Almost every tree with m edges decomposes $K_{2m,2m}$

M. Drmota* and A. Lladó †

Institute of Discrete Mathematics and Geometry
Vienna University of Technology, Vienna, Austria

Dept. Matemàtica Aplicada 4
Universitat Politècnica de Catalunya-BarcelonaTech, Barcelona, Spain

Abstract

We show that asymptotically almost surely a tree with m edges decomposes the complete bipartite graph $K_{2m,2m}$, a result connected to a conjecture of Graham and Häggkvist. The result also implies that asymptotically almost surely a tree with m edges decomposes the complete graph with $O(m^2)$ edges. An ingredient of the proof consists in showing that the bipartition classes of the base tree of a random tree have roughly equal size.

MSC2010: 05C51, 05C80

1 Introduction

Given two graphs H and G we say that H decomposes G if G is the edge–disjoint union of isomorphic copies of H. The following is a well–known conjecture of Ringel.

Conjecture 1 (Ringel [17]) *Every tree with m edges decomposes the complete graph K_{2m+1}.*

The conjecture has been verified by a number of particular classes of trees, see the dynamic survey of Gallian [9]. By using the polynomial method, the conjecture was verified by Kézdy [12] for the more general class of so–called *stunted* trees. As mentioned by the author, this class is still small among the set of all trees.

The following bipartite version of the conjecture was formulated by Graham and Häggkvist.

*E-mail address: `michael.drmota@tuwien.ac.at`
†E-mail address: `allado@ma4.upc.edu`

Conjecture 2 (Graham and Häggkvist [10]) *Every tree with m edges decomposes the complete bipartite graph $K_{m,m}$.*

Again the conjecture has been verified by a number of cases; see e.g. [14]. Approximate versions of the two conjectures have been also proved. Häggkvist [10] proved that every tree with m edges and at least $(m+1)/2$ leaves decomposes $K_{2m,2m}$. The authors of [15] proved that every tree with m edges is a subtree of a tree with $m' < 2m$ edges which decomposes $K_{m',m'}$ and $K_{2m'+1}$. It is also known that every tree with m edges and radius r decomposes $K_{m'}$ and $K_{m'',m''}$ where $m' \leq 32(2r+4)m^2+1$ (Kézdy and Snevily [13]) and $m'' \leq rm$ [14]. However, to our knowledge, there are no results stating that every tree with m edges decomposes K_{cm+1} or $K_{cm,cm}$ for some absolute constant c of reasonable size. The purpose of this paper is to show such a result for almost all trees.

Let $\mathcal{T}$ denote the class of (unlabelled) trees and let $\mathcal{T}_m$ be the class of trees with m edges. By a random tree with m edges we mean a tree chosen from $\mathcal{T}_m$ with the uniform distribution. We say that a random tree satisfies a property $\mathcal{P}$ asymptotically almost surely (a.a.s) if the probability that a random tree with m edges satisfies $\mathcal{P}$ tends to one with $m \to \infty$. Our main result is the following theorem.

Theorem 1 *Asymptotically almost surely a tree with m edges decomposes $K_{2m,2m}$.*

Robinson and Schwenk [18] proved that the average number of leaves in an (unlabelled) random tree with m edges is asymptotically cm with $c \approx 0.438$. Drmota and Gittenberger [6] showed that the distribution of the number of leaves in a random tree with m edges is asymptotically normal with variance $c_2 m$ for some positive constant c_2. Thus, asymptotically almost surely a random tree with m edges has more than $2m/5$ leaves. When $m = p$ is a prime, it was proved in [3] that a tree with at least $p/3$ leaves decomposes $K_{2p,2p}$, thus providing a proof of Theorem 1 for primes. The primality of the number of edges is related to an application of Alon's Combinatorial Nullstellensatz. In fact a version of the main result in Alon [2] is implicitly used in the proof, a result which may fail to hold for m nonprime. In the present paper we use further properties of random trees and substitute the polynomial method by a combinatorial argument due to Häggkvist [10] to prove our main result.

We note an application of Theorem 1 to Ringel's conjecture. Let $g(m)$ be the smallest integer n such that any tree with m edges decomposes the complete graph K_n. It was shown by Yuster [20] that $g(m) = O(m^{10})$ and the upper bound was reduced by Kézdy and Snevily [13] to $g(m) = O(m^3)$. Since $K_{2m,2m}$ decomposes the complete graph K_{8m^2+1} (see Snevily [19]), Theorem 1 shows that $g(m) = O(m^2)$ asymptotically almost surely.

The paper is organized as follows. In Section 2 we give some additional properties of random trees which will be used in the proof of the main result. Section 3 describes rainbow embeddings and cyclic decompositions and Section 4 contains the proof of Theorem 1. The paper concludes with some final comments.

2 Bipartition classes of random trees

The recursive definition of a tree as a collection of subtrees hanging from a root usually translates to functional equations for the generating functions counting parameters of the trees. This is the case for the so–called additive parameters, for instance the number of leaves: the total number of leaves of a tree is the sum of the leaves of each of the subtrees hanging from a root. For such additive parameters the general results on generating functions show that their probability distribution is asymptotically normal. For the degrees of nodes in a random tree this is explicitly done by Drmota and Gitterberger. In particular the following statement is a specialisation of [6, Theorem 2.1].

Theorem 2 *The number $X_{m,1}$ of vertices of degree one in a random tree with m edges is asymptotically normal, with expected value $\mathbb{E}(X_{m,1}) = cm+O(1)$ and variance $\mathbb{V}\mathrm{ar}(X_{m,1}) = c_1 m + O(1)$ for some constants c, c_1 which can be computed with arbitrary precision.*

It had been already proved by Robinson and Schwenk [18] that the constant c in the above theorem is $c \approx 0.438$. Since the variance is linear in the expected value, there is concentration of $X_{m,1}$ around its mean. In particular, for every $\epsilon > 0$,

$$\Pr(X_{m,1}) < (c - \epsilon)m \to 0 \ (m \to \infty). \tag{1}$$

Thus, a random tree with m edges has a.a.s. more than $2m/5$ leaves, a fact which is good enough for our present purposes.

We will use another property of a random tree T. For short we denote by *bipartition classes* the vertex classes of the bipartition of a tree. We are interested in the size of a bipartition class in the base tree of T (the tree obtained from T by deleting its leaves.) Unfortunately this is not a parameter whose analysis can be explicitly found in the literature, although the techniques to study it are well established. Our goal is to prove the following theorem.

Theorem 3 *The bipartition classes A, B of the base tree of a random tree with m edges satisfy a.a.s.*

$$||A| - |B|| \le \epsilon m,$$

for every fixed $\epsilon > 0$.

Before giving the proof of Theorem 3 let us recall some basic facts for the asymptotic analysis of unlabelled trees (see e.g. Drmota [5] for details). The generating function

$$t(x) = \sum_{m\ge 0} t_m x^{m+1}$$

of the number t_m of rooted unlabelled trees with m edges satisfies the functional equation

$$t(x) = x \exp\left(\sum_{i\ge 1} \frac{1}{i} t(x^i) \right), \tag{2}$$

obtained by describing an element of $\mathcal{T}$ as a node together with a multiset of elements of $\mathcal{T}$. It is known that the radius x_0 of convergence of $t(x)$ satisfies $0 < x_0 < 1$ and that $t(x)$ has a local representation of the form

$$t(x) = g(x) - h(x)\sqrt{1 - \frac{x}{x_0}}, \tag{3}$$

where $g(x)$ and $h(x)$ are analytic in a neighbourhood of x_0 and satisfy $g(x_0) = 1$ and $h(x_0) > 0$. Furthermore x_0 is the only singularity on the radius of convergence $|x| = x_0$ and $t(x)$ can be analytically continued to $\{x \in \mathbb{C} : |x| < x_0 + \eta\} \setminus [x_0, \infty)$ for some $\eta > 0$. By singularity analysis (see [8] or [5]) this leads to an asymptotic expansion for t_m of the form

$$t_m = c_1 m^{-3/2} x_0^{-m} \left(1 + O\left(\frac{1}{m}\right)\right)$$

with some constant $c_1 > 0$.

The generating function $\tilde{t}(x)$ of unrooted unlabelled trees is then given by

$$\tilde{t}(x) = t(x) - \frac{t(x)^2}{2} + \frac{t(x^2)}{2}.$$

This follows from the fact that rooted trees are in bijection with the union of unrooted trees and (unordered) pairs of different rooted trees (where the two roots are joint by an additional edge to recover again a tree) see [16] or [5]. Since $t(x_0) = g(x_0) = 1$ the dominant squareroot singularity cancels and one obtains a local representation of the form

$$\tilde{t}(x) = \tilde{g}(x) + \tilde{h}(x) \left(1 - \frac{x}{x_0}\right)^{3/2}$$

(where $\tilde{g}(x)$ and $\tilde{h}(x)$ are analytic in a neighbourhood of x_0 and satisfy $\tilde{g}(x_0) > 0$ and $\tilde{h}(x_0) > 0$) which leads to

$$\tilde{t}_m \sim c_2 m^{-5/2} x_0^{-m} \left(1 + O\left(\frac{1}{m}\right)\right) \tag{4}$$

for some constant $c_2 > 0$.

We can take also into account the size of the base tree (or equivalently the number of inner vertices.) Let

$$t(x,u) = \sum_{m,k} t_{m,k} x^{m+1} u^k,$$

where $t_{m,k}$ denotes the number of rooted trees with m edges and k inner vertices (including the root if the tree has at least one edge even if the root has degree one). Then we have

$$t(x,u) = xu \exp\left(\sum_{i\ge 1} \frac{1}{i} t(x^i,u^i)\right) - x(u-1).$$

Note that it is necessary to include (formally) the root into the set of inner vertices in order to have a recursive description. Next the corresponding generating function $\tilde{t}(x,u)$ is given by

$$\tilde{t}(x,u) = t(x,u) - x(u-1)t(x,u) - \frac{t(x,u)^2}{2} + \frac{t(x^2,u^2)}{2}.$$

At this stage we have to disregard the root of those rooted trees as an inner vertex if the root has degree one. This is done by substracting $x(u-1)t(x,u)$.

Next we distinguish between the bipartition classes A and B of the base tree. We have to be careful since the distinction between A and B is not unique. Only in the rooted case we can define A as the set of inner vertices with even distance to the root and B as the set of inner vertices with odd distance to the root. More precisely, let

$$t(x,w_0,w_1) = \sum_{m,k_0,k_1} t_{m,k_0,k_1} x^{m+1} w_0^{k_0} w_1^{k_1},$$

where t_{m,k_0,k_1} denotes the number of rooted trees with m edges and k_0 inner vertices (including the root if the tree has at least one edge even if the root has degree one) with even distance to the root and k_1 inner vertices with odd distance to the root. Then we have

$$t(x,w_0,w_1) = xw_0 \exp\left(\sum_{i\ge 1} \frac{1}{i} t(x^i,w_1^i,w_0^i)\right) - x(w_0-1).$$

Note that this equation is not an equation for $t(x,w_0,w_1)$ since the right hand side involves $t(x,w_1,w_0)$. However, by using another iteration we are led to the slightly more involved equation

$$t(x,w_0,w_1) = xw_0 \exp\left(\sum_{i\ge 1} \frac{1}{i}\left(x^i w_1^i \exp\left(\sum_{j\ge 1} \frac{1}{j} t(x^{ij},w_0^{ij},w_1^{ij})\right) - x^i(w_1^i-1)\right)\right) - x(w_0-1).$$

Before we study unrooted trees let us discuss the case of rooted trees in more detail. Recall that we are interested in the difference $|A| - |B|$ which we can do by setting $w_0 = w$ and $w_1 = w^{-1}$. Hence, if $T(x,w) = \sum_{m,\ell} T_{m,\ell} x^{m+1} w^\ell$ denotes the generating function, where $T_{m,\ell}$ denotes the number of rooted trees with m edges and $|A| - |B| = \ell$ (where ℓ is some – possibly negative – integer and the root is contained in A even if the root has degree one) then $T(x,w) = t(x,w,w^{-1})$ and we have

$$T(x,w) = xw \exp\left(\sum_{i\ge 1} \frac{1}{i}\left(x^i w^{-i} \exp\left(\sum_{j\ge 1}\frac{1}{j} T(x^{ij}, w^{ij})\right) - x^i(w^{-i}-1)\right)\right) - x(w-1).$$

As usual we denote by $a_n = [x^n]\, a(x)$ the n-th coefficient of a power series $a(x) = \sum_{n\ge 0} a_n x^n$. With the help of this notation it follows that

$$\mathbb{E}\, w^{|A|-|B|} = \frac{[x^{m+1}]\, T(x,w)}{[x^{m+1}]\, T(x,1)},$$

where the expectation is taken over all trees with m edges. This magnitude can be determined asymptotically if w is close to 1 with the help of standard singularity analysis tools.

Lemma 1 *Let A and B be the two bipartition classes in the base tree of a rooted unlabelled tree with m edges. Then there exists $\eta > 0$ such that uniformly for complex w with $|w-1| \le \eta$ we have*

$$\mathbb{E}\, w^{|A|-|B|} = A(w)B(w)^m \left(1 + O\left(\frac{1}{m}\right)\right) \tag{5}$$

for some analytic functions $A(w)$ and $B(w)$ that satisfy $A(1) = B(1) = 1$ and $B'(1) = 0$.

Proof. If we set $w = 1$ then the $T(x,1) = t(x)$ and we already know that $t(x)$ has a singular expansion of squareroot type, see (3). The idea is to show that we can obtain a similar singular expansion for $T(x,w)$ if w is close to 1:

$$T(x,w) = g(x,w) - h(x,w)\sqrt{1 - \frac{x}{x_0(w)}}, \tag{6}$$

and that there is an analytic continuation of $T(x,w)$ for $\{x \in \mathbb{C} : |x| < |x_0(w)| + \eta\} \setminus [x_0(w), \infty)$ for some $\eta > 0$. Of course, if we can verify these properties then standard singularity analysis (see [5]) leads to

$$[x^{m+1}]\, T(x,w) = \frac{h(x_0(w), w)}{2\sqrt{\pi}} m^{-3/2} x_0(w)^{-m-1}\left(1 + O\left(\frac{1}{m}\right)\right)$$

and consequently to (5) with

$$A(w) = \frac{h(x_0(w), w)}{h(x_0(1), 1)} \quad \text{and} \quad B(w) = \frac{x_0}{x_0(w)}.$$

It remains then to check that $x_0'(1) = 0$ which is equivalent to $B'(1) = 0$.

In order to prove (6) we just have to adapt the methods of [6] (see also [5]). Since $x_0 < 1$ it follows that $T(x^i, 1) = t(x^i)$ is analytic for $|x| < \sqrt{x_0}$ if $i \ge 2$. Furthermore, since $||A| - |B|| \le m$, it follows that $|T(x^i, w^i)| \le T(|x\overline{w}|^i, 1) = t(|x\overline{w}|^i)$, where $\overline{w} = \max\{|w|, |w|^{-1}\}$. Hence there exist $\eta > 0$ such that all functions $T(x^i, w^i)$ with $i \ge 2$ are analytic for $|x| < x_0 + \eta$ and $|w - 1| < \eta$. Furthermore, since $t(0) = 0$ we also have the upper bound $T(x^i, w^i) = O(|x\overline{w}|^i)$ for $i \ge 2$. Hence, we can assume that $T(x^i, w^i)$ (for $i \ge 2$) are already known functions when we are searching for the solution $y = T(x, w)$ of the equation

$$\begin{aligned}
y = xw \exp\Bigg(& xw^{-1} \exp\left(y + \sum_{j \ge 2} \frac{1}{j} T(x^j, w^j) \right) - x(w^{-1} - 1) \\
& + \sum_{i \ge 2} \frac{1}{i} \left(x^i w^{-i} \exp\left(\sum_{j \ge 1} \frac{1}{j} T(x^{ij}, w^{ij}) \right) - x^i (w^{-i} - 1) \right) \Bigg) \\
& - x(w-1) \\
= xw \exp\Bigg(& xw^{-1} \exp\left(y + \sum_{j \ge 2} \frac{1}{j} T(x^j, w^j) \right) - x(w^{-1} - 1) \\
& + \sum_{i \ge 2} \frac{1}{i} T(x^i, w^{-i}) \Bigg) - x(w-1).
\end{aligned}$$

This can be rewritten as $y = F(x, y, w)$, where F is a power series with non-negative coefficients, namely by expanding the exponential function the potential negative terms on the right hand side disappear. Hence, we can apply [5, Theorem 2.21] and obtain (6) locally around $(x, w) = (x_0, 1)$. (Note that the property $t_m > 0$ is sufficient to provide analytic continuation as required, hence, this is automatically satisfied.)

Finally, $x_0'(1)$ is given by $x_0'(1) = -F_w(x_0, t(x_0), 1)/F_x(x_0, t(x_0), 1)$. Thus, we only have to check that $F_w(x_0, t(x_0), 1) = 0$. Recall that, from (3), we have $t(x_0) = 1$ and that

$$\exp\left(1 + \sum_{i \ge 2} \frac{1}{i} t(x_0^i) \right) = \exp\left(\sum_{i \ge 1} \frac{1}{i} t(x_0^i) \right) = \frac{1}{x_0}.$$

Consequently, if we represent F as $F(x,y,w) = xwe^{G(x,y,w)} - x(w-1)$ with

$$G(x,y,w) = xw^{-1}\exp\left(y + \sum_{j\ge 2}\frac{1}{j}T(x^j,w^j)\right) - x(w^{-1}-1) + \sum_{i\ge 2}\frac{1}{i}T(x^i,w^{-i})$$

then we have

$$G(x_0,1,1) = 1 + \sum_{i\ge 2}\frac{1}{i}t(x_0^i) = \log\frac{1}{x_0}$$

and

$$G_w(x_0,1,1) = -1 + \sum_{j\ge 2}T_w(x_0^j,1) + x_0 - \sum_{i\ge 2}T_w(x_0^i,1) = -1 + x_0$$

which gives

$$F_w(x_0,1,1) = 1 + G_w(x_0,1,1) - x_0 = 0$$

as proposed. This completes the proof of the lemma. □

Lemma 1 has two immediate consequences that can be deduced from the following version of Hwang's Quasi-Power-Theorem (see [5, Theorem 2.22] and [11] for the original statement).

Lemma 2 *Let X_n be a sequence of random variables with the property that*

$$\mathbb{E}\, w^{X_n} = A(w)B(w)^{\lambda_n}\left(1 + O\left(\frac{1}{\phi_n}\right)\right) \tag{7}$$

holds uniformly in a complex neighborhood of $w=1$, where λ_n and ϕ_n are sequences of positive real numbers with $\lambda_n\to\infty$ and $\phi_n\to\infty$, and $A(w)$ and $B(w)$ are analytic functions in this neighbourhood of $w=1$ with $A(1)=B(1)=1$. Then X_n satisfies a central limit theorem of the form

$$\frac{1}{\sqrt{\lambda_n}}\left(X_n - \mathbb{E}\,X_n\right) \to N\left(0,\sigma^2\right) \tag{8}$$

and we have

$$\mathbb{E}\,X_n = \lambda_n\mu + O\left(1 + \lambda_n/\phi_n\right)$$

and

$$\mathbb{V}\mathrm{ar}\,X_n = \lambda_n\sigma^2 + O\left((1+\lambda_n/\phi_n)^2\right),$$

where $\mu = B'(1)$ and $\sigma^2 = B''(1) + B'(1) - B'(1)^2$. Finally there exist positive constants c_1, c_2, c_3 such that

$$\Pr\left(|X_n - \mathbb{E}\,X_n| \ge \varepsilon\lambda_n\right) \le c_1 e^{-c_2\varepsilon^2\lambda_n} \tag{9}$$

uniformly for $\varepsilon \le c_3$.

In particular it follows (because of $B'(1) = 0$)

$$\Pr\left(||A| - |B|| \geq \varepsilon m\right) \leq C e^{-c\varepsilon^2 m}$$

for some positive constants c and C and for sufficiently small $\varepsilon > 0$. Of course this is precisely the statement that we want to prove for unlabelled trees.

Remark 1 *It should be noted that the linear behaviour of the variance is sufficient to provide (with the help of Chebyshev's inequality) a bound of the form*

$$\Pr\left(||A| - |B|| \geq \varepsilon m\right) \leq \frac{\mathbb{V}\mathrm{ar}(|A| - |B|)}{\varepsilon^2 m^2} = \frac{\sigma^2 m + O(1)}{\varepsilon^2 m^2} = O\left(\frac{1}{m}\right)$$

for every $\varepsilon > 0$.

Furthermore we note that the linear behaviour of the variance can be directly checked with the help of the squareroot expansion (6). Actually since the mean value is bounded (due to the property $x_0'(1) = 0$) the variance and the second moment are almost the same:

$$\mathbb{V}\mathrm{ar}(|A| - |B|) \leq \mathbb{E}(|A| - |B|)^2 = \frac{[x^{m+1}]\, T_w(x,1) + T_{ww}(x,1)}{[x^{m+1}]\, T(x,1)}.$$

By using the property $x_0'(1) = 0$ we have

$$T_w(x,1) = g_w(x,1) - h_w(x,1)\sqrt{1 - \frac{x}{x_0}}.$$

and

$$T_{ww}(x,1) = g_{ww}(x,1) - h_w(x,1)\sqrt{1 - \frac{x}{x_0}} - \frac{x_0''(1)x}{x_0^2}\frac{h(x,1)}{2\sqrt{1 - \frac{x}{x_0}}}.$$

so that (by another application of the singulartiy analysis) the linear behaviour of

$$\mathbb{E}(|A| - |B|)^2 = -(x_0''(1)/x_0)m + O(1)$$

follows. Actually we will use this kind of approach for unrooted trees. □

In a final step we deal with unrooted trees. As mentioned above it is not possible to distinguish between the sets A and B in unrooted trees. This will be also reflected in the combinatorial construction that we use. Namely if we use Otter's bijection between rooted trees and the union of unrooted trees and unordered pairs of different rooted trees then we would obtain the generating function

$$\tilde{t}(x, w_0, w_1) = t(x, w_0, w_1) - x(w_0 - 1)t(x, w_1, w_0) - \frac{t(x, w_0, w_1)t(x, w_1, w_0)}{2} + \frac{t(x^2, 1, 1)}{2}$$

or the generating function

$$\tilde{T}(x,w) = T(x,w) - x(w-1)T(x,w^{-1}) - \frac{T(x,w)T(x,w^{-1})}{2} + \frac{T(x^2,1)}{2}.$$

The problem with these expressions is that we (have to) lose track of the distribution of $|A|-|B|$. Actually the term $T^{(2)}(x,w) = \frac{1}{2}T(x,w)T(x,w^{-1}) - \frac{1}{2}T(x^2,1)$ (that encodes unordered pairs of different rooted trees that are linked by an additional edge) takes only care of the absolute value $||A|-|B||$ in the following way. The number of resulting trees with m edges and with $||A|-|B|| = \ell$ (where $\ell \ge 0$) is given by

$$[x^{m+1}w^{\ell}]\, T^{(2)}(x,w) + [x^{m+1}w^{-\ell}]\, T^{(2)}(x,w).$$

The reason is that every possible unordered pair is counted twice as ordered pairs, one with $|A|-|B| = \ell$ and one with $|A|-|B| = -\ell$. The factor $\frac{1}{2}$ discounts this overcounting to the right value, however, the symmetrized distribution of $|A|-|B|$ persists.

This means that the Laurent series

$$f_m(w) = \frac{[x^{m+1}]\, \tilde{T}(x,w)}{[x^{m+1}]\, \tilde{T}(x,1)}$$

encodes the distribution of the absolute value $||A|-|B||$ of the form

$$\Pr(||A|-|B|| = \ell) = [w^{\ell}]\, f_m(w) + [w^{-\ell}]\, f_m(w). \tag{10}$$

Unfortunately we cannot prove something like a central limit theorem for $|A|-|B|$ but it is still possible to keep track of the second moment.

Lemma 3 *Let A and B denote the two bipartition classes in unrooted unlabelled trees with m edges. Then*

$$\mathbb{E}(|A|-|B|)^2 = O(m)$$

Proof. We first note that

$$\begin{aligned}
\mathbb{E}(|A|-|B|)^2 &= \sum_{\ell>0} \Pr(||A|-|B|| = \ell)\, \ell^2 \\
&= \sum_{\ell>0} \Big([w^{\ell}]\, f_m(w) + [w^{-\ell}]\, f_m(w) \Big)\, \ell^2 \\
&= \sum_{\ell\in\mathbb{Z}} [w^{\ell}]\, f_m(w)\, \ell^2 \\
&= \frac{[x^{m+1}]\, (\tilde{T}_w(x,1) + \tilde{T}_{ww}(x,1))}{[x^{m+1}]\, \tilde{T}(x,1)}.
\end{aligned}$$

Furthermore, by using (6) and the property that $x_0'(1) = 0$ we have

$$\begin{aligned}\tilde{T}_w(x,1) &= T_w(x,1) - xT(x,1) \\ &= (g_w(x,1) - xg(x,1)) - (h_w(x,1) - xh(x,1))\sqrt{1 - \frac{x}{x_0}}\end{aligned}$$

and

$$\begin{aligned}\tilde{T}_{ww}(x,1) &= T_{ww}(x,1)(1 - T(x,1)) + 2xT_w(x,1) + T(x,1)T_w(x,1) \\ &= -\frac{x_0''(1)x}{x_0^2}\frac{h(x,1)(1-g(x,1))}{2\sqrt{1-\frac{x}{x_0}}} + g_2(x) - h_2(x)\sqrt{1 - \frac{x}{x_0}}\end{aligned}$$

for some functions g_2, h_2 that are analytic at x_0. Now we use that property that $g(x_0, 1) = t(x_0) = 1$ so that we also have

$$-\frac{x_0''(1)x}{x_0^2}\frac{h(x,1)(1-g(x,1))}{2\sqrt{1-\frac{x}{x_0}}} = g_3(x) - h_3(x)\sqrt{1 - \frac{x}{x_0}}$$

for some functions g_2, h_2 that are analytic at x_0. Summing up we have

$$\tilde{T}_w(x,1) + \tilde{T}_{ww}(x,1)) = g_3(x) - h_3(x)\sqrt{1 - \frac{x}{x_0}}$$

(for some functions g_3, h_3 that are analytic at x_0) and by singularity analysis it follows that

$$[x^{m+1}]\,(\tilde{T}_w(x,1) + \tilde{T}_{ww}(x,1)) = c_3 m^{-3/2} x_0^{-m}\left(1 + O\left(\frac{1}{m}\right)\right)$$

for some constant $c_3 \geq 0$. Since the asymptotic expansion of $\tilde{t}_m = [x^{m+1}]\,\tilde{T}(x,1)$ is given by (4) we finally obtain $\mathbb{E}(|A| - |B|)^2 = O(m)$ as proposed. □

As in Remark 1 this property implies

$$\Pr(||A| - |B|| \geq \varepsilon m) = O\left(\frac{1}{\varepsilon^2 m}\right)$$

for every $\varepsilon > 0$ and, thus, completes the proof of Theorem 3.

3 Rainbow embeddings

The general approach to show that a tree T decomposes a complete graph or a complete bipartite graph consists in showing that T cyclically decompose the corresponding graphs, namely, that the decomposition is given by the orbit of a tree by a cyclic authomorphism group of the graph. We next recall the basic principle behind this approach in slightly different terminology.

A rainbow embedding of a graph H into an oriented arc–colored graph X is an injective homomorphism f of some orientation $\vec{H}$ of H in X such that no two arcs of $f(\vec{H})$ have the same color.

Let $X = Cay(G, S)$ be a Cayley digraph of an abelian group G with respect to an antisymmetric subset $S \subset G$ (that is, $S \cap -S = \emptyset$). We consider X as an arc–colored oriented graph, by giving to each arc $(x, x+s)$, $x \in G, s \in S$, the color s. Suppose that H admits a rainbow embedding f in X. For each $a \in G$ the translation $x \to x + a$, $x \in G$, is an automorphism of X which preserves the colors and has no fixed points. Therefore, each translation sends $f(\vec{H})$ to an isomorphic copy which is edge–disjoint from it. Thus the sets of translations for all $a \in G$ give rise to $n := |G|$ edge–disjoint copies of $\vec{H}$ in X. By ignoring orientations and colors, we thus have n edge disjoint copies of H in the underlying graph of X. The same is true if the values of a are restricted to a subgroup (or even a subset) of G. The term *cyclic decomposition* refers to the situation where the values of a are restricted to a cyclic subgroup of G.

We will use the above approach with the Cayley graph $X = Cay(\mathbb{Z}_m \times \mathbb{Z}_4, \mathbb{Z}_m \times \{1\})$. We note that the underlying graph of X is isomorphic to $K_{2m,2m}$. The strategy of the proof is to show first that the base tree T_0 of a random tree with m edges admits a rainbow embedding f into X in such a way that $f(T_0) \subset \mathbb{Z}_m \times \{1, 2\}$. This can actually be achieved greedily as shown in the proof of next lemma.

Lemma 4 *Let m be a positive integer. Let T be a tree with $n < 3m/5$ edges and bipartition classes A, B. If $||A| - |B|| \leq m/10$ then there is a rainbow embedding f of T into $X = Cay(\mathbb{Z}_m \times \mathbb{Z}_4, \mathbb{Z}_m \times \{1\})$ such that $f(V(T)) \subset (\mathbb{Z}_m \times \{1\}) \cup (\mathbb{Z}_m \times \{2\})$.*

Proof. We may assume that $|B| \geq |A|$. Let x be an endvertex of T in B (there is at least one such endvertex since $|B| \geq |A|$.) Suppose that f is a rainbow embedding $T' = T - x$ such that $f(V(T')) \subset \mathbb{Z}_m \times \{1, 2\}$. Denote by $A' = f(V(T')) \cap (\mathbb{Z}_m \times \{1\})$, $B' = f(V(T')) \cap (\mathbb{Z}_m \times \{2\})$. We may assume that f sends the vertex y of T' adjacent to x in T to A', say $f(y) = (a_y, 1)$.

Let $C' \subset \mathbb{Z}_m \times \{1\}$ be the set of colors not used by the rainbow embedding f. Since

$$|C'| > m - n \geq 2m/5 = 3m/10 + m/10 > |V(T')|/2 + ||B'| - |A'|| \geq |B'|,$$

there is $(z, 1) \in C'$ such that $(a_y + z, 2) \notin B'$. Therefore one can extend f to a rainbow embedding of T by defining $f(x) = (a_y + z, 2)$: the new leaf is colored by the unused color $(z, 1)$ and the embedding is rainbow. Since the statement of the lemma trivially holds for $n = 1$, it also holds for every $n < 3m/5$. □

4 Completing the decomposition

The second step involves a proper embedding of the leaves of T. For this we use the following result, which is a specialisation of Häggkvist [10, Corolary 2.8].

Theorem 4 (Häggkvist,[10]) *Let G be a d–regular bipartite graph with bipartition $A = \{a_1, \dots, a_n\}$ and B. Let $C = (c_{ij})$ be an $n \times n$ matrix with nonnegative integer entries such that, for each $i = 1, \dots, n$, we have*

$$\sum_k c_{ik} = \sum_k c_{ki} = d. \tag{11}$$

Then there is an edge–coloring of G with n colors such that, for each pair $i, j \in \{1, 2, \dots n\}$, the vertex a_i is incident with c_{ij} edges of color j and every vertex in B is incident with at most one edge of color j. □

Under its assumptions, the above theorem ensures that G admits an edge decomposition into n forests (each one identified by one color class of the edge–coloring) such that the centers of the stars of each copy belong to the same bipartition class and the degrees of the stars in the i–th forest are the entries of the i–th row of the matrix C. In particular, if C is a circulant matrix (each row is a cyclic shift of the preceding one), then G admits a decomposition into n isomorphic copies of a forest.

The next Lemma isolates the application Theorem 4 for our decomposition purposes.

Lemma 5 *Let T be a tree with m edges. If the base tree T_0 of T admits a rainbow embedding f in $X = Cay(\mathbb{Z}_m \times \mathbb{Z}_4, \mathbb{Z}_m \times \{1\})$ such that $f(V(T_0)) \subset (\mathbb{Z}_m \times \{1\}) \cup (\mathbb{Z}_m \times \{2\})$ then T decomposes $K_{2m,2m}$.*

Proof. For a subgraph H of X and an element $(i, j) \in \mathbb{Z}_m \times \mathbb{Z}_4$ we shall denote by $H + (i, j)$ the image of H by the automorphism $\phi_{i,j} : (x, y) \to (x + i, y + j)$ of X given by the translation (i, j).

Denote by $S \subset \mathbb{Z}_m \times \{1\}$ the set of colors used by the rainbow embedding f of the base tree T_0. As described in the beginning of Section 3, the set of translations of $f(T_0)$ by elements of $\mathbb{Z}_m \times \mathbb{Z}_4$ is an edge–decomposition of the Cayley graph $Cay(\mathbb{Z}_m \times \mathbb{Z}_4, S \times \{1\})$.

Let A_0 and B_0 denote the bipartition classes of T_0 so that $f(A_0) \subset \mathbb{Z}_m \times \{1\}$ and $f(B_0) \subset \mathbb{Z}_m \times \{2\}$. For each $(i, 2) \in f(B_0)$ denote by d_i the number of endvertices of T adjacent to that vertex of $f(T_0)$, and define $d_i = 0$ whenever $(i, 2) \notin f(B_0)$.

Choose a subset $S_1 \subset \mathbb{Z}_m \times \{1\} \setminus S$ with cardinality $m_1 = \sum_{i \in \mathbb{Z}_m} d_i$ and consider the subgraph $\vec{G_1}$ of $Cay(\mathbb{Z}_m \times \mathbb{Z}_4, S_1 \times \{1\})$ induced by $(\mathbb{Z}_m \times \{2\}) \cup (\mathbb{Z}_m \times \{3\})$. The underlying graph G_1 of $\vec{G_1}$ is a m_1–regular bipartite graph.

Consider the circulant $(m\times m)$ matrix C with entries indexed by the elements in $\mathbb{Z}_m\times\mathbb{Z}_m$ where the first column is defined as $c_{0,i}=d_i$. The rowsums and columnsums of C are all equal to $\sum_{i\in\mathbb{Z}_m} d_i=m_1$. Thus the matrix C satisfies the hypothesis of Theorem 4. Hence, the bipartite graph G_1 admits a decomposition

$$G_1=F_0\oplus F_1\oplus\cdots\oplus F_{m-1},$$

where each F_j is a forest of stars in which the degree of the vertex $(i,2)$ is $c_{j,i}$. We observe that, since C is a circulant matrix, $(f(T_0)+(i,0))\oplus F_i$ gives an isomorphic copy of the subtree $T_1\subset T$ obtained from T by removing the end vertices in the bipartition class which contains B_0. Moreover, for each $(i,j)\in\mathbb{Z}_m\times\mathbb{Z}_4$, the subgraph $(f(T_0)+(i,j))\oplus(F_i+(0,j))$ is also isomorphic to T_1. For different (i,j) and (i',j') the corresponding pair of trees are edge–disjoint because they arise from different translations of T_0 and of F_i and the colors of the colors of $f(T_0)$ and of F_i do not overlap. Moreover, the set of these trees for for all (i,j) cover all the edges of $Cay(\mathbb{Z}_m\times\mathbb{Z}_4,(S\cup S_1)\times\{1\})$. Therefore, this set of trees decomposes $Cay(\mathbb{Z}_m\times\mathbb{Z}_4,(S\cup S_1)\times\{1\})$.

In order to obtain a decomposition of X by the whole of T it only remains to embed the endvertices of the tree which belong to the bipartition class which contains B_0 (which are adjacent to vertices in A_0.) This is done in the same way as for the endvertices in the bipartition class which contains A_0 with the obvious modifications. The number of these remaining endvertices is $m-|S|-|S_1|$. By setting $S_2=\mathbb{Z}_m\setminus(S\cup S_1)$, we can obtain as before a decomposition of the underlying bipartite graph G_2 of the subgraph of $Cay(\mathbb{Z}_m\times\mathbb{Z}_4,S_2\times\{1\})$ induced by $(\mathbb{Z}_m\times\{0\})\cup(\mathbb{Z}_m\times\{1\})$ of the form

$$G_2=F_0'\oplus F_1'\oplus\cdots\oplus F_{m-1}',$$

in such a way that $(f(T_0)+(i,j))\oplus((F_i'\oplus F_i)+(0,j))$ is isomorphic to T for each $(i,j)\in\mathbb{Z}_m\times\mathbb{Z}_4$, and the set of all these copies of T decomposes $Cay(\mathbb{Z}_m\times\mathbb{Z}_4,\mathbb{Z}_m\times\{1\})$, a directed graph whose underlying graph is isomorphic to $K_{2m,2m}$. This completes the proof. □

The placement of the tree T in the above Lemma is illustrated in Figure 4.

The proof of Theorem 1 follows now directly from Lemma 4 and Lemma 5 and the results on random trees from Section 2.

Proof of Theorem 1. By Theorem 2 and the remarks following it a random tree with m edges has a.a.s. more than $2m/5$ leaves and, by Theorem 3, the cardinalities of the bipartition classes of the base tree of T differ less than $m/10$ in absolute value a.a.s. By Lemma 4, the base tree of T admits a.a.s. a rainbow embedding in $Cay(\mathbb{Z}_m\times\mathbb{Z}_4,\mathbb{Z}_m\times\{1\})$ in such a way that the image of the embedding sits in $\mathbb{Z}_m\times\{1,2\}$. In that case, Lemma 5 ensures that the tree T decomposes $K_{2m,2m}$. □

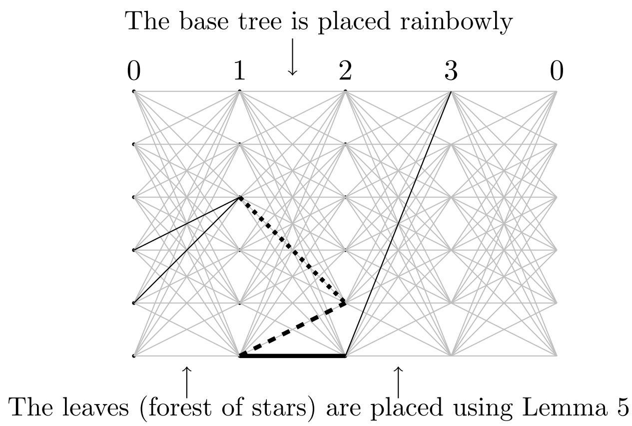

Figure 1: An illustration of the procedure in Lemma 5

5 Final comments

As mentioned in the Introduction, Theorem 1 follows from results in [3] when m is a prime. Actually the conclusion there is stronger, as it follows from these results that almost every tree with a prime number p of edges admits a rainbow embedding in $Cay(\mathbb{Z}_p \times \mathbb{Z}_4, \mathbb{Z}_p \times \{1\})$. In other words, almost every tree with p edges decomposes *cyclically* the complete bipartite graph $K_{2p,2p}$. Although it seems quite unlikely to us that the methods provide a result concerning the decomposition of $K_{m,m}$ by almost all trees of m edges, the technique used in [3] seems to be close to show that almost all trees with m edges decompose cyclically $K_{2m,2m}$.

The solution of the following problem, which may have its own interest, would provide the desired result:

Problem 1 *Let* $\mathbf{V}(a_1, \ldots, a_k)$ *denote the Vandermonde matrix*

$$\begin{pmatrix} 1 & a_1 & a_1^2 & \cdots & a_1^{k-1} \\ \vdots & & & & \vdots \\ 1 & a_k & a_k^2 & \cdots & a_k^{k-1} \end{pmatrix}.$$

Its permanent is

$$per\mathbf{V}(a_1, \ldots, a_k) = \sum_{\sigma \in Sym(k)} a_{\sigma(1)}^{k-1} a_{\sigma(2)}^{k-2} \cdots a_{\sigma(k)}^{0}.$$

Is it true that every sequence $a_1, \ldots, a_k$*,* $k < m$*, of* m*–th roots of unity (repetitions allowed) contains a subsequence* $a_{i_1}, \ldots, a_{i_t}$ *of length* $t \geq k/2$ *such that* $Per\mathbf{V}(a_{i_1}, \ldots, a_{i_t}) \neq 0$*?*

We note that the answer to the above question is obviously negative when the permanent is replaced by its close relative the determinant. A positive answer to the above problem would be relevant to prove the following proposition which is enough to show that almost every tree with m edges decomposes cyclically $K_{2m,2m}$ by replacing Lemma 5 in the proof of Theorem 1.

Proposition 1 *Let $a_1, \ldots, a_k$ be a sequence of elements (repetitions allowed) in the cyclic group $\mathbb{Z}_m$. If the answer to Problem 1 is positive, then for every pair of subsets $C, D \subset \mathbb{Z}_m$ with $|C| \geq |D| + k$ there are distinct elements $c_1, \ldots, c_k \in C$ such that the sums $a_1 + c_1, \ldots, a_k + c_k$ are pairwise distinct and none belongs to D.*

Proof. We identify the cyclic group $\mathbb{Z}_m$ with the multiplicative subgroup of the m–th roots of unity in the field $\mathbb{C}$ of complex numbers. Thus, if $\omega = e^{2\pi i/m}$, the elements of the sequence are $\omega^{a_1}, \ldots, \omega^{a_k}$. Consider the following polynomial in $\mathbb{C}[X_1, \ldots, X_k]$:

$$P = V(X_1, \ldots, X_k)V(\omega^{a_1}X_1, \ldots, \omega^{a_1}X_k) \prod_{i=1}^{k} \prod_{d \in D} (\omega^{a_i}X_i - \omega^d),$$

where

$$V(z_1, \ldots, z_k) = \prod_{1 \leq i < j \leq k} (z_i - z_j) = \sum_{\sigma \in Sym(k)} (-1)^{sign(\sigma)} z_{\sigma(1)}^{k-1} z_{\sigma(2)}^{k-2} \cdots z_{\sigma(k)}^{0},$$

denotes the Vandermonde determinant. By the positive answer to Problem 1, we may assume that $PerV(a_1, \ldots, a_t) \neq 0$ for some $t \geq k/4$.

The polynomial $V(X_1, \ldots, X_k)V(\omega^{a_1}X_1, \ldots, \omega^{a_1}X_k)$ has a term

$$X_1^{k-1} X_2^{k-1} \cdots X_t^{k-1} X_{t+1}^{2(k-t-1)} X_{t+2}^{2(k-t-2)} \cdots X_{k-1}^{2} X_k^{0} \tag{12}$$

of maximum degree with coefficient

$$Per\mathbf{V}(a_1, \ldots, a_t)\omega^{\sum_{i=1}^{k-t} a_{t+i}2(k-t-i)} \neq 0.$$

Indeed, suppose that the product of $X_{\sigma(1)}^{k-1} X_{\sigma(2)}^{k-2} \cdots X_{\sigma(k)}^{0}$ from the expansion of $V(X_1, \ldots, X_k)$ and $\omega^{a_{\tau_1}} X_{\tau(1)}^{k-1} \omega^{a_{\tau_2}} X_{\tau(2)}^{k-2} \cdots \omega^{a_{\tau_k}} X_{\tau(k)}^{0}$ gives rise to the monomial (12). Then the powers of X_i for $i = t+1, \ldots, k$ in (12) require that $\sigma(k) = \tau(k) = k, \sigma(k-1) = \tau(k-1) = k-1, \ldots, \sigma(t+1) = \tau(t+1) = k-t-1$. On the other hand, the powers of X_i for $i = 1, \ldots, t$ require that the restrictions of σ and τ to $\{1, 2, \ldots, t\}$ are mutually reverse: $\sigma(i) + \tau(i) = t+1$. This provides the stated value of the coefficient (for the appearance of the term $Per\mathbf{V}(a_1, \ldots, a_t)$ a detailed computation can be found in [4].)

Thus the polynomial P has a term

$$X_1^{(k-1)+|D|}X_2^{(k-1)+|D|}\cdots X_t^{(k-1)+|D|}X_{t+1}^{2(k-t-1)+|D|}\cdots X_k^{|D|}$$

of maximum degree with nonzero coefficient.

Therefore, since $2(k-t-1) < k-1$ and $|C| \geq k+|D|$, it follows from the Combinatorial Nullstellensatz that there are elements $c_1, \ldots, c_k \in C$ such that $P(\omega^{c_1}, \ldots, \omega^{c_k}) \neq 0$.

In particular, as $V(\omega^{c_1}, \ldots, \omega^{c_k}) \neq 0$, the elements $c_1, \ldots, c_k$ are pairwise distinct. Moreover, since $V(\omega^{a_1+c_1}, \ldots, \omega^{a_k+c_k}) \neq 0$ the sums $a_1 + c_1, \ldots, a_k + c_k$ are pairwise distinct. Finally, since $\prod_{i=1}^k \prod_{d\in D}(\omega^{a_i+c_i} - \omega^d) \neq 0$, none of the sums belong to D. This completes the proof. □

In order to prove Theorem 1 we can now use Lemma 4 and Proposition 1 to conclude that a.a.s. every tree admits a rainbow embedding in the directed Cayley graph X defined in section 3. In one step, by Lemma 4, the base tree T_0 of T can be rainbowly embedded in X, because the number of leaves of T is a.a.s. at least $2m/5$, by using some set D of colors. In the second step the leaves of T can be rainbowly embedded in X as in the proof of Lemma 5 by using Proposition 1 instead of Theorem 4: we have $k = m - |D|$ leaves to allocate and we can choose $C = \mathbb{Z}_m \times \{1\}$. By extending the rainbow embedding of the base tree with the appropriate assignment of the end vertices to the values $c_1, \ldots, c_k$, none of the colors of the leaves belong to the set D of colors used by the embedding of the base tree T_0.

Acknowledgements

The authors are grateful to the careful reading of the referee by pointing out some inaccuracies and suggesting many improvements to the readability of the paper. The first author acknowledges the support of the Austrian Science Fund FWF, Project F5002. The second author acknowledges the support of the Ministry of Science and Technology of Spain, and the European Regional Development Fund (ERDF) under project MTM2011-28800-C02-01 and by the Catalan Research Council under grant 2009SGR1387.

References

[1] N. Alon, Combinatorial Nullstellensatz, *Combinatorics, Probability and Computing* **8** (1999), 7–29.

[2] N. Alon, Additive Latin Transversals, *Israel J. Math.* **117** (2000), 125–130.

[3] M. Càmara, A. Lladó, J. Moragas. On a conjecture of Graham and Haggkvist with the polynomial method, *European J. Combin.* 30 (7) (2009) 1585–1592.

[4] S. Dasgupta, G. Károlyi, O. Serra, B. Szegedy, Transversals of additive Latin squares. *Israel J. Math.* 126 (2001), 17–28.

[5] M. Drmota. Random Trees. Springer-Verlag, 2009.

[6] M. Drmota, B. Gittenberger. The distribution of nodes of given degree in random trees, *J. of Graph Theory* 31 (1999) 227–253.

[7] M. Drmota, B. Gittenberger. The shape of unlabeled rooted random trees. *European J. Combin.* 31 (2010) 2028–2063.

[8] P. Flajolet and A. Odlyzko, Singularity analysis of generating functions. *SIAM J. Discrete Math.* 3 (1990), no. 2, 216–240.

[9] J. A. Gallian, A Dynamic Survey of Graph Labeling, *The Electronic Journal of Combinatorics* 5 (2007) # DS6.

[10] R. L. Häggkvist, Decompositions of Complete Bipartite Graphs, *Surveys in Combinatorics*, Johannes Siemons Ed., Cambridge University Press (1989) 115–146.

[11] H. K. Hwang. Théoremes limites pour les structures combinatoires et les fonctions arithmetiques. PhDThesis, Ecole Polytechnique, New York, 1994.

[12] A.E. Kézdy. ρ–valuations for some stunted trees. *Discrete Math.* 306 (21) (2006) 2786–2789.

[13] A. E. Kézdy, H. S. Snevily. Distinct Sums Modulo n and Tree Embeddings. *Combinatorics, Probability and Computing*, 11, (2002) 35–42.

[14] A. Lladó, S.C. López. Edge-Decompositions of $K_{n,n}$ Into Isomorphic Copies of a Given Tree, *J. of Graph Theory* 48 (2005), 1–18.

[15] A. Lladó, S.C. López, J. Moragas. Every tree is a large subtree of a tree that decomposes K_n or $K_{n,n}$, *Discrete Math.* 310 (2010) 838–842.

[16] R. Otter, The number of trees. *Ann. of Math.* (2) 49, (1948). 583–599.

[17] G. Ringel, Problem 25, Theory of Graphs and its Applications, Nakl. CSAV, Praha, (1964) pp. 162.

[18] R.W. Robinson, A.J. Schwenk. The distribution of degrees in a large random tree. *Discrete Math.* 12(4) (1975) 359-372.

[19] H. Snevily. New families of graphs that have α–labelings. *Discrete Math.* 170 (1997) 185–194.

[20] R. Yuster. Packing and decomposition of graphs with trees. *J. Combin. Theory Ser. B* 78 (2000) 123–140.