On the Local Profiles of Trees

Sébastien Bubeck[1] and Nati Linial[2]

[1]DEPARTMENT OF OPERATIONS RESEARCH AND FINANCIAL ENGINEERING
PRINCETON UNIVERSITY
PRINCETON 08540, NJ
E-mail: sbubeck@princeton.edu

[2]SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
THE HEBREW UNIVERSITY OF JERUSALEM
JERUSALEM 91904, ISRAEL
E-mail: nati@cs.huji.ac.il

Received October 20, 2013; Revised November 20, 2014

Published online in Wiley Online Library (wileyonlinelibrary.com).
DOI 10.1002/jgt.21865

Abstract: We study the local profiles of trees. We show that in contrast with the situation for general graphs, the limit set of k-profiles of trees is convex. We initiate a study of the defining inequalities of this convex set. Many challenging problems remain open. © 2015 Wiley Periodicals, Inc. J. Graph Theory 00: 1–11, 2015

1. INTRODUCTION

For (unlabeled) trees T, S, we denote by $c(S, T)$ the number of copies of S in T, or in other words the number of injective homomorphism from S to T. Let $T_1^k, \ldots, T_{N_k}^k$ be a list of all (isomorphism types of) k-vertex trees,[1] where T_1^k, T_2^k are the k-vertex path and the k-vertex star, respectively. *The k-profile of a tree T* is the vector $p^{(k)}(T) \in \mathbb{R}^{N_k}$ whose

Contract grant sponsor: ISF; Contract grant sponsor: I-Core.

[1]Recall that the sequence $(N_k)_{k\geq 1}$ starts with $1, 1, 1, 2, 3, 6 \ldots$

ith coordinate is

$$\left(p^{(k)}(T)\right)_i = \frac{c(T_i^k, T)}{Z_k(T)}, \text{ where } Z_k(T) = \sum_{j=1}^{N_k} c(T_j^k, T).$$

In other words, the k-profile is the induced density vector of k-vertex trees. We are interested in understanding the limit set of k-profiles:

$$\Delta_(k) = \left\{ p \in \mathbb{R}^{N_k} : \exists (T_n), |T_n| \xrightarrow[n\to\infty]{\infty}, \text{ and } p^{(k)}(T_n) \xrightarrow[n\to\infty]{} p \right\},$$

where $|T|$ denotes the number of vertices in T.

Our main result, proved in Section 2, is as follows.

Theorem 1. *The set $\Delta(k)$ is convex.*

This property of profiles of trees is in sharp contrast with what happens for general graphs. Let $\Delta(k)$ be the k-profiles limit set of general graphs (which is defined as $\Delta(k)$ with a list of all k-vertex graphs rather than k-vertex trees). The first and second coordinates in $p \in \Delta(k)$ correspond to k-anticliques and k-cliques, respectively. Clearly $e_1 = (1, 0, \dots, 0), e_2 = (0, 1, 0, \dots, 0) \in \Delta(k)$ but $\frac{1}{2}e_1 + \frac{1}{2}e_2 \notin \Delta(k)$. Not only is $\Delta(k)$ nonconvex, it is even computationally infeasible to derive a description of its convex hull (see [1]). Our understanding of the sets $\Delta(k)$ is rather fragmentary (e.g., [2]). Flag algebras [3] are a major tool in such investigations. The convexity of $\Delta_(k)$ suggests that we may have a better chance understanding profiles of trees by deriving the linear inequalities that define these sets. We take some steps in this direction. Concretely, we prove the following result in Section 3.

Theorem 2. *Let $p \in \Delta(k)$, then*

$$p_1 + p_2 \geq \frac{1}{2N_k k^{2k}}.$$

We suspect that a stronger lower bound holds here. In Section 3, we give examples that show that $p_1 + p_2$ can be exponentially small in k.

For 5-profiles, we get a better inequality. In Section 4, we prove the following theorem.

Theorem 3. *Let $p \in \Delta(5)$, then*

$$p_2 \geq \frac{1 - 2p_1}{37}.$$

The above inequality holds with equality at the point $(1/2, 0) \in \Delta(5)$, but we believe that it is not tight for $p \in \Delta(5)$ such that $p_2 > 0$. We discuss tightness in more detail in Section 4. We end the article with a list of open problems in Section 5.

2. CONVEXITY OF THE *k*-PROFILES LIMIT SET

In this section, we prove Theorem 2. We first explain how to "glue" two trees, and then show how gluing allows us to generate convex combinations of tree profiles.

Step 1: The gluing operation. If T and S are trees, we define $T \boxtimes_k S$ as follows. This is a tree that consists of a copy of T, a copy of S, and a $(k-1)$-vertex path that connects some arbitrary leaf x in T to an arbitrary leaf y in S. In other words, we add to S and T a path $x = z_0, \ldots, z_k = y$, where $z_1, \ldots, z_{k-1}$ are new vertices. The resulting tree depends of course on the choice of the two leaves x and y, but we ignore this issue, since this will not affect anything that is said below.

We denote by $D(K)$ the largest vertex degree in a given tree K. The following inequalities are easy to verify:

$$c\left(T_i^k, T\right) + c\left(T_i^k, S\right) \le c\left(T_i^k, T \boxtimes_k S\right) \le c\left(T_i^k, T\right) + c\left(T_i^k, S\right) + kD(T)^{k-2} + kD(S)^{k-2}, \tag{1}$$

and consequently

$$Z_k(T) + Z_k(S) \le Z_k(T \boxtimes_k S) \le Z_k(T) + Z_k(S) + kN_kD(T)^{k-2} + kN_kD(S)^{k-2}. \tag{2}$$

We define by induction $T^{\boxtimes_k \ell} = T^{\boxtimes_k (\ell-1)} \boxtimes_k T$ (with $T^{\boxtimes_k 1} = T$). Observe that $D(T^{\boxtimes_k \ell}) = D(T)$ and thus using (1) and (2) one has

$$\ell c\left(T_i^k, T\right) \le c(T_i^k, T^{\boxtimes_k \ell}) \le \ell c\left(T_i^k, T\right) + 2k(\ell-1)D(T)^{k-2}, \tag{3}$$

$$\ell Z_k(T) \le Z_k(T^{\boxtimes_k \ell}) \le \ell Z_k(T) + 2kN_k(\ell-1)D(T)^{k-2}. \tag{4}$$

Step 2: Convex combinations by gluing. Let $p, q \in \Delta(k)$. Namely, there exists two sequences of trees T_n and S_n such that

$$|T_n|, |S_n| \underset{n\to\infty}{\longrightarrow} \infty, \text{ and } \left(p^{(k)}(T_n), p^{(k)}(S_n)\right) \underset{n\to\infty}{\longrightarrow} (p, q).$$

Now, given $\lambda \in (0, 1)$, we want to construct a sequence of trees R_n such that

$$|R_n| \underset{n\to\infty}{\longrightarrow} \infty, \text{ and } p^{(k)}(R_n) \underset{n\to\infty}{\longrightarrow} \lambda p + (1-\lambda)q.$$

First, let α_n/β_n be a sequence of rational numbers that converges to λ. We correspondingly define the sequence of trees R_n via:

$$R_n = T_n^{\boxtimes_k [\alpha_n Z_k(S_n)]} \boxtimes_k S_n^{\boxtimes_k [(\beta_n - \alpha_n) Z_k(T_n)]}.$$

Using (2) and (4) one immediately obtains

$$\begin{aligned}&\beta_n Z_k(T_n) Z_k(S_n)\\ &\le Z_k(R_n)\\ &\le \beta_n Z_k(T_n)Z_k(S_n) + 2kN_k\alpha_n Z_k(S_n)D(T_n)^{k-2} + 2kN_k(\beta_n - \alpha_n)Z_k(T_n)D(S_n)^{k-2}.\end{aligned} \tag{5}$$

Now the key observation is that

$$D(T_n)^{k-2} = o(Z_k(T_n)). \tag{6}$$

Indeed, $Z_k(T_n) \ge \binom{D(T_n)}{k-1}$ follows by counting k-vertex stars rooted at the highest degree vertex in T_n, which yields Equation (6) if $D(T_n) \to \infty$. On the other hand, if $D(T_n)$ is bounded then (6) is also clearly true since $Z_k(T_n) \to \infty$.

Using (6), one can rewrite (5) as

$$Z_k(R_n) = \beta_n Z_k(T_n) Z_k(S_n) + o(\beta_n Z_k(T_n) Z_k(S_n)).$$

Similarly, using (1) and (3) we obtain

$$c\left(T_i^k, R_n\right) = \alpha_n Z_k(S_n) c\left(T_i^k, T_n\right) + (\beta_n - \alpha_n) Z_k(T_n) c\left(T_i^k, S_n\right) + o(\beta_n Z_k(T_n) Z_k(S_n)).$$

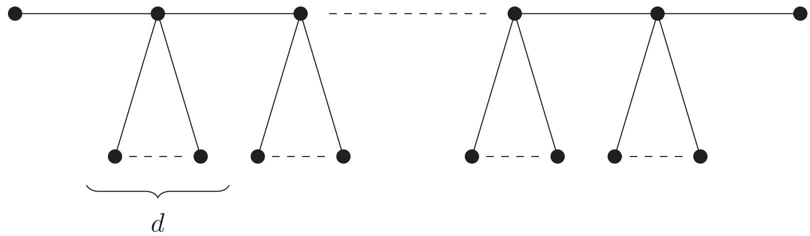

FIGURE 1. A d-millipede.

We combine these two identities and conclude that

$$\left(p^{(k)}(R_n)\right)_i = \frac{c\left(T_i^k, R_n\right)}{Z_k(R_n)} = (1+o(1))\left[\frac{\alpha_n}{\beta_n}\frac{c\left(T_i^k, T_n\right)}{Z_k(T_n)} + \left(1-\frac{\alpha_n}{\beta_n}\right)\frac{c\left(T_i^k, S_n\right)}{Z_k(S_n)}\right]$$
$$+\, o(1) \to \lambda p_i + (1-\lambda)q_i,$$

as claimed.

3. ON STARS AND PATHS

In this section, we prove Theorem 2. We use the shorthand $P_k(T) = c(T_1^k, T)$ and $S_k(T) = c(T_2^k, T)$, and also omit the reference to T whenever it is clear from context. Before delving into the proof, let us show why the exponential decrease in k is unavoidable. A *d-millipede* is a tree where all nonleaf vertices reside on a single path and they have degree $d+2$ each. See Figure 1 for an illustration. The number of nonleaves is called the d-millipede's *length*.

We denote by T_n the $(k-4)$-millipede of length n with k even. Also, R_k is the $\frac{k-4}{2}$-millipede of length 2. It is easy to see that for $k \geq 6$,

$$Z_k(T_n) \geq c(R_k, T_n) \geq 2(n-2)\binom{k-3}{(k-2)/2} \geq (n-2)(3/2)^{k/2},$$

and

$$S_k(T_n) = 0, \qquad P_k(T_n) \leq n(k-3)^2.$$

Thus, the limiting profile that corresponds to the sequence (T_n) satisfies

$$p_1 + p_2 \leq \frac{(k-3)^2}{(3/2)^{k/2}}. \tag{7}$$

For $k > 3$, let (k) be the projection of $\Delta(k)$ on the first two coordinates, that is,

$$(k) = \{((p_1, p_2), p \in \Delta(k)\}.$$

As a side note, we also observe that the above inequality yields

$$\overline{\cup_k (k)} = \{x \in_+^2 : x_1 + x_2 \leq 1\},$$

where $\overline{A}$ denotes the closure of a set A. Indeed $(1, 0)$ and $(0, 1)$ are always in (k), (7) shows that for k large enough one can find a point arbitrarily close to $(0, 0)$, and thus using the convexity of (k) (Theorem 1) one obtains the above set equality.

We now turn to the proof of Theorem 2. We repeatedly use the following obvious result that we state without a proof.

Lemma 1. *A tree with maximal degree D has at most kN_kD^{k-1} k-vertex subtrees that contain a given vertex.*

Lemma 2 is an enumerative analog of the probabilistic statement of Theorem 2, which applies when $S_k = 0$. In Lemma 3 we deal with the case of $S_k \geq 0$, which then yields Theorem 2.

Lemma 2. *If $D(T) \leq k-2$ for some tree T, then*

$$Z_k \leq kN_k(k-2)^{k-1}P_k + kN_k(k-2)^{2k-2}.$$

Proof. For trees with $n \leq (k-2)^{k-1}$ vertices, this inequality clearly follows from Lemma 1. For $n > (k-2)^{k-1}$, we proceed by induction. Clearly for this range of n, the tree's diameter must be at least $2(k-2)$. In other words, it must contain a copy P of $P_{2(k-2)+1}$. Let the tree T' be obtained by removing a leaf x from T. This eliminates at least one k-vertex path, namely the path from x toward P possibly proceeding toward P's furthest end. In other words:

$$P_k(T) \geq P_k(T') + 1.$$

Furthermore by Lemma 1,

$$Z_k(T) \leq Z_k(T') + kN_k(k-2)^{k-1}.$$

Applying the induction hypothesis to T' yields

$$Z_k(T') \leq kN_k(k-2)^{k-1}P_k(T') + kN_k(k-2)^{2k-2},$$

together with the two above inequalities this gives the same inequality for T. ■

Lemma 3. *Every tree satisfies*

$$Z_k \leq N_k k^{2k}(P_k + 2S_k + 1).$$

Proof. First, observe that if $n \leq k^k$ then by (a variant of) Lemma 1:

$$\begin{aligned} Z_k &\leq \sum_{u:d(u)\leq k-2} kN_k d(u)^{k-1} + \sum_{u:d(u)\geq k-1} kN_k d(u)^{k-1} \\ &\leq N_k k^{2k} + \sum_{u:d(u)\geq k-1} kN_k(k-1)^{k-1}\binom{d(u)}{k-1} \\ &\leq N_k k^{2k} + N_k k^k S_k, \end{aligned}$$

as needed. For larger trees, we prove the following stronger inequality by induction on the number of vertices:

$$Z_k \leq N_k k^{2k}\,(P_k+1)\,1\{P_k \geq 1\} + 2N_k k^{2k} S_k.$$

Clearly, the expression $1\{P_k \geq 1\}$ captures the information whether or not T's diameter is at least $k-1$. The base case $n = k^k$ follows since necessarily $P_k \geq 1$ or $S_k \geq 1$. The induction step has two cases:

Case 1: If $D(T) \leq k-2$, then Lemma 2 yields the inequality, since $P_k \geq 1$.

Case 2: Let v be the vertex of largest degree $d \geq k-1$, and let $T_1, \ldots, T_d$ be the trees of the forest $T \setminus \{v\}$. By Lemma 1

$$Z_k(T) \leq \sum_{i=1}^{d} Z_k(T_i) + kN_k d^{k-1}.$$

Furthermore,

$$S_k(T) \geq \sum_{i=1}^{d} S_k(T_i) + \binom{d}{k-1} \geq \sum_{i=1}^{d} S_k(T_i) + \left(\frac{d}{k-1}\right)^{k-1}$$

and

$$(1 + P_k(T))1\{P_k(T) \geq 1\} \geq \sum_{i=1}^{d} (1 + P_k(T_i))1\{P_k(T_i) \geq 1\}.$$

To see why the last inequality holds true, observe first that it is trivial if $\sum_{i=1}^{d} 1\{P_k(T_i) \geq 1\} \in \{0, 1\}$. Furthermore if $\sum_{i=1}^{d} 1\{P_k(T_i) \geq 1\} \geq 2$, then for each i such that $P_k(T_i) \geq 1$, one can find a path in T containing both v and vertices from T_i, which means that in this case one even has $P_k(T) \geq \sum_{i=1}^{d} (1 + P_k(T_i))1\{P_k(T_i) \geq 1\}$.

Combine the three above displays and apply induction to the T_i's to conclude:

$$\begin{aligned} Z_k(T) &\leq \sum_{i=1}^{d} Z_k(T_i) + kN_k d^{k-1} \\ &\leq N_k k^{2k} \sum_{i=1}^{d} (P_k(T_i) + 1)\, 1\{P_k(T_i) \geq 1\} + 2N_k k^{2k} \sum_{i=1}^{d} S_k(T_i) + kN_k d^{k-1} \\ &\leq N_k k^{2k} (1 + P_k(T))1\{P_k(T) \geq 1\} + 2N_k k^{2k} S_k(T), \end{aligned}$$

which concludes the proof. ■

4. 5-PROFILES

Clearly, $\Delta(5)$ is entirely determined by (5). In this section, we prove Theorem 1 that improves Theorem 2 for $k = 5$.

Before we embark on the proof, we show that millipedes generate a "large" set of points in (5). To simplify notation, let $P(T) = c(T_1^5, T)$, $S(T) = c(T_2^5, T)$, and $Y(T) = c(T_3^5, T)$ (note that T_3^5 has the Y-shape). We also omit the dependency on T whenever it is clear from context. For a d-millipede of length n, we get the following expressions:

$$\begin{aligned} S &= n\binom{d+2}{4}, \\ P &= (n-2)(d+1)^2, \\ Y &= 2(n-2)\binom{d+1}{2}(d+1) + 2\binom{d+1}{2}(d+1) = (n-1)(d+1)^2 d, \\ S + Y + P &= n\binom{d+2}{4} + (n-2)(d+1)^3 + (d+1)^2 d. \end{aligned}$$

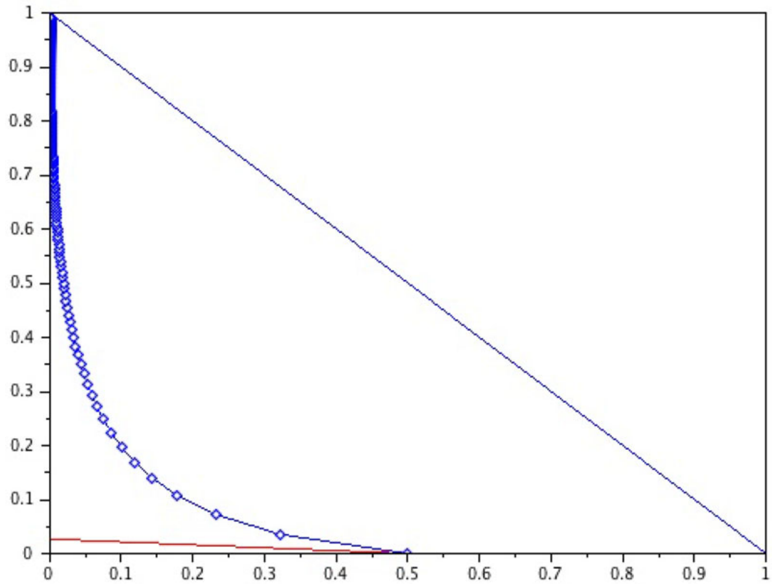

FIGURE 2. The equation of the red line is $y = \frac{1-2x}{37}$. In blue, the polygonal curve connecting consecutive m_d, $d \geq 1$ of Equation (8) as well as (0, 1) to (1, 0). By Theorem 3, the set (5) lies above the red line and by Theorem 1 it contains the convex domain bounded by the blue lines.

In particular for fixed d and $n \to \infty$, we get the following point in (5):

$$m_d = \left(\frac{(d+1)^2}{\binom{d+2}{4} + (d+1)^3}, \frac{\binom{d+2}{4}}{\binom{d+2}{4} + (d+1)^3} \right). \tag{8}$$

Thus, by convexity we have

$$(5) \supseteq \text{conv}(\{(0,1)\} \cup \{m_d, d \geq 0\}). \tag{9}$$

We cannot rule out the possibility that this is, in fact an equality. This inclusion and the inequality from Theorem 3 are illustrated in Figure 2.

Our proof of Theorem 3 proceeds along the route that we took in proving Theorem 2. Now, however, we are much more careful with the details. Lemma 4, a counterpart of Theorem 3, gives an inequality on the unnormalized quantities when $S = 0$. The general case $S \geq 0$ is handled in Lemma 5 that yields Theorem 3.

Lemma 4. *If $D(T) \leq 3$, then*

$$Y \leq P + 4,$$

with equality if and only if T is a 1-millipede.

Note that to prove Theorem 3, we will only need the inequality provided by Lemma 4.

Proof. It is immediate that a 1-millipede satisfies $Y = P + 4$. We prove the inequality in two steps. A third step shows that only 1-millipedes satisfy $Y = P + 4$.

Step 1: **A formula for $P - Y$.** We say that a vertex of degree 3 has *type xyz* with $x, y, z \in \{0, 1, 2\}$ if its three neighbors have degree $x + 1, y + 1$, and $z + 1$, respectively. The number of vertices of type xyz is denoted as n_{xyz}. Similarly, we define for degree-2 vertices the quantity n_{xy}.
A straightforward (but slightly painful) calculation yields

$$P = 12n_{222} + 8n_{221} + 4n_{220} + 5n_{211} + 2n_{210} + 3n_{111} + n_{110} + 4n_{22} + 2n_{21} + n_{11},$$

and

$$Y = 6n_{222} + 5n_{221} + 4n_{220} + 4n_{211} + 3n_{210} + 2n_{200} + 3n_{111} + 2n_{110} + n_{100}.$$

Hence,

$$P - Y = 6n_{222} + 3n_{221} + n_{211} - n_{210} - 2n_{200} - n_{110} - n_{100} + 4n_{22} + 2n_{21} + n_{11}. \tag{10}$$

Step 2: **Double counting**. Let n_x be the number of degree-x vertices. Clearly $n_1 + n_2 + n_3 = n$, and by double counting of edges, also $n_1 + 2n_2 + 3n_3 = 2(n - 1)$. In particular,

$$n_1 - n_3 = 2. \tag{11}$$

Next, observe that n_1 and n_3 can easily be expressed in terms of the parameters n_{xy} and n_{xyz}. Namely,

$$n_3 = n_{222} + n_{221} + n_{220} + n_{211} + n_{210} + n_{200} + n_{111} + n_{110} + n_{100},$$
$$n_1 = n_{220} + n_{210} + 2n_{200} + n_{110} + 2n_{100} + n_{20} + n_{10}.$$

Together with (11) we find

$$-n_{222} - n_{221} - n_{211} + n_{200} - n_{111} + n_{100} + n_{20} + n_{10} = 2. \tag{12}$$

Next, adding (10) to twice (12) one gets

$$P - Y + 4 = 4n_{222} + n_{221} - n_{211} - n_{210} - 2n_{111} - n_{110} + n_{100} + 4n_{22} + 2n_{21} + n_{11} + 2n_{20} + 2n_{10}.$$

It only remains to show that the right-hand side term is nonnegative. To this end, we count edges between a degree-2 vertex and degree-3 vertex in two ways: once from the degree-3 side and once from the degree-2 side

$$n_{221} + 2n_{211} + n_{210} + 3n_{111} + 2n_{110} + n_{100} = 2n_{22} + n_{21} + n_{20}.$$

This concludes the proof of the inequality stated in the theorem. Note that we have, in fact, shown a more precise statement:

$$P - Y + 4 = 4n_{222} + 2n_{221} + n_{211} + n_{111} + n_{110} + 2n_{100} + 2n_{22} + n_{21} + n_{11} + n_{20} + 2n_{10}. \tag{13}$$

Step 3: **The equality case**. Equation (13) shows that if $P - Y + 4 = 0$ then

$$4n_{222} + 2n_{221} + n_{211} + n_{111} + n_{110} + 2n_{100} + 2n_{22} + n_{21} + n_{11} + n_{20} + 2n_{10} = 0.$$

In particular the tree contains no degree-2 vertices, and no degree-3 vertices of type 222. In other words, it has only leaves and degree-3 vertices of types 220 and 200. Moreover, by (12) in this case $n_{200} = 2$. A straightforward inductive proof shows that the tree must be a 1-millipede. ■

We now adapt Lemma 4 to the case where $S > 0$. This more general inequality directly implies Theorem 3.

Lemma 5. *All trees satisfy*

$$Y \leq 36S + P + 4.$$

Proof. First observe the following expressions:

$$Y = \sum_{\{u,v\}\in E} \left(\binom{d(v)-1}{2}(d(u)-1) + \binom{d(u)-1}{2}(d(v)-1) \right).$$

We split $Y = Y_s + Y_\ell$, where

$$Y_s = \sum_{\{u,v\}\in E:\max(d(u),d(v))\leq 3} \left(\binom{d(v)-1}{2}(d(u)-1) + \binom{d(u)-1}{2}(d(v)-1) \right),$$

and

$$Y_\ell = \sum_{\{u,v\}\in E:\max(d(u),d(v))\geq 4} \left(\binom{d(v)-1}{2}(d(u)-1) + \binom{d(u)-1}{2}(d(v)-1) \right).$$

The proof deals separately with Y_s and Y_ℓ.

Step 1: We prove that $Y_\ell \leq 36S$ by observing

$$S = \sum_{u\in V} \binom{d(u)}{4} = \frac{1}{4} \sum_{u,v:\{u,v\}\in E} \binom{d(u)-1}{3}$$
$$= \frac{1}{4} \sum_{\{u,v\}\in E} \left(\binom{d(u)-1}{3} + \binom{d(v)-1}{3} \right),$$

and making a term-by-term comparison with the expression for Y_ℓ. We use the fact that for any nonnegative integers $x \neq 2$, $y \geq 3$

$$yx(x-1) + xy(y-1) \leq x^2(x-1) + y^2(y-1) \leq 3(x(x-1)(x-2) + y(y-1)(y-2)),$$

and furthermore for $x = 2$ this inequality (without the intermediate step) is also true.

Step 2: We prove by induction on the size of the tree that $Y_s \leq P + 4$. The base case is trivial. The induction step has three cases:

Case 1: $D(T) \leq 3$. The inequality follows readily from Lemma 4.

Case 2: There are two neighbors u, v in T, where $d(u) \geq 4$ and v is a leaf. Clearly,

$$Y_s(T) \leq Y_s(T') \text{ and } P(T') \leq P(T),$$

where $T' := T \setminus \{v$. By applying the induction hypothesis to T', we see that $Y_s(T') \leq P(T') + 4$, which implies $Y_s \leq P + 4$.

Case 3: There is a vertex u in T with $d(u) \geq 4$, and no neighbor of u is a leaf. Let v be a neighbor of u and let T_1, T_2 be the two trees of the forest obtained by removing the edge uv and adding a new edge to v, where u is in T_1 and v in T_2. As in Case 2,

$$Y_s(T) \leq Y_s(T_1) + Y_s(T_2).$$

Observe that we can assume that v was selected such that T_2 has at least three edges, for otherwise $Y_s(T) = 0$ and thus the inequality would trivially hold. Indeed if T_2 had two edges for all neighbors of u, then $T \setminus \{u$ would be a matching, and thus any copy of T_3^5 in T would have u in its "middle edge," which implies $Y_s(T) = 0$.

Now clearly if T_2 has at least three edges,

$$P(T) \geq P(T_1) + P(T_2) + 2(d(u) - 1) \geq P(T_1) + P(T_2) + 4.$$

Applying the induction hypothesis to T_1 and T_2 and using the above inequalities yield $Y_s \leq P + 4$ in this case as well. ■

5. OPEN PROBLEMS

(1) Is the blue curve in Figure 2 tight? That is, is (9) in fact an equality? Less ambitiously, can the bound in Lemma 5 be improved to $Y \leq 9S + P + K$, for some universal $K \geq 0$? If true, this shows that the first segment of the polygonal curve is tight.

(2) Recall that (k) is the projection of the limit set of k-profiles to the first two coordinates. Are these sets increasing, that is, is it true that

$$(k) \subset (k+1)$$

for all integer k?

(3) Let $p \in \Delta(k)$. Does $p_1 = 0$ imply $p_2 = 1$?

(4) Imitating a concept from graph theory we define the *inducibility* of a tree T to be $\limsup \frac{c(T,\mathcal{T})}{Z_{|T|}(\mathcal{T})}$, where the lim sup is over trees $\mathcal{T}$ of size tending to infinity. By gluing many copies of T as in Section 2, it is easy to show that every T has positive inducibility. By Theorem 2 paths and stars are the only trees with inducibility 1, but are there other trees with inducibility arbitrarily close to 1? If such trees do not exist, is it nonetheless possible to find infinitely many trees of inducibility $\geq \epsilon$ for some $\epsilon > 0$? Note that in the realm of graphs, there are infinitely many distinct graphs with inducibility $> \frac{1}{10}$, for example, the complete bipartite graphs $H = K_{3,r}$ with $r > 10$. It can be easily verified that randomly chosen set of $r + 3$ vertices in $K_{3n,rn}$ for n large spans a copy of H with probability >0.1.

(5) Call a sequence of trees (T_n) *k-universal* if

$$\liminf_{n\to\infty} \left(p^{(k)}(T_n)\right)_i > 0$$

for every $i \in [N_k]$. The convexity of $\Delta(k)$ and the fact that every tree has positive inducibility implies that k-universal sequences exist. But does there exist a sequence of trees which is k-universal simultaneously for every k? For general graphs the answer is positive, for example, using $G(n, p)$ graphs.

(6) Is there a probabilistic interpretation to the profile of a tree?

(7) In this article, we found only linear inequalities satisfied by the sets $\Delta\mathcal{T}(k)$. We wonder if higher order inequalities can be derived as well. Is there a framework similar to flag algebras that applies to trees?

ACKNOWLEDGMENTS

This research was carried out at the Simons Institute for the Theory of Computing. We are grateful to the Simons Institute for offering us such a wonderful research environment. We also thank an anonymous referee for fixing a mistake in the first version of this article.

REFERENCES

[1] H. Hatami and S. Norine, Undecidability of linear inequalities in graph homomorphism densities, J Am Math Soc 24(2) (2011), 547–565.
[2] H. Huang, N. Linial, H. Naves, Y. Peled, and B. Sudakov, On the 3-local profiles of graphs, arXiv preprint, arXiv:1211.3106 (2012).
[3] A. Razborov, Flag algebras, J Symbol Logic (2007), 1239–1282.