_{G'}(v)) = \frac{4}{3(d_{G'}(v)+1)}$ and hence $D(v) \geqslant \frac{4}{3(d_{G'}(v)+1)} - \frac{4}{3(d_G(v)+1)} > 0$.

If $v \in C'$, then $\ell(v) = k$, and thus

$$D(v) = f_C(d_{G'}(v)) - \frac{2}{d_G(v) + 1} \frac{1}{k + 1} \geqslant f_C(d_{G'}(v)) - \frac{2}{3(d_G(v) + 1)} \geqslant f_C(d_{G'}(v)) - \frac{1}{6}.$$

If $d_G(v) \leqslant k + 2$, then $d_{G'}(v) \leqslant 2$ and hence $f_C(d_{G'}(v)) \geqslant \frac{1}{6}$, so $D(v) \geqslant 0$. If $d_G(v) \geqslant k + 3$, then $D(v) \geqslant f_C(d_{G'}(v)) - \frac{2}{3(d_G(v)+1)} \geqslant \frac{2}{3(d_{G'}(v)+1)} - \frac{2}{3(d_G(v)+1)} > 0$.

Therefore, $D(v) \geqslant 0$ holds in all possible cases for vertex v, and the claim holds.

Using our claim, we may lower bound the number of vertices in F as follows

$$\big|V(F)\big| = |L| + \big|V(F')\big| \geqslant |L| + f(G') = h_{k,G}(G) + \sum_{v \in V(G')} D(v) \geqslant h_{k,G}(G),$$

and this concludes the proof. $\square$

5. Proof of the ABC Lemma

In this section, we prove Lemma 11.

Proof of Lemma 11. Given an induced linear forest F of a graph G, we say that F *respects* a partition (A, B, C) of $V(G)$ if F satisfies the degree bounds in the lemma w.r.t. the partition (A, B, C).

The proof of the lemma is by contradiction. Suppose thus that the lemma is false, and let G be a graph and let (A, B, C) be a partition of its vertex set that together form a counterexample to the lemma with a minimum number of vertices. Thus, G does not contain an induced linear forest F respecting (A, B, C) that has at least $f(G, A, B, C)$ vertices. Clearly, the minimality of G implies that G is connected. Furthermore, the lemma is easily seen to hold for $|V(G)| = 1, 2$, so we also have $|V(G)| \geqslant 3$.

Let us introduce the following notations: Given a nonnegative integer i, we let A_i denote the set of vertices in A having degree i in G. The sets B_i and C_i are defined similarly w.r.t. B and C.

In the proof, we will move around vertices between the three sets A, B, C. We see these sets as 'ranks', with A being the highest rank and C the lowest. *Promoting* a vertex v consists in moving v from B to A, or from C to B. Likewise, *demoting* v consists in moving v from A to B, or from B to C.

Note that every vertex v of G has degree at least 1. We define the *gain* $\gamma(v)$ of a vertex v to be the increase of $f(v; G)$ when $d(v)$ is decreased by one, that is,

$$\gamma(v) := \begin{cases} f_A(d(v)-1) - f_A(d(v)) & \text{if } v \in A \\ f_B(d(v)-1) - f_B(d(v)) & \text{if } v \in B \\ f_C(d(v)-1) - f_C(d(v)) & \text{if } v \in C. \end{cases}$$

Observe that $\gamma(v) \geqslant 0$ always holds, since $f_A(d)$, $f_B(d)$, and $f_C(d)$ are nonincreasing functions. Given that the values of the gains for vertices with small degrees will be repeatedly used in the proofs, we summarize these values in Table 1 for reference.

	$v \in A$	$v \in B$	$v \in C$
$d = 1$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{5}{6}$
$d = 2$	$\frac{1}{6}$	$\frac{1}{2}$	0
$d = 3$	$\frac{1}{6}$	0	0
$d = 4$	$\frac{1}{10}$	$\frac{1}{15}$	$\frac{1}{30}$
$d \geqslant 5$	$\frac{2}{d(d+1)}$	$\frac{4}{3d(d+1)}$	$\frac{2}{3d(d+1)}$

Table 1. Value of $\gamma(v)$ for a vertex v of degree d.

Similarly, we define the *loss* $\lambda(v)$ of vertex v to be the decrease in $f(v; G)$ when $d(v)$ is increased by one, that is,

$$\lambda(v) := \begin{cases} f_A(d(v)) - f_A(d(v)+1) & \text{if } v \in A \\ f_B(d(v)) - f_B(d(v)+1) & \text{if } v \in B \\ f_C(d(v)) - f_C(d(v)+1) & \text{if } v \in C. \end{cases}$$

Here also, $\lambda(v) \geqslant 0$ always holds. Table 2 gives a summary of the possible values for $\lambda(v)$.

With these notations in hand, we may now turn to the proof, which is split into several claims. We remark that in the proofs we often consider other (smaller) graphs derived from G; let us

	$v \in A$	$v \in B$	$v \in C$
$d = 1$	$\frac{1}{6}$	$\frac{1}{2}$	0
$d = 2$	$\frac{1}{6}$	0	0
$d = 3$	$\frac{1}{10}$	$\frac{1}{15}$	$\frac{1}{30}$
$d \geqslant 4$	$\frac{2}{(d+1)(d+2)}$	$\frac{4}{3(d+1)(d+2)}$	$\frac{2}{3(d+1)(d+2)}$

TABLE 2. Value of $\lambda(v)$ for a vertex v of degree d.

emphasize that the gain $\gamma(v)$ and loss $\lambda(v)$ of a vertex v must always be interpreted w.r.t. the original graph G and partition (A, B, C) (that is, as they are defined above).

Here is a quick outline of the proof. Ideally, we would have liked to use the same approach as in the proof of Theorem 4. However, this is not possible here because of the existence of vertices v not satisfying $f(v; G) = d(v)\gamma(v)$. Instead, we will show that there is a special vertex v^* such that all its neighbors have gain at least that of v^*, except perhaps for one neighbor that has zero gain. We will then use this vertex v^* to derive a contradiction regarding the minimality of our counterexample. The proof is organized as follows:

- Claim 1 to Claim 12 establish general properties of the graph G,
- these claims are then used to define v^*,
- Claim 13 to Claim 19 show various properties of v^*,
- a final contradiction is derived using all these properties.

Claim 1. *$f(v; G) > \sum_{w\in N(v)} \gamma(w)$ holds for every vertex $v \in V(G)$. In particular, $f(v; G) > \gamma(w)$ holds for every edge $vw \in E(G)$.*

Proof. Let $v \in V(G)$. The fact that G and (A, B, C) form a minimum counterexample implies that $f(G, A, B, C) > f(G', A', B', C')$, where $G' := G - v$ and (A', B', C') is the restriction of (A, B, C) to $V(G')$. In particular,

$$f(G, A, B, C) > f(G', A', B', C') = f(G, A, B, C) - f(v; G) + \sum_{w\in N(v)} \gamma(w).$$

Adding $f(v; G) - f(G, A, B, C)$ on both sides yields the required inequality. ■

Recall that we use the notation $N[S]$ for the set of vertices at distance at most 1 from a given set S of vertices in G, and $N^i(S)$ for the set of vertices at distance exactly i from S.

Claim 2. *Let F be an induced linear forest in G respecting (A, B, C) with $|V(F)| \geqslant 1$. Let $G' := G - N[V(F)]$ and let (A', B', C') be the restriction of (A, B, C) to $V(G')$. Then*

$$\begin{aligned}
|V(F)| &< f(G, A, B, C) - f(G', A', B', C') \\
&\leqslant \sum_{v\in N[V(F)]} f(v; G) - \sum_{w\in N^2(V(F))} \gamma(w) \\
&\leqslant \sum_{v\in N[V(F)]} f(v; G).
\end{aligned}$$

Proof. By the minimality of our counterexample, we know that G' contains an induced linear forest F' with at least $f(G', A', B', C')$ vertices. (Note that possibly $V(G') = \varnothing$, in which case

$V(F') = \varnothing$ and $f(G', A', B', C') = 0$.) Since $F \cup F'$ is an induced linear forest in G respecting (A, B, C), it follows that

$$f(G, A, B, C) > \left|V(F \cup F')\right| = \left|V(F)\right| + \left|V(F')\right| \geqslant \left|V(F)\right| + f(G', A', B', C').$$

Subtracting $f(G', A', B', C')$ from both sides yields the first inequality of the claim.

The second inequality of the claim follows from the fact that the three functions $f_A(d)$, $f_B(d)$, and $f_C(d)$ are non increasing and that $d_{G'}(v) \leqslant d_G(v) - 1$ holds for every vertex $v \in N_G^2(V(F))$, while $d_{G'}(v) = d_G(v)$ for every $v \in V(G') - N_G^2(V(F))$.

The last inequality follows from the fact that $\gamma(w) \geqslant 0$ holds for every vertex w of G. ■

Claim 3. *Let $v \in V(G)$. Then $1 - f(v; G) < \sum_{w \in N(v)} f(w; G)$, and thus in particular $\max_{w \in N(v)} f(w; G) > \frac{1 - f(v;G)}{d(v)}$.*

Proof. Immediate by Claim 2 with $F = G[\{v\}]$. ■

Claim 4. *Let $v \in V(G)$. Then decreasing $d(v)$ by one and then demoting v lowers $f(v; G)$ by at most $\frac{1}{6}$.*

Proof. For $d \geqslant 1$, define $D_A(d) := f_A(d) - f_B(d-1)$ and $D_B(d) := f_B(d) - f_C(d-1)$. Proving the claim amounts to showing that $D_A(d) \leqslant \frac{1}{6}$ and $D_B(d) \leqslant \frac{1}{6}$ for every $d \geqslant 1$.

We have $D_A(1) = D_A(2) = -\frac{1}{6} < 0$, $D_A(3) = \frac{1}{6}$ and $D_A(d) = \frac{2(d-2)}{3d(d+1)} < \frac{2}{3d} \leqslant \frac{1}{6}$ for $d \geqslant 4$. Thus, $D_A(d) \leqslant \frac{1}{6}$ for every $d \geqslant 1$.

We have $D_B(1) = -\frac{1}{6} < 0$, $D_B(2) = D_B(3) = \frac{1}{6}$ and $D_B(d) = \frac{2(d-1)}{3d(d+1)} < \frac{2}{3d} \leqslant \frac{1}{6}$ for $d \geqslant 4$. Thus, $D_B(d) \leqslant \frac{1}{6}$ for every $d \geqslant 1$. ■

Let $\delta(G)$ denote the minimum degree of a vertex in G.

Claim 5. $\delta(G) \geqslant 2$.

Proof. We already know that $\delta(G) \geqslant 1$ since G is connected and has at least three vertices.

Let $vw \in E(G)$. Using Claim 1 and the fact that $d(v) \geqslant 1$, we deduce that $\gamma(w) < f(v; G) \leqslant \frac{5}{6}$. This implies in particular that $C_1 = \varnothing$. (This is because it can be checked from the definition of $\gamma(w)$ that $\gamma(w) \leqslant \frac{5}{6}$, with equality if and only if $w \in C_1$.)

Now, suppose that there is some vertex v with degree 1 in G. Then $v \in A \cup B$ (as discussed above), and thus $f(v; G) = \frac{5}{6}$. Let w be the unique neighbor of v. Claim 1 ensures that $f(w; G) > \gamma(v) = \frac{1}{6}$, thus $w \notin C$. Let $G' := G - v$, and let (A', B', C') be the partition of $V(G')$ obtained by restricting (A, B, C) to $V(G')$ and demoting w. It follows from Claim 4 that

$$\begin{aligned} f(G', A', B', C') &= f(G, A, B, C) - f(v; G) - (f(w; G) - f(w; G')) \\ &\geqslant f(G, A, B, C) - f(v; G) - \frac{1}{6} \\ &\geqslant f(G, A, B, C) - 1. \end{aligned}$$

By the minimality of our counterexample, we know that G' contains an induced linear forest F' respecting (A', B', C') with at least $f(G', A', B', C')$ vertices. Then, $F := G[V(F') \cup \{v\}]$ is an induced linear forest of G respecting (A, B, C) and having at least $\left|V(F')\right| + 1 \geqslant f(G', A', B', C') + 1 \geqslant f(G, A, B, C)$ vertices, a contradiction.

Therefore, $\delta(G) \geqslant 2$, as claimed. ■

Claim 6. $B_2 = C_2 = \varnothing$.

Proof. Suppose that v is a degree-2 vertex of G. Denote by u and w its two neighbors, in such a way that $f(u;G) \geqslant f(w;G)$.

If $v \in B_2$, then by Claim 1, we know in particular $f(w;G) > \gamma(v) = \frac{1}{2}$ and $f(u;G) > \gamma(v) = \frac{1}{2}$. Since $\delta(G) \geqslant 2$, it follows that $u, w \in A_2$, and thus $\gamma(u) = \gamma(w) = \frac{1}{6}$. But then Claim 1 also states that $\frac{1}{3} = f(v;G) > \gamma(u) + \gamma(w) = \frac{1}{3}$, which is a contradiction.

If $v \in C_2$, then by Claim 3 we know that $f(u;G) + f(w;G) > 1 - f(v;G) = \frac{5}{6}$, which implies that $f(u;G) > \frac{5}{12}$, which implies in turn that $u \in A_2 \cup A_3$ and $\gamma(u) = \frac{1}{6}$. However, it follows then from Claim 1 that $\frac{1}{6} = f(v;G) > \gamma(u) = \frac{1}{6}$, a contradiction.

We deduce that v is neither in B_2 nor in C_2 (and thus is in A_2), as claimed. ■

Given two sets X and Y of vertices of G, we call X–Y *edge* an edge of G with one endpoint in X and the other endpoint in Y.

Claim 7. *G contains no A_2–B_3 edge.*

Proof. Suppose that v is a vertex in B_3. It follows from Claim 1 that $\left|N(v) \cap (A_2 \cup A_3)\right| \leqslant 1$. Thus, denoting x, y, z the three neighbors of v in such a way that $f(x;G) \geqslant f(y;G) \geqslant f(z;G)$, it follows that at most one of them is in A_2, and if there is one, it must be x. Arguing by contradiction, let us suppose that $x \in A_2$. Then $f(y;G) \leqslant \frac{2}{5}$ and $f(z;G) \leqslant \frac{2}{5}$.

Let $G' := G - \{v, y, z\}$. By the minimality of our counterexample, there is an induced linear forest F' in G' of size at least $f(G', A', B', C')$ respecting the partition (A', B', C') obtained by restricting (A, B, C) to $V(G')$. Now, observe that $F := G[V(F') \cup \{v\}]$ is an induced linear forest in G respecting (A, B, C) whose number of vertices is at least

$$\begin{aligned} f(G', A', B', C') + 1 &\geqslant f(G, A, B, C) + 1 - f(v;G) - f(y;G) - f(z;G) + \gamma(x) \\ &\geqslant f(G, A, B, C) + \frac{1}{30} \\ &\geqslant f(G, A, B, C), \end{aligned}$$

which contradicts the fact that G and (A, B, C) form a counterexample. ■

Claim 8. $\min_{w \in N(v)} f(w;G) > \frac{1}{6}$ *for every vertex $v \in B_3$. In particular, G contains no B_3–C edge.*

Proof. Suppose that v is a vertex in B_3 and denote by x, y, z its three neighbors in such a way that $f(x;G) \geqslant f(y;G) \geqslant f(z;G)$. Claim 7 implies that $x \notin A_2$, thus $d(x) \geqslant 3$, and $\lambda(x) \leqslant \frac{1}{10}$. The same holds true for vertex y.

By Claim 1, $\gamma(x) + \gamma(y) < f(v;G) = \frac{1}{3}$. Thus, x and y cannot be both in A_3 (as they would then each have a gain of $\frac{1}{6}$), and it follows that $y \notin A_3$ since $f(x;G) \geqslant f(y;G)$. This implies in turn that $\lambda(y) \leqslant \frac{1}{15}$, and thus

$$\lambda(x) + \lambda(y) \leqslant \frac{1}{10} + \frac{1}{15} = \frac{1}{6}.$$

Now, let $G' := G - z + xy$ (note that the edge xy might already be in $G - z$, in which case it is not added). Let (A', B', C') be the partition of $V(G')$ obtained by restricting (A, B, C) to $V(G')$ and promoting v (thus, $v \in A'_2$). Observe that every induced linear forest F' in G' respecting (A', B', C') is also an induced linear forest in G respecting (A, B, C), since the edge xy ensures that F' cannot contain all three vertices v, x, y. By the minimality of our counterexample, F' can be chosen so that it contains at least $f(G', A', B', C')$ vertices. Thus, we must have

$f(G', A', B', C') < f(G, A, B, C)$. Observe that $d_G(x) - 1 \leqslant d_{G'}(x) \leqslant d_G(x) + 1$ (depending on whether the edges xy and xz are in G or not), and thus in particular

$$f(x; G) - f(x; G') \leqslant \lambda(x).$$

The same holds for vertex y. Hence, letting $Z \coloneqq N_G(z) - \{v, x, y\}$, we obtain

$$\begin{aligned}
0 &< f(G, A, B, C) - f(G', A', B', C') \\
&= f(z; G) + \underbrace{f(v; G) - f(v; G')}_{=\frac{1}{3}-\frac{2}{3}=-\frac{1}{3}} + \underbrace{f(x; G) - f(x; G')}_{\leqslant \lambda(x)} + \underbrace{f(y; G) - f(y; G')}_{\leqslant \lambda(y)} - \underbrace{\sum_{w \in Z} \gamma(w)}_{\geqslant 0} \\
&\leqslant f(z; G) - \frac{1}{3} + \lambda(x) + \lambda(y) - 0 \\
&\leqslant f(z; G) - \frac{1}{3} + \frac{1}{6} = f(z; G) - \frac{1}{6}.
\end{aligned}$$

Therefore, $f(z; G) > \frac{1}{6}$. ■

Claim 9. *If $v \in B_3$, then $|N(v) \cap B_3| \leqslant 1$.*

Proof. Suppose that v is a vertex in B_3 and denote by x, y, z its three neighbors. Arguing by contradiction, suppose that at least two of them are in B_3, say $x, y \in B_3$. From Claim 8, we know that z cannot be in C. Furthermore, by Claim 6 and Claim 7, we know that $d(z) \geqslant 3$.

Let $G' \coloneqq G - v + xy + yz + xz$. (Note that some of the edges xy, yz, xz might already be in G.) Let (A', B', C') be the partition of $V(G')$ obtained by restricting (A, B, C) to $V(G')$ and demoting x, y, and z. Since $|V(G')| < |V(G)|$, the graph G' contains an induced linear forest F' respecting (A', B', C') with at least $f(G', A', B', C')$ vertices. Now, let $F \coloneqq G[V(F') \cup \{v\}]$. Observe that F is an induced linear forest in G respecting (A, B, C), since F' contains at most one of x, y, z. Observe also that $d_{G'}(x) \leqslant d_G(x) + 1$, and thus $f(x; G) - f(x; G') \leqslant f_B(3) - f_C(4) = \frac{1}{5}$. The same holds for vertex y. Moreover,

$$f_A(d) - f_B(d+1) = \frac{2(d+4)}{3(d+1)(d+2)} \qquad \text{and} \qquad f_B(d) - f_C(d+1) = \frac{2(d+3)}{3(d+1)(d+2)}$$

for $d \geqslant 3$. Hence, since $d_{G'}(z) \leqslant d_G(z) + 1$, we deduce that $f(z; G) - f(z; G') \leqslant \frac{2(3+4)}{3(3+1)(3+2)} = \frac{7}{30}$. Therefore,

$$\begin{aligned}
|V(F)| &\geqslant f(G', A', B', C') + 1 \\
&= f(G, A, B, C) - f(v; G) - \sum_{w \in N(v)} (f(w; G) - f(w; G')) + 1 \\
&\geqslant f(G, A, B, C) - f(v; G) - 2\frac{1}{5} - \frac{7}{30} + 1 \\
&= f(G, A, B, C) + \frac{1}{30} \\
&\geqslant f(G, A, B, C),
\end{aligned}$$

contradicting the fact that G and (A, B, C) form a counterexample. ■

Claim 10. *If $v \in A_3$, then $|N(v) \cap B_3| \leqslant 1$.*

Proof. Suppose that v is a vertex in A_3 and denote by x, y, z its three neighbors. Arguing by contradiction, suppose that at least two of them are in B_3, say $x, y \in B_3$. Note that by Claim 1, we know that $f(z; G) > \gamma(v) = \frac{1}{6}$, thus $z \notin C$, and hence z can be demoted. Let $G' \coloneqq G - v + xy$, and let (A', B', C') be the partition of $V(G')$ obtained by restricting (A, B, C) to $V(G')$ and demoting x, y, z (thus, x, y are both in C'_2 or both in C'_3, depending on whether the edge xy exists in G or not). Since $|V(G')| < |V(G)|$, the graph G' contains an induced linear forest F' respecting (A',B',C') with at least $f(G',A',B',C')$ vertices. Let $F := G[V(F')\cup\{v\}]$. Observe that F is an induced linear forest in G respecting (A,B,C). Indeed, $d_F(w) \leqslant d_{F'}(w)+1$ for every $w \in N(v)$, and $d_F(v) \leqslant 2$ since F' contains at most one of x,y (because of the edge xy), and furthermore v is not in a cycle in F since $x,y \in C'$. But then

$$\begin{aligned}\left|V(F)\right| &= \left|V(F')\right| + 1 \geqslant f(G',A',B',C') + 1\\ &\geqslant f(G,A,B,C) - f(v;G) - 3\frac{1}{6} + 1\\ &= f(G,A,B,C),\end{aligned}$$

where the second inequality is obtained by Claim 4, which ensures that $f(z;G) - f(z;G') \leqslant \frac{1}{6}$, and by the fact that $f_C(2) = f_C(3) = \frac{1}{6}$, hence for $w \in \{x,y\}$, we know that $f(w;G)-f(w;G') = \frac{1}{3} - \frac{1}{6} = \frac{1}{6}$. But this contradicts the fact that G and (A,B,C) form a counterexample. ■

Claim 11. *If u,v are two distinct vertices in B_3, then $\left|N(u)\cap N(v)\right| \leqslant 1$.*

Proof. Arguing by contradiction, suppose that u,v are two distinct vertices in B_3 with two common neighbors x,y. We know from Claim 7 and Claim 10 that neither x nor y can be in $A_2 \cup A_3$. In particular $f(x;G) < \frac{1}{2}$ and $f(y;G) < \frac{1}{2}$. Let $G' := G - \{x,y\}$, and let F' be an induced linear forest in G' respecting the restriction of (A,B,C) to $V(G')$ with at least $f(G',A',B',C')$ vertices. Obviously, F' is an induced linear forest in G respecting (A,B,C) as well. Then,

$$\begin{aligned}\left|V(F')\right| &\geqslant f(G',A',B',C')\\ &\geqslant f(G,A,B,C) - f(x;G) - f(y;G) + 2\left(f_B(1) - f_B(3)\right)\\ &> f(G,A,B,C) - 2\frac{1}{2} + 2\left(\frac{5}{6} - \frac{1}{3}\right)\\ &= f(G,A,B,C),\end{aligned}$$

which is a contradiction. ■

Claim 12. *G does not contain any C_3–C_3 edge.*

Proof. Arguing by contradiction, suppose that vw is an edge of G with $v,w \in C_3$. Denote by x,y and s,t the other two neighbors of respectively v and w, in such a way that $f(x;G) \geqslant f(y;G)$ and $f(s;G) \geqslant f(t;G)$. By Claim 3 we know that $f(x;G)+f(y;G) > 1-f(v;G)-f(w;G) = \frac{2}{3}$, and in particular $f(x;G) > \frac{1}{3}$. Furthermore, Claim 1 implies that $\gamma(x) < \frac{1}{6}$, and thus $x \in A_4$. From Claim 1, we know that $\gamma(y) < f(v;G) - \gamma(x) = \frac{1}{6} - \frac{1}{10} = \frac{1}{15}$. Finally, since $f(y;G) > \frac{2}{3} - f(x;G) = \frac{4}{15}$, we deduce that either $y \in A_6$ or $y \in B_3$. However, Claim 8 implies that $y \notin B_3$, hence $y \in A_6$. A symmetric argument shows that $s \in A_4$ and $t \in A_6$.

Let $S := N^2(w) \cap \{x,y\}$. Let $G' := G - N[w]$ and let (A',B',C') be the restriction of (A,B,C) to $V(G')$. Since $\gamma(z) \geqslant \frac{1}{21}$ holds for every vertex $z \in S$, applying Claim 2 with $F := G[\{w\}]$ we obtain

$$\begin{aligned}1 &< f(G,A,B,C) - f(G',A',B',C')\\ &= f(v;G) + f(w;G) + f(s;G) + f(t;G) + \sum_{z\in N^2(w)} \left(f(z;G) - f(z;G')\right)\\ &\leqslant f(v;G) + f(w;G) + f(s;G) + f(t;G) + \sum_{z\in S}\left(f(z;G) - f(z;G')\right)\\ &\leqslant f(v;G) + f(w;G) + f(s;G) + f(t;G) - \sum_{z\in S}\gamma(z)\\ &\leqslant \frac{107}{105} - \frac{1}{21}|S|,\end{aligned}$$

implying that $|S| < \frac{2}{5}$, and thus $|S| = 0$. Hence, $x, y \notin N^2(w)$, and therefore $x, y \in N(w)$, implying that $x = s$ and $y = t$.

Let $G'' := G - \{v, w\}$, and let (A'', B'', C'') be the restriction of (A, B, C) to $V(G'')$. By the minimality of our counterexample, we have $f(G, A, B, C) > f(G'', A'', B'', C'')$. However,

$$\begin{aligned} f(G'', A'', B'', C'') &= f(G, A, B, C) - 2f_C(3) + \big(f_A(2) - f_A(4)\big) + \big(f_A(4) - f_A(6)\big) \\ &= f(G, A, B, C) + \frac{1}{21} \\ &\geqslant f(G, A, B, C), \end{aligned}$$

which is a contradiction. ∎

Let us show that some vertex of G has (strictly) positive gain. Observe that if $\gamma(v) = 0$ holds for a vertex $v \in V(G)$ then v must belong to one of the three sets B_3, C_2, C_3. Moreover, $C_2 = \varnothing$ by Claim 6. Thus, if all vertices of G have a gain of zero, then from the connectedness of G and Claim 8 we deduce that either all vertices are in B_3, or they are in all C_3. However, Claim 9 rules out the former case, while Claim 12 rules out the latter. Therefore, some vertex of G has positive gain.

For the rest of the proof, we fix some vertex v^* with $\gamma(v^*) > 0$ and minimizing $\gamma(v^*)$ among all such vertices. Furthermore, in case $\gamma(v^*) = \frac{1}{6}$ and $A_3 \neq \varnothing$, we choose v^* so that $v^* \in A_3$.

Observe that $d(v^*) \geqslant 2$ (since $\delta(G) \geqslant 2$ by Claim 5) and $d(v^*) \geqslant 4$ in case $v^* \in B \cup C$.

Claim 13. *The vertex v^* satisfies the following three properties:*

(1) $v^* \notin A_2$,

(2) $f(v^*; G) = d(v^*)\gamma(v^*)$, and

(3) $N(v^*) \cap (B_3 \cup C_3) \neq \varnothing$.

Proof. Let us first show that (1) implies (2). We already know that $d(v^*) \geqslant 4$ in case $v^* \in B \cup C$. If $v^* \notin A_2$, then it can be checked that $f(v^*; G) = d(v^*)\gamma(v^*)$ holds in all possible cases for v^*. Thus, (1) implies (2).

From (2) and Claim 1, we obtain that $d(v^*)\gamma(v^*) = f(v^*; G) > \sum_{w \in N(v^*)} \gamma(w)$, implying that $\gamma(w) < \gamma(v^*)$ holds for some $w \in N(v^*)$, and thus $\gamma(w) = 0$ by our choice of v^*, implying in turn that $w \in B_3 \cup C_3$. Thus, (2) implies (3).

Hence, it only remains to prove (1). Arguing by contradiction, suppose that $v^* \in A_2$. Then, it follows from our choice of v^* that $V(G) = A_2 \cup B_3 \cup C_3$.

If $B_3 \neq \varnothing$, then by Claim 7 and Claim 8 every vertex in B_3 has all its neighbors in B_3. However, this contradicts Claim 9. Thus, $B_3 = \varnothing$.

If $C_3 \neq \varnothing$, then every vertex in C_3 has all its neighbors in A_2 by Claim 12. In particular, there is an edge $xy \in E(G)$ with $x \in C_3$ and $y \in A_2$. But then, $f(x; G) > \gamma(y)$ by Claim 1, which is not possible since $f(x; G) = \gamma(y) = \frac{1}{6}$. Thus, $C_3 = \varnothing$.

We conclude that $V(G) = A_2$, and thus that G is a cycle. Then $F := G - v^*$ is an induced linear forest in G that respecting (A, B, C) and with $|V(G)| - 1 \geqslant \frac{2}{3}|V(G)| = f(G, A, B, C)$ vertices, contradicting the fact G and (A, B, C) form a counterexample. Therefore, $v^* \notin A_2$, and (1) holds. ∎

Claim 14. *$d(v^*) \geqslant 4$, and thus $f(v^*; G) \leqslant \frac{2}{5}$ and $\gamma(v^*) \leqslant \frac{1}{10}$.*

Proof. If $d(v^*) \geqslant 4$, then it is easily checked that $f(v^*; G) \leqslant \frac{2}{5}$ (with equality only if $v^* \in A_4$), and thus $\gamma(v^*) \leqslant \frac{1}{10}$ since $\gamma(v^*) = f(v^*; G)/d(v^*)$ by Claim 13.

Let us show that $d(v^*) \geqslant 4$. Arguing by contradiction, suppose this is not the case. Then it follows from Claim 13 that $v^* \in A_3$. Furthermore, by our choice of v^*, every vertex of G with nonzero gain is in $A_2 \cup A_3$, and it follows that $V(G) = A_2 \cup A_3 \cup B_3 \cup C_3$. By Claim 1, we know that $f(w; G) > \gamma(v^*) = \frac{1}{6}$ holds for every vertex $w \in N(v^*)$. Since $f_C(3) = \frac{1}{6}$, it follows that $N(v^*) \cap C_3 = \varnothing$. Claim 13 implies then that there is some vertex $w \in N(v^*) \cap B_3$. By Claim 8, we know that $N(w) \subseteq B_3 \cup A_2 \cup A_3$. Claim 9 implies that w has a neighbor x that is distinct from v^* and not in B_3. Thus, $x \in A_2 \cup A_3$. Claim 7 implies then that $x \in A_3$. But then $\gamma(x) + \gamma(v^*) = f(w; G)$, which contradicts Claim 1. ■

Claim 15. *If $C_3 \neq \varnothing$ then $v^* \notin A_4 \cup A_5 \cup B_4$, and in particular, $f(v^*; G) \leqslant \frac{2}{7}$ and $\gamma(v^*) \leqslant \frac{1}{21}$.*

Proof. Suppose v is a vertex in C_3. Then it follows from Claim 8 and Claim 12 that $\gamma(w) > 0$ for every neighbor w of v. Using Claim 1, we deduce then that

$$\frac{1}{6} = f(v; G) > \sum_{w \in N(v)} \gamma(w) \geqslant 3\gamma(v^*),$$

and thus $\gamma(v^*) < \frac{1}{18}$. This implies that $v^* \notin A_4 \cup A_5 \cup B_4$, and in particular $f(v^*; G) \leqslant \frac{2}{7}$ and $\gamma(v^*) \leqslant \frac{1}{21}$. ■

Claim 16. $\left|N(v^*) \cap C_3\right| \leqslant 1$.

Proof. Arguing by contradiction, suppose that $u, w \in N(v^*) \cap C_3$ with $u \neq w$. Then $uw \notin E(G)$ by Claim 12. Denote by x, y the two neighbors of u distinct from v^*, and by s, t the two neighbors of w distinct from v^*. We may assume without loss of generality that $f(x; G) \geqslant f(y; G)$, $f(s; G) \geqslant f(t; G)$ and $f(x; G) \geqslant f(s; G)$. Claim 1 implies that, for every vertex $z \in N(u) \cup N(w)$, we have $\gamma(z) < \frac{1}{6}$, and thus $z \notin A_2 \cup A_3$, which implies in turn that $f(z; G) \leqslant \frac{2}{5}$. Since $uw \notin E(G)$, using Claim 2 with $F = G[\{u, w\}]$ we obtain

$$2 < f(u; G) + f(w; G) + \sum_{z \in N(u) \cup N(w)} f(z; G) \leqslant 2\frac{1}{6} + \left|N(u) \cup N(v)\right| \frac{2}{5},$$

and thus $\left|N(u) \cup N(w)\right| \geqslant \left\lceil \frac{25}{6} \right\rceil = 5$. Hence, v^* is the only common neighbor of u and w, and x, y, s, t are all distinct. Furthermore, Claim 8 and Claim 12 imply that $\{x, y, s, t\} \cap (B_3 \cup C_3) = \varnothing$. In particular, x, y, s, t all have non-zero gain. Since $C_3 \neq \varnothing$, by Claim 15 we know that $v^* \notin A_4$. Using Claim 2 with $F = G[\{u, w\}]$ again, we obtain that

$$\begin{aligned}
2 &< 2 \cdot \frac{1}{6} + f(x; G) + f(y; G) + f(s; G) + f(t; G) + f(v^*; G) \\
&\leqslant \frac{1}{3} + 5 \max\{f(x; G), f(y; G), f(s; G), f(t; G), f(v^*; G)\} \\
&= \frac{1}{3} + 5 \max\{f(x; G), f(v^*; G)\}.
\end{aligned}$$

(To see the last equality, recall that $f(x; G) \geqslant f(y; G)$ and $f(x; G) \geqslant f(s; G) \geqslant f(t; G)$). We deduce that $\max\{f(x; G), f(v^*; G)\} > \frac{1}{3}$, thus $\{x, v^*\} \cap A_4 \neq \varnothing$, but since $v^* \notin A_4$ we must have $x \in A_4$ and thus $f(x; G) = \frac{2}{5}$ and $\gamma(x) = \frac{1}{10}$. Applying Claim 1 on u again, we obtain

$$\frac{1}{6} = f(u; G) > \gamma(x) + \gamma(y) + \gamma(v^*) \geqslant \gamma(x) + 2\gamma(v^*) = \frac{1}{10} + 2\gamma(v^*).$$

Thus, $\gamma(v^*) < \frac{1}{2}(\frac{1}{6} - \frac{1}{10}) = \frac{1}{30}$. It follows that v^* does not belong to any of the sets $A_4, A_5, A_6, A_7, B_4, B_5, B_6, C_4$, and hence $f(v^*; G) \leqslant \frac{2}{9}$. Since Claim 1 implies in particular that $\frac{1}{6} = f(u; G) > \gamma(x) + \gamma(y)$, it follows that $\gamma(y) < \frac{1}{6} - \gamma(x) = \frac{1}{6} - \frac{1}{10} = \frac{1}{15}$, from which we deduce that y is not in A_4, A_5, B_4, and hence $f(y;G) \leqslant \frac{2}{7}$. Now, observe that

$$\begin{aligned}
\frac{2}{9} &\geqslant f(v^*;G) \\
&> 2 - 2\frac{1}{6} - f(x;G) - f(y;G) - (f(s;G) + f(t;G)) \\
&\geqslant \frac{5}{3} - \frac{2}{5} - \frac{2}{7} - \big(f(s;G) + f(t;G)\big) \\
&= \frac{103}{105} - \big(f(s;G) + f(t;G)\big),
\end{aligned}$$

which implies

$$f(s;G) + f(t;G) > \frac{103}{105} - \frac{2}{9} = \frac{239}{315} > \frac{2}{3}.$$

In particular, $f(s;G) > \frac{1}{3}$, and thus $s \in A_4$ (since $s \notin A_2 \cup A_3$), and $f(s;G) = \frac{2}{5}$. Furthermore, $f(t;G) > \frac{239}{315} - \frac{2}{5} = \frac{113}{315} > \frac{1}{3}$ implying $t \in A_4$. However, applying Claim 1 on w we obtain

$$\frac{1}{6} = f(w;G) > \gamma(s) + \gamma(t) = 2\frac{1}{10} = \frac{1}{5},$$

which is a contradiction. ■

Claim 17. *If $v^* \in A_4$ and v^* is in an triangle u, v^*, w in G, then at least one of u, w is not in B_3.*

Proof. Arguing by contradiction, suppose that $v^* \in A_4$ and that u, v^*, w is a triangle in G with $u, w \in B_3$. Let x be the neighbor of u distinct from v^* and w, and let y be the neighbor of w distinct from v^* and u. By Claim 11, we know that $x \neq y$. Claim 7 implies that $x, y \notin A_2$. It follows that $f(x;G) \leqslant \frac{1}{2}$ and $f(y;G) \leqslant \frac{1}{2}$.

Since $f(v^*;G) = \frac{2}{5}$, using Claim 2 with $F = G[\{u,w\}]$, we obtain that

$$\begin{aligned}
2 &< f(u;G) + f(w;G) + f(x;G) + f(y;G) + f(v^*;G) \\
&= \frac{2}{3} + f(x;G) + f(y;G) + f(v^*;G) \\
&= \frac{16}{15} + f(x;G) + f(y;G),
\end{aligned}$$

and thus $f(x;G) + f(y;G) > \frac{14}{15}$. Since $x, y \notin A_2$, we thus must have $x, y \in A_3$.

Reapplying Claim 2 with $F = G[\{u,w\}]$, and letting $Z := N^2(V(F))$, we obtain $2 < 2 + \frac{1}{15} - \sum_{z \in Z} \gamma(z)$, that is,

$$\sum_{z \in Z} \gamma(z) < \frac{1}{15}.$$

Since every vertex with non-zero gain has gain at least $\gamma(v^*) = \frac{1}{10}$, we deduce that Z is a (possibly empty) subset of $B_3 \cup C_3$. Since $x \in A_3$, Claim 10 implies that $(N(x) - \{u\}) \cap (B_3 \cup C_3) = \varnothing$. Since $N(x) \subseteq N[V(F)] \cup Z$, and since v^*, x, y are the only vertices of $N[V(F)] \cup Z$ not in $B_3 \cup C_3$, it follows that $N(x) = \{u, v^*, y\}$. By symmetry, we deduce that $N(y) = \{w, v^*, x\}$. But then $V(G) = \{u, v^*, w, x, y\}$, and

$$\begin{aligned}
f(G,A,B,C) &= f(u;G) + f(w;G) + f(v^*;G) + f(x;G) + f(y;G) \\
&= 2f_B(3) + f_A(4) + 2f_A(3) \\
&= 2 + \frac{1}{15} \\
&< 3.
\end{aligned}$$

However, $G[\{x,y,w\}]$ is an induced linear forest of G with three vertices respecting (A,B,C), contradicting the fact that G and (A,B,C) form a counterexample. ■

Claim 18. $\left|N(v^*) \cap B_3\right| \leqslant 1.$

Proof. Arguing by contradiction, suppose that $u, w \in N(v^*) \cap B_3$ with $u \neq w$.

First, suppose that $uw \in E(G)$. Let x be the neighbor of u distinct from v^* and w, and let y be the neighbor of w distinct from v^* and u. By Claim 11, we know that $x \neq y$. Also, $f(v^*; G) \leqslant \frac{1}{3}$ since $d(v^*) \geqslant 4$ (by Claim 14) and $v^* \notin A_4$ (by Claim 17). Using Claim 2 with $F = G[\{u, w\}]$ we then obtain that

$$\begin{aligned} 2 &< f(u;G) + f(w;G) + f(x;G) + f(y;G) + f(v^*;G) \\ &= \frac{2}{3} + f(x;G) + f(y;G) + f(v^*;G) \\ &\leqslant 1 + f(x;G) + f(y;G), \end{aligned}$$

and thus $f(x;G) + f(y;G) > 1$. However, this implies that at least one of x, y is in A_2, which contradicts Claim 7. Therefore, we conclude that $uw \notin E(G)$.

Let x, y be the two neighbors of u distinct from v^*, and let s, t be the two neighbors of w distinct from v^*. We may assume without loss of generality that $f(x;G) \geqslant f(y;G)$ and $f(s;G) \geqslant f(t;G)$, and also $\lambda(x)+\lambda(y) \geqslant \lambda(s)+\lambda(t)$. Claim 11 implies that x, y, s, t are all distinct. Claim 7 implies that $x, y \notin A_2$. It follows that $f(x;G) \leqslant \frac{1}{2}$ and $\gamma(x) \leqslant \frac{1}{6}$, and these two inequalities are strict in case $x \notin A_3$. The same holds for vertex y. By Claim 1, we know that $\gamma(x)+\gamma(y) < f(u;G) = \frac{1}{3}$, thus we cannot have both x and y in A_3, and hence $y \notin A_3$ (since $f(y;G) \leqslant f(x;G)$). It follows that $\lambda(x) + \lambda(y) \leqslant \frac{1}{10} + \frac{1}{15} = \frac{1}{6}$, and also $\lambda(s) + \lambda(t) \leqslant \lambda(x) + \lambda(y) \leqslant \frac{1}{6}$.

Let $G' := G - v^* + xy + st$. (Note that the edges xy and st might already be in G.) Let (A', B', C') be obtained from the restriction of (A, B, C) to $V(G')$ by promoting u and w (thus $u, w \in A_2'$). Observe that every induced linear forest F' in G' respecting (A', B', C') is also an induced linear forest in G respecting (A, B, C). (Indeed, the triangle u, x, y in G' ensures that F' misses at least one of these three vertices, and same for the triangle w, s, t in G'.) Since the lemma holds true for G' and (A', B', C'), it follows that

$$f(G, A, B, C) > f(G', A', B', C') \geqslant f(G, A, B, C) - f(v^*; G) - \lambda(x) - \lambda(y) - \lambda(s) - \lambda(t) + \frac{2}{3},$$

where the rightmost inequality holds with equality in case $xy, st \notin E(G)$. That is,

$$\frac{2}{3} - f(v^*; G) < \lambda(x) + \lambda(y) + \lambda(s) + \lambda(t).$$

Using that $\lambda(s) + \lambda(t) \leqslant \lambda(x) + \lambda(y) \leqslant \frac{1}{6}$, we deduce that $f(v^*; G) > \frac{1}{3}$, and thus $v^* \in A_4$ and $f(v^*; G) = \frac{2}{5}$. Also,

$$\frac{4}{15} = \frac{2}{3} - f(v^*; G) < \lambda(x) + \lambda(y) + \lambda(s) + \lambda(t) \leqslant 2\left(\lambda(x) + \lambda(y)\right),$$

implying that $\lambda(x) + \lambda(y) > \frac{2}{15}$. In particular, $\lambda(z) > \frac{1}{15}$ holds for some $z \in \{x, y\}$, and it can then be checked that $z \in A_3$ is the only possibility for z since $x, y \notin A_2$. Since $y \notin A_3$, we must have $z = x$. Since $v^* \in A_4$, we know from Claim 1 that $\frac{1}{3} = f(u;G) > \gamma(v^*) + \gamma(x) + \gamma(y) = \frac{1}{10} + \frac{1}{6} + \gamma(y)$, and thus $\gamma(y) < \frac{1}{3} - \frac{1}{6} - \frac{1}{10} = \frac{1}{15} < \gamma(v^*) = \frac{1}{10}$. By our choice of v^*, this implies that $\gamma(y) = 0$, and thus $y \in B_3 \cup C_3$. Since Claim 8 ensures $y \notin C$, we deduce that $y \in B_3$. Claim 10 implies in turn that $xy \notin E(G)$. Moreover, $v^*y \notin E(G)$ by Claim 17. It follows that y has exactly two neighbors outside of $\{x, u, v^*\}$. Denote these two neighbors by q and r in such a way that $f(q) \geqslant f(r)$. (Note that $\{q, r\} \cap \{s, t\}$ is not necessarily empty.)

Since $u \in B_3$, Claim 9 applied on y implies that $q, r \notin B_3$. Also, $q, r \notin C$ by Claim 8. Thus, $\gamma(q) > 0$ and $\gamma(r) > 0$, and hence $\gamma(q) \geqslant \gamma(v^*) = \frac{1}{10}$ and $\gamma(r) \geqslant \gamma(v^*) = \frac{1}{10}$ by our choice of v^*. This implies in turn that $q, r \notin B$, and hence $q, r \in A$. Since $q, r \notin A_2$ by Claim 7, we must have $q, r \in A_3 \cup A_4$. In particular, $\lambda(q) + \lambda(r) \leqslant 2\frac{1}{10}$.

Now, let $G'' := G - u + qr$ (note that the edge qr may already exist in G) and let (A'', B'', C'') be the partition (A, B, C) restricted to $V(G'')$ where the vertex y is promoted. Since $|V(G'')| < |V(G)|$, there is an induced linear forest F'' in G'' respecting (A'', B'', C'') with at least $f(G'')$ vertices. Observe that F'' is also and induced linear forest in G respecting (A, B, C), thanks to the triangle y, q, r in G''. It follows that

$$\begin{aligned}|V(F'')| &\geqslant f(G'')\\ &= f(G, A, B, C) - f(u; G) + \gamma(v^*) + \gamma(x) + (f(y; G'') - f(y; G)) - (\lambda(q) + \lambda(r))\\ &\geqslant f(G, A, B, C) - \frac{1}{3} + \frac{1}{10} + \frac{1}{6} + \left(\frac{2}{3} - \frac{1}{3}\right) - 2\frac{1}{10}\\ &\geqslant f(G, A, B, C),\end{aligned}$$

contradicting the fact that G and (A, B, C) form a counterexample. This concludes the proof of the claim. ■

Claim 19. $N(v^*) \cap B_3 = \varnothing$.

Proof. Arguing by contradiction, suppose that $N(v^*) \cap B_3$ is non empty. By Claim 18, there exists a unique vertex w in $N(v^*) \cap B_3$. Let x, y denote the neighbors of w distinct from v^* in such a way that $\lambda(x) \geqslant \lambda(y)$. Let $Z := N(v^*) - (B_3 \cup C_3)$ be the set of neighbors of v^* with non-zero gain. Let $G' := G - v^* + xy$ (note that the edge xy may already exist in G), and let (A', B', C') be obtained from the restriction of (A, B, C) to $V(G')$ by promoting w (thus $w \in A'$). Then, there is an induced linear forest F' in G' respecting (A', B', C') with at least $f(G', A', B', C')$ vertices. Observe that $G[V(F')]$ is an induced linear forest in G respecting (A, B, C) (thanks to the edge xy in G'). It follows that $f(G, A, B, C) > f(G', A', B', C')$.

By Claim 16, we know that v^* has a most one neighbor in C_3. Thus, $|Z| \geqslant d(v^*) - 2$. Let $Z' := Z - \{x, y\}$. Then, it can be checked that the following holds

$$\begin{aligned}0 &> f(G', A', B', C') - f(G, A, B, C)\\ &= -f(v^*; G) + \big(f(w; G') - f(w; G)\big) + \sum_{z \in Z \cup \{x,y\}} \big(f(z; G') - f(z; G)\big)\\ &= -f(v^*; G) + \big(f_A(2) - f_B(3)\big) + \sum_{z \in Z'} \big(f(z; G') - f(z; G)\big) + \sum_{z \in \{x,y\}} \big(f(z; G') - f(z; G)\big)\\ &= -f(v^*; G) + \frac{1}{3} + \sum_{z \in Z'} \gamma(z) + \sum_{z \in \{x,y\}} \big(f(z; G') - f(z; G)\big).\end{aligned}$$

By Claim 8 and Claim 12, we know that $x, y \notin B_3 \cup C_3$. For $z \in \{x, y\} \cap Z$, observe that $d_{G'}(z) \leqslant d_G(z)$ since $v^* \in N_G(z) - N_{G'}(z)$, and hence $f(z; G') \geqslant f(z; G)$. For $z \in \{x, y\} - Z$, observe that $d_{G'}(z) \leqslant d_G(z) + 1$ (with equality in case $xy \notin E(G)$). Thus, we deduce that

$$\begin{aligned}0 &> -f(v^*; G) + \frac{1}{3} + \sum_{z \in Z'} \gamma(z) + \sum_{z \in \{x,y\} - Z} \big(f(z; G') - f(z; G)\big)\\ &\geqslant -f(v^*; G) + \frac{1}{3} + \sum_{z \in Z'} \gamma(z) - \sum_{z \in \{x,y\} - Z} \lambda(z)\end{aligned}$$

and hence

$$f(v^*; G) > \frac{1}{3} + |Z'|\, \gamma(v^*) - \sum_{z \in \{x,y\} - Z} \lambda(z). \tag{3}$$

Let us show that $\{x, y\} \cap Z = \varnothing$. If x and y both are in Z, then it follows from (3) that $f(v^*; G) > \frac{1}{3} + |Z'|\, \gamma(v^*) \geqslant \frac{1}{3}$. Thus $v^* \in A_4$, and $f(v^*; G) = \frac{2}{5}$. Claim 15 implies then that C_3

must be empty, and hence $|Z'| = |Z| - 2 = d(v^*) - 1 - 2 = 1$ and

$$\frac{2}{5} = f(v^*;G) > \frac{1}{3} + |Z'|\,\gamma(v^*) = \frac{1}{3} + \frac{1}{10} = \frac{13}{30} > \frac{2}{5},$$

a contradiction. Therefore, x and y cannot both be in Z.

Now assume that either x or y is in Z but not both. Observe that $\lambda(y) \leqslant \lambda(x) \leqslant \frac{1}{10}$. Combining this observation with (3) and the fact that $|Z'| = |Z| - 1 \geqslant d(v^*) - 3 \geqslant 1$, it follows that

$$f(v^*;G) > \frac{1}{3} + |Z'|\,\gamma(v^*) - \frac{1}{10} = \frac{7}{30} + |Z'|\,\gamma(v^*) > \frac{7}{30}.$$

We know that $v^* \notin C$ since $f(v^*;G) > \frac{7}{30} > \frac{1}{6}$. If $v^* \in B$, then $d(v^*) = 4$ (otherwise $f(v^*;G)$ would be smaller than $\frac{7}{30}$), and thus $\gamma(v^*) = \frac{1}{15}$ and $f(v^*;G) = \frac{4}{15}$. But then

$$\frac{4}{15} = f(v^*;G) > \frac{7}{30} + |Z'|\,\frac{1}{15} \geqslant \frac{7}{30} + \frac{1}{15},$$

which is a contradiction. Hence, $v^* \notin B$, and we we deduce that $v^* \in A$. Moreover, the vertex v^* must satisfy

$$\frac{2}{d(v^*)+1} = f(v^*;G) > \frac{7}{30} + |Z'|\,\gamma(v^*) = \frac{7}{30} + \frac{2|Z'|}{d(v^*)(d(v^*)+1)},$$

or equivalently

$$2d(v^*) > \frac{7}{30}d(v^*)(d(v^*)+1) + 2|Z'| \geqslant \frac{7}{30}d(v^*)(d(v^*)+1) + 2(d(v^*)-3),$$

where the second inequality comes from the fact that $|Z'| = |Z| - 1 \geqslant d(v^*) - 3$. In particular,

$$d(v^*)(d(v^*)+1) < 6\frac{30}{7} = \frac{180}{7}.$$

Hence, we deduce that

$$d(v^*) \leqslant \left\lfloor \frac{-1 + \sqrt{1 + 4\frac{180}{7}}}{2} \right\rfloor = 4,$$

and by Claim 14, we deduce that $v^* \in A_4$ and thus $f(v^*;G) = \frac{2}{5}$. In particular, $C_3 = \varnothing$ by Claim 15, and thus $|Z'| = |Z| - 1 = d(v^*) - 2 = 2$. Hence,

$$\frac{2}{5} = f(v^*;G) > \frac{7}{30} + |Z'|\,\gamma(v^*) = \frac{7}{30} + 2\frac{1}{10} = \frac{13}{30} > \frac{2}{5},$$

which is a contradiction. We conclude that neither x nor y are in Z, as claimed.

Since $x, y \notin Z$, it from follows from Claim 13 and (3) that

$$d(v^*)\gamma(v^*) = f(v^*;G) > \frac{1}{3} + |Z|\,\gamma(v^*) - \lambda(x) - \lambda(y) \geqslant \frac{1}{3} + (d(v^*)-2)\gamma(v^*) - 2\lambda(x)$$

and thus

$$2\gamma(v^*) + 2\lambda(x) > \frac{1}{3}.$$

Since $x \notin A_2$ by Claim 7, we know that $\lambda(x) \leqslant \frac{1}{10}$. Thus, we deduce that $\gamma(v^*) > \frac{1}{6} - \lambda(x) \geqslant \frac{1}{6} - \frac{1}{10} = \frac{1}{15}$. Using Claim 14, it follows that $v^* \in A_4$ and $\gamma(v^*) = \frac{1}{10}$. Using the same inequality, we find $\lambda(x) > \frac{1}{6} - \gamma(v^*) = \frac{1}{6} - \frac{1}{10} = \frac{1}{15}$. This implies that $x \in A_3$ and $\gamma(x) = \frac{1}{6}$. Let us also recall that $\gamma(v^*) \leqslant \gamma(y)$ by our choice of v^*, since $\gamma(y) > 0$. Now, applying Claim 1 on w, we obtain

$$\gamma(v^*) \leqslant \gamma(y) < f(w;G) - \gamma(x) - \gamma(v^*) = \frac{1}{3} - \frac{1}{6} - \frac{1}{10} = \frac{1}{15} < \frac{1}{10} = \gamma(v^*),$$

which is a contradiction. ∎

Claim 13, Claim 16, and Claim 19 together imply that v^* has no neighbor in B_3 and exactly one neighbor in C_3, let us call it w. Let x, y be the two neighbors of w that are distinct from v^* in such a way that $f(x;G) \geqslant f(y;G)$, and let $Z := N(v^*) - N[w]$. Observe that for every $z \in Z$, we have $\gamma(z) > 0$, and thus $\gamma(v^*) \leqslant \gamma(z)$. Using Claim 2 with $F = G[\{w\}]$, we obtain

$$\begin{aligned}1 &< f(w;G) + f(v^*;G) + f(x;G) + f(y;G) - \sum_{z\in Z}\gamma(z)\\ &\leqslant \frac{1}{6} + f(v^*;G) + f(x;G) + f(y;G) - |Z|\,\gamma(v^*).\end{aligned}$$

Since $C_3 \neq \varnothing$, Claim 15 implies that $\gamma(v^*) \leqslant \frac{1}{21}$. By Claim 13, we know that $f(v^*;G) = d(v^*)\gamma(v^*)$. Since $1 \leqslant d(v^*) - |Z| \leqslant 3$, it follows that

$$\begin{aligned}\frac{5}{6} &< d(v^*)\gamma(v^*) + f(x;G) + f(y;G) - (d(v^*) - 3)\gamma(v^*)\\ &= f(x;G) + f(y;G) + 3\gamma(v^*)\\ &\leqslant f(x;G) + f(y;G) + 3\frac{1}{21},\end{aligned}$$

and thus

$$f(x;G) = \max\{f(x;G), f(y;G)\} \geqslant \frac{1}{2}\left(f(x;G) + f(y;G)\right) > \frac{1}{2}\left(\frac{5}{6} - \frac{1}{7}\right) = \frac{29}{84} > \frac{1}{3}.$$

Since $\gamma(x) < \frac{1}{6}$ (by Claim 1), we deduce that $x \in A_4$ and thus $f(x;G) = \frac{2}{5}$ and $\gamma(x) = \frac{1}{10}$. Then,

$$f(y;G) > \frac{29}{42} - f(x;G) = \frac{61}{210} > \frac{2}{7},$$

and since $\gamma(y) < \frac{1}{6}$ (by Claim 1), it follows that $y \in A_4 \cup A_5$. But now $\gamma(x) + \gamma(y) \geqslant \frac{1}{10} + \frac{1}{15} = \frac{1}{6}$, which contradicts Claim 1.

This final contradiction shows that G and (A, B, C) do not form a counterexample. This concludes the proof of Lemma 11. □

6. Forests of stars

As discussed in the introduction, a special kind of forests of caterpillars—or equivalently, graphs of pathwidth at most 1—are forests of stars, where every connected component is a star. (We remark that an isolated vertex is considered to be a star.) Forests of stars correspond to graphs of treedepth at most 2. The aim of this section is to study the impact of restricting caterpillars to be stars. The lower bound in Corollary 5 is no longer true with this extra requirement, and in fact there are infinitely many extremal lower bounds for finding induced forests of stars in graphs as a function of their degree sequence. Theorem 7 gives a characterization of all these lower bounds, which we restate here for convenience.

Theorem 7. *For every $\varepsilon \in \mathbb{R}$ with $0 \leqslant \varepsilon \leqslant \frac{1}{6}$, the following function provides a lower bound for induced forests of stars:*

$$f_\varepsilon : \mathbb{N} \to [0,1] : d \mapsto \begin{cases} 1 & \text{if } d = 0\\ 1-\varepsilon & \text{if } d = 1\\ \min\left\{\frac{3}{5}, \frac{1}{2}+\varepsilon\right\} & \text{if } d = 2\\ \min\left\{\frac{2}{d+1}, \frac{1}{d}+\varepsilon\right\} & \text{if } d \geqslant 3.\end{cases}$$

That is, every graph G contains an induced forest of stars with at least $\sum_{v\in V(G)} f_\varepsilon(d(v))$ vertices. Moreover, this lower bound is extremal, and every lower bound for induced forests of stars is dominated by some extremal lower bound of this form.

The proof of Theorem 7 relies on the following lemma.

Lemma 12 (AB lemma for forests of stars)**.** *Let G be a graph and let (A, B) be a partition of its vertex set (where parts are possibly empty). Then G contains an induced forest of stars F satisfying*

- *every edge $vw \in E(F)$ with $w \in B$ satisfies $v \in A$ and $d_F(v) = 1$, and*
- $\left|V(F)\right| \geqslant \sum_{v\in A} f_A(d_G(v)) + \sum_{v\in B} f_B(d_G(v))$

where

$$f_A(d) := \begin{cases} 1 & \text{if } d = 0 \\ \frac{5}{6} & \text{if } d = 1 \\ \frac{3}{5} & \text{if } d = 2 \\ \frac{2}{d+1} & \text{if } d \geqslant 3 \end{cases} \qquad \text{and} \qquad f_B(d) := \frac{1}{d+1}.$$

Before proving Lemma 12, let us show that it implies Theorem 7. In the subsequent proofs, an induced forest of stars F in a graph G is said to *respect* a partition (A, B) of $V(G)$ if F satisfies the first condition in Lemma 12 w.r.t. the partition (A, B), namely, that every edge $vw \in E(F)$ with $w \in B$ satisfies $v \in A$ and $d_F(v) = 1$.

Proof of Theorem 7 assuming Lemma 12. Let $\varepsilon \in \mathbb{R}$ with $0 \leqslant \varepsilon \leqslant \frac{1}{6}$. We first show that f_ε is a lower bound for $\alpha_{\mathcal{S}}$.

Let G be a graph. We need to show that $\alpha_{\mathcal{S}}(G) \geqslant \sum_{v\in V(G)} f_\varepsilon(d_G(v))$. Clearly, it is enough to prove this in the case where G is connected, so let us assume that G is connected. Also, we may assume that $|V(G)| \geqslant 3$, since the inequality is easily seen to hold in case G is isomorphic to K_1 or K_2. This implies in particular that no two leaves of G are adjacent.

Let $L := \{v \in V(G) \text{ s.t. } d_G(v) = 1\}$ be the set of leaves of G (which is thus an independent set). Let $G' := G - L$. Let (A', B') be the partition of $V(G')$ defined as follows

$$A' := \big\{v \in V(G) - L \text{ s.t. } N_G(v) \cap L = \varnothing\big\},$$
$$B' := \big\{v \in V(G) - L \text{ s.t. } N_G(v) \cap L \neq \varnothing\big\}.$$

By Lemma 12, there exists an induced forest of stars F' respecting (A', B') in G' with at least $\sum_{v\in A'} f_A(d_{G'}(v)) + \sum_{v\in B'} f_B(d_{G'}(v))$ vertices. Let $F := G[V(F') \cup L]$. We claim that F is an induced forest of stars in G. It is clear that F is a forest, since it is obtained from F' by (possibly) adding some leaves. Now, suppose that some connected component of F is not a star. Then, one can find an edge vw in that component where both v and w have degree at least 2 in F. These two vertices (and the edge vw) must be in F' as well, since all vertices in $F - V(F')$ are leaves of F. At least one of v, w is a leaf in F' (since every component of F' is a star), say w is a leaf in F'. Since w is no longer a leaf in F, it follows that $w \in B'$. By the conditions of Lemma 12, this implies in turn that $v \in A'$ and v is a leaf of F'. But then, v has no leaf neighbor in G, and hence v is a leaf of F as well, a contradiction. Therefore, F is forest of stars, as claimed.

Next, we will show that F has at least $\sum_{v\in V(G)} f_\varepsilon(d_G(v))$ vertices. First, observe that

$$f_A(d_{G'}(v)) = f_A(d_G(v)) \geqslant f_\varepsilon(d_G(v))$$

for every vertex $v \in A'$, since v has degree at least 2 in G.

For $v \in B'$, let $\ell(v) := \left|N_G(v) \cap L\right|$, and let

$$\begin{aligned} D(v) :=& \big(f_B(d_{G'}(v)) + \ell(v)\big) - \left(f_\varepsilon(d_G(v)) + \sum_{w\in N_G(v)\cap L} f_\varepsilon(d_G(w))\right) \\ =& f_B(d_{G'}(v)) + \ell(v)\varepsilon - f_\varepsilon(d_G(v)). \end{aligned}$$

Observe that $D(v) > 0$ in case $d_{G'}(v) = 0$. If $d_{G'}(v) \geqslant 1$, since $\ell(v) \geqslant 1$ and $d_G(v) \geqslant d_{G'}(v) + 1$, we know that $d_G(v) \geqslant 2$, and in particular $f_\varepsilon(d_G(v)) \leqslant \frac{1}{d_G(v)} + \varepsilon$. Therefore,

$$D(v) \geqslant \frac{1}{d_{G'}(v)+1} + \varepsilon - \frac{1}{d_G(v)} - \varepsilon \geqslant \frac{1}{d_{G'}(v)+1} - \frac{1}{d_{G'}(v)+1} = 0.$$

Thus, $D(v) \geqslant 0$ holds for every vertex $v \in B'$. It follows that

$$\begin{aligned}
\big|V(F)\big| &= \big|V(F')\big| + |L| \\
&\geqslant \sum_{v \in A'} f_A(d_{G'}(v)) + \sum_{v \in B'} f_B(d_{G'}(v)) + |L| \\
&\geqslant \sum_{v \in A'} f_\varepsilon(d_G(v)) + \sum_{v \in B'} f_B(d_{G'}(v)) + |L| \\
&= \sum_{v \in A'} f_\varepsilon(d_G(v)) + \sum_{v \in B'} \big(f_\varepsilon(d_G(v)) + D(v)\big) + \sum_{v \in L} (1-\varepsilon) \\
&\geqslant \sum_{v \in V(G)} f_\varepsilon(d_G(v)),
\end{aligned}$$

as desired. This concludes the proof that f_ε is a lower bound for $\alpha_{\mathcal{S}}$.

Now, it remains to show that (i) every lower bound for $\alpha_{\mathcal{S}}$ is dominated by f_ε for some ε satisfying $0 \leqslant \varepsilon \leqslant \frac{1}{6}$, and (ii) that these bounds are all extremal. We will proceed as in the proof of Theorem 8. Let us first show (i). Let $\varphi : \mathbb{N} \to \mathbb{R}$ be a lower bound for $\alpha_{\mathcal{S}}$, and define $\varepsilon := 1 - \varphi(1)$. Since φ is a lower bound for $\alpha_{\mathcal{S}}$, we know that

$$\varphi(0) \leqslant \alpha_{\mathcal{S}}(K_1) = 1 = f_\varepsilon(0)$$

and $5\varphi(2) \leqslant \alpha_{\mathcal{S}}(C_5)$, and thus

$$\varphi(2) \leqslant \frac{1}{5}\alpha_{\mathcal{S}}(C_5) = \frac{3}{5}.$$

Furthermore, ε was chosen so that $\varphi(1) = f_\varepsilon(1)$, and $\varphi(d) \leqslant \frac{2}{d+1}$ holds for all $d \geqslant 3$, as witnessed by the complete graph K_{d+1}.

Thus, it only remains to show that $\varphi(d) \leqslant \frac{1}{d} + \varepsilon$ for all $d \geqslant 2$. To do so, let $n \geqslant 2$, and define the graph K'_n as follows:

$$V(K'_n) = \{(v, i) : 1 \leqslant v \leqslant n, i \in \{0, 1\}\}$$

and there is an edge between (v, i) and (w, j) if and only if $i = j = 0$ and $v \neq w$, or $i = 0$, $j = 1$ and $v = w$. Informally, K'_n is the graph obtained by adding a single leaf to every vertex of the complete graph K_n.

We claim that $\alpha_{\mathcal{S}}(K'_n) = n + 1$. One can observe that $\alpha_{\mathcal{S}}(K'_n) \geqslant n + 1$ since $\{(v, 1) : 1 \leqslant v \leqslant n\} \cup \{(1, 0)\}$ induces a forest of stars with exactly $n + 1$ vertices. To see that $\alpha_{\mathcal{S}}(K'_n) \leqslant n + 1$, observe that every induced forest of stars contains at most two vertices from $\{(v, 0) : 1 \leqslant v \leqslant n\}$; furthermore, if it contains two such vertices, then it contains at most $n - 1$ vertices from $\{(v, 1) : 1 \leqslant v \leqslant n\}$, as otherwise it would contain an induced P_4.

It follows that

$$n - n\varepsilon + n\varphi(n) = n\varphi(1) + n\varphi(n) \leqslant \alpha_{\mathcal{S}}(K'_n) = n + 1$$

and in particular

$$\varphi(n) \leqslant \frac{1}{n} + \varepsilon,$$

as desired. We conclude that $\varphi \leqslant f_\varepsilon$.

Again, similarly to the argument in the proof of Theorem 8, the functions f_ε are pairwise incomparable for different values of ε, and thus must all be extremal. □

In order to prove Lemma 12, we need the following easy lemma on cubic graphs. (A *cubic* graph is a graph where every vertex has degree 3.)

Lemma 13. *Every cubic graph G admits a partition (V_1, V_2) of $V(G)$ such that $\Delta(G[V_i]) \leqslant 1$ for $i \in \{1, 2\}$. In particular, G contains an induced forest with at least $\frac{1}{2}\big|V(G)\big|$ vertices, where every connected component is isomorphic to K_1 or K_2.*

Proof. Given a partition (V_1, V_2) of $V(G)$, we let $E(V_1, V_2)$ denote the set of V_1–V_2 edges in G. Observe that, if (V_1, V_2) is a partition of $V(G)$, and if v is a vertex with at least two neighbors in its own part, then moving v to the other part increases $|E(V_1, V_2)|$ by at least 1. It follows that if we let (V_1, V_2) be a partition of $V(G)$ maximizing $|E(V_1, V_2)|$, then every vertex has at most one neighbor in its part, as desired. $\square$

We remark that Lemma 13 is a special case of a theorem of Lovász [4] about (non-proper) vertex colorings of graphs where each color class induces a graph whose maximum degree satisfies a prescribed upper bound.

Now, we prove Lemma 12.

Proof of Lemma 12. We use similar notations as in Lemma 11: Given a vertex v of a graph G, and a partition (A, B) of $V(G)$, we let $f(v; G, A, B)$ be $f_A(d_G(v))$ or $f_B(d_G(v))$, depending on whether $v \in A$ or $v \in B$. We abbreviate $f(v; G, A, B)$ into $f(v; G)$ when the partition (A, B) is clear from the context. Also, we let $f(G, A, B) \coloneqq \sum_{v \in V(G)} f(v; G, A, B)$, and for $i \geqslant 0$, we let A_i denote the set of vertices in A with degree at most i, and we define B_i similarly w.r.t. B.

To prove the lemma, we argue by contradiction. Suppose the lemma is false, and let G and (A, B) be a counterexample minimizing $\big|V(G)\big|$. Clearly, G is connected, and $\big|V(G)\big| \geqslant 3$.

Observe that every vertex v of G has degree at least 1. As in the proof of Lemma 11, we define the *gain* $\gamma(v)$ of a vertex v to be the increase of $f(v; G)$ when $d(v)$ is decreased by one, that is,

$$\gamma(v) \coloneqq \begin{cases} f_A(d(v) - 1) - f_A(d(v)) & \text{if } v \in A \\ f_B(d(v) - 1) - f_B(d(v)) & \text{if } v \in B. \end{cases}$$

Observe that, contrary to the proof of Lemma 11, here $\gamma(v) > 0$ always holds (that is, a gain of zero is not possible). If $v \in B$, then $\gamma(v) = \frac{1}{d(v)(d(v)+1)}$, and if $v \in A$:

$$\gamma(v) = \begin{cases} \frac{1}{6} & \text{if } d(v) = 1 \\ \frac{7}{30} & \text{if } d(v) = 2 \\ \frac{1}{10} & \text{if } d(v) = 3 \\ \frac{2}{d(v)(d(v)+1)} & \text{if } d(v) \geqslant 4 \end{cases}$$

The proof is split in a number of claims. Our first two claims mirror respectively Claim 1 and Claim 2 in the proof of Lemma 11. The proofs are verbatim the same (up to replacing (A, B, C) with (A, B) and 'linear forest' by 'forest of stars'), thus we do not repeat them here.

Claim 20. $f(v; G) > \sum_{w \in N(v)} \gamma(w)$ *holds for every vertex* $v \in V(G)$.

Claim 21. *Let F be an induced forest of stars in G respecting (A, B) with $\big|V(F)\big| \geqslant 1$. Let $G' \coloneqq G - N[V(F)]$ and let (A', B') be the restriction of (A, B) to $V(G')$. Then*

$$\big|V(F)\big| < f(G, A, B) - f(G', A', B') \leqslant \sum_{v \in N[V(F)]} f(v; G) - \sum_{w \in N^2(V(F))} \gamma(w) \leqslant \sum_{v \in N[V(F)]} f(v; G).$$

In the next claim we show that G has minimum degree at least 2. Thus, this claim parallels Claim 5 in the proof of Lemma 11. However, the proofs of the two claims are different, because each depends on the particular values of $f(v;G)$ for a vertex v in G, which are not the same in the two lemmas.

Claim 22. $\delta(G) \geqslant 2$.

Proof. Since G is connected with at least three vertices, we already know that $\delta(G) \geqslant 1$. Arguing by contradiction, suppose that v is a vertex of G with degree 1. Let w be its neighbor. Note that $d(w) \geqslant 2$, since G is not isomorphic to K_2.

Case 1: $v \in A_1$. Let $G' := G - v$ and let (A', B') be obtained by restricting (A, B) to $V(G')$ and moving w to B' in case it is not already there. By the minimality of our counterexample, we know that G' contains an induced forest of stars F' respecting (A', B') with at least $f(G', A', B')$ vertices. Observe that $F := G[V(F') \cup \{v\}]$ is an induced forest of stars of G respecting (A, B) (thanks to the fact that $w \in B'$). Observe also that $f(w;G) - f(w;G') \leqslant \frac{1}{6}$ holds in all possible cases for the vertex w (with equality if $w \in A_3$). It follows that

$$\begin{aligned}
|V(F)| &= |V(F')| + 1 \\
&\geqslant f(G', A', B') + 1 \\
&= f(G, A, B) - f(v;G) - \big(f(w;G) - f(w;G')\big) + 1 \\
&\geqslant f(G, A, B) - f(v;G) - \frac{1}{6} + 1 \\
&= f(G, A, B),
\end{aligned}$$

a contradiction. Hence, $v \notin A_1$.

Case 2: $v \in B_1$. Claim 20 implies that $f(w;G) > \gamma(v) = \frac{1}{2}$. It follows that $w \in A$ and $d(w) \leqslant 2$, and hence $w \in A_2$ since $d(w) \geqslant 2$. Let u be the neighbor of w distinct from v. Note that $f(u;G) \leqslant \frac{5}{6}$ since $d(u) \geqslant 1$. Using Claim 21 with $F = G[\{v, w\}]$, we obtain

$$2 < f(v;G) + f(w;G) + f(u;G) \leqslant \frac{1}{2} + \frac{3}{5} + \frac{5}{6} = \frac{29}{15},$$

a contradiction. This concludes the proof. ■

Claim 23. $A = A_2 \cup A_3$ *and* $B = B_2$.

Proof. Let $v \in V(G)$ be a vertex of minimum gain. By Claim 20 and our choice of v,

$$f(v;G) > \sum_{w \in N(v)} \gamma(w) \geqslant d(v)\gamma(v),$$

that is, $\gamma(v) < \frac{1}{d(v)} f(v;G)$. This implies that $v \notin B$ and $v \notin A_i$ for $i \geqslant 4$. Recalling that $\delta(G) \geqslant 2$ (by Claim 22), it follows that $v \in A_2 \cup A_3$, and in particular, $\gamma(v) \geqslant \frac{1}{10}$. By our choice of vertex v, this implies in turn that every vertex of G has gain at least $\frac{1}{10}$, and thus belongs to A_2, A_3, or B_2. ■

Claim 24. *There is no B_2–B_2 edge in G.*

Proof. Arguing by contradiction, suppose that vw is an edge of G with $v, w \in B_2$. Let x be the neighbor of v distinct from w. By Claim 20,

$$\frac{1}{3} = f(v;G) > \gamma(w) + \gamma(x) = \frac{1}{6} + \gamma(x),$$

that is, $\gamma(x) < \frac{1}{6}$. In particular, x cannot be in A_2 nor in B_2, and thus is in A_3 by Claim 23.

Case 1: $xw \in E(G)$**.** Let z be the neighbor of x distinct from v and w. Note that $f(z;G) \leqslant \frac{3}{5}$ by Claim 23. Using Claim 21 with $F = G[\{v,x\}]$, we obtain

$$2 < f(v;G) + f(w;G) + f(x;G) + f(z;G) = \frac{7}{6} + f(z;G) \leqslant \frac{7}{6} + \frac{3}{5} = \frac{53}{30} < 2,$$

a contradiction.

Case 2: $xw \notin E(G)$**.** Let y, z denote the two neighbors of x distinct from v. Using Claim 21 with $F = G[\{v\}]$, we obtain

$$1 < f(v;G) + f(w;G) + f(x;G) - \gamma(y) - \gamma(z) \leqslant 2\frac{1}{3} + \frac{1}{2} - 2\frac{1}{10} = \frac{29}{30} < 1,$$

a contradiction.

Thus, both cases lead to a contradiction, which concludes the proof. ■

Claim 25. $B_2 = \varnothing$.

Proof. Arguing by contradiction, suppose that v is a vertex in B_2. Denote by u, w the two neighbors of v in such a way that $f(u;G) \geqslant f(w;G)$. Then, $u, w \in A_2 \cup A_3$ by Claim 23 and Claim 24. Furthermore, using Claim 21, we obtain that $\frac{1}{3} = f(v;G) > \gamma(u) + \gamma(w)$, and hence $u, w \in A_3$.

Let $W := (N(u) \cup N(w)) - \{v\}$. It follows from Claim 20 that if a vertex in A_3 has a neighbor in B_2, then both its remaining neighbors must be in A_3. This implies that $W \subseteq A_3$.

Case 1: $uw \in E(G)$**.** Let x be the neighbor of u distinct from v and w. Thus, $x \in W$, and hence $x \in A_3$. Using Claim 21 with $F = G[\{u,v\}]$, we obtain

$$2 < f(u;G) + f(v;G) + f(w;G) + f(x;G) = \frac{11}{6} < 2,$$

a contradiction.

Case 2: $uw \notin E(G)$**.** Using Claim 21 with $F = G[\{v\}]$, we obtain

$$1 < f(v;G) + f(u;G) + f(w;G) - \frac{|W|}{10} = \frac{4}{3} - \frac{|W|}{10},$$

that is, $|W| < \frac{10}{3}$. Using the same claim with $F = G[\{v,u,w\}]$, we obtain

$$3 < f(v;G) + f(u;G) + f(w;G) + \frac{|W|}{2} = \frac{4}{3} + \frac{|W|}{2},$$

that is, $|W| > \frac{10}{3}$, which is impossible.

Thus, both cases lead to a contradiction, as desired. ■

It follows from Claim 23 and Claim 25 that $V(G) = A_2 \cup A_3$.

Claim 26. *If* $v \in A_3$*, then* $\left|N(v) \cap A_2\right| \leqslant 1$.

Proof. Let $v \in A_3$ and let x, y, z be the three neighbors of v. We may assume that $f(x;G) \geqslant f(y;G) \geqslant f(z;G)$. Recall that $A = A_2 \cup A_3$ and $B = \varnothing$, by Claim 23 and Claim 25. If $x \notin A_2$, then $N(v) \subseteq A_3$, and we are done. If $x \in A_2$, let us show that $y, z \in A_3$. By Claim 20, we know that

$$\frac{1}{2} = f(v;G) > \gamma(x) + \gamma(y) + \gamma(z) = \frac{7}{30} + \gamma(y) + \gamma(z) \geqslant \frac{7}{30} + 2\gamma(z).$$

Thus, $\gamma(z) < \frac{2}{15} < \frac{7}{30}$, and $z \notin A_2$, hence $z \in A_3$. Finally, we also know that

$$\gamma(y) < f(v;G) - \gamma(x) - \gamma(z) = \frac{1}{2} - \frac{7}{30} - \frac{1}{10} = \frac{1}{6} < \frac{7}{30},$$

Therefore, $y \notin A_2$, and it follows that $y \in A_3$. ■

Claim 27. *Every 3-vertex path in G contains at least one vertex in A_3.*

Proof. Arguing by contradiction, suppose that xyz is a 3-vertex path in G avoiding A_3. Then $x, y, z \in A_2$.

If $xz \in E(G)$ then G is isomorphic to K_3. Then, $f(G, A, B) = \frac{9}{5} \leqslant 2$ and $F := G[\{x, y\}]$ is an induced forest of stars respecting (A, B), showing that G and (A, B) do not form a counterexample, a contradiction. Hence, $xz \notin E(G)$.

Let v be the neighbor of x distinct from y, and let w be the neighbor of z distinct from y. If $v = w$ then by Claim 26, we know that $v \in A_2$, and thus G is isomorphic to a cycle of length 4. Then, $f(G, A, B) = \frac{12}{5} \leqslant 3$ and $F := G[\{x, y, z\}]$ is an induced forest of stars respecting (A, B), showing that G and (A, B) do not form a counterexample, a contradiction. Hence, $v \neq w$.

Since $v, w \in A_2 \cup A_3$, using Claim 21 with $F = G[\{x, y, z\}]$ we obtain

$$3 < f(x; G) + f(y; G) + f(z; G) + f(v; G) + f(w; G) \leqslant 5f_A(2) = 3,$$

a contradiction. ■

Claim 28. *There is no A_2–A_2 edge in G.*

Proof. Arguing by contradiction, suppose that vw is an A_2–A_2 edge in G. Let x be the neighbor of v distinct from w, and let y be the neighbor of w distinct from v.

Case 1: $x = y$. By Claim 26, we know that $x \in A_2$, but then G is isomorphic to K_3, which we know is not possible (as discussed in the proof of Claim 27).

Case 2: $x \neq y$. Applying Claim 27 on the two paths xvw and vwy, we deduce that $x, y \in A_3$. Let z be a neighbor of y outside $\{w, x\}$.

First, suppose that $xy \in E(G)$. Using Claim 21 with $F = G[\{v, w, y\}]$, we obtain

$$3 < f(v; G) + f(w; G) + f(x; G) + f(y; G) + f(z; G) = \frac{11}{5} + f(z; G) \leqslant \frac{11}{5} + \frac{3}{5} < 3,$$

a contradiction. Thus $xy \notin E(G)$.

Now, let $F := G[\{v, w\}]$. Then $N[V(F)] = \{v, w, x, y\}$, and $\left|N^2(V(F))\right| \geqslant 2$ since x and y have degree 3 and are not adjacent. But then, using Claim 21 on F, we obtain

$$2 < \frac{11}{5} - \frac{1}{10}\left|N^2(V(F))\right| \leqslant 2,$$

a contradiction. This concludes the proof of the claim. ■

Claim 29. *If $\{x, y, z\}$ induces a triangle in G, then $x, y, z \in A_3$, and $\left|N_G(\{x, y, z\}) \cap A_2\right| \leqslant 1$.*

Proof. Suppose that $\{x, y, z\}$ induces a triangle in G. We may assume that $f(x; G) \geqslant f(y; G) \geqslant f(z; G)$.

First, we prove that $x, y, z \in A_3$. Arguing by contradiction, suppose this is not the case. Then, $x \in A_2$, since $V(G) = A_2 \cup A_3$ and $f(x; G) \geqslant f(y; G) \geqslant f(z; G)$. By Claim 28, we know that $y, z \notin A_2$, and thus $y, z \in A_3$. Let $G' := G - x$ and let (A', B') be the restriction of (A, B) to $V(G')$ with y, z moved to B'. Observe that

$$\begin{aligned}
f(G', A', B') &= f(G, A, B) - f(x; G) - f(y; G) - f(z; G) + f(y; G') + f(z; G') \\
&= f(G, A, B) - \frac{3}{5} - 2\left(f_A(3) - f_B(2)\right) \\
&= f(G, A, B) - \frac{3}{5} - 2\frac{1}{6} \\
&= f(G, A, B) - \frac{14}{15}.
\end{aligned}$$

By the minimality of our counterexample, there exists an induced forest of stars F' in G' respecting (A', B') and containing at least $f(G', A', B')$ vertices. Let $F := G[V(F') \cup \{x\}]$. Observe that F is an induced forest of stars in G. Indeed, since $y, z \in B'$ and $yz \in E(G')$, they cannot both be in $V(F')$. Furthermore, if one of them is in $V(F')$, say $y \in V(F')$, then $N_F(y) = N_{F'}(y) \cup \{x\} \subseteq A' \cup \{x\}$, and it follows that all vertices in $N_F(y)$ have degree 1 in F, as desired. Note also that F (trivially) respects (A, B), since all vertices of F are in A. The number of vertices in F is then

$$\left|V(F)\right| = \left|V(F')\right| + 1 \geqslant f(G', A', B') + 1 = f(G, A, B) + \frac{1}{15} \geqslant f(G, A, B),$$

contradicting the fact that G and (A, B) form a counterexample. Therefore, $x, y, z \in A_3$, as claimed.

Now, let us show that $\left|N_G(\{x, y, z\}) \cap A_2\right| \leqslant 1$. Let u (resp. v, w) be the neighbor of x (resp. y, z) not in the triangle xyz. Arguing by contradiction, suppose that $\left|N_G(\{x, y, z\}) \cap A_2\right| \geqslant 2$. Without loss of generality, we may assume that $u \neq v$ and $u, v \in A_2$, thus $f(u; G) = f(v; G) = \frac{3}{5}$. Let t be the neighbor of v distinct from y. By Claim 28, we know that $t \notin A_2$, and thus $t \in A_3$. In particular, $t \neq u$. By the first part of this proof, we also know that $t \neq x$ and $t \neq z$, otherwise there would be a triangle containing a degree-2 vertex.

Case 1: $t = w$**.** Observe that $N_G[\{v, y, z\}] = \{x, y, z, v, w\}$, and $w \neq u$ (since $w = t \in A_3$ and $u \in A_2$). Applying Claim 21 on $F := G[\{v, y, z\}]$, we obtain

$$3 < \underbrace{f(x; G) + f(y; G) + f(z; G) + f(w; G)}_{=4 \cdot f_A(3) = 4\frac{1}{2}} + \underbrace{f(v; G)}_{=f_A(2)} = 2 + \frac{3}{5} < 3,$$

which is a contradiction.

Case 2: $t \neq w$**.** Let $S := \{x, y, v\}$, let $W := N(S) = \{z, u, t\}$ and let $Z := N^2(S)$. Applying Claim 21 on $F := G[S]$, we obtain

$$3 < f(x; G) + f(y; G) + f(v; G) + f(z; G) + f(u; G) + f(t; G) - \frac{1}{10}|Z| = 4\frac{1}{2} + 2\frac{3}{5} - \frac{1}{10}|Z|.$$

In particular, this implies that $|Z| \leqslant 1$. But since $w \in Z$, we deduce that $Z = \{w\}$. This implies however that $N(t) - \{v\} \subseteq \{x, y, z, w\}$. But we know that $t \neq u$, and hence $x \notin N(t)$, and also $y \notin N(t)$ by the first part of this proof. It follows that $N(t) = \{v, z, w\}$, and in particular $z \in N(t)$, therefore $t = w$, which is a contradiction.

Both cases lead to a contradiction, which concludes the proof of the claim. ■

Claim 30. *If* $v \in A_2$*, then* $\left|N^2(v)\right| = 4$.

Proof. Suppose that $v \in A_2$ and let u, w be the two neighbors of v. By Claim 29, we know that $uw \notin E(G)$, and by Claim 28, we know that $u, w \in A_3$. Let $W := N^2(v) = N(\{u, w\}) - \{v\}$. We know that $W \subseteq A_3$ by Claim 26. Let $F := G[\{v, u, w\}]$ (thus $W = N(V(F))$). Applying Claim 21 on F, we obtain

$$3 < f(v; G) + f(u; G) + f(w; G) + \sum_{x \in W} f(x; G) = \frac{3}{5} + 2\frac{1}{2} + |W|\,\frac{1}{2}.$$

Thus $|W| > 2\left(3 - \frac{3}{5}\right) - 2 = \frac{14}{5}$, and hence $|W| \geqslant 3$. Furthermore, we know that $|W| \leqslant 4$, and hence $|W| \in \{3, 4\}$.

To prove the lemma, we must show that $|W| = 4$. Arguing by contradiction, suppose not, thus $|W| = 3$. Denote the three vertices in W as x, y, z in such a way that $N(u) = \{v, x, y\}$ and $N(w) = \{v, x, z\}$. Let $Z := N^2(V(F))$. Applying again Claim 21 on F, we obtain

$$3 < \frac{3}{5} + \frac{5}{2} - \sum_{t \in Z} \gamma(t) \leqslant \frac{31}{10} - |Z|\,\frac{1}{10}.$$

Thus $|Z| < 1$, that is, $Z = \varnothing$, and hence $N[W] = W \cup \{u, w\}$. Since $w \notin N(y)$, we deduce that $N(y) = \{u, x, z\}$. Similarly, since $u \notin N(z)$, we deduce that $N(z) = \{w, x, y\}$. But then $N(x) = \{u, w, y, z\}$, which is a contradiction since x has degree 3 (and u, w, y, z are all distinct). We conclude that $|W| = 4$, as desired. ■

Claim 31. *If v and w are two distinct vertices in A_2, then* $\mathrm{dist}_G(v, w) \geqslant 4$.

Proof. Arguing by contradiction, suppose that there are two distinct vertices v, w in A_2 such that $\mathrm{dist}_G(v, w) \leqslant 3$. Claim 28 ensures that $\mathrm{dist}_G(v, w) > 1$, thus $\mathrm{dist}_G(v, w) \in \{2, 3\}$.

If $\mathrm{dist}_G(v, w) = 2$, then there is some vertex $u \in N(v) \cap N(w)$. By Claim 28, we know that $u \notin A_2$, thus $u \in A_3$ since $V(G) = A_2 \cup A_3$. But then u contradicts Claim 26. It follows that $\mathrm{dist}_G(v, w) = 3$.

Consider a length-3 path $vxyw$ from v to w in G. Let s be the neighbor of v distinct from x and let t be that of w distinct from y. Observe that $s \neq t$, since otherwise vsw would be a length-2 path from v to w, contradicting $\mathrm{dist}_G(v, w) = 3$. For the same reason, $s \neq y$ and $t \neq x$. Thus x, y, s, t are all distinct. Note also that $x, y, s, t \in A_3$ by Claim 28. By Claim 29, we know that $sx \notin E(G)$ and $yt \notin E(G)$. By Claim 30, we also know that $xt \notin E(G)$ and $sy \notin E(G)$.

Case 1: $st \in E(G)$**.** Let x' (resp. y', s', t') be the neighbor of x (resp. y, s, t) not in the cycle $vxywts$. We know that $x', y', s', t' \in A_3$ since no vertex in A_3 can have more than one neighbor in A_2 by Claim 26. Let us show that these four vertices are all distinct. By Claim 29, we know that $x' \neq y'$ and $s' \neq t'$. By Claim 30, we also know that $x' \neq s'$ and $y' \neq t'$. If $x' = t'$, then let $G' := G - \{x, t\}$ and let (A', B') be the restriction of (A, B) to $V(G')$. Note that $f(x'; G) = \frac{1}{2}$ and $f(x'; G') = \frac{5}{6}$. By the minimality of our counterexample, we know that G' admits an induced forest of stars F' respecting (A', B') whose number of vertices is at least

$$\begin{aligned}
f(G', A', B') &= f(G, A, B) - f(x; G) - f(t; G) + \gamma(v) + \gamma(w) + f(x'; G') - f(x'; G) + \gamma(y) + \gamma(s) \\
&= f(G, A, B) - 2\frac{1}{2} + 2\frac{7}{30} + \frac{5}{6} - \frac{1}{2} + 2\frac{1}{10} \\
&= f(G, A, B).
\end{aligned}$$

But F' is also a induced forest of stars in G respecting (A, B), contradicting the fact that G and (A, B) form a counterexample. We deduce that $x' \neq t'$, and by a similar argument, one can show that $y' \neq s'$. It follows that x', y', s', t' are all distinct, as claimed. But then applying Claim 21 on $F := G[\{v, w, x, t\}]$, we get

$$4 < 6 f_A(3) + 2 f_A(2) - \gamma(y') - \gamma(s') = 3 + \frac{6}{5} - 2\frac{1}{10} = 4,$$

which is a contradiction.

Case 2: $st \notin E(G)$**.** Let x' be the neighbor of x distinct from v and y. Let $G' := G - \{v, w, s, x, y, x'\}$ and let (A', B') be the restriction of (A, B) to $V(G')$ where t is moved to B'. By minimality of our counterexample, we know that there exists an induced forest of stars F' in G' respecting (A', B') and whose number of vertices is at least $f(G', A', B')$. Define $F := G[V(F') \cup \{v, x, w\}]$. Observe that F is an induced forest of stars in G respecting (A, B) (thanks in part to the fact that $t \in B'$). Let $Z := N(\{v, w, s, x, y, x'\}) - \{t\}$. Note that $\big|V(F)\big| < f(G, A, B)$ since G and (A, B) form a counterexample. Thus

$$\begin{aligned}
f(G, A, B) &> \big|V(F)\big| = \big|V(F')\big| + 3 \\
&\geqslant f(G', A', B') + 3 \\
&= f(G, A, B) + 3 - 2\frac{3}{5} - 5\frac{1}{2} + f(t; G') + \sum_{z \in Z} \big(f(z; G') - f(z; G)\big)
\end{aligned}$$

and hence

$$\frac{7}{10} > f(t;G') + \sum_{z\in Z}\big(f(z;G') - f(z;G)\big) \geqslant f(t;G') + |Z|\,\frac{1}{10}. \tag{4}$$

Observe that $d_{G'}(t) \leqslant d_G(t) - 1 = 2$ since $tw \in E(G)$.

Let $Z_1 := Z \cap N(x')$ and let $Z_2 := Z \cap N(s)$. We claim that $|Z_2| = 2$. To show this observe that:

- $sw \notin E(G)$, since $s \neq t$;
- $sx \notin E(G)$ by Claim 29;
- $sy \notin E(G)$ by Claim 30 on v;
- $sx' \notin E(G)$ by Claim 30 on v as well;
- $st \notin E(G)$ by the hypothesis of Case 2.

Altogether, this implies that $|Z_2| = 2$. Now observe that by (4):

$$f(t;G') < \frac{1}{10}\big(7 - |Z|\big) \leqslant \frac{1}{10}\big(7 - |Z_2|\big) = \frac{1}{2}.$$

Thus, we deduce that $f(t;G') = \frac{1}{3}$ and $d_{G'}(t) = 2$. In particular, $N(t) \cap \{v, s, x, y, x'\} = \varnothing$.

Now we show that $|Z_1| = 2$. Observe that:

- $x'v \notin E(G)$ by Claim 29;
- $x'w \notin E(G)$, since otherwise $x' = t$, which is not possible since $xt \notin E(G)$;
- $x's \notin E(G)$ by Claim 30 on v;
- $x'y \notin E(G)$ by Claim 29;
- $x't \notin E(G)$ by the remark hereabove.

This implies that $|Z_1| = 2$.

Furthermore, by (4), we know that

$$\frac{1}{10}|Z| \leqslant \sum_{z\in Z}\gamma(z) \leqslant \sum_{z\in Z}\big(f(z;G') - f(z;G)\big) < \frac{7}{10} - f(t;G') = \frac{7}{10} - \frac{1}{3} = \frac{11}{30}.$$

In particular, $|Z| < \frac{110}{30}$, and thus $|Z| \leqslant 3$. Since Z_1 and Z_2 are 2-element subsets of Z, we deduce that there exists $z' \in Z_1 \cap Z_2$. Thus, $z' \in Z$ and $x', s \in N(z')$, and hence $f(z';G') - f(z';G) \geqslant \frac{1}{3}$. Using the fact that $|Z| \geqslant 2$, it follows that

$$\sum_{z\in Z}\big(f(z;G') - f(z;G)\big) \geqslant \frac{1}{3} + \frac{1}{10} > \frac{11}{30},$$

contradicting the inequality above. This concludes the proof of the claim. ■

We are now ready for finishing the proof of the lemma. Clearly, $A_3 \neq \varnothing$ (possibly $A_2 = \varnothing$). For every vertex $v \in A_2$, let P_v be the path uvw where u, w are the two neighbors of v in G. Observe that $u, w \in A_3$, and $uw \notin E(G)$ by Claim 29. Furthermore, P_v and $P_{v'}$ are vertex disjoint for every two distinct vertices $v, v' \in A_2$ by Claim 31. Let G' be obtained from $G - A_2$ by adding an edge between the two endpoints of P_v for every $v \in A_2$. (Equivalently, G' can be seen as being obtained from G by contracting one edge of each path P_v.) Observe that G' is a cubic graph.

Applying Lemma 13 on G' gives an induced forest F' in G' with at least $\frac{1}{2}\big|V(G')\big| = \frac{1}{2}|A_3|$ vertices, and whose components are all isomorphic to either K_1 or K_2. Let $F := G[V(F') \cup A_2]$. We claim that F is an induced forest of stars in G. Clearly, F is a forest, as follows from the definition of G'. Moreover, no connected component of F can be incident to more than one vertex from A_2 (otherwise, two vertices of A_2 would be at distance less than 4 in G, contradicting Claim 31). By the definition of G', it follows then that every connected component of F containing a vertex v of A_2 is a path on at most 3 vertices (note that this could possibly be the path P_v in case F' contains the edge linking both endpoints of P_v in G'). Furthermore, since $V(G) = A_2 \cup A_3$, F trivially respects (A, B). Finally, observe that

$$\big|V(F)\big| = \big|V(F')\big| + |A_2| \geqslant \frac{1}{2}|A_3| + |A_2| \geqslant \frac{1}{2}|A_3| + \frac{3}{5}|A_2| = f(G, A, B),$$

which contradicts the fact that G and (A, B) form a counterexample. This concludes the proof of the lemma. □

Acknowledgments

We thank Louis Esperet for bringing Conjecture 3 to our attention and for his comments on an earlier version of the paper.

References

[1] S. Akbari, A. Amanihamedani, S. Mousavi, H. Nikpey, and S. Sheybani. On the maximum order of induced paths and induced forests in regular graphs. *arXiv preprint arXiv:1911.02332*, 2019.

[2] N. Alon, J. Kahn, and P. D. Seymour. Large induced degenerate subgraphs. *Graphs and Combinatorics*, 3:203–211, 1987.

[3] Y. Caro. New results on the independence number. Technical report, Tel-Aviv University, 1979.

[4] L. Lovász. On decomposition of graphs. *Studia Scientiatium Mathematicarum Hugarica*, 1:237–238, 1966.

[5] J. Nešetřil and P. Ossona de Mendez. *Sparsity*, volume 28 of *Algorithms and Combinatorics*. Springer, Heidelberg, 2012, pages xxiv+457. Graphs, structures, and algorithms.

[6] N. Punnim. Forests in random graphs. *Southeast Asian Bulletin of Mathematics*, 27(2), 2003.

[7] V. Wei. A lower bound on the stability number of a simple graph. Technical report, Bell Laboratories Technical Memorandum New Jersey, 1981.