Graph homomorphisms between trees

Péter Csikvári

Department of Mathematics, Massachusetts Institute of Technology
Cambridge, MA 02139

Department of Computer Science, Eötvös Loránd University
H-1117 Budapest, Pázmány Péter sétány 1/C, Hungary

peter.csikvari@gmail.com

Zhicong Lin

Department of Mathematics and Statistics, Lanzhou University
Lanzhou 730000, P.R. China

Institut Camille Jordan, Université Claude Bernard Lyon 1
43 boulevard du 11 novembre 1918, F-69622 Villeurbanne, France

lin@math.univ-lyon1.fr

Submitted: Feb 8, 2014; Accepted: Sep 26, 2014; Published: Oct 9, 2014
Mathematics Subject Classifications: 05C05, 05C30, 05C35

Abstract

In this paper we study several problems concerning the number of homomorphisms of trees. We begin with an algorithm for the number of homomorphisms from a tree to any graph. By using this algorithm and some transformations on trees, we study various extremal problems about the number of homomorphisms of trees. These applications include a far reaching generalization and a dual of Bollobás and Tyomkyn's result concerning the number of walks in trees.

Some other main results of the paper are the following. Denote by $\hom(H,G)$ the number of homomorphisms from a graph H to a graph G. For any tree T_m on m vertices we give a general lower bound for $\hom(T_m,G)$ by certain entropies of Markov chains defined on the graph G. As a particular case, we show that for any graph G,

$$\exp(H_\lambda(G))\lambda^{m-1} \leqslant \hom(T_m,G),$$

where λ is the largest eigenvalue of the adjacency matrix of G and $H_\lambda(G)$ is a certain constant depending only on G which we call the spectral entropy of G. We also show that if T_m is any fixed tree and

$$\hom(T_m,P_n) > \hom(T_m,T_n),$$

for some tree T_n on n vertices, then T_n must be the tree obtained from a path P_{n-1} by attaching a pendant vertex to the second vertex of P_{n-1}.

All the results together enable us to show that among all trees with fixed number of vertices, the path graph has the fewest number of endomorphisms while the star graph has the most.

Keywords: trees; walks; graph homomorphisms; adjacency matrix; extremal problems; KC-transformation; Markov chains

1 Introduction

We use standard notations and terminology of graph theory, see for instance [2, 4]. The graphs considered here are finite and undirected without multiple edges and loops. Given a graph G, we write $V(G)$ for the vertex set and $E(G)$ for the edge set. A *homomorphism* from a graph H to a graph G is a mapping $f : V(H) \to V(G)$ such that the images of adjacent vertices are adjacent. Let $\mathrm{Hom}(H, G)$ denote the set of homomorphisms from H to G and by $\mathrm{hom}(H, G)$ the number of homomorphisms from H to G. Throughout this article, we write P_n and S_n for the path and the star on n vertices, respectively. The length of a path is the number of its edges. The *union of graphs* G and H is the graph $G \cup H$ with vertex set $V(G) \cup V(H)$ and edge set $E(G) \cup E(H)$. A tree T together with a root vertex v will be denoted by $T(v)$.

The problem of computing $\mathrm{hom}(H, G)$ is difficult in general. However, there has been recent interest in counting homomorphisms between special graphs. In particular, formulas for computing the number of homomorphisms between two different paths were given in [1, 16]. But even for these special trees, the formulas are bulky and inelegant. In Section 2, we shall give an algorithm for computing the number of homomorphisms from trees to any graph. This algorithm will be called *Tree-walk algorithm*.

Recently, the first author proved a conjecture of Nikiforov concerning the number of closed walks on trees. He proved in [6] that, for a fixed integer m, the number of closed walks of length m on trees of order n attains its maximum at the star S_n and its minimum at the path P_n. In other words,

$$\mathrm{hom}(C_m, P_n) \leqslant \mathrm{hom}(C_m, T_n) \leqslant \mathrm{hom}(C_m, S_n), \tag{1.1}$$

where T_n is a tree on n vertices and C_m is the cycle on m vertices.

Bollobás and Tyomkyn [3] gave a variant of the first author's result by replacing the number of closed walks by the number of all walks, that is

$$\mathrm{hom}(P_m, P_n) \leqslant \mathrm{hom}(P_m, T_n) \leqslant \mathrm{hom}(P_m, S_n), \tag{1.2}$$

where T_n is a tree on n vertices. In both [3] and [6], the authors use a certain transformation of trees. In [6], it is called the *generalized tree shift*, whereas in [3], it is renamed to *KC-transformation*.

To define this transformation, let x and y be two vertices of a tree T such that every interior vertex of the unique x–y path P in T has degree two, and write z for the neighbor of y on this path. Let $N(v)$ denote the set of neighbors of a vertex v. The

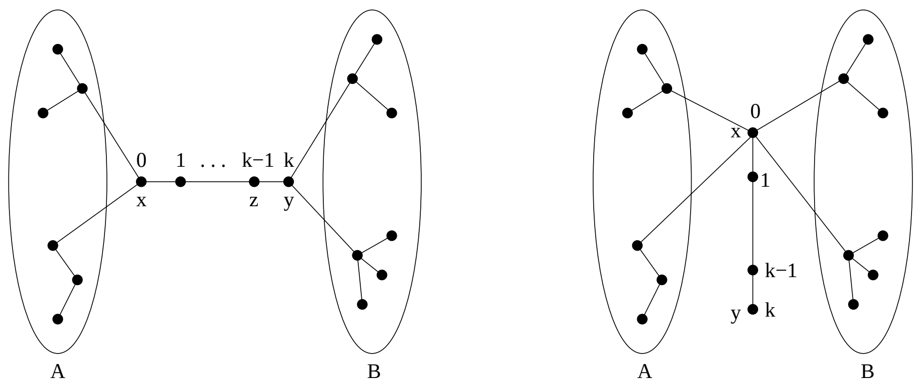

Figure 1: The KC-transformation.

KC-transformation, $KC(T, x, y)$, of the tree T with respect to the path P is obtained from T by deleting all edges between y and $N(y) \setminus z$ and adding the edges between x and $N(y) \setminus z$ instead (See Fig. 1). Note that $KC(T, x, y)$ and $KC(T, y, x)$ are isomorphic.

The following property of KC-transformation was proved in [6].

Proposition 1.1. *The KC-transformation gives rise to a graded poset of trees on* n *vertices (graded by the number of leaves) with the star as the largest and the path as the smallest element. See Figure 2.*

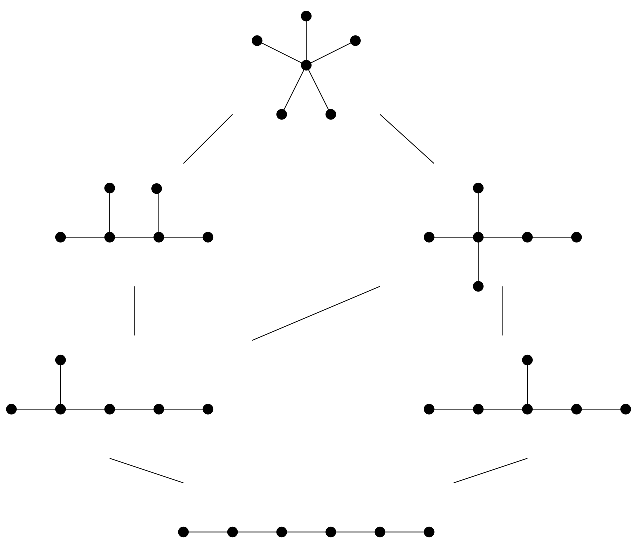

Figure 2: The induced poset of KC-transformation on trees of 6 vertices.

In [6] the first author proved that the KC-transformation increases the number of closed walks of fixed length in trees. By Proposition 1.1, this leads to the proof of inequality (1.1).

In the very same spirit, Bollobás and Tyomkyn [3] showed that the KC-transformation increases the number of walks of fixed length in trees. In the language of graph homomorphism, their result can be restated as follows.

Theorem 1.2 (Bollobás–Tyomkyn)**.** *Let T be a tree and let T' be obtained from T by a KC-transformation. Then*

$$\hom(P_m, T') \geqslant \hom(P_m, T) \tag{1.3}$$

for any $m \geqslant 1$.

Now a natural question arises: does inequality (1.3) still hold when P_m is replaced by an arbitrary fixed tree? A tree is called *starlike* if it has at most one vertex of degree greater than two. Note that paths are starlike. We answer this question in the affirmative for starlike trees.

Theorem 1.3. *Let T be a tree and T' the KC-transformation of T with respect to a path of length k. Then the inequality*

$$\hom(H, T') \geqslant \hom(H, T) \tag{1.4}$$

holds when k is even and H is any tree, or k is odd and H is a starlike tree.

Moreover, we find a counterexample for inequality (1.4) when k is odd and H is not a starlike tree (see the end of Section 3).

Another extremal problem concerning the number of homomorphisms between trees that worth considering is to find the extremal trees for $\hom(\cdot\,, P_n)$ over all trees on m vertices. We address this question in a follow-up paper [8], here we only mention the main result of this paper. This result can be considered as a dual of inequality (1.2).

Theorem 1.4. *Let T_m be a tree on m vertices. Furthermore, let $\operatorname{diam}(T_m)$ denote the diameter of T_m.*

(i) Let T'_m be obtained from T_m by a KC-transformation. If n is even, or n is odd and $\operatorname{diam}(T_m) \leqslant n-1$, then

$$\hom(T_m, P_n) \leqslant \hom(T'_m, P_n). \tag{1.5}$$

(ii) For any m, n,

$$\hom(P_m, P_n) \leqslant \hom(T_m, P_n) \leqslant \hom(S_m, P_n).$$

As we mentioned, the proof of Theorem 1.4 will be given in [8], it only builds on the algorithm of Section 2. Note that inequality (1.5) is not true in general when n is odd and $\operatorname{diam}(T_m)$ is greater than $n-1$.

For the sake of keeping this paper self-contained, we will also give a new proof for the following theorem of Sidorenko [19] concerning the extremal property of the stars among trees. Note that Fiol and Garriga [10] proved the special case of this theorem when $T_m = P_m$, clearly, they were not aware of the work of Sidorenko.

Theorem 1.5 (Sidorenko)**.** *Let G be an arbitrary graph and let T_m be a tree on m vertices. Then*

$$\hom(T_m, G) \leqslant \hom(S_m, G).$$

After all, it is a natural question whether it is true or not that

$$\hom(P_m, G) \leqslant \hom(T_m, G)$$

for any tree T_m on m vertices. Surprisingly, the answer is no! It was already known to A. Leontovich [14]. It turns out that even if one restricts G to be a tree there is a counterexample (see Remark 4.13).

We have already seen a few examples to the phenomenon that in many extremal problems concerning trees it turns out that the maximal (minimal) value of the examined parameter is attained at the star and the minimal (maximal) value is attained at the path among trees on n vertices (cf. [7,17]). In what follows we will show that this phenomenon occurs quite frequently if one studies homomorphisms of trees.

Let $Y_{a,b,c}$ be the starlike tree on $a+b+c+1$ vertices which has exactly 3 leaves and the vertex of degree 3 has distance a, b, c from the leaves, respectively.

Theorem 1.6. *Let T_n be a tree on n vertices. Assume that for a tree T_m we have*

$$\hom(T_m, T_n) < \hom(T_m, P_n).$$

Then $T_n = Y_{1,1,n-3}$ and n is even.

In fact, we conjecture that we only have to exclude the case $n = 4$ and $T_4 = S_4$.

Conjecture 1.7. Let T_n be a tree on n vertices, where $n \geqslant 5$. Then for any tree T_m we have

$$\hom(T_m, P_n) \leqslant \hom(T_m, T_n).$$

An *endomorphism* of a graph is a homomorphism from the graph to itself. For a graph G, denote by $\mathrm{End}(G)$ the set of endomorphisms of G. We remark that $\mathrm{End}(G)$ forms a monoid with respect to the composition of mappings. One of the main results of this paper is the following extremal property about the number of endomorphisms of trees.

Theorem 1.8. *For all trees T_n on n vertices we have*

$$|\mathrm{End}(P_n)| \leqslant |\mathrm{End}(T_n)| \leqslant |\mathrm{End}(S_n)|.$$

Both the proofs of Theorem 1.6 and the first part of Theorem 1.8 require a crucial lower bound involving Markov chains for the number of graph homomorphisms from trees (see Theorem 4.1). Our lower bound generalizes a recent result due to Dellamonica et al. [9] (by choosing the classical Markov chain on graphs) and is also closely related to works of Kopparty and Rossman [13] and Rossman and Vee [21]. The idea of studying homomorphisms via entropies of Markov chains on graphs seems new.

The rest of this paper is organized as follows. In Section 2, we state the tree-walk algorithm. Section 3 is devoted to the proof of Theorem 1.3. In Section 4, we prove some lower bounds involving Markov chains and an upper bound (Theorem 1.5) for the number of homomorphisms from trees to an arbitrary graph. The proofs of Theorem 1.6 and Theorem 1.8 are given in Section 5, where some lower bounds concerning the homomorphisms of arbitrary trees are also proved.

In order to make our paper transparent, we offer the following two tables, Figure 3 and 4, which summarize our results. In both tables, the first row follows from Theorem 1.2 or its generalization Corollary 3.3. The last row is obvious since $\hom(S_m, G)$ is the sum of degree powers of G and it also follows from Corollary 3.3. The first, second and third columns follow from Theorem 1.4, Theorem 1.5 and Corollary 3.5 respectively. The "X" means that there is no inequality between the two expressions in general and the "?" means that we do not know whether the statement is true or not.

$$
\begin{array}{ccccc}
\hom(P_m, P_n) & \leqslant & \hom(P_m, T_n) & \leqslant & \hom(P_m, S_n) \\
\rotatebox[origin=c]{90}{$\geqslant$} & & X & & \rotatebox[origin=c]{90}{$\geqslant$} \\
\hom(T_m, P_n) & \overset{(*)}{\leqslant} & \hom(T_m, T_n) & X & \hom(T_m, S_n) \\
\rotatebox[origin=c]{90}{$\geqslant$} & & \rotatebox[origin=c]{90}{$\geqslant$} & & \rotatebox[origin=c]{90}{$\geqslant$} \\
\hom(S_m, P_n) & \leqslant & \hom(S_m, T_n) & \leqslant & \hom(S_m, S_n)
\end{array}
$$

Figure 3: The number of homomorphisms between trees of sizes m and n. The $(*)$ means that there are some well-determined (possible) counterexamples which should be excluded.

$$
\begin{array}{ccccc}
\hom(P_n, P_n) & \leqslant & \hom(P_n, T_n) & \leqslant & \hom(P_n, S_n) \\
\rotatebox[origin=c]{90}{$\geqslant$} & & ? & & \rotatebox[origin=c]{90}{$\geqslant$} \\
\hom(T_n, P_n) & \leqslant & \hom(T_n, T_n) & X & \hom(T_n, S_n) \\
\rotatebox[origin=c]{90}{$\geqslant$} & & \rotatebox[origin=c]{90}{$\geqslant$} & & \rotatebox[origin=c]{90}{$\geqslant$} \\
\hom(S_n, P_n) & \leqslant & \hom(S_n, T_n) & \leqslant & \hom(S_n, S_n)
\end{array}
$$

Figure 4: The number of homomorphisms and endomorphisms of trees of size n.

2 The Tree-walk algorithm

In this section we shall state an algorithm for the number of homomorphisms from a tree to any graph. As a generalized concept of walks in graphs, we call a homomorphism from a tree to a graph a *tree-walk* on this graph.

Let $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_n)$ be two vectors. We usually denote by $\|\mathbf{a}\| = a_1 + a_2 + \cdots + a_n$ the *norm* of $\mathbf{a}$ and by $\mathbf{a} * \mathbf{b} = (a_1b_1, \ldots, a_nb_n)$ the *Hadamard product* of $\mathbf{a}$ and $\mathbf{b}$. Denote by $\mathbf{1_n}$ the n-dimensional row vector with all entries are equal to 1. Let G be a graph with n vertices. The *adjacency matrix* of G is the $n \times n$ matrix $A_G := (a_{uv})_{u,v \in V(G)}$, where $a_{uv} = 1$ when $uv \in E(G)$, otherwise 0. We begin with a fundamental lemma about the number of walks in a graph.

Lemma 2.1. *Let G be a labeled graph and $A = A_G$ the adjacency matrix of G. Then the (i, j)-entry of the matrix A^n counts the number of walks in G from vertex i to vertex j with length n.*

Proof. By easy induction on n. See for example [20, Theorem 4.7.1]. ☐

Definition 2.2 (hom-vector)**.** Let T be a tree and G be a graph with vertices labeled by $1, 2, \ldots, n$. Let $v \in V(T)$ be any vertex of T. The n-dimensional vector

$$\mathbf{h}(T, v, G) := (h_1, h_2, \ldots, h_n)$$

where

$$h_i = |\{f \in \mathrm{Hom}(T, G) \mid f(v) = i\}|,$$

is called the *hom-vector* at v from T to G. Clearly, $\hom(T, G) = \|\mathbf{h}(T, v, G)\|$. Sometimes, we also call $\mathbf{h}(T, v, G)$ the hom-vector from the rooted tree $T(v)$ to the graph G and use the more compact notation $\mathbf{h}(T(v), G)$.

The following Tree-walk algorithm can be viewed as a generalization of Lemma 2.1 for computing the number of tree-walks in graphs.

The Tree-walk algorithm. Let $A = A_G$ be the adjacency matrix of the labeled graph G. Let v be a vertex of the tree T. We now give the algorithm to compute $\mathbf{h}(T, v, G)$. We consider two type of recursion steps.

Recursion 1. If v is a non-leaf vertex of T, then we can decompose T to $T_1 \cup T_2$ such that $V(T_1) \cap V(T_2) = \{v\}$, and T_1 and T_2 are strictly smaller than T. In this case

$$\mathbf{h}(T, v, G) = \mathbf{h}(T_1, v, G) * \mathbf{h}(T_2, v, G).$$

Recursion 2. If v is a leaf with the unique neighbor u in T, then

$$\mathbf{h}(T, v, G) = \mathbf{h}(T - v, u, G)A.$$

Hence we use Recursion 1 or Recursion 2 according to the vertex v is a non-leaf or a leaf. In most of the proofs we simply check whether some property of the vector $\mathbf{h}(T, v, G)$ remains valid after applying Recursion 1 and Recursion 2.

Note that if a leaf has distance d from the closest vertex of degree at least 3 then we can execute a sequence of Recursion 2 in one step by simply multiplying the corresponding hom-vector by A^d. This way we can speed up the algorithm a bit. For the sake of convenience, we include an example.

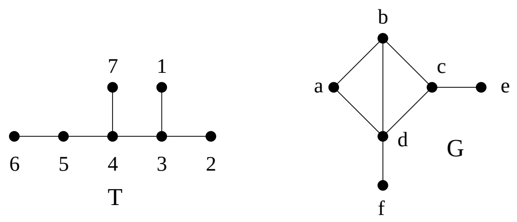

Figure 5: A labeled tree T and a graph G.

Example 2.3. Let T and G be the tree and the graph depicted in Fig. 5. Denote by $T[V]$ the induced subtree on the vertex set $V \subseteq V(T)$. Let us compute $\mathbf{h}(T,7,G)$ by the Tree-walk algorithm. First, we compute $\mathbf{h}(T[1,3],3,G)$ by using Recursion 2:

$$\mathbf{h}(T[1,3],3,G) = \mathbf{1_7}A_G = (2,3,3,4,1,1).$$

Next we compute $\mathbf{h}(T[1,2,3],3,G)$ by using Recursion 1:

$$\mathbf{h}(T[1,2,3],3,G) = \mathbf{h}(T[1,3],3,G) * \mathbf{h}(T[2,3],3,G) = (4,9,9,16,1,1).$$

Now let us compute $\mathbf{h}(T[1,2,3,4],4,G)$ by using Recursion 2 again:

$$\mathbf{h}(T[1,2,3,4],4,G) = \mathbf{h}(T[1,2,3],3,G)A_G = (25,29,26,23,9,16).$$

As a next step we determine $\mathbf{h}(T[6,5,4],4,G)$:

$$\mathbf{h}(T[6,5,4],4,G) = \mathbf{1_7}A_G^2 = (7,9,8,9,3,4).$$

Hence

$$\begin{aligned}\mathbf{h}(T[1,2,3,4,5,6],4,G) =& \mathbf{h}(T[1,2,3,4],4,G) * \mathbf{h}(T[6,5,4],4,G)\\ =&(175,261,208,207,27,64).\end{aligned}$$

Finally,

$$\mathbf{h}(T,7,G) = \mathbf{h}(T[1,2,3,4,5,6],4,G)A_G = (468,590,495,708,208,207).$$

Thus $\hom(T,G) = \|\mathbf{h}(T,7,G)\| = 2676$.

3 Proof of Theorem 1.3

The main purpose of this section is to prove Theorem 1.3. We shall give an inductive proof of Theorem 1.2 which can be generalized to tree-walks by the tree-walk algorithm.

We first need some notations. Let T be a tree and $T' = KC(T,p_0,p_k)$ its KC-transformation with respect to a path P of length k, a path with vertices labeled consecutively with $p_0,p_1,\ldots,p_k$. We denote by A and B the components of p_0 and p_k in the subgraph of T by deleting all the edges of P. Let A', B' and P' be the components of T' corresponding with components A, B and P under the KC-transformation, respectively. The vertices of the path P' will be labeled consecutively with $p'_0, p'_1, \ldots, p'_k$, where p'_i is corresponding to p_i for $0 \leqslant i \leqslant k$. So $p_0 \in A, p_k \in B$ in T, and $p'_0 \in A', B'$ in T'.

Lemma 3.1. *Let $a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2$ be positive numbers satisfying the inequalities: $a_i \geqslant \max(c_i, d_i)$, $a_i + b_i \geqslant c_i + d_i$ for $i = 1, 2$. Then $a_1a_2 \geqslant \max(c_1c_2, d_1d_2)$ and $a_1a_2 + b_1b_2 \geqslant c_1c_2 + d_1d_2$.*

Proof. Clearly, we only have to prove that $a_1a_2 + b_1b_2 \geqslant c_1c_2 + d_1d_2$, the other inequality is trivial. Note that $b_i \geqslant \max(0, c_i + d_i - a_i)$. If one of $c_i + d_i - a_i < 0$, say $c_1 + d_1 < a_1$ then

$$a_1a_2 \geqslant (c_1 + d_1)a_2 \geqslant c_1c_2 + d_1d_2.$$

If both $c_i + d_i - a_i \geqslant 0$ for $i = 1, 2$, then

$$\begin{aligned} & a_1a_2 + b_1b_2 \geqslant a_1a_2 + (c_1 + d_1 - a_1)(c_2 + d_2 - a_2) \\ =& c_1c_2 + d_1d_2 + (a_1 - c_1)(a_2 - d_2) + (a_1 - d_1)(a_2 - c_2) \geqslant c_1c_2 + d_1d_2. \end{aligned}$$

Hence we are done. ☐

We first treat the case of k being even.

Proof of first part of Theorem 1.3. In this proof k is even: $k = 2t$. We label $V(A)\backslash p_0$ with $\{a_m \mid 1 \leqslant m \leqslant M\}$, $V(A') \setminus p'_0$ with $\{a'_i \mid 1 \leqslant m \leqslant M\}$, $V(B) \setminus p_k$ with $\{b_n \mid 1 \leqslant n \leqslant N\}$ and $V(B) \setminus p'_0$ with $\{b'_n \mid 1 \leqslant n \leqslant N\}$, where a_m (resp. b_n) is corresponding to a'_m (resp. b'_n) under the KC-transformation. For $v \in H$, we always write

$$\begin{aligned} \mathbf{h}(H, v, T) &= (a_1, a_2, \ldots, a_M, p_0, p_1, \ldots, p_k, b_1, b_2, \ldots, b_N) \text{ and} \\ \mathbf{h}(H, v, T') &= (a'_1, a'_2, \ldots, a'_M, p'_0, p'_1, \ldots, p'_k, b'_1, b'_2, \ldots, b'_N), \end{aligned}$$

where we use the labels of vertices of T and T' to index the parameters of the hom-vectors to T and T' respectively. We hope that it will not cause any confusion. We shall prove by induction on the steps of tree-walk algorithm that

$$\begin{aligned} a'_m \geqslant a_m,\ b'_n &\geqslant b_n \\ p'_i + p'_{k-i} &\geqslant p_i + p_{k-i} \\ p'_i \geqslant p_i,\ p'_i &\geqslant p_{k-i} \end{aligned}$$

for $1 \leqslant m \leqslant M, 1 \leqslant n \leqslant N, 0 \leqslant i \leqslant t$.

It is easy to verify that all these inequalities are satisfied after applying any recursion step of the tree-walk algorithm. When v is a leaf of H then it is trivial that these inequalities are preserved. If v is not a leaf then we use Lemma 3.1 to see that the Hadamard-product preserves these inequalities. ☐

Lemma 3.2. *Let k be odd and assume that A and B have at least two vertices. Let $a_m(r), b_n(r), p_i(r), a'_m(r), b'_n(r), p'_i(r)$ denote the number of homomorphism of P_r into T and T', respectively, such that the endvertex of P_r goes to the vertices $a_m, b_n, p_i, a'_m, b'_n$ and p'_i, respectively. Then the following inequalities hold for every r:*

$$a'_m(r) \geqslant a_m(r),\ b'_n(r) \geqslant b_n(r) \tag{3.1}$$
$$p'_i(r) + p'_j(r) \geqslant p_i(r) + p_j(r) \tag{3.2}$$
$$p'_i(r) + p'_j(r) \geqslant p_{k-i}(r) + p_{k-j}(r) \tag{3.3}$$

for $1 \leqslant m \leqslant M, 1 \leqslant n \leqslant N$ and $i + j \leqslant k$.

Proof. We prove the claim by induction on r. For $r = 1, 2$, the claim is trivial. Note that we only have to prove that

$$a'_m(r) \geqslant a_m(r),\ b'_n(r) \geqslant b_n(r)$$
$$p'_i(r) + p'_j(r) \geqslant p_i(r) + p_j(r)$$

for $1 \leqslant m \leqslant M, 1 \leqslant n \leqslant N$ and $i + j \leqslant k$. We obtain the inequality

$$p'_i(r) + p'_j(r) \geqslant p_{k-i}(r) + p_{k-j}(r)$$

by simply exchanging the role of A and B. Also note that if we put $i = j$ in the inequality (3.2) and (3.3) we obtain that $p'_i(r) \geqslant p_i(r), p_{k-i}(r)$ for $i < k/2$.

Observe that for any vertex v we have:

$$v(r) = \sum_{u \in N(v)} u(r-1).$$

We will treat the cases $k = 1$ and $k \geqslant 3$ separately.

Case 1: $k = 1$. In this case, we have to prove the inequalities:

$$a'_m(r) \geqslant a_m(r),\ b'_n(r) \geqslant b_n(r), p'_0(r) \geqslant \max(p_0(r), p_1(r)), p'_0(r) + p'_1(r) \geqslant p_0(r) + p_1(r).$$

The inequalities $a'_m(r) \geqslant a_m(r),\ b'_n(r) \geqslant b_n(r)$ simply follow from the inequalities $a'_m(r-1) \geqslant a_m(r-1),\ b'_n(r-1) \geqslant b_n(r-1)$, and $p'_0(r-1) \geqslant p_0(r-1), p_1(r-1)$.

Observe that

$$\begin{aligned}
p'_0(r) &= \sum_{a'_m \in N(p'_0)} a'_m(r-1) + \sum_{b'_n \in N(p'_0)} b'_n(r-1) + p'_1(r-1) \\
&= \sum_{a'_m \in N(p'_0)} a'_m(r-1) + \sum_{b'_n \in N(p'_0)} b'_n(r-1) + p'_0(r-2) \\
&\geqslant \sum_{a_m \in N(p_0)} a_m(r-1) + \sum_{b_n \in N(p_1)} b_n(r-1) + p_0(r-2) \\
&\geqslant \sum_{a_m \in N(p_0)} a_m(r-1) + \sum_{b_n \in N(p_1)} b_n(r-2) + p_0(r-2)
\end{aligned}$$

$$= \sum_{a_m \in N(p_0)} a_m(r-1) + p_1(r-1) = p_0(r).$$

We used the induction hypothesis and that $b_m(r-1) \geqslant b_m(r-2)$. In general, $u(r) \geqslant u(r-1)$ since any homomorphism of P_{r-1} starting at the vertex u can be extended to a homomorphism of P_r starting at u. Clearly, we can get $p_0'(r) \geqslant p_1(r)$ similarly, or we just switch the role of A and B.

Finally,

$$\begin{aligned}
&p_0'(r) + p_1'(r) \\
&= \sum_{a_m' \in N(p_0')} a_m'(r-1) + \sum_{b_n' \in N(p_0')} b_n'(r-1) + p_0'(r-1) + p_1'(r-1) \\
&\geqslant \sum_{a_m \in N(p_0)} a_m(r-1) + \sum_{b_n \in N(p_1)} b_n(r-1) + p_0(r-1) + p_1(r-1) = p_0(r) + p_1(r).
\end{aligned}$$

Hence we are done in this case.

Case 2: $k \geqslant 3$. Clearly, the inequalities $a_m'(r) \geqslant a_m(r)$, $b_n'(r) \geqslant b_n(r)$ simply follow from the inequalities $a_m'(r-1) \geqslant a_m(r-1)$, $b_n'(r-1) \geqslant b_n(r-1)$, and $p_0'(r-1) \geqslant p_0(r-1), p_k(r-1)$ as before.

So we only have to prove the inequality $p_i'(r) + p_j'(r) \geqslant p_i(r) + p_j(r)$ for $i + j \leqslant k$. We can assume that $i \leqslant j$. If $i \geqslant 1$, then $j \leqslant k-1$ and

$$\begin{aligned}
p_i'(r) + p_j'(r) &= (p_{i-1}'(r-1) + p_{j+1}'(r-1)) + (p_{i+1}'(r-1) + p_{j-1}'(r-1)) \\
&\geqslant (p_{i-1}(r-1) + p_{j+1}(r-1)) + (p_{i+1}(r-1) + p_{j-1}(r-1)) \\
&= p_i(r) + p_j(r).
\end{aligned}$$

So we only have to consider the case $i = 0$. In this case we consider the cases $j = 0, j = 1, 2 \leqslant j \leqslant k-2, j = k-1, j = k$ separately. Unfortunately, all of them behaves a bit differently.

Subcase $j = 0$:

$$\begin{aligned}
2p_0'(r) &= 2\Bigg(\sum_{a_m' \in N(p_0')} a_m'(r-1) + \sum_{b_n' \in N(p_0')} b_n'(r-1) + p_1'(r-1) \Bigg) \\
&\geqslant 2\Bigg(\sum_{a_m \in N(p_0)} a_m(r-1) + p_1(r-1) \Bigg) = 2p_0(r),
\end{aligned}$$

since $p_1'(r-1) \geqslant p_1(r-1)$, because $1 < k/2$.

Subcase $j = 1$:

$$p_0'(r) + p_1'(r) = \sum_{a_m' \in N(p_0')} a_m'(r-1) + \sum_{b_n' \in N(p_0')} b_n'(r-1) + p_1'(r-1) + p_0'(r-1) + p_2'(r-1)$$

$$\geqslant \sum_{a_m \in N(p_0)} a_m(r-1) + p_1(r-1) + p_0(r-1) + p_2(r-1) = p_0(r) + p_1(r),$$

since $p'_1(r-1) \geqslant p_1(r-1)$ and $p'_0(r-1) + p'_2(r-1) \geqslant p_0(r-1) + p_2(r-1)$.

Subcase $2 \leqslant j \leqslant k-2$: Here we jump back from r to $r-2$, so we need a few notations. Let d_A and d_B denote the degree of p'_0 in A and B, respectively. Furthermore, let $d(v,u)$ denote the distance of the vertices u and v. Then

$$p'_0(r) + p'_j(r) = \sum_{a'_m : d(a'_m, p'_0)=2} a'_m(r-2) + \sum_{b'_n : d(b'_n, p'_0)=2} b'_n(r-2) + (d_A + d_B + 1)p'_0(r-2)$$

$$+ p'_2(r-2) + p'_{j-2}(r-2) + 2p'_j(r-2) + p'_{j+2}(r-2) \geqslant$$

$$\geqslant \sum_{a_m : d(a_m, p_0)=2} a_m(r-2) + (d_A+1)p_0(r-2) + p_2(r-2) +$$

$$+ p_{j-2}(r-2) + 2p_j(r-2) + p_{j+2}(r-2) = p_0(r) + p_j(r),$$

since the inequality follows from the following inequalities:

$$\begin{aligned}
a'_m(r-2) &\geqslant a_m(r-2) \\
b'_n(r-2) &\geqslant 0 \\
(d_B - 1)p'_0(r-2) &\geqslant -p_0(r-2) \\
(d_A - 1)p'_0(r-2) &\geqslant (d_A - 1)p_0(r-2) \\
p'_2(r-2) + p'_{j-2}(r-2) &\geqslant p_2(r-2) + p_{j-2}(r-2) \\
2(p'_0(r-2) + p'_j(r-2)) &\geqslant 2(p_0(r-2) + p_j(r-2)) \\
p'_0(r-2) + p'_{j+2}(r-2) &\geqslant p_0(r-2) + p_{j+2}(r-2).
\end{aligned}$$

Subcase $j = k-1$:

$$\begin{aligned}
&p'_0(r) + p'_{k-1}(r) \\
&= \sum_{a'_m \in N(p'_0)} a'_m(r-1) + \sum_{b'_n \in N(p'_0)} b'_n(r-1) + p'_1(r-1) + p'_{k-2}(r-1) + p'_k(r-1) \\
&= \sum_{a_m : d(a'_m, p'_0)=2} a'_m(r-2) + d_A p'_0(r-2) \\
&\quad + \sum_{b'_n \in N(p'_0)} b'_n(r-1) + p'_0(r-2) + p'_2(r-2) + p'_{k-3}(r-2) + 2p'_{k-1}(r-2).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&p_0(r) + p_{k-1}(r) \\
&= \sum_{a_m \in N(p_0)} a_m(r-1) + p_1(r-1) + p_{k-2}(r-1) + p_k(r-1)
\end{aligned}$$

$$= \sum_{a_m : d(a_m,p_0)=2} a_m(r-2) + d_A p_0(r-2) + p_0(r-2) + p_2(r-2) + p_{k-3}(r-2) + 2p_{k-1}(r-2)$$
$$+ \sum_{b_n \in N(p_k)} b_n(r-2).$$

The inequality $p'_0(r) + p'_{k-1}(r) \geqslant p_0(r) + p_{k-1}(r)$ follows from

$$\begin{aligned} a'_m(r-2) &\geqslant a_m(r-2) \\ b'_n(r-1) &\geqslant b_n(r-1) \geqslant b_n(r-2) \\ (d_A-1)p'_0(r-2) &\geqslant (d_A-1)p_0(r-2) \\ p'_2(r-2) + p'_{k-3}(r-2) &\geqslant p_2(r-2) + p_{k-3}(r-2) \\ 2(p'_0(r-2) + p'_{k-1}(r-2)) &\geqslant 2(p_0(r-2) + p_{k-1}(r-2)). \end{aligned}$$

Subcase $j = k$:

$$\begin{aligned} p'_0(r) + p'_k(r) &= \sum_{a'_m \in N(p'_0)} a'_m(r-1) + \sum_{b'_n \in N(p'_0)} b'_n(r-1) + p'_1(r-1) + p'_{k-1}(r-1) \\ &\geqslant \sum_{a_m \in N(p_0)} a_m(r-1) + \sum_{b_n \in N(p_k)} b_n(r-1) + p_1(r-1) + p_{k-1}(r-1) \\ &= p_0(r) + p_k(r). \end{aligned}$$

□

Proof of the second part of Theorem 1.3. From Lemma 3.2 we only keep the inequalities

$$\begin{aligned} a'_m(r) &\geqslant a_m(r) \\ b'_n(r) &\geqslant b_n(r) \\ p'_i(r) &\geqslant p_i(r), p_{k-i}(r) \\ p'_i(r) + p'_{k-i}(r) &\geqslant p_i(r) + p_{k-i}(r) \end{aligned}$$

for $1 \leqslant m \leqslant M, 1 \leqslant n \leqslant N, 0 \leqslant i \leqslant k/2$.

For a tree H and $v \in H$, let us write

$$\begin{aligned} \mathbf{h}(H,v,T) &= (a_1, a_2, \ldots, a_M, p_0, p_1, \ldots, p_k, b_1, b_2, \ldots, b_N) \text{ and} \\ \mathbf{h}(H,v,T') &= (a'_1, a'_2, \ldots, a'_M, p'_0, p'_1, \ldots, p'_k, b'_1, b'_2, \ldots, b'_N), \end{aligned}$$

where we use the labels of vertices of T and T' to index the parameters of the hom-vectors to T and T', respectively. We say that $\mathbf{h}(H,v,T) \leqslant \mathbf{h}(H,v,T')$ if the following inequalities hold

$$\begin{aligned} a'_m &\geqslant a_m \\ b'_n &\geqslant b_n \end{aligned}$$

$$p'_i + p'_{k-i} \geqslant p_i + p_{k-i}$$
$$p'_i \geqslant p_i$$
$$p'_i \geqslant p_{k-i}$$

for $1 \leqslant m \leqslant M, 1 \leqslant n \leqslant N, 0 \leqslant i \leqslant k/2$. As we have seen these inequalities hold for a path P_r and its endvertex. Since these inequalities are preserved for Hadamard-product by Lemma 3.1, we see that $\mathbf{h}(H, v, T) \leqslant \mathbf{h}(H, v, T')$ for starlike trees H, where v is the center of the starlike tree. This implies that

$$\hom(H, T') \geqslant \hom(H, T). \qquad \square$$

The following generalization of inequality (1.2) follows immediately from Proposition 1.1 and Theorem 1.3.

Corollary 3.3. *Let H be a starlike tree and let T_n be a tree on n vertices. Then*

$$\hom(H, P_n) \leqslant \hom(H, T_n) \leqslant \hom(H, S_n).$$

The reader may wonder that if inequality (1.4) holds when k is odd and H is not a starlike tree. This is not true in general. A counterexample will be constructed in the following, which also shows that

$$\hom(H, T_n) \leqslant \hom(H, S_n)$$

is not true for any tree H.

Proposition 3.4. *Let T be a tree with color classes A and B considered as a bipartite graph. Then*

$$\hom(T, S_n) = (n-1)^{|A|} + (n-1)^{|B|}.$$

Corollary 3.5. *Let T_m be a tree on m vertices, then*

$$\hom(P_m, S_n) \leqslant \hom(T_m, S_n) \leqslant \hom(S_m, S_n).$$

If $T \neq S_m$ then the second inequality is strict.

Proof of Proposition 3.4. Since T and S_n are bipartite graphs, a color class of T have to go into a color class of S_n. If the color class A goes to the center of S_n, then any vertex belonging to the color class of B can go to any leaf of the star, so it provides $(n-1)^{|B|}$ homomorphisms. The other case provides $(n-1)^{|A|}$ homomorphisms. $\square$

This simple proposition also shows us how to construct a tree T_n for which $\hom(T_n, T_n) > \hom(T_n, S_n)$.

Let $T_n = S^*_{2k}$ be the doublestar on $2k$ vertices with $2k-2$ leaves and two vertices of degree k. Then it is easy to see that

$$\hom(S^*_{2k}, S^*_{2k}) > 2(k-1)^{2(k-1)} = 2(k^2 - 2k + 1)^{k-1},$$

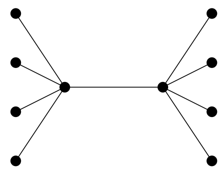

Figure 6: The doublestar S^*_{10}.

while

$$\hom(S^*_{2k}, S_{2k}) = 2(2k-1)^k.$$

Hence for $k \geqslant 5$ we have

$$\hom(S^*_{2k}, S^*_{2k}) > \hom(S^*_{2k}, S_{2k}).$$

Note that S_{2k} can be obtained from S^*_{2k} by a KC-transformation.

4 Graph homomorphisms from trees

4.1 Markov chains and homomorphisms

Theorem 4.1. *Let G be a graph and let $P = (p_{ij})$ be a Markov chain on G:*

$$\sum_{j \in N(i)} p_{ij} = 1 \quad \textit{for all} \;\; i \in V(G),$$

where $p_{ij} \geqslant 0$ and $p_{ij} = 0$ if $(i,j) \notin E(G)$. Let $Q = (q_i)$ be the stationary distribution of P:

$$\sum_{j \in N(i)} q_j p_{ji} = q_i \quad \textit{for all} \;\; i \in V(G).$$

Let us define the following entropies:

$$H(Q) = \sum_{i \in V(G)} q_i \log \frac{1}{q_i},$$

and

$$H(D|Q) = \sum_{i \in V(G)} q_i \log d_i,$$

where d_i is the degree of the vertex i, and let

$$H(P|Q) = \sum_{i \in V(G)} q_i \left(\sum_{j \in N(i)} p_{ij} \log \frac{1}{p_{ij}} \right).$$

Let T_m be a tree with ℓ leaves on m vertices, where $m \geqslant 3$. Then

$$\hom(T_m, G) \geqslant \exp\Big(H(Q) + \ell H(D|Q) + (m-1-\ell) H(P|Q) \Big).$$

Proof. Let v be a root of T. Let a_i be the number of homomorphisms of T_m into G such that the root vertex v goes into the vertex $i \in V(G)$. Let

$$F(T_m(v), G) = \prod_{i=1}^{n} a_i^{q_i}.$$

We will show by induction on m that

$$F(T_m(v), G) \geqslant \exp\left(\ell^* H(D|Q) + (m - 1 - \ell^*) H(P|Q)\right),$$

where ℓ^* is the number of leaves different from v, so it is ℓ if v is not a leaf and $\ell - 1$ if v is a leaf. Note that

$$F(K_2(v), G) = \exp H(D|Q).$$

If v is not a leaf of T_m, then we can decompose T_m to $T_1(v)$ and $T_2(v)$. Then

$$F(T_m(v), G) = F(T_1(v), G) F(T_2(v), G)$$

because of the Hadamard-products of the hom-vectors. From this the claim follows immediately by induction.

If v is a leaf of T_m with the unique neighbor u, then let

$$\mathbf{h}(T_m - v, u, G) = (b_1, \ldots, b_n).$$

So $a_i = \sum_{j \in N(i)} b_j$.

For positive numbers $r_1, \ldots, r_t$ and positive weights $w_1, \ldots, w_t$ with $\sum_{i=1}^{t} w_i = 1$, the weighted AM-GM inequality says that

$$\begin{aligned} r_1 + \cdots + r_t &= w_1 \left(\frac{r_1}{w_1}\right) + \cdots + w_t \left(\frac{r_t}{w_t}\right) \\ &\geqslant \left(\frac{r_1}{w_1}\right)^{w_1} \cdots \left(\frac{r_t}{w_t}\right)^{w_t} \\ &= \exp\left(\sum_{i=1}^{t} w_i \log \frac{1}{w_i}\right) \prod_{i=1}^{t} r_i^{w_i}. \end{aligned}$$

Hence

$$\begin{aligned} F(T_m(v), G) = \prod_{i=1}^{n} a_i^{q_i} &= \prod_{i=1}^{n} \left(\sum_{j \in N(i)} b_j\right)^{q_i} \\ &\geqslant \prod_{i=1}^{n} \left(\prod_{j \in N(i)} \left(\frac{b_j}{p_{ij}}\right)^{p_{ij}}\right)^{q_i} \\ &= \prod_{i=1}^{n} \left(\prod_{j \in N(i)} \left(\frac{1}{p_{ij}}\right)^{p_{ij} q_i}\right) \prod_{i=1}^{n} b_i^{\sum_{j \in N(i)} p_{ji} q_j} \end{aligned}$$

$$= \prod_{i=1}^{n} \left(\prod_{j \in N(i)} \left(\frac{1}{p_{ij}} \right)^{p_{ij} q_i} \right) \prod_{i=1}^{n} b_i^{q_i}.$$

In the last step we used that Q is a stationary distribution with respect to P. Hence

$$F(T_m(v), G) \geqslant \exp(H(P|Q)) F((T_m - v)(u), G).$$

Now the claim follows by induction.

To finish the proof of the theorem, we only have to choose a nonleaf root and use that

$$\hom(T_m, G) = \sum_{i=1}^{n} a_i \geqslant \exp(H(Q)) F(T_m(v), G). \qquad \square$$

Remark 4.2. Note that the inequality $H(D|Q) \geqslant H(P|Q)$ always holds. Consequently,

$$\hom(T_m, G) \geqslant \exp(H(Q) + (m-1)H(P|Q)).$$

As Theorem 4.1 suggests, this is an inequality for entropies and indeed, it can be proved in this way. By P and Q, we defined a distribution on the set of homomorphisms: we choose a root according to Q, then we choose every nonleaf new vertex according to P and finally we choose the leaves uniformly. The entropy of this distribution is exactly $H(Q) + \ell H(D|Q) + (m - 1 - \ell)H(P|Q)$ since every nonleaf vertex has distribution Q. Note that this entropy is smaller than the entropy of the uniform distribution, that is, $\log \hom(T_m, G)$. For basic facts about entropy, see for example [5].

Theorem 4.3. *Let G be a connected graph on the vertex set $\{1, 2, \dots, n\}$ and let λ be the largest eigenvalue of the adjacency matrix of the graph G. Let $\underline{y}$ be a positive eigenvector of unit length corresponding to λ. Let $q_i = y_i^2$. Then for any rooted tree T_m on m vertices we have*

$$\hom(T_m, G) \geqslant \exp(H_\lambda(G)) \lambda^{m-1},$$

where

$$H_\lambda(G) = \sum_{i=1}^{n} q_i \log \frac{1}{q_i}$$

is the spectral entropy of the graph G.

Proof. We will use Theorem 4.1. Let $p_{ij} = \frac{y_j}{\lambda y_i}$. Since $\underline{y}$ is a positive eigenvector, we have $p_{ij} > 0$. For all i we have $\lambda y_i = \sum_{j \in N(i)} y_j$, thus $\sum_{j \in N(i)} p_{ij} = 1$. For $q_i = y_i^2$ we have

$$q_i p_{ij} = y_i^2 \frac{y_j}{\lambda y_i} = \frac{1}{\lambda} y_i y_j = q_j p_{ji}.$$

Hence

$$\sum_{i \in N(j)} q_j p_{ji} = \sum_{i \in N(j)} q_i p_{ij} = q_i.$$

This means that $P = (p_{ij})$ is a Markov chain with stationary distribution $Q = (q_i)$. The conditional entropy

$$\begin{aligned} H(P|Q) &= \sum_{i \in V(G)} q_i \left(\sum_{j \in N(i)} p_{ij} \log \frac{1}{p_{ij}} \right) \\ &= \sum_{i \in V(G)} y_i^2 \left(\sum_{j \in N(i)} \frac{y_j}{\lambda y_i} \log \frac{\lambda y_i}{y_j} \right) \\ &= \sum_{\{i,j\} \in E(G)} \frac{y_i y_j}{\lambda} \left(2 \log \lambda + \log \frac{y_i}{y_j} + \log \frac{y_j}{y_i} \right) \\ &= \log(\lambda) \frac{1}{\lambda} \sum_{(i,j) \in E(G)} y_i y_j = \log \lambda. \end{aligned}$$

Hence the result follows from Theorem 4.1. □

Remark 4.4. A Markov chain is called reversible if $q_i p_{ij} = q_j p_{ji}$ for all $i, j \in V(G)$. As we have seen, the Markov chain constructed in the previous proof is reversible. It is not hard to show that on trees every Markov chain is reversible.

Remark 4.5. Theorem 4.3 is the best possible in the sense that there cannot be a larger number than λ in such a statement since

$$\hom(P_m, G) \leqslant n\lambda^{m-1}.$$

Indeed,

$$\frac{\hom(P_m, G)}{n} = \frac{\mathbf{1_n}^T A^{m-1} \mathbf{1_n}}{\mathbf{1_n}^T \mathbf{1_n}} \leqslant \max_{v \neq \underline{0}} \frac{v^T A^{m-1} v}{v^T v} = \lambda_{\max}(A^{m-1}) = \lambda^{m-1}.$$

Note that we can deduce that if $(T_m)_{m=1}^{\infty}$ is a sequence of trees such that T_m has m vertices then

$$\liminf_{m \to \infty} \hom(T_m, G)^{1/m} \geqslant \liminf_{m \to \infty} \hom(P_m, G)^{1/m} = \lambda.$$

This result could have been deduced as well from a theorem of B. Rossman and E. Vee [21] claiming that

$$\hom(T_m, G) \geqslant \hom(C_m, G),$$

where C_m is the cycle on m vertices. In fact, this was proved for directed trees and cycles, but it implies the inequality for undirected tree and cycle. This result can also be deduced from Theorem 3.1 of [13].

The following special case of Theorem 4.1, involving the degree sequence of graphs, is Theorem 3 in the paper [9].

Theorem 4.6 (Dellamonica et al.)**.** *Let G be a graph on the vertex set $\{1, 2, \ldots, n\}$ with $e(G)$ edges and with degree sequence $(d_1, \ldots, d_n)$. Then for any tree T_m on m vertices we have*

$$\hom(T_m, G) \geqslant 2e(G) \cdot C^{m-2},$$

where

$$C = \left(\prod_{i=1}^{n} d_i^{d_i}\right)^{1/2e(G)}.$$

Proof. Let us consider the following classical Markov chain: $p_{ij} = \frac{1}{d_i}$ if $j \in N(i)$. The stationary distribution is $q_i = \frac{d_i}{2e(G)}$. Note that

$$H(P|Q) = \sum_{i \in V(G)} q_i \left(\sum_{j \in N(i)} p_{ij} \log \frac{1}{p_{ij}}\right) = \sum_{i \in V(G)} q_i \log d_i = \frac{1}{2e(G)} \sum_{i \in V(G)} d_i \log d_i = \log C$$

and

$$H(Q) + H(P|Q) = \sum_{(i,j) \in E(G)} q_i p_{ij} \log \frac{1}{q_i p_{ij}} = \sum_{(i,j) \in E(G)} \frac{1}{2e(G)} \log(2e(G)) = \log(2e(G)).$$

Hence the result follows from Theorem 4.1. □

Definition 4.7. The homomorphism density $t(H, G)$ is defined as follows:

$$t(H, G) = \frac{\hom(H, G)}{|V(G)|^{|V(H)|}}.$$

This is the probability that a random map is a homomorphism.

Sidorenko's conjecture says that

$$t(H, G) \geqslant t(K_2, G)^{e(H)}$$

for every bipartite graph H with $e(H)$ edges. It is known that Sidorenko's conjecture [18] is true for trees. By now, there are many proofs for this particular case of Sidorenko's conjecture: see [11, 15] and it can be deduced as well from Theorem 3.1 of [13]. Below we give a new proof for this fact.

Theorem 4.8. *For any tree T_m on m vertices and a graph G we have*

$$t(T_m, G) \geqslant t(K_2, G)^{m-1}.$$

Proof. Let $|V(G)| = n$. The theorem will immediately follows form Theorem 4.6. By convexity of the function $x \log x$ we have

$$\frac{1}{2e(G)} \sum_{i \in V(G)} d_i \log d_i \geqslant \frac{1}{2e(G)} n \left(\frac{2e(G)}{n} \log \frac{2e(G)}{n}\right) = \log \frac{2e(G)}{n}.$$

Hence

$$t(T_m, G) = \frac{\hom(T_m, G)}{n^m} \geqslant \frac{1}{n^m} 2e(G) \left(\frac{2e(G)}{n}\right)^{m-2} = \left(\frac{2e(G)}{n^2}\right)^{m-1} = t(K_2, G)^{m-1}. \quad \square$$

4.2 Sidorenko's theorem on extremality of stars

The objective of this section is to give a new proof for Theorem 1.5 in order to keep this paper self-contained. This was proved originally by Sidorenko [19]. Our proof is very similar to the original one, but it is slightly more elementary.

Before we start the proof we will need two definitions and two lemmas.

Definition 4.9. Let M_u and N_v be two rooted graphs with root vertices u and v, respectively. Then $M_u \circ_{u=v} N_v$ denotes the graph obtained from $M_u \cup N_v$ by identifying the vertices u and v.

Lemma 4.10. *Let $R_{u,v}$ be a graph with specified (not necessarily distinct) vertices u and v. Let $J_{u'}$ and $K_{v'}$ be two graphs with root vertices u' and v'. Finally, let the graphs A, B and C be obtained from $R_{u,v}, J_{u'}, K_{v'}$ as follows:*

$$A = (R_{u,v} \circ_{u=u'} J_{u'}) \circ_{v=v'} K_{v'},$$
$$B = (R_{u,v} \circ_{u=u'} J_{u'}) \circ_{u=u'} J_{u'},$$
$$C = (R_{u,v} \circ_{v=v'} K_{v'}) \circ_{v=v'} K_{v'}.$$

(In other words, in B and C we attach two copies of the same graph at the specified vertex.) Then for any graph G we have

$$2\hom(A,G) \leqslant \hom(B,G) + \hom(C,G).$$

Proof. Let $i, j \in V(G)$ and let $h(R_{u,v}, i, j)$ denote the number of homomorphisms of $R_{u,v}$ to G where u goes to i and v goes to j. We similarly define $h(J_{u'}, i)$ and $h(K_{v'}, j)$. Then

$$\hom(A,G) = \sum_{i,j\in V(G)} h(R_{u,v}, i, j)h(J_{u'}, i)h(K_{v'}, j).$$

Similarly,

$$\hom(B,G) = \sum_{i,j\in V(G)} h(R_{u,v}, i, j)h(J_{u'}, i)^2,$$

and

$$\hom(C,G) = \sum_{i,j\in V(G)} h(R_{u,v}, i, j)h(K_{v'}, j)^2.$$

Hence

$$\hom(B,G) + \hom(C,G) - 2\hom(A,G) = \sum_{i,j\in V(G)} h(R_{u,v}, i, j)(h(J_{u'}, i) - h(K_{v'}, j))^2 \geqslant 0.$$

We are done. □

Definition 4.11. Let $d(u,v)$ be the distance of the vertices $u, v \in V(G)$. Then the *Wiener-index* $W(G)$ of a graph G is defined as

$$W(G) := \sum_{u,v \in V(G)} d(u,v).$$

In our application $R_{u,v}$ will be a tree and $J_{u'}$ and $K_{v'}$ be the trees on 2 vertices. The following lemma about the Wiener-index is trivial.

Lemma 4.12. *Let $R_{u,v}$ be a tree with distinct vertices u and v. Let $J_{u'}$ and $K_{v'}$ be two copies of the two-node trees with root vertices u' and v', respectively. Finally, let the graphs A, B and C be obtained from $R_{u,v}, J_{u'}, K_{v'}$ as in the former lemma. Then $2W(A) > W(B) + W(C)$.*

Proof of Theorem 1.5. Let $\mathcal{T}_G$ be the set of those trees F on m vertices for which $\hom(F, G)$ is maximal. Let $T \in \mathcal{T}_G$ be the tree for which $W(T)$ is minimal. We show that $T = S_m$. Assume for contradiction that $T \neq S_m$. Then T has two leaves, a and b such that $d(a,b) \geqslant 3$. Let u and v be the unique neighbors of a and b, respectively. Then $u \neq v$. Let $R_{u,v} = T - \{a, b\}$, $J_{u'} = \{u', a\}$ and $K_{v'} = \{v', b\}$. Then

$$A = (R_{u,v} \circ_{u=u'} J_{u'}) \circ_{v=v'} K_{v'} = T.$$

As in the lemmas, let

$$B = (R_{u,v} \circ_{u=u'} J_{u'}) \circ_{u=u'} J_{u'},$$
$$C = (R_{u,v} \circ_{v=v'} K_{v'}) \circ_{v=v'} K_{v'}.$$

Note that B and C are also trees on m vertices. By the Lemma we have

$$2\hom(A, G) \leqslant \hom(B, G) + \hom(C, G).$$

Since $A = T \in \mathcal{T}_G$, then $\hom(B, G) + \hom(C, G) \leqslant 2\hom(A, G)$. So $\hom(A, G) = \hom(B, G) = \hom(C, G)$ implying that $B, C \in \mathcal{T}_G$ as well. But then $2W(T) > W(B) + W(C)$, so one of them has strictly smaller Wiener-index than T, this contradicts the choice of T. Hence T must be S_m. ☐

Remark 4.13. Let E_7 be the tree obtained from P_6 by putting a pendant edge to the third vertex of the path. Then there is a tree T for which

$$\hom(P_7, T) > \hom(E_7, T).$$

The following tree T is suitable: let $T = T(k_1, k_2, k_3)$ be the tree where the root vertex v_0 have k_1 neighbors, all of its neighbors has $k_2 + 1$ neighbors and the vertices having distance 2 from v_0 have $k_3 + 1$ neighbors. If we choose k_1, k_2, k_3 such that $k_2 \ll k_1 \ll k_3 \ll k_1 k_2$ (for instance $k_i = k^{\alpha_i}$, where $\alpha_2 < \alpha_1 < \alpha_3 < \alpha_1 + \alpha_2$ and k is large), then

$$\hom(P_7, T) - \hom(E_7, T) = k_1^2 k_2^2 k_3^2 + o(k_1^2 k_2^2 k_3^2).$$

5 Proofs of Theorems 1.6 and 1.8

In this section we give the proof of Theorem 1.6 and Theorem 1.8. As we will see, Theorem 1.6 with some additional observations implies Theorem 1.8.

To prove Theorem 1.6 we will build on the fact that there are not many homomorphisms into a path. Indeed, by Theorem 1.5 we have

$$\hom(T_m, P_n) \leqslant \hom(S_m, P_n) = (n-2)2^{m-1} + 2.$$

So for a particular tree T_n, it is enough to prove that for every tree T_m we have

$$\hom(T_m, T_n) \geqslant (n-2)2^{m-1} + 2. \tag{5.1}$$

This would immediately imply that

$$\hom(T_m, T_n) \geqslant \hom(T_m, P_n). \tag{5.2}$$

We will prove that inequality 5.1 is indeed true for all trees T_n with at least four leaves and for a large class of trees with three leaves. For the remaining trees with three leaves we use Theorem 1.3.

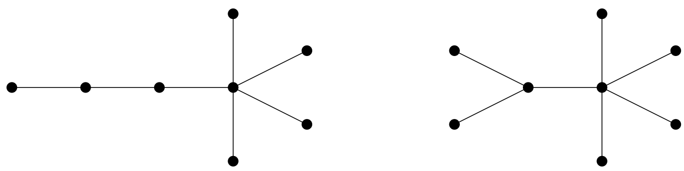

Figure 7: The trees T_8 (left) and T'_8 (right).

Remark 5.1. To prove Theorem 1.6 and Theorem 1.8 we cannot rely entirely on the use of KC-transformation. That is why we had to find another strategy to prove these theorems.

Indeed, KC-transformation does not always increase the number of endomorphisms of trees. The first counterexample is the two trees on 8 vertices in Fig. 7. The tree T'_8 is the KC-transformation of T_8, but $|\mathrm{End}(T'_8)| = 10430 < 17190 = |\mathrm{End}(T_8)|$.

5.1 The extremality of the star

Note that Theorem 1.5 and Theorem 1.3 together implies the following chain of inequalities:

$$|\mathrm{End}(T_n)| = \hom(T_n, T_n) \leqslant \hom(S_n, T_n) \leqslant \hom(S_n, S_n) = |\mathrm{End}(S_n)|,$$

since S_n is a starlike tree. In this section, we will also give a direct proof for it.

Theorem 1.8 (Second part). Let T_n be a tree on n vertices. Then

$$|\mathrm{End}(T_n)| \leqslant |\mathrm{End}(S_n)|.$$

If $T_n \neq S_n$ then strict inequality holds.

Proof. For the sake of simplicity we prove the statement for $n \geqslant 17$. The same proof applies to $n < 17$, we only need to compute a bit more carefully. In the end of the proof we will give the details of this more precise calculation.

Note that $|\text{End}(S_n)| = (n-1)^{n-1} + (n-1)$.

Let T_n be a tree on n vertices and let $d = d_1 \geqslant d_2 \geqslant \ldots \geqslant d_n$ be its degree sequence. Note that $d_1 + d_2 \leqslant n$, since the tree has only $n-1$ edges and the stars corresponding to the first two largest degrees can share at most one common edge.

First we prove that $|\text{End}(T_n)| \leqslant nd^{n-1}$. To see it, let $u_1, \ldots u_n$ be the vertices of the tree T_n such that $u_1, \ldots, u_k$ induces a tree for every k. Then we can chose the image of u_1 by n ways, and if we have already chosen the image of $u_1, \ldots, u_{k-1}$, then we can chose the image of u_k in at most d ways, since it must be the neighbor of some previous vertex. This means that $|\text{End}(T_n)| \leqslant nd^{n-1}$.

If $d \leqslant 2n/3$ then

$$nd^{n-1} \leqslant n \left(\frac{2n}{3}\right)^n \leqslant (n-1)^{n-1}$$

if $n \geqslant 17$, since then

$$\left(\frac{3}{2}\right)^n \geqslant en^2 \geqslant n^2 \left(1 + \frac{1}{n-1}\right)^{n-1}.$$

So we can assume that $d \geqslant \frac{2n}{3}$. Set $d = n - k$. We can assume that $T_n \neq S_n$, consequently $k \geqslant 2$. Let v_1 be the vertex having the largest degree and $v_2, \ldots, v_{d+1}$ its neighbors. Now we can decompose the set of endomorphisms according to the image of v_1 is v_1 or not. If it is v_1 then there can be at most d^{n-1} such endomorphisms. If the image of v_1 is not v_1, then we can chose that image in at most $(n-1)$ ways and the image of $v_2, v_3, \ldots, v_{d+1}$ can be chosen at most d_2 times and the image of all other vertices can be chosen in at most d ways. Hence

$$|\text{End}(T_n)| \leqslant d^{n-1} + (n-1)d_2^d d^{n-1-d}.$$

All we need to prove is that if $d \leqslant n-2$ then

$$d^{n-1} + (n-1)d_2^d d^{n-1-d} \leqslant (n-1)^{n-1} + (n-1).$$

With the notations $d = n - k$ we have

$$d^{n-1} + (n-1)d_2^d d^{n-1-d} \leqslant (n-k)^{n-1} + (n-1)k^{n-k}(n-k)^{k-1}.$$

By the binomial theorem we have

$$(n-1)^{n-1} = (n-k+k-1)^{n-1} \geqslant (n-k)^{n-1} + (n-1)(n-k)^{n-2}(k-1).$$

It is enough to prove that $(n-k)^{n-2} \geqslant k^{n-k}(n-k)^{k-1}$. This is equivalent with

$$(n-k)^{n-k-1} \geqslant k^{n-k}$$

and it is true since it is equivalent with

$$\left(\frac{n}{k}-1\right)^{n-k} \geqslant 2^{n-k} \geqslant n-k.$$

In the last step we have used that $n/k \geqslant 3$.

It is clear from the proof that we only have to check whether one of the inequalities hold for some d:

$$n \leqslant \left(\frac{n-1}{d}\right)^{n-1} \quad \text{or} \quad \left(\frac{n}{k}-1\right)^{n-k} \geqslant n-k.$$

For $8 \leqslant n \leqslant 16$ it is easy to see that if $d \leqslant n-4$ then the first inequality holds and if $d > n-4$, equivalently $k \leqslant 3$ then the second inequality holds. For $n = 5, 6, 7$ the first inequality holds if $d \leqslant n-3$, and the second inequality holds if $d > n-3$, equivalently $k \leqslant 2$. For $n = 4$ the claim is trivial $30 = |\mathrm{End}(S_4)| > |\mathrm{End}(P_4)| = 16$. □

5.2 The extremality of the path

Theorem 5.2. *Let T_m and T_n be trees on m and n vertices, respectively. If the tree T_n has at least four leaves, then*

$$\hom(T_m, T_n) \geqslant (n-2)2^{m-1} + 2.$$

An easy consequence of this theorem is the following.

Corollary 5.3. *If T_n is a tree on n vertices with at least 4 leaves, then*

$$\hom(T_m, T_n) \geqslant \hom(T_m, P_n).$$

Proof. Indeed,

$$\hom(T_m, T_n) \geqslant (n-2)2^{m-1} + 2 = \hom(S_m, P_n) \geqslant \hom(T_m, P_n),$$

where the second inequality follows from Theorem 1.5. □

A consequence of this theorem and Theorem 1.4 (the proof of which will be given in the follow-up paper [8]) is that path has the minimal number of endomorphisms.

Theorem 1.8(First part). For all trees T_n on n vertices we have

$$|End(T_n)| \geqslant |End(P_n)|.$$

Proof. If T_n has at least four leaves, then

$$\hom(T_n, T_n) \geqslant \hom(T_n, P_n) \geqslant \hom(P_n, P_n),$$

where the first inequality follows from Corollary 5.3, while the second inequality follows from Theorem 1.4. If the tree T_n has exactly three leaves, then it is star-like. Hence we can use Theorem 1.3 to prove the first inequality:

$$\hom(T_n, T_n) \geqslant \hom(T_n, P_n) \geqslant \hom(P_n, P_n).$$

□

The proof of Theorem 5.2 will be given next, which would complete the proof of Theorem 1.8.

First, we prove a reduction lemma which says that we only have to prove Theorem 5.2 for trees with exactly 4 leaves.

Lemma 5.4 (Reduction lemma). *Let T_m be a tree on m vertices and let n be fixed. Assume that for any tree T_k we have*

$$\hom(T_m, T_k) \geqslant (k-2)2^{m-1} + 2,$$

where $k < n$ and T_k has at least four leaves, or $k = n$ and T_k has exactly four leaves. Then for any tree T_n on n vertices with at least 4 leaves we have

$$\hom(T_m, T_n) \geqslant (n-2)2^{m-1} + 2.$$

In the proof of this lemma we will subsequently use the following very simple fact.

Fact. If G is a graph and G_1, G_2 are induced subgraphs of G with possible intersection, then for any graph H we have

$$\hom(H, G) \geqslant \hom(H, G_1) + \hom(H, G_2) - \hom(H, G_1 \cap G_2).$$

Proof of the lemma. We can assume that $m \geqslant 2$. Assume that T_n is a tree with at least 5 leaves. Otherwise we have nothing to prove.

Let us call a path maximal in T_n if it connects leaves. If a maximal path contains k vertices of degree at least 3, then we say that the maximal path has k branches.

Case 1: T_n contains a maximal path with at least 3 branches. Let v_0Pv_r be a maximal path with vertices $u_1, \dots, u_k$ having degree at least 3. Let $B_1, \dots, B_k$ be the branches which we get if we delete all vertices and edges of the path v_0Pv_r except $u_1, \dots, u_k$. So B_i is a rooted tree with root u_i. Let u_2^- and u_2^+ be the two neighbors of u_2 on the path v_0Pv_r. Let $T^{(2)}$ be the tree induced by the vertices $V(B_2) \cup \{u_2^-, u_2^+\}$. Let $|V(T^{(2)})| = t$. We distinguish two cases.

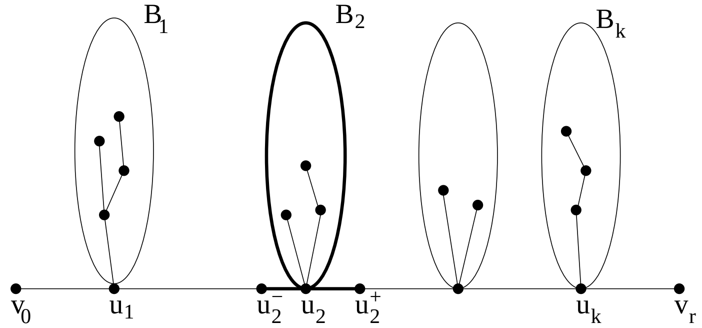

Figure 8: A path with branches.

Subcase 1.1: $\hom(T_m, T^{(2)}) < (t-2)2^{m-1} + 2$. Consider the following trees G_1 and G_2. G_1 is the tree spanned by the vertices $v_0 P u_2^+$ and the branches B_1, B_2. G_2 is the tree spanned by the vertices $u_2^- P v_r$ and the branches $B_2, \ldots, B_k$. Note that $G_1 \cup G_2 = T_n$, $G_1 \cap G_2 = T^{(2)}$ and G_1, G_2 contains at least 4 leaves, because $k \geqslant 3$. By the hypothesis of the lemma we have

$$\hom(T_m, G_i) \geqslant (|V(G_i)| - 2)2^{m-1} + 2$$

for $i = 1, 2$. Hence

$$\hom(T_m, T_n) \geqslant \hom(T_m, G_1) + \hom(T_m, G_2) - \hom(T_m, G_1 \cap G_2) \geqslant$$

$$\geqslant (|V(G_1)| - 2)2^{m-1} + 2 + (|V(G_2)| - 2)2^{m-1} + 2 - ((|V(G_1 \cap G_2)| - 2)2^{m-1} + 2) =$$

$$= (((|V(G_1 \cup G_2)| - 2)2^{m-1} + 2 = (n-2)2^{m-1} + 2.$$

In this case we are done.

Case 1.2: $\hom(T_m, T^{(2)}) \geqslant (t-2)2^{m-1} + 2$. Consider the following trees G_1 and G_2. G_1 is the tree spanned by the vertices $(V(T_n) \setminus V(T^{(2)})) \cup \{u_2^-, u_2, u_2^+\}$. G_2 is simply $T^{(2)}$. Note that $G_1 \cup G_2 = T_n$, $G_1 \cap G_2 = \{u_2^-, u_2, u_2^+\} = P_3$ and G_1 contains at least 4 leaves. By the hypothesis of the lemma we have

$$\hom(T_m, G_1) \geqslant (|V(G_1)| - 2)2^{m-1} + 2.$$

We also know that in this case

$$\hom(T_m, G_2) \geqslant (|V(G_2)| - 2)2^{m-1} + 2.$$

Note that

$$\hom(T_m, P_3) \leqslant \hom(S_m, P_3) = 2^{m-1} + 2.$$

Then

$$\hom(T_m, T_n) \geqslant \hom(T_m, G_1) + \hom(T_m, G_2) - \hom(T_m, G_1 \cap G_2) \geqslant$$

$$\geqslant (|V(G_1)| - 2)2^{m-1} + 2 + (|V(G_2)| - 2)2^{m-1} + 2 - ((|V(G_1 \cap G_2)| - 2)2^{m-1} + 2) =$$

$$= (((|V(G_1 \cup G_2)| - 2)2^{m-1} + 2 = (n-2)2^{m-1} + 2.$$

In this case we are done too.

Case 2: All maximal paths of T_n have at most 2 branches. In the following we show that they have quite simple structure: they are starlike or double starlike trees, see Figure 9.

Let v_1 be a vertex of T_n of degree at least 3. Let us decompose T_n to the branches $B_1', B_2', \ldots B_k'$ at v_1. So v_1 is a leaf in the trees $B_1', B_2', \ldots B_k'$. We show that all except at most one of $B_1', B_2', \ldots, B_k'$ are paths. Assume that, for instance, B_1', B_2' are not paths. Then they contains at least two leaves of T_n: B_1' contains u_1, u_2, B_2' contains u_3, u_4. Then the maximal path $u_1 P u_3$ has at least three branches: one-one inside the branches B_1' and B_2' and B_3' at the vertex v. If all branches are paths, then we are done: T_n is starlike. If

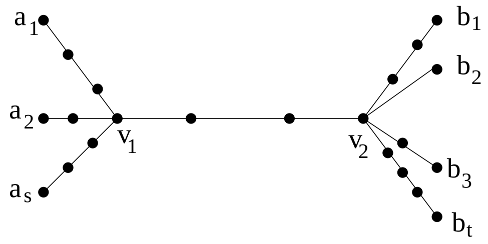

Figure 9: A double starlike tree.

one of them is not path, say B_1', then let us consider the vertex $v_2 \in V(B_1')$ having degree at least 3 which is closest to v_1. Repeating the previous argument to v_2 instead of v_1, all except one branches at v_2 must be path and we also know that the branch containing v_1 is not path. Hence the tree is double starlike, where the middle path is v_1Pv_2.

We can consider a starlike tree as a double starlike tree, where $v_1 = v_2$. Let $a_1, \dots, a_s$ and $b_1, \dots, b_t$ be the leaves of T_n, where $a_1, \dots, a_s$ are closer to v_1 than to v_2, while $b_1, \dots, b_t$ are closer to v_2 than to v_1. If $v_1 = v_2$ we just decompose the set of leaves into two sets of (almost) equal size. Note that $s, t \geqslant 2$. Since we can assume that there are at least 5 leaves, we assume that $s + t \geqslant 5$.

If $u_1, \dots, u_\ell$ are some vertices of a tree, then we say that the tree spanned by $u_1, \dots, u_\ell$ is the smallest subtree which contains the vertices $u_1, \dots, u_\ell$. It is

$$span(u_1, \dots, u_\ell) = \cup_{1 \leqslant i,j \leqslant \ell} u_i P u_j.$$

Subcase 2.1: If $s \geqslant 3$ and $t \geqslant 3$ both hold, then let G_1 be the tree spanned by the vertices a_1, a_2, b_1, b_2, and let G_2 be the tree spanned by the vertices $a_2, \dots a_s, b_2, \dots, b_t$. Then $G_1 \cup G_2 = T_n$, $G_1 \cap G_2 = a_2Pb_2$ and both G_1, G_2 have at least 4 leaves. Since

$$\hom(T_m, G_1 \cap G_2) \leqslant \hom(S_m, G_1 \cap G_2) =$$

$$= \hom(S_m, a_2Pb_2) = (|V(G_1 \cap G_2)| - 2)2^{m-1} + 2,$$

we have

$$\hom(T_m, T_n) \geqslant \hom(T_m, G_1) + \hom(T_m, G_2) - \hom(T_m, G_1 \cap G_2) \geqslant$$

$$\geqslant (|V(G_1)| - 2)2^{m-1} + 2 + (|V(G_2)| - 2)2^{m-1} + 2 - ((|V(G_1 \cap G_2)| - 2)2^{m-1} + 2) =$$
$$= (|V(G_1 \cup G_2)| - 2)2^{m-1} + 2 = (n-2)2^{m-1} + 2.$$

Hence we are done in this case.

Subcase 2.2: If $s \geqslant 4$ and $t = 2$ then let G_1 be the tree spanned by a_1, a_2, b_1, b_2 and let G_2 be the tree spanned by $a_2, \dots, a_k, b_1$. Then $G_1 \cup G_2 = T_n$, $G_1 \cap G_2 = a_2Pb_1$ and both G_1, G_2 have at least 4 leaves. In this case we are done as before. Clearly, the case $s \geqslant 2$ and $t \geqslant 4$ is completely similar.

Subcase 2.3: The last case is $s = 3, t = 2$ (and $s = 2, t = 3$). Let $G_1 = span(a_1, a_2, b_1, b_2)$, $G_2 = span(a_2, a_3, b_1, b_2)$, $G_3 = span(a_2, b_1, b_2)$, $G_4 = (a_1, a_2, a_3, v_2)$. Then $G_1 \cap G_2 = G_3$, $G_3 \cap G_4 = a_2 P v_2$. Note that G_1, G_2, G_4 has 4 leaves, thus

$$\hom(T_m, G_i) \geqslant (|V(G_i)| - 2)2^{m-1} + 2$$

for $i = 1, 2, 4$. If $\hom(T_m, G_3) \leqslant (|V(G_3)| - 2)2^{m-1} + 2$, then from $T_n = G_1 \cup G_2$, $G_3 = G_1 \cap G_2$ we obtain that $\hom(T_m, T_n) \geqslant (n-2)2^{m-1} + 2$. If $\hom(T_m, G_3) \geqslant (|V(G_3)| - 2)2^{m-1} + 2$, then from $T_n = G_3 \cap G_4$ we obtain that $\hom(T_m, T_n) \geqslant (n-2)2^{m-1} + 2$. Hence we are done in this case as well. □

Proof of Theorem 5.2. The result immediately follows from Lemma 5.4 and Proposition 5.5 below. □

Proposition 5.5. *Let T_n be a tree on n vertices with exactly four leaves. Then for any tree T_m on m vertices we have*

$$\hom(T_m, T_n) \geqslant (n-2)2^{m-1} + 2,$$

where n is the number of vertices of T_n.

Proof. If T_n has a vertex of degree 4 then by Theorem 4.6 we have

$$\hom(T_m, T_n) \geqslant 2(n-1)C^{m-2},$$

where

$$C = \left(\prod_{i=1}^{n} d_i^{d_i}\right)^{1/2e(T_n)} = 2.$$

Hence

$$\hom(T_m, T_n) \geqslant (n-1)2^{m-1} \geqslant (n-2)2^{m-1} + 2.$$

For the case when T_n has two vertices of degree 3, we need more preparation.

Lemma 5.6. *Let T_n be a tree with exactly 4 leaves and two vertices of degree 3. Let x and y be the vertices of T_n with degree 3. Assume that there are at most 3 vertices of T_n which have degree 2 and not on the path xPy. Then for any tree T_m on m vertices we have*

$$\hom(T_m, T_n) \geqslant (n-2)2^{m-1} + 2,$$

where n is the number of vertices of T_n.

Proof. We can assume that $m \geqslant 4$, otherwise the statement is trivial. We prove the slightly stronger inequality

$$\hom(T_m, T_n) > \left(n - 2 + \frac{1}{8}\right) 2^{m-1}.$$

If $m \geqslant 4$, then this implies that

$$\hom(T_m, T_n) > (n-2)2^{m-1} + 1$$

or equivalently,

$$\hom(T_m, T_n) \geqslant (n-2)2^{m-1} + 2.$$

To prove this statement we use Theorem 4.1 with a suitable Markov chain. Let $p_{ij} = \frac{1}{2}$ for $(i,j) \in E(T_n)$ if i has degree 2. Naturally, $p_{ij} = 1$ if $(i,j) \in E(G)$ and i is a leaf. Finally, if $i \in \{x,y\}$, $j \in xPy$ then $p_{ij} = \frac{1}{2}$ and if $i \in \{x,y\}$, $j \notin xPy$ then $p_{ij} = \frac{1}{4}$.

Let r be the number of vertices of xPy. Then $n = r + t + 4$ for an integer $t \geqslant 0$. Let $N = 4r + 2t + 4$. Then the stationary distribution is the following: $q_i = \frac{4}{N}$ if $i \in xPy$, $q_i = \frac{2}{N}$ if $i \notin xPy$, but has degree 2 and finally, $q_i = \frac{1}{N}$ if i is a leaf.

Then

$$H(P|Q) = \frac{N-12}{N}\log 2 + \frac{8}{N}\left(\frac{1}{2}\log 2 + \frac{1}{2}\log 4\right) = \log 2.$$

On the other hand,

$$\begin{aligned}
&H(Q) + 2(H(D|Q) - H(P|Q)) \\
&= \left(\frac{4r}{N}\log\frac{N}{4} + \frac{2t}{N}\log\frac{N}{2} + \frac{4}{N}\log\frac{N}{1}\right) + 2\frac{8}{N}\left(\log 3 - \frac{3}{2}\log 2\right) \\
&= \log\frac{N}{4} + \frac{2t}{N}\log 2 + \frac{16(\log 3 - \log 2)}{N}.
\end{aligned}$$

Note that

$$\begin{aligned}
\log\left(n - 2 + \frac{1}{8}\right) - \log\frac{N}{4} &\leqslant \int_{N/4}^{n-2}\frac{1}{x}dx \\
&\leqslant \frac{n - 2 + \frac{1}{8} - N/4}{N/4} \\
&= \frac{4}{N}\left(\frac{t}{2} + 1 + \frac{1}{8}\right) \\
&= \frac{1}{N}\left(2t + \frac{9}{2}\right).
\end{aligned}$$

Hence if

$$\frac{1}{N}\left(2t + \frac{9}{2}\right) \leqslant \frac{2t}{N}\log 2 + \frac{16(\log 3 - \log 2)}{N},$$

then

$$\log\left(n - 2 + \frac{1}{8}\right) \leqslant H(Q) + 2(H(D|Q) - H(P|Q)),$$

consequently

$$\hom(T_k, G) \geqslant \exp(H(Q) + 2H(D|Q) + (m-3)H(P|Q)) > \left(n - 2 + \frac{1}{8}\right)2^{m-1}.$$

The above inequality is satisfied if

$$t \leqslant \frac{8\log\frac{3}{2} - \frac{9}{4}}{1 - \log 2} \approx 3.238.$$

This proves the statement of the theorem. $\square$

Lemma 5.7. *Let T_n be a tree obtained from a path on $n-8$ vertices by gluing one-one P_5 at the middle vertices to both ends of the path P_{n-8} (see Fig. 10). Then for any tree T_m on m vertices we have*

$$\hom(T_m, T_n) \geqslant (n-1)2^{m-1}.$$

Proof. We will show by induction on m that

$$\hom(T_m, T_n) \geqslant (n-1)2^{m-1}. \tag{5.3}$$

Let v be any leaf of T_m with unique neighbor u and let $T_{m-1} = T_m - v$ be a rooted tree with root u.

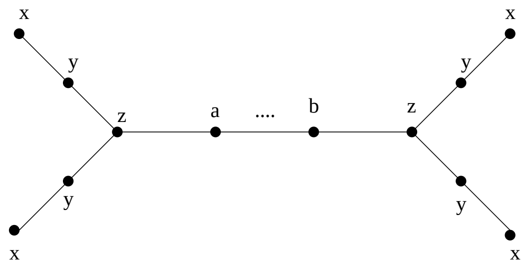

Figure 10: Special double starlike trees.

Let us use the hom-vectors of the Fig. 10, that is

$$\mathbf{h}(T_{m-1}, u, T_n) = (x, x, y, y, z, a, \ldots, b, z, y, y, x, x).$$

Now suppose that

$$\hom(T_{m-1}, T_n) \geqslant (n-1)2^{m-2}.$$

It is easy to see by induction that $z > 2x$ if T_{m-1} has at least two vertices. By tree-walk algorithm, we have

$$\begin{aligned}\hom(T_m(v), T_n) &= 4x + 8y + 6z + 2(a + \cdots + b)\\ &\geqslant 8x + 8y + 4z + 2(a + \cdots + b)\\ &= 2\hom(T_{m-1}, G)\\ &\geqslant (n-1)2^{m-1},\end{aligned}$$

which shows (5.3). $\square$

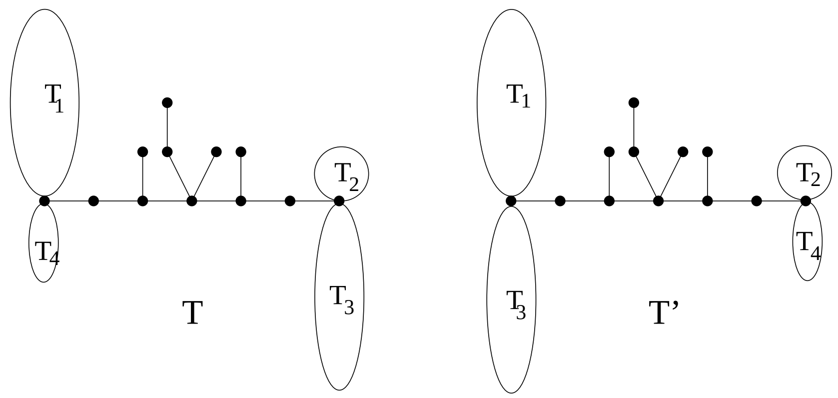

Figure 11: LS-switch.

Next we introduce a transformation which we will call LS-switch (Large-Small switch).

Definition 5.8 (LS-switch)**.** Let $R(u, v)$ be a tree with specified vertices u and v such that the distance of u and v is even and R has an automorphism of order 2 which exchanges the vertices u and v. Let $T_1(x), T_2(x), T_3(y), T_4(y)$ be rooted trees such that $T_2(x)$ is the rooted subtree of $T_1(x)$ and $T_4(y)$ is the rooted subtree of $T_3(y)$. Let the tree T be obtained from the trees $R(u, v), T_1(x), T_2(x), T_3(y), T_4(y)$ by attaching a copy of $T_1(x), T_4(y)$ to $R(u, v)$ at vertex u and a copy of $T_2(x), T_3(y)$ at vertex v. Assume that the tree T' is obtained from the trees $R(u, v), T_1(x), T_2(x), T_3(y), T_4(y)$ by attaching a copy of $T_1(x), T_3(y)$ to $R(u, v)$ at vertex u and a copy of $T_2(x), T_4(y)$ at vertex v. Then T' is the LS-switch of T. Observe that there is a natural bijection between the color classes of T' and T.

A particular case of the LS-switch is when $R(u, v)$ is a path of even length with end vertices u and v, $T_2(x)$ and $T_4(y)$ are one-vertex rooted trees, then T' is obtained from T by an even-KC-transformation, i.e., KC-transformation according to a path of even length. Another useful special case is when $R(u, v)$ is a tree where we attach an arbitrary tree to the middle vertex of the path on 3 vertices and u and v are the end vertices of the path (in this case the automorphism simply switches u and v), and $T_2(x), T_4(y)$ are the rooted trees with 1 vertex, in this case we get back to a particular case of the original *Kelmans-transformation* [12].

The following theorem with respect to the LS-switch is just an extension of the even case of Theorem 1.3.

Theorem 5.9. *Let T' be the LS-switch of T. Let H be an arbitrary tree. Then*

$$\hom(H, T) \leqslant \hom(H, T').$$

Proof. (Sketch.) The unique shortest path connecting all T_i's in T (or T') will be denoted by P_{2k}, a path of even length with vertices labeled consecutively by $0, 1, \ldots, 2k$. Without loss of generality, we can assume that $0 \in V(T_1)$. For $1 \leqslant j \leqslant 2k-1$, let A_j denote the component of T that contains the vertex j when we delete all edges of P_{2k}. By the definition of LS-switch, the subtrees A_j and A_{2k-j} are isomorphic, so we can identify

$V(A_j)\setminus\{j\}$ with $V(A_{2k-j})\setminus\{2k-j\}$. We will also consider $V(T_2)\setminus\{0,2k\}$ as the subset of $V(T_1)\setminus\{0,2k\}$ and $V(T_4)\setminus\{0,2k\}$ as the subset of $V(T_3)\setminus\{0,2k\}$.

Let v be a vertex of H. For $0\leqslant s\leqslant 2k$, $u\in V(T_i)\setminus\{0,2k\}$ $(1\leqslant i\leqslant 4)$ and $a\in V(A_j)\setminus\{j\}$ $(1\leqslant j\leqslant 2k-1)$, we define

$$p_s := |\{\{f\in \mathrm{Hom}(H,T): f(v)=m\}|,\ p'_s := |\{\{f\in \mathrm{Hom}(H,T'): f(v)=m\}|,$$

$$t_i(u) := |\{f\in \mathrm{Hom}(H,T): f(v)=u\}|,\ t'_i(u) := |\{f\in \mathrm{Hom}(H,T'): f(v)=u\}|,$$

and

$$p_j(a) := |\{f\in \mathrm{Hom}(H,T): f(v)=a\}|,\ p'_j(a) := |\{f\in \mathrm{Hom}(H,T'): f(v)=a\}|.$$

We prove by induction that the following inequalities are preserved by the steps of the tree-walk algorithm. For any $0\leqslant s\leqslant k$, $a\in V(A_j)\setminus\{j\}$ $(1\leqslant j\leqslant k)$, $u\in V(T_2)\setminus\{0,2k\}$, $w\in V(T_4)\setminus\{0,2k\}$, $x\in V(T_1)\setminus V(T_2)$ and $y\in V(T_3)\setminus V(T_4)$ we have

$$p'_{k-s}+p'_{k+s}\geqslant p_{k-s}+p_{k+s} \text{ and } p'_{k-s}\geqslant p_{k+s},p_{k-s} \tag{5.4}$$
$$p'_j(a)+p'_{2k-j}(a)\geqslant p_j(a)+p_{2k-j}(a) \text{ and } p'_j(a)\geqslant p_{2k-j}(a),p_j(a) \tag{5.5}$$
$$t'_1(u)+t'_2(u)\geqslant t_1(u)+t_2(u) \text{ and } t'_1(u)\geqslant t_1(u),t_2(u) \tag{5.6}$$
$$t'_3(w)+t'_4(w)\geqslant t_3(w)+t_4(w) \text{ and } t'_3(w)\geqslant t_3(w),t_4(w) \tag{5.7}$$
$$t'_1(x)\geqslant t_1(x) \text{ and } t'_3(y)\geqslant t_3(y). \tag{5.8}$$

We only need to check that the two operations in the tree-walk algorithm preserve all the above inequalities, which is routine and left to the reader. ☐

Now assume that T_n has two vertices, x and y, of degree 3. Among these trees (n vertices, 4 leaves, two vertices of degree 3) let us choose $\overline{T_n}$ to be the one for which $\hom(T_m,\overline{T_n})$ is minimal and among these trees the length of the path is maximal.

Let the four leaves of $\overline{T_n}$ denoted by z_1,z_2,z_3,z_4 such that z_1,z_2 are closer to x than y, and z_3,z_4 are closer to y than x. Let the number of edges of $xPy, xPz_1, xPz_2, yPz_3, yPz_4$ be a,b,c,d,e, respectively. We show that $\max(b,c,d,e)\leqslant 2$. Indeed, if say $b>2$ then $\overline{T_n}$ can be obtained by an LS-switch from a graph T_n^* as follows.

If b is even, then let u be the unique vertex such that $d(z_1,u)=2$. Then $uPx=R(u,x)$ is a path of even length. Let $T_2=z_1Pu$. Furthermore, let T_1 be the tree spanned by the vertices x,z_3,z_4. $T_3=xPz_2$ and $T_4=\{u\}$. Then T_2 is a rooted subtree of T_1 and T_4 is a rooted subtree of T_3. Now making an inverse LS-swith we obtain T_n^*. By Theorem 5.9, we know that

$$\hom(T_m,\overline{T_n})\geqslant \hom(T_m,T_n^*)$$

and in T_n^*, the vertices of degree 3, y and u, has distance $a+b-2>a$ contradicting the choice of $\overline{T_n}$.

If b is odd, then let u be the unique neighbor of z_1 and we repeat the previous argument. The distance of u and x is even again.

Hence we can assume that $\max(b,c,d,e) \leqslant 2$. If not all of them are 2, then we can use Lemma 5.6 to get that

$$\hom(T_m, \overline{T_n}) \geqslant (n-2)2^{m-1} + 2.$$

If $b = c = d = e = 2$, then we use Lemma 5.7 to obtain that

$$\hom(T_m, \overline{T_n}) \geqslant (n-1)2^{m-1} \geqslant (n-2)2^{m-1} + 2.$$

This completes the proof of Proposition 5.5. □

5.3 Trees with 3 leaves

Lemma 5.10. *(a) Let $n = a+b+c+1$, and $\min(a,b,c) \geqslant 2$. Then for any tree T_m on m vertices we have*

$$\hom(T_m, Y_{a,b,c}) \geqslant (n-2)2^{m-1} + 2.$$

(b) Let $n = a+b+2$, and $\min(a,b) \geqslant 3$. Then for any tree T_m on m vertices we have

$$\hom(T_m, Y_{a,b,1}) \geqslant (n-2)2^{m-1} + 2.$$

Proof. (a)

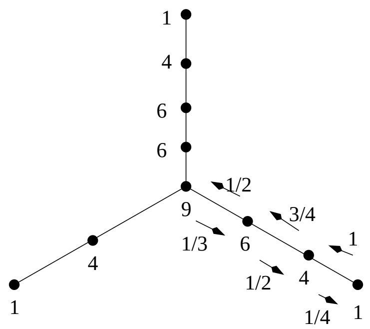

Figure 12: $Y_{a,b,c}$ where $\min(a,b,c) \geqslant 2$ with a special Markov chain.

We can think to $Y_{a,b,c}$ with $\min(a,b,c) \geqslant 2$ as follows: we consider $Y_{2,2,2}$ and we subdivide the edges between the vertex of degree 3 and its neighbors a few times. Let us write the weights $1,4,1,4,1,4,9$ to the vertices of $Y_{2,2,2}$ according to the figure and then let us write weights 6 on the new vertices obtained by subdivision. It is easy to check that there is a unique Markov chain on $Y_{a,b,c}$, where the stationary distribution is proportional to the weights. (In fact, we write a few transition probabilities on the figure.)

It is easy to check that $H(P|Q) = \log 2$ and if $N = 24 + 6(n-7) = 6(n-3)$, then

$$H(Q) + 2(H(D|Q) - H(P|Q))$$

$$= \left(\frac{9}{N}\log\frac{N}{9} + \frac{12}{N}\log\frac{N}{4} + \frac{3}{N}\log\frac{N}{1} + \frac{6(n-7)}{N}\log\frac{N}{6}\right) +$$
$$+ 2\cdot\frac{12}{N}\left(\log 2 - \left(\frac{1}{4}\log 4 + \frac{3}{4}\log\frac{4}{3}\right)\right)$$
$$= \log\frac{N}{6} + \frac{9}{N}\log\frac{6}{9} + \frac{12}{N}\log\frac{6}{4} + \frac{3}{N}\log\frac{6}{1} + \frac{24}{N}\left(\log 2 - \left(\frac{1}{4}\log 4 + \frac{3}{4}\log\frac{4}{3}\right)\right)$$
$$= \log(n-3) + \frac{24}{N}\log\frac{3}{2}.$$

Since

$$\log(n-2+\varepsilon) - \log(n-3) = \int_{n-3}^{n-2+\varepsilon}\frac{dx}{x} \leqslant \frac{1+\varepsilon}{n-3} = \frac{6}{N}(1+\varepsilon)$$

we can choose $\varepsilon = 4\log\frac{3}{2} - 1 > \frac{1}{2}$ to deduce that

$$\hom(T_m, Y_{a,b,c}) \geqslant (n-2+\varepsilon)2^{m-1}.$$

This is already greater than $(n-2)2^{m-1}+2$ for $m \geqslant 3$. The statement is trivial for $m \leqslant 2$.

(b)

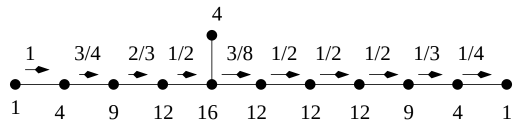

Figure 13: $Y_{a,b,1}$ where $\min(a,b) \geqslant 3$ with a special Markov chain.

We use completely the same argument as in part (a). We think to $Y_{a,b,1}$ as a subdivision of $Y_{3,3,1}$ and use the Markov chain on the figure. Again we have $H(P|Q) = \log 2$ and the sum of the weights is $N = 48 + 12(n-8) = 12(n-4)$. Hence

$$H(Q) + 2(H(D|Q) - H(P|Q)) = \log\frac{N}{12} + \frac{12}{N}\left(\frac{11}{3}\log 3 - \frac{1}{3}\log 2\right).$$

Since

$$\log(n-2+\varepsilon) - \log(n-4) = \int_{n-4}^{n-2+\varepsilon}\frac{dx}{x} \leqslant \frac{2+\varepsilon}{n-4} = \frac{12}{N}(2+\varepsilon)$$

we can choose $\varepsilon = \frac{11}{3}\log 3 - \frac{1}{3}\log 2 - 2 > 1.79$ to deduce that

$$\hom(T_m, Y_{a,b,c}) \geqslant (n-2+\varepsilon)2^{m-1}.$$

This is already greater than $(n-2)2^{m-1}+2$ for $m \geqslant 2$. The statement is trivial for $m = 1$. □

Remark 5.11. Since every Markov chains is reversible on a tree, there is a natural way to define a new Markov chain on a subdivided edge. Assume that the probabilities of the stationary distribution were q_i, q_j and p_{ij}, p_{ji} were the transition probabilities at the vertices i, j. Then $q_i p_{ij} = q_j p_{ji}$ (reversibility) and we can put a vertex r with weight $2q_i p_{ij}$ and $p_{ri} = p_{rj} = 1/2$ on the edge (i, j). Then the new stationary distribution will be proportional to the weights $\{q_i \mid i \in V(T)\} \cup \{2q_i p_{ij}\}$.

Theorem 1.6. Let T_n be a tree on n vertices. Assume that for a tree T_m we have

$$\hom(T_m, T_n) < \hom(T_m, P_n).$$

Then $T_n = Y_{1,1,n-3}$ and n is even.

Proof. Note that if T_n has at least 4 leaves then Theorem 5.5 implies that

$$\hom(T_m, T_n) \geqslant (n-2)2^{m-1} + 2 = \hom(S_m, P_n) \geqslant \hom(T_m, P_n)$$

contradicting to the condition of the theorem. Hence $T_n = Y_{a,b,c}$ for some a, b, c. Observe that if one of a, b, c is even then $Y_{a,b,c}$ can be obtained from P_n by an even-KC-transformation and then Theorem 1.3 implies

$$\hom(T_m, T_n) \geqslant \hom(T_m, P_n)$$

contradicting to the condition of the theorem. Note that if n is odd, then one of a, b, c is necessarily even and so we are done. From Lemma 5.10 we also know that $\min(a, b, c) = 1$, say $c = 1$ and $\min(a, b) \leqslant 2$. But then $\min(a, b) = 1$, because it must be odd. Hence $T_n = Y_{1,1,n-3}$ and n is even. □

Remark 5.12. There is a tree T_m for which $\hom(T_m, S_4) < \hom(T_m, P_4)$. On Fig. 14 one

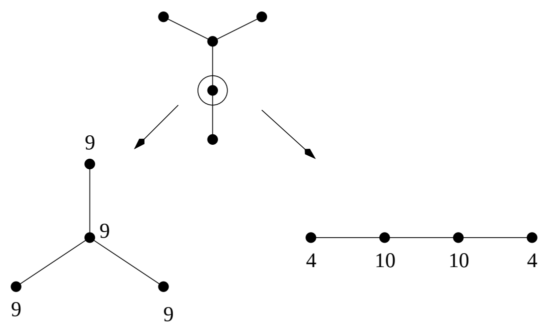

Figure 14: An example.

can see a rooted tree and its homomorphism vectors to S_4 and P_4. Now if we attach k copies of this rooted tree at the root then for the obtained tree T_m we have

$$\hom(T_m, P_4) = 2 \cdot 4^k + 2 \cdot 10^k > 4 \cdot 9^k = \hom(T_m, S_4)$$

for large enough k.

On the other hand, it seems that $\hom(T_m, Y_{1,1,n-3}) \geqslant \hom(T_m, P_n)$ if $n \geqslant 6$ and even.

6 Open problems

We collected a few open problems and conjectures in this section.

We first recall a conjecture from the Introduction, namely that there is no exceptional case in Theorem 1.6 if $n \geqslant 5$.

Conjecture 1.7 Let T_n be a tree on n vertices, where $n \geqslant 5$. Then for any tree T_m we have

$$\hom(T_m, P_n) \leqslant \hom(T_m, T_n).$$

Note that to prove Conjecture 1.7, one only needs to prove that for any tree T_m we have

$$\hom(T_m, P_n) \leqslant \hom(T_m, Y_{1,1,n-3})$$

for $n \geqslant 6$, where n is even.

There is also an open problem in Figure 4, if true, would provide an alternative proof of the first part of Theorem 1.8 (through Theorem 1.2).

Problem 6.1. Is it true that

$$\hom(P_n, T_n) \leqslant \hom(T_n, T_n)$$

for every tree T_n on n vertices?

We believe that the answer is affirmative for this question. This question naturally leads to the following problem.

Problem 6.2. Characterize all graphs G for which

$$\hom(P_m, G) \leqslant \hom(T_m, G)$$

for all m and all trees T_m on m vertices.

Note that if G is d-regular, then $\hom(P_m, G) = \hom(T_m, G) = |V(G)|d^{m-1}$. We have also seen that the inequality of Problem 6.2 is satisfied if $G = P_n$ or S_n. Probably, it is hard to characterize these graphs. Maybe, it is easier to describe those graphs G for which the inequality of Problem 6.2 is satisfied for large enough m.

The dual of Problem 6.2 is also natural:

Problem 6.3. Characterize all trees T_m on m vertices for which

$$\hom(P_m, G) \leqslant \hom(T_m, G)$$

for all graph G.

Probably, this is an easier problem than Problem 6.2. Note that already Sidorenko [19] achieved nice results on this problem. Still the problem is far from being solved.

In light of the tree-walk algorithm, it would be interesting to develop an algorithm for computing the number of homomorphisms from bipartite graphs to any graph.

Acknowledgments

The first author is very grateful to László Lovász and Miklós Simonovits for suggesting to study the papers of Alexander Sidorenko. He also thanks Benjamin Rossman for the comments on the paper [21]. He is also very grateful to Balázs Gerencsér for the useful discussion on Markov chains and entropies.

This research was partly done while the second author was visiting the Alfréd Rényi Institute of Mathematics. He is very grateful to Miklós Abért and the Institute for their hospitality and support. He also would like to thank Yanfeng Luo for introducing the graph homomorphisms and Jiang Zeng for all his encouragement during this work.

Special thanks goes to Masao Ishikawa for suggesting the name tree-walk algorithm.

References

[1] Sr. Arworn and P. Wojtylak, An algorithm for the number of path homomorphisms, Discrete Math., **309** (2009), 5569–5573.

[2] B. Bollobás, *Modern graph theory*, Graduate Texts in Mathematics, 184. Springer-Verlag, New York, 1998.

[3] B. Bollobás and M. Tyomkyn, Walks and paths in trees, J. Graph Theory, **70** (2012), 54–66.

[4] J.A. Bondy and U.S.R. Murty, *Graph Theory*, Springer-Verlag, New York, 2008.

[5] T.M. Cover and J.A. Thomas, *Elements of information theory*, 2nd ed., John Wiley & Sons, New Jersey, 2006.

[6] P. Csikvári, On a poset of trees, Combinatorica, **30** (2010), 125–137.

[7] P. Csikvári, On a poset of trees II, J. Graph Theory, **74** (2013), 81–103.

[8] P. Csikvári and Z. Lin, Graph homomorphisms between trees, II: Path coloring trees, preprint, 2014.

[9] D. Dellamonica Jr., P. Hacell, T. Łuczak, D. Mubayi, B. Nagle, Y. Person, V. Rödl, and M. Schacht, Tree-minimal graphs are almost regular, J. Comb., **3** (2012), 49–62.

[10] M.A. Fiol and E. Garriga, Number of walks and degree powers in a graph, Discrete Math., **309** (2009), 2613–2614.

[11] C. Jagger, P. Stovicek and A. Thomason, Multiplicities of subgraphs, Combinatorica, **16** (1996), 123–141.

[12] A.K. Kelmans, On graphs with randomly deleted edges, Acta. Math. Acad. Sci. Hung., **37** (1981), 77–88.

[13] S. Kopparty and B. Rossman, The homomorphism domination exponent, European J. Combin., **32** (2011), 1097–1114.

[14] A.M. Leontovich, The number of mappings of graphs, the ordering of graphs and the Muirhead theorem, Problemy Peredachi Informatsii 25 (1989), no. 2, 91–104; translation in Problems Inform. Transmission 25 (1989) 154-165.

[15] J.L.X. Li and B. Szegedy, On the logarithmic calculus and Sidorenko's conjecture, Combinatorica, to appear.

[16] Z. Lin and J. Zeng, On the number of congruence classes of paths, Discrete Math., **312** (2012), 1300–1307.

[17] L. Lovász and J. Pelikán, On the eigenvalues of trees, Per. Mathematica Hungarica, **3** (1973), 175–182.

[18] A. Sidorenko, A correlation inequality for bipartite graph, Graph Combin., **9** (1993), 201–204.

[19] A. Sidorenko, A partially ordered set of functionals corresponding to graphs, Discrete Math., **131** (1994), 263–277.

[20] R.P. Stanley, *Enumerative combinatorics*, Vol. 1, Cambridge University Press, Cambridge, 1997.

[21] B. Rossman and E. Vee, Counting homomorphisms from trees and cycles, 2006, unpublished manuscript.