On the Independence Number of Random Trees via Tricolourations

Etienne Bellin
Ecole Polytechnique, Palaiseau, France

Abstract

We are interested in the independence number of large random simply generated trees and related parameters, such as their matching number or the kernel dimension of their adjacency matrix. We express these quantities using a canonical tricolouration, which is a way to colour the vertices of a tree with three colours. As an application we obtain limit theorems in L^p for the renormalised independence number in large simply generated trees (including large size-conditioned Bienaymé-Galton-Watson trees).

2012 ACM Subject Classification Mathematics of computing → Random graphs

Keywords and phrases Independence number, simply generated tree, Galton-Watson tree, tricolouration

Digital Object Identifier 10.4230/LIPIcs.AofA.2022.2

Acknowledgements I am grateful to Igor Kortchemski for his careful reading of the manuscript and for telling me Frederic Chapoton's suggestion to consider canonical tricolourations of random trees. I am also grateful to the anonymous referees and their useful remarks.

1 Introduction

A subset S of vertices of a finite graph G is called an *independent set* if there is no pair of connected vertices in S. The *independence number* of G, denoted by $I(G)$, is the biggest cardinal of an independent set of G. The independence number is a well studied quantity in computational complexity theory. It is known that computing the independence number is NP-hard in general (see e.g. [10, Sec. 3.1.3]). A lot of work has been carried out to describe algorithms computing the independence number in general graphs [19, 22] and in special classes of graphs where the computational time can be decreased (e.g. cubic graphs [21], claw-free graphs [17], P_5-free graphs [15]). The independence number has also received interest in combinatorics and in probability. Upper bounds have been found using probabilistic methods for cubic graphs [3]. Asymptotics have been found in certain classes of random trees (e.g. conditioned Bienaymé-Galton-Watson trees [8], simply generated trees [2], random recursive trees and binary search trees [9], and a wider class of random trees constructed from a Crump–Mode–Jagers branching process [12]). Finally we mention three articles giving applications of the independence number in scheduling theory [13], coding theory [4] and collusion detection in voting pools [1].

The goal of this article is to study the independence number of large simply generated trees, generalising some results of [7] and [2]. Simply generated trees are a wide class of random plane trees (i.e. rooted and ordered trees) introduced in [16] and encompass Bienaymé-Galton-Watson trees (BGW trees for short) conditioned to have a fixed number of vertices. Informally, a BGW tree with offspring distribution μ is a plane tree where vertices have an i.i.d. number of children with law μ. Various natural models of random trees are obtained with appropriate choices of the offspring distribution: e.g., uniform plane trees, uniform plane d-ary trees and uniform Cayley trees (see [11, Sec. 10]). In order to study the independence number, we will use a particular tricolouration of trees introduced in [23] and

33rd International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2022).
Editor: Mark Daniel Ward; Article No. 2; pp. 2:1–2:14
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

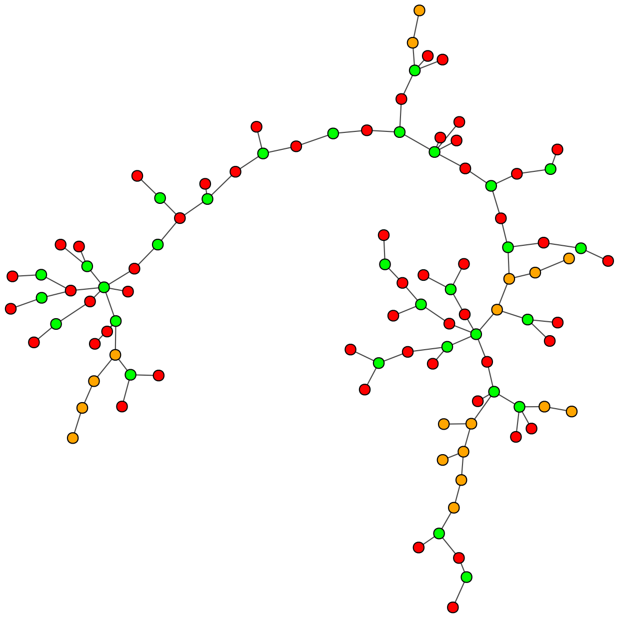

Figure 1 Tricolouration of a BGW tree with 100 vertices and a Poisson offspring distribution of parameter 1. The algorithm used to tricolour this tree can be found in [5, Appendix A].

later studied in [6], [7] and [5]. This colouring is based on the notion of covering. A *covering* of a finite tree T is a subset of vertices S of T such that every edge of T is adjacent to a vertex of S. A *smallest covering* of T is a covering with minimal cardinality. In general, a tree has more than one smallest covering. For every vertex v of T we colour v in the following way:

- If v belongs to every smallest covering, we colour v in green.
- If v belongs to no smallest covering, we colour it in red.
- If v belongs to some smallest coverings but not all, we colour it in orange.

For a tree T, denote by $n_g(T)$, $n_o(T)$ and $n_r(T)$, respectively, the number of green, orange and red vertices in T. It has been noticed in [7] that the size of a smallest covering of a tree T is equal to $n_g(T) + n_o(T)/2$. Since the complementary of a smallest covering is an independent set of maximal size, the independence number of T is $n_r(T)+n_o(T)/2$. Actually, other statistics of the tree T can be expressed as a linear combination of $n_g(T)$, $n_o(T)$ and $n_r(T)$. For instance the *matching number* $M(T)$ (i.e. the maximum size of a partial vertex matching) is equal to the size of a smallest covering which is $n_g(T) + n_o(T)/2$. The *nullity* $N(T)$ (i.e. the kernel dimension of the adjacency matrix) is $n_r(T) - n_g(T)$. The *edge cover number* and the *clique cover number* also coincide with the independence number on trees (see [9]). Our main result (Theorem 4) concerns simply generated trees, but, to keep this introduction short, let us state here a particular case for critical BGW trees.

▶ **Theorem 1.** *Let T_n be a Bienaymé-Galton-Watson tree with reproduction law μ, conditioned on having n vertices. Denote by $G(t) := \sum_{k=0}^{\infty} \mu_k t^k$ the generating function of μ and let q be the unique solution of $G(1-q) = q$ in $[0,1]$. Suppose that μ has mean 1, then, the following convergences hold in L^p for every $p > 0$:*

$$\frac{n_g(T_n)}{n} \xrightarrow[n\to\infty]{L^p} \frac{1-q+(1-2q)G'(1-q)}{1+G'(1-q)}, \qquad \frac{n_o(T_n)}{n} \xrightarrow[n\to\infty]{L^p} \frac{2\,q\,G'(1-q)}{1+G'(1-q)},$$

$$\frac{n_r(T_n)}{n} \xrightarrow[n\to\infty]{L^p} \frac{q}{1+G'(1-q)}.$$

Explicit computations of the expected number of green, orange and red vertices have been carried out in the case of a uniform Cayley tree with fixed size in [7] using generating functions. This confirms the convergence, in mean value, of Theorem 1. Indeed, it is well known that a BGW tree with Poisson distribution of parameter 1 conditioned to have n vertices has the same law as a uniform Cayley tree with n vertices. To our knowledge, there is no estimates or asymptotics, other than [7], for the expected number of green, orange and red vertices in simply generated trees. As we said before, different quantities such as the independence number $I(T)$, the matching number $M(T)$ and the nullity $N(T)$ of a tree T can be expressed in terms of $n_g(T)$, $n_o(T)$ and $n_r(T)$, so Theorem 1 yields limit theorems for these in the case of critical BGW trees.

▶ **Corollary 2.** *With the same notation and hypothesis as in Theorem 1, the following convergences hold in L^p for every $p > 0$:*

$$\frac{I(T_n)}{n} \xrightarrow[n\to\infty]{L^p} q, \qquad \frac{M(T_n)}{n} \xrightarrow[n\to\infty]{L^p} 1-q, \qquad \frac{N(T_n)}{n} \xrightarrow[n\to\infty]{L^p} 2q-1.$$

With the same hypothesis, the authors of [8] show the convergence of $I(T_n)/n$ in probability towards q. Moreover, in [2] the convergence of the first and second moment of $I(T_n)/n$ is studied. More precisely, it is shown that $\mathbb{E}\left[I(T_n)\right] = nq + O(1)$ and $V(I(T_n)) = \nu n + O(1)$ for some constant ν. Therefore, the convergence of the independence number, in the settings of Corollary 2 is not new, but Corollary 5 generalises this convergence for simply generated trees that are not equivalent to conditioned critical BGW trees. The main tool to prove Theorem 4 is the use of limit theorems for uniformly pointed simply generated trees found in [20] (see Section 4). In the first section we introduce simply generated trees. In the next section we state our main result in its most general form. To prove our main result, we first explain the limit theorems of [20] in the third section. The next two sections give properties of the tricolouration and describe how to colour the limiting trees. Finally, in the last section we prove Theorem 4.

2 Simply generated trees

Let $\mathbf{w} := (w_i)_{i\geq 0}$ be a sequence of nonnegative weights. A *simply generated tree* having n vertices with weight sequence $\mathbf{w}$ is a random plane tree T_n such that for every finite plane tree T,

$$\mathbb{P}\left(T_n = T\right) = \frac{1}{Z_n}\left(\prod_{v\in T} w_{k_v}\right)\mathbb{1}_{|T|=n}$$

where k_v is the outdegree of the vertex v in T, $|T|$ is the number of vertices in T and Z_n is the normalising constant defined by

$$Z_n := \sum_{|T|=n}\prod_{v\in T} w_{k_v}.$$

Notice that, when the weight sequence $\mathbf{w}$ is actually a probability sequence (i.e. the sum of the weights is equal to 1) then we recover the class of BGW trees. For T_n to be well defined, one needs Z_n to be nonzero. First of all, suppose that $w_0 > 0$ and $w_k > 0$ for some $k \geq 1$ otherwise $Z_n = 0$ for all $n \geq 1$. Let $\text{span}(\mathbf{w}) := \gcd\{i \geq 0 \,|\, w_i > 0\}$ (since $w_k > 0$ this quantity is well defined). The following result, found for instance in [11, Cor. 15.6], characterises the n's such that $Z_n > 0$:

▶ **Lemma 3** (Janson 2012). *If $Z_n > 0$ then $n \equiv 1 \mod \text{span}(\mathbf{w})$. Conversely, there exists n_0 such that for all $n \geq n_0$ satisfying $n \equiv 1 \mod \text{span}(\mathbf{w})$, $Z_n > 0$.*

Throughout this document and in Theorem 4, we suppose that $w_0 > 0$, $w_k > 0$ for some $k \geq 1$ and that all the n's appearing satisfy $n \geq n_0$ and $n \equiv 1 \mod \text{span}(\mathbf{w})$

Let $\rho \in [0, +\infty]$ be the radius of convergence of the generating series

$$\phi(x) := \sum_{i \geq 0} w_i x^i.$$

It is shown in [11, Lemma 3.1] that, if $\rho > 0$, then the function defined by

$$\psi(x) := \frac{x\phi'(x)}{\phi(x)}$$

is increasing on $[0, \rho)$ and we can define $\nu := \lim_{x \to \rho} \psi(x) \in (0, +\infty]$. We distinguish three different regimes:

- Regime 1 when $\rho > 0$ and $\nu \geq 1$. In this case there is a unique $\tau \in [0, \rho]$ such that $\tau < +\infty$ and $\psi(\tau) = 1$.
- Regime 2 when $\rho > 0$ and $0 < \nu < 1$. In this case $\rho < +\infty$ and we set $\tau := \rho$.
- Regime 3 when $\rho = 0$.

In regime 1 and 2 we can define a probability function given by

$$\pi_k := \frac{\tau^k w_k}{\phi(\tau)}. \tag{1}$$

The associated mean and generating function are respectively given by

$$m := \min(1, \nu) \qquad \text{and} \qquad G(x) := \frac{\phi(\tau x)}{\phi(x)}. \tag{2}$$

An important result of [11] is that, T_n, in regime 1 or 2, has the same law as a BGW tree with reproduction law π conditioned to have n vertices. In regime 3, T_n is not distributed like a conditioned BGW tree. Note that a critical or super-critical BGW tree or a BGW tree with a reproduction law with infinite mean, conditioned to have n vertices, always lays in regime 1. Moreover for a critical BGW tree with reproduction law μ, the probability π is the same as μ. A sub-critical BGW tree, conditioned to have n vertices, is either in regime 1 or in regime 2. We define *complete condensation* to be the condition:

$$\Delta(T_n) = (1 - m)n + n\mathcal{E}_n \tag{3}$$

where $\Delta(T_n)$ is the maximum degree of a vertex of T_n and $\mathcal{E}_n$ is a random variable converging in probability towards 0. For instance, complete condensation happens in regime 2 when there exists $\theta > 1$ and a slowly varying function ℓ such that $\pi_k = \ell(k)k^{-(1+\theta)}$ (see [14]). Complete condensation also happens in regime 3 for the weight sequence $w_k = k!^\alpha$ for $\alpha > 0$ (see [11, Ex. 19.36]).

3 Main results

In this section we state our main results. We keep all the notations and assumptions of Section 2.

▶ **Theorem 4.** *Let T_n be a simply generated tree with n vertices according to the weight sequence $\mathbf{w} = (w_i)_{i\geq 0}$. Recall that, in regime 1 and 2, G denotes the generating function of π defined in (1). Let q be the unique solution of $q = G(1-q)$ in $[0,1]$.*

1. *In regime 1 and in regime 2 with complete condensation (meaning that (3) is satisfied), the following convergences hold in L^p for every $p > 0$:*

$$\frac{n_g(T_n)}{n} \xrightarrow[n\to\infty]{L^p} \frac{1-q+(1-2q)G'(1-q)}{1+G'(1-q)}, \qquad \frac{n_o(T_n)}{n} \xrightarrow[n\to\infty]{L^p} \frac{2\,q\,G'(1-q)}{1+G'(1-q)},$$

$$\frac{n_r(T_n)}{n} \xrightarrow[n\to\infty]{L^p} \frac{q}{1+G'(1-q)}.$$

2. *In regime 3 with complete condensation, the following convergences hold in L^p for every $p > 0$:*

$$\frac{n_g(T_n)}{n} \xrightarrow[n\to\infty]{L^p} 0, \qquad \frac{n_o(T_n)}{n} \xrightarrow[n\to\infty]{L^p} 0, \qquad \frac{n_r(T_n)}{n} \xrightarrow[n\to\infty]{L^p} 1.$$

▶ **Corollary 5.** *We keep the same notation and hypothesis as in Theorem 4. Recall that $I(T_n)$, $M(T_n)$ and $N(T_n)$ are, respectively, the independence number, the matching number and the nullity of T_n.*

1. *In regime 1 and in regime 2 with complete condensation, the following convergences hold in L^p for every $p > 0$:*

$$\frac{I(T_n)}{n} \xrightarrow[n\to\infty]{L^p} q, \qquad \frac{M(T_n)}{n} \xrightarrow[n\to\infty]{L^p} 1-q, \qquad \frac{N(T_n)}{n} \xrightarrow[n\to\infty]{L^p} 2q-1.$$

2. *In regime 3 with complete condensation, the following convergences hold in L^p for every $p > 0$:*

$$\frac{I(T_n)}{n} \xrightarrow[n\to\infty]{L^p} 1, \qquad \frac{M(T_n)}{n} \xrightarrow[n\to\infty]{L^p} 0, \qquad \frac{N(T_n)}{n} \xrightarrow[n\to\infty]{L^p} 1.$$

4 Limit theorems for uniformly pointed simply generated trees

In this section we explain the results proved in [20] which will be our basic tool to prove Theorem 4. All the proofs and details of this section can be found in the above mentioned article. As said in the introduction, these results are limit theorems for uniformly pointed simply generated trees. A *pointed tree* is simply a couple (T, v) with a plane tree (i.e. rooted and ordered tree) T and a distinguished vertex v of T. A uniformly pointed simply generated tree is a couple (T_n, v_n) where T_n is a simply generated tree with n vertices and v_n is a distinguished vertex chosen uniformly at random among the n vertices of T_n. Basically, in regime 1 and in regime 2 and 3 with complete condensation, the local tree structure around v_n converges towards an infinite random tree which depends only on the regime. To formally define the notion of convergence used here, one needs to consider T_n to be a subtree of a big ambient tree denoted by $\mathcal{U}_\infty^\bullet$. Every plane tree (e.g. T_n) is considered, by definition, to be rooted, however we will encounter some infinite trees without any root which is unusual in the classic framework of plane trees. Let

$$\mathcal{V}_\infty := \{\emptyset\} \cup \bigcup_{n\geq 1} (\mathbb{N}^*)^n$$

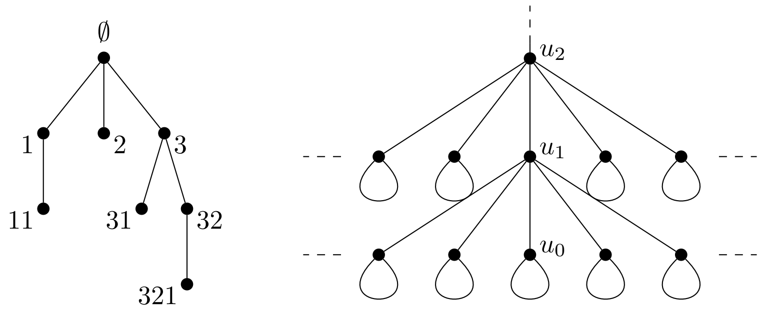

Figure 2 On the left, a subtree of the Ulam-Harris tree $\mathcal{U}_\infty$. On the right, the tree $\mathcal{U}_\infty^\bullet$. Each loop represents a copy of the Ulam-Harris tree.

be the set of words (empty word included) formed in the alphabet $\mathbb{N}^* = \{1, 2, 3, \dots\}$. Usually, plane trees are defined as subtrees of the so-called *Ulam-Haris tree*, denoted here by $\mathcal{U}_\infty$, which is the tree with vertex set $\mathcal{V}_\infty$ and edge set $\{(a_1 \dots a_{n-1}; a_1 \dots a_n) \,|\, \forall n, a_1, \dots, a_n \in \mathbb{N}^*\}$. With this definition, all the plane trees have a root which is a common ancestor to every vertex of the tree (it is the vertex designated by the empty word $\emptyset$). However, in regime 1, the root of T_n is, in a local point of view, at infinite distance from the distinguished vertex v_n. It suggests that the local limit of T_n around v_n has an infinite spine of ancestors and therefore, has no root. This is why we need a more general framework than the usual one for plane trees. Here we explain informally the construction of $\mathcal{U}_\infty^\bullet$. Let $u_0, u_1, \dots$ be vertices, in the plane, lined up to form an infinite connected spine. Each vertex u_i with $i > 0$ gets an infinite countable number of children on the left and on the right of its child u_{i-1}. Then, all the leaves of the current tree (u_0 included) give birth to the Ulam-Harris tree $\mathcal{U}_\infty$. The tree we obtain from this construction is denoted by $\mathcal{U}_\infty^\bullet$ and its set of vertices is denoted by $\mathcal{V}_\infty^\bullet$ (see Figure 2). To formally define this tree one could start by creating the vertex set $\mathcal{V}_\infty^\bullet$ as a subset of $\mathbb{N} \times \mathbb{Z} \times \mathcal{V}_\infty$ and then describing the edge set. However we think that the above informal construction is enough for our purpose. As we said, we want to represent T_n as a subtree of $\mathcal{U}_\infty^\bullet$. First we make clear what we call a subtree of $\mathcal{U}_\infty$ and $\mathcal{U}_\infty^\bullet$. Denote by $\mathcal{E}_\infty$ the edge set of $\mathcal{U}_\infty$ and $\mathcal{E}_\infty^\bullet$ the edge set of $\mathcal{U}_\infty^\bullet$.

▶ **Definition 6.** *A subtree t of $\mathcal{U}_\infty$ is a tree with vertex set included in $\mathcal{V}_\infty$ and edge set included in $\mathcal{E}_\infty$, such that the vertex $\emptyset$ belongs to t and such that there are no holes in t, meaning that: if $v = (a_1 \dots a_n)$ is a vertex of t with $a_n > 0$ then $(a_1 \dots a_n - 1)$ is also a vertex of t. Similarly a subtree t of $\mathcal{U}_\infty^\bullet$ is a tree with vertex set included in $\mathcal{V}_\infty^\bullet$ and edge set included in $\mathcal{E}_\infty^\bullet$, such that the vertex u_0 belongs to t and such that there are no holes in t (see Figure 5). A subtree of $\mathcal{U}_\infty^\bullet$ is rooted if the set $\{k \geq 0 \,|\, u_k \in t\}$ is finite. In this case the vertex u_k with maximal k in t is called the root of t.*

The representation of T_n as a subtree of $\mathcal{U}_\infty^\bullet$ will obviously depend on the distinguished vertex v_n since we want to look at the local structure around this vertex. More precisely, let T be a plane tree and v be a vertex of T. We identify the distinguished vertex v with the element u_0 and the root of T is identified with the element u_h where h is the graph distance between the root and v in T (it is the height of v). All the other vertices of T are identified such that the plane order is preserved (See Figure 3). We denote by (T, v) the subtree of $\mathcal{U}_\infty^\bullet$ obtained this way. We can now formally define the notion of convergence we use.

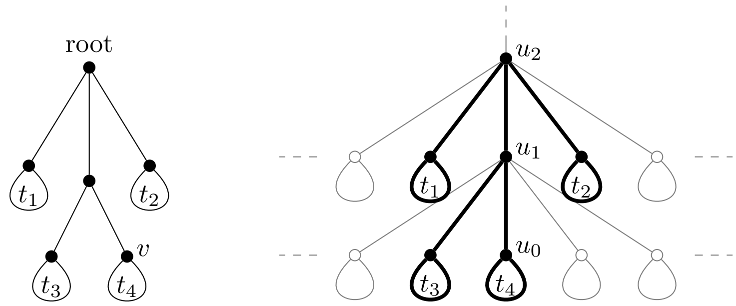

Figure 3 On the left, a finite tree T with a distinguished vertex v at distance 2 from the root. The loops t_1, t_2, t_3 and t_4 represent subtrees of T which can be seen as subtrees of the Ulam-Harris tree $\mathcal{U}_\infty$. On the right, the representation of T as a subtree of $\mathcal{U}_\infty^\bullet$.

▶ **Definition 7.** *Let t_n and t be subtrees of $\mathcal{U}_\infty^\bullet$ for all n. We say that (t_n) converges towards t and write $t_n \to t$ if for all $v \in \mathcal{V}_\infty^\bullet$,*

$$\mathbb{1}_{v \in t_n} \xrightarrow[n\to\infty]{} \mathbb{1}_{v\in t}.$$

This notion of convergence induces a topology that is metrizable and compact over the set of subtrees of $\mathcal{U}_\infty^\bullet$. Before stating the limit theorems, one needs to define the limiting trees T_1^*, T_2^* and T_3^* (seen as subtrees of $\mathcal{U}_\infty^\bullet$) that correspond, respectively, to regime 1, 2 and 3. Let T be a BGW tree with reproduction law π, given by (1), in regime 1 or 2. Let $\hat{\pi}$ be the probability measure on $\mathbb{N} \cup \{\infty\}$ given by

$$\hat{\pi}_k = k\pi_k \quad \forall k \in \mathbb{N} \quad \text{and} \quad \hat{\pi}_\infty = 1 - m$$

where $m := \min(1, \nu)$ is the mean of π.

- First we define T_1^* in regime 1. Notice that in this case $\hat{\pi}_\infty = 0$. We attach to u_0 an independent copy of T. For $k \geq 1$, u_k receives offspring according to an independent copy of $\hat{\pi}$. Then u_{k-1} is identified with a child of u_k chosen uniformly at random. Finally, we attach an independent copy of T to all the children of u_k, except u_{k-1} (see Figure 4).
- Now we define T_2^* in regime 2. In this case $\hat{\pi}_\infty > 0$. We attach to u_0 an independent copy of T. For $k \geq 1$, u_k receives offspring according to an independent copy of $\hat{\pi}$. Notice that almost surely, there exist $1 \leq i < j$ two integers such that $u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_{j-1}$ have a finite number of children and u_i and u_j have an infinite number of children. For every $k \in \{1, \ldots, i-1, i+1, \ldots, j-1\}$, u_{k-1} is identified with a child of u_k chosen uniformly at random, while u_i gets infinitely many children on the left and the right of its child u_{i-1}. Finally, for all $k \geq 1$, we attach an independent copy of T to all the children of u_k, except u_{k-1}. The tree T_2^* is the tree obtained by keeping all the descendants of u_{j-1} (see Figure 4).
- Finally, T_3^* is simply composed of the vertex u_1 having infinitely many children on the left and on the right of u_0, all of them, including u_0, being leaves (see Figure 4).

▶ **Theorem 8** (Stufler 2018). *Let (T_n, v_n) be a uniformly pointed simply generated tree with n vertices. Suppose that we are in regime $i = 1$ or in regime $i \in \{2, 3\}$ with complete condensation (meaning that (3) is satisfied). Then the convergence*

$$(T_n, v_n) \xrightarrow[n\to\infty]{(d)} T_i^*$$

holds in distribution for the topology induces by the convergence of Definition 7.

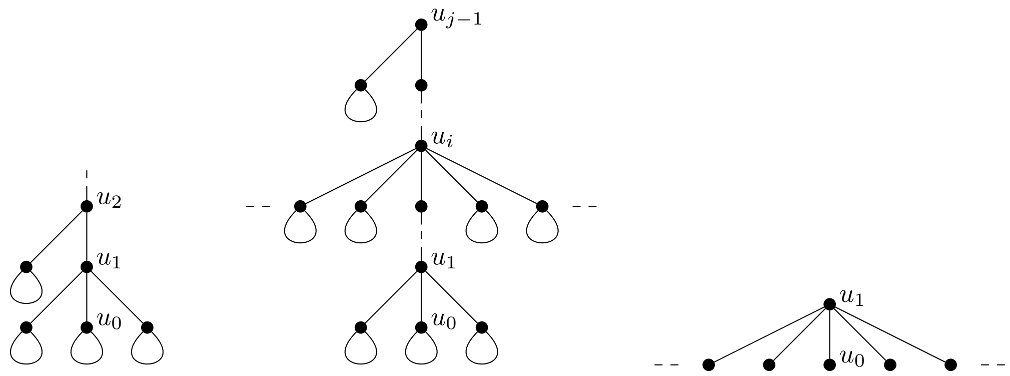

Figure 4 Representation of the trees T_1^*, T_2^* and T_3^*, respectively, from left to right. Each loop represents a copy of a Bienaymé-Galton-Watson tree of reproduction law π.

5 Properties of the tricolouration

In this section, we look at some general properties of the tricolouration defined in the introduction that will be essential to prove our main result. Let $T_1, \dots, T_n$ be n rooted finite trees. We define $T_1 * \dots * T_n$ the rooted tree obtained by creating an edge between each root of $T_1, \dots, T_n$ and a new vertex which will be the root of $T_1 * \dots * T_n$. In particular the number of vertices $\#V(T_1 * \dots * T_n)$ equals $1 + \#V(T_1) + \dots + \#V(T_n)$. And the number of edges $\#E(T_1 * \dots * T_n)$ equals $n + \#E(T_1) + \dots + \#E(T_n)$. We say that a rooted tree has colour c if the root has colour c.

▶ **Lemma 9.** *Let $T_1, \dots, T_n$ be n rooted finite trees. Set $T := T_1 * \dots * T_n$, then*
1. *T is red if $T_1, \dots, T_n$ are all non-red.*
2. *T is orange if exactly one tree among $T_1, \dots, T_n$ is red.*
3. *T is green if two or more trees among $T_1, \dots, T_n$ are red.*

Proof. Denote by $C(T)$ the size of a smallest covering of T. Notice that $C(T)$ is equal to $C(T_1) + \dots + C(T_n) + \Delta$ with $\Delta \in \{0, 1\}$. More precisely, $C(T) = C(T_1) + \dots + C(T_n)$ if and only if all the T_i's are non-red.
1. Suppose that $T_1, \dots, T_n$ are all non-red. We can take a smallest covering for each T_i such that the root of T_i is included in the covering. Then the union of these coverings gives a smallest covering of T. Moreover we can see that all the smallest coverings of T are obtained this way. Thus T is red.
2. Suppose than T_1 is red and $T_2, \dots, T_n$ are all non-red. We can take a smallest covering for each T_i in addition to the root of T. This gives a smallest covering of T, so T is either green or orange. We can also take a smallest covering for each T_i, $i > 1$, such that the root of T_i is included in the covering, a smallest covering of T_1 and the root of T_1. This also gives a smallest covering of T. Thus T is orange.
3. Suppose that T_1 and T_2 are red. As for the previous case, we can take a smallest covering for each T_i in addition to the root of T. This gives a smallest covering for T. But, as opposed to the previous case, all the smallest coverings of T are obtained this way. Thus T is green. ◀

If T_1, T_2 are finite trees and v_1, v_2 are vertices of, respectively, T_1 and T_2, then we denote by $(T_1, v_1) * (T_2, v_2)$ the tree obtained from T_1 and T_2 by drawing an edge between v_1 and v_2.

▶ **Lemma 10.** *With the same notation as above, set* $T := (T_1, v_1) * (T_2, v_2)$*. If* v_1 *is green in the tricolouration of* T_1 *then, the colour of every vertex* v *in the tricolouration of* T *is just the same as its colour in the tricolouration of* T_1 *(if* v *is a vertex of* T_1*) or* T_2 *(if* v *is a vertex of* T_2*). In other words, the tricoloured tree* T *is simply obtained by drawing an edge between* v_1 *and* v_2 *and keeping the colours of* T_1 *and* T_2*.*

Proof. Notice that, since v_1 is green in the tricolouration of T_1, a smallest covering of T_1 combined with a smallest covering of T_2 gives a smallest covering of T. Conversely a smallest covering of T is necessarily obtained by combining a smallest covering of T_1 and T_2. ◀

6 Tricolouration of the infinite limiting trees

In this section, we extend the definition of the tricolouration given in the introduction to the random infinite limiting trees T_1^*, T_2^* and T_3^* defined in Section 4. The initial definition applies only to finite trees since a covering of smallest size only makes sense in this context. Even though it seems not obvious to find a satisfactory definition of "smallest covering" for an infinite tree, it is still possible to describe a canonical way to tricolour the trees T_1^*, T_2^* and T_3^* using the properties found in Section 5.

Let t be a subtree of $\mathcal{U}_\infty^\bullet$ (finite or not) such that for every $k \geq 0$ and for every child v of u_k, distinct from u_{k-1}, v has a finite number of descendants. In other words, for all $k \geq 0$, all the children of u_k, distinct from u_{k-1}, are roots of finite trees. Since those trees are finite, it makes sense to consider their tricolouration. A *good* vertex of t is a vertex u_k with $k \geq 0$ such that, at least two of its children, distinct from u_{k-1}, are red in the tricolouration of the finite subtree they produce. If t is finite, then, from Lemma 9, a good vertex is a green vertex for the tricolouration of t. Notice that T_1^*, T_2^* and T_3^* satisfy the same hypothesis as t almost surely.

- We begin with the definition of the tricolouration of T_1^*. Almost surely, there exists an increasing sequence $(k_i)_i$ such that for all i, u_{k_i} is good. For all i, all the vertices below u_{k_i} (u_{k_i} included) get the same colour in T_1^* as their colour in the tricolouration of the subtree rooted at u_{k_i} (which is a finite tree). Notice that, from Lemma 9, u_{k_i} gets necessarily the colour green. Lemma 10 ensures that this way of colouring is consistent when taking larger i.
- For T_2^* we colour the unique vertex with infinite degree in green. Then, by cutting this vertex from T_2^* we obtain a (infinite) forest of finite trees who gets their induced tricolouration. Notice that, almost surely, the vertex with infinite degree is good.
- Lastly, all the leaves of T_3^* are coloured in red and the root u_1 is coloured in green.

We finish this section with the following lemma which explicitly gives the colour distribution of the vertex u_0 in T_i^*. This lemma will be useful when proving Theorem 4.

▶ **Lemma 11.** *Let* $p_i(c)$ *be the probability that* u_0 *has colour* c *in* T_i^**.*

1. *In regime* $i = 1$ *and* $i = 2$ *with complete condensation we have*

$$p_i(green) = \frac{1 - q + (1 - 2q)G'(1 - q)}{1 + G'(1 - q)}, \qquad p_i(orange) = \frac{2\,q\,G'(1 - q)}{1 + G'(1 - q)},$$

$$p_i(red) = \frac{q}{1 + G'(1 - q)}.$$

2. *In regime 3 with complete condensation we have that* $p_3(red) = 1$*.*

Proof. The case of regime 3 is obvious, let us focus on regime 1 and 2. Let T be a BGW tree with reproduction law π. Denote by q the probability that the root of T is red. From Lemma 9 we deduce that

$$q = \sum_{k\geq 0} \pi_k (1-q)^k = G(1-q).$$

Let $\widetilde{T}$ be the tree obtained from T_i^* by cutting the edge between u_0 and u_1 and keeping the component containing u_1. Let $\widetilde{q}$ be the probability that u_1 is red in $\widetilde{T}$. Then, from Lemma 9 again,

$$\widetilde{q} = \sum_{k\geq 1} k\pi_k (1-q)^{k-1}(1-\widetilde{q}) = (1-\widetilde{q})G'(1-q).$$

Finally,

$$p_i(\text{red}) = \sum_{k\geq 0} \pi_k (1-q)^k (1-\widetilde{q}) = \frac{q}{1+G'(1-q)}.$$

And

$$p_i(\text{orange}) = \sum_{k\geq 0} \pi_k (1-q)^k \widetilde{q} + \sum_{k\geq 1} k\pi_k (1-q)^{k-1} q(1-\widetilde{q}) = \frac{2qG'(1-q)}{1+G'(1-q)}.$$

Finally we deduce the value $p_i(\text{green})$ by the law of total probability. ◀

7 Proof of Theorem 4

All this section is devoted to the proof of Theorem 4. We keep all the notation of Theorem 4 and suppose that we are in regime $i = 1$ or in regime $i \in \{2, 3\}$ with complete condensation. Let c be a colour in {red, green, orange}. Recall that $p_i(c)$ is the probability that the vertex u_0 has colour c in the tree T_i^* in regime i. The idea is to prove the convergence of the first two moments of $n_c(T_n)/n$, namely

$$\frac{1}{n}\mathbb{E}\left[n_c(T_n)\right] \xrightarrow[n\to\infty]{} p_i(c) \quad \text{and} \quad \frac{1}{n^2}\mathbb{E}\left[n_c(T_n)^2\right] \xrightarrow[n\to\infty]{} p_i(c)^2.$$

Then, using Lemma 12, we will conclude that $n_c(T_n)/n$ converges in L^p towards $p_i(c)$ for all $p > 0$. Actually, the convergence of the second moment won't be required in regime 3. Recall that the explicit computation of $p_i(c)$ can be found in Lemma 11.

▶ **Lemma 12.** *Let (X_n) be a sequence of random variables with values in $[0, 1]$, and $\alpha \in [0, 1]$. Suppose that one of the following condition is satisfied.*

- *The convergences $\mathbb{E}[X_n] \to \alpha$ and $\mathbb{E}\left[X_n^2\right] \to \alpha^2$ hold when $n \to \infty$.*
- *The convergence $\mathbb{E}[X_n] \to \alpha$ holds when $n \to \infty$ and $\alpha = 1$.*

Then for all $p > 0$, (X_n) converges towards α in L^p.

Proof of Lemma 12. First, we show that (X_n) converges towards α in probability. Let $\varepsilon > 0$. In the first case we use Markov's inequality which gives

$$\mathbb{P}\left(|X_n - \alpha| \geq \varepsilon\right) \leq \frac{\mathbb{E}\left[(X_n-\alpha)^2\right]}{\varepsilon^2} \xrightarrow[n\to\infty]{} 0.$$

In the second case we notice that

$$\mathbb{E}[X_n] \leq (1-\varepsilon)\mathbb{P}(X_n \leq 1-\varepsilon) + \mathbb{P}(X_n > 1-\varepsilon) = 1 - \varepsilon\mathbb{P}(X_n \leq 1-\varepsilon).$$

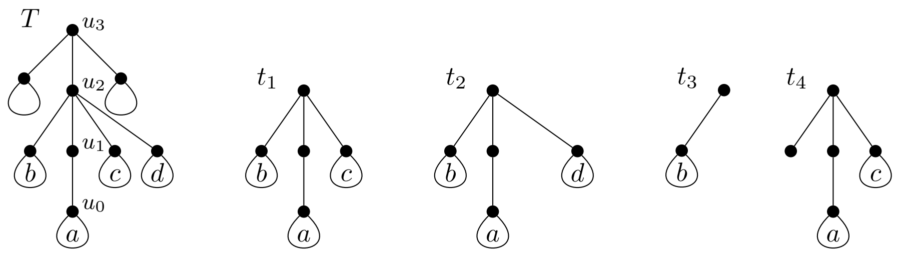

Figure 5 Illustration of Definitions 6 and 13. Only the tree t_1 satisfies $t_1 \preceq T$. The tree t_3 is not even a subtree of $\mathcal{U}_\infty^\bullet$ since it doesn't contain u_0. The tree t_2 is not a subtree of $\mathcal{U}_\infty^\bullet$ either since it has a hole between u_1 and the rightmost child of u_2. Finally t_4 is a subtree of $\mathcal{U}_\infty^\bullet$ but doesn't satisfy $t_4 \preceq T$ because the descendants of the left child of u_2 are missing.

Consequently $\lim_{n\to\infty} \mathbb{P}(X_n \leq 1-\varepsilon) = 0$ and the convergence in probability is shown in both cases. Second, let ℓ be an accumulation point of the sequence $(\mathbb{E}[|X_n-\alpha|^p])_n$ and $(n_k)_k$ be an extraction such that the convergence to ℓ occurs. From $(X_{n_k})_k$ we can extract a subsequence that converges almost surely to α. From the dominated convergence theorem we deduce that $\ell = 0$. ◀

Let v_n, v_n' be vertices chosen independently and uniformly in T_n. Notice that

$$\frac{1}{n}\mathbb{E}[n_c(T_n)] = \mathbb{P}(v_n \text{ has colour } c \text{ in } T_n) \quad \text{and}$$

$$\frac{1}{n^2}\mathbb{E}\left[n_c(T_n)^2\right] = \mathbb{P}(v_n \text{ and } v_n' \text{ have colour } c \text{ in } T_n).$$

Convergence of the first moment

First we prove the convergence of the first moment. Recall the notation of Section 6 when defining the tricolouration of the infinite trees T_1^*, T_2^* and T_3^*. Denote by $k \geq 0$ the first positive integer such that u_k is good ($k = 1$ almost surely in regime 3). Let τ_i^* be the subtree of $\mathcal{U}_\infty^\bullet$ obtained from T_i^* by cutting the edge between u_k and u_{k+1} and keeping the component containing u_0 ($\tau_3^* = T_3^*$ in regime 3). Note that, by construction, the tricolouration of τ_i^* is the restriction of its tricolouration in T_i^* and that u_k is green. The following definition introduces a useful order relation between trees.

▶ **Definition 13.** *Let T and t be subtrees of $\mathcal{U}_\infty^\bullet$ such that t is rooted at u_j for some $j \geq 0$. Suppose that u_j is also a vertex of T and denote by E_j the set of edges of T adjacent to u_j. We write $t \preceq T$ if there exists a subset $e_j \subset E_j$ such that t is the tree obtained from T by cutting all the edges from e_j and keeping the component containing u_j (see Figure 5).*

Let $\mathcal{F}$ be the set of rooted subtrees t of $\mathcal{U}_\infty^\bullet$ such that the root $u_j \in t$ is the only good vertex of t. For $t \in \mathcal{F}$ such that the root u_j of t has finite degree, denote by $v_\ell(t)$ (resp. $v_r(t)$) the leftmost (resp. rightmost) child of the root u_j of t. Let $\mathcal{F}_0$ be the set of elements $t \in \mathcal{F}$ such that: the root u_j of t has finite degree ; u_j has exactly two red neighbors distinct from u_{j-1} ; and ($v_\ell(t) = u_{j-1}$ or $v_\ell(t)$ is red) and ($v_r(t) = u_{j-1}$ or $v_r(t)$ is red). Notice that if t_1 and t_2 are distinct elements of $\mathcal{F}_0$, then we can't have $t_1 \preceq t_2$ nor $t_2 \preceq t_1$. Moreover for every $t_1 \in \mathcal{F}$ there exists a unique $t_2 \in \mathcal{F}_0$ such that $t_2 \preceq t_1$. In other words $\mathcal{F}_0$ is the set of equivalence classes for the equivalence relation $t_1 \sim t_2$ iff $t_1 \preceq t_2$ or $t_2 \preceq t_1$. Notice that almost surely $\tau_i^* \in \mathcal{F}$.

Fix $\varepsilon > 0$. Let $\mathcal{T}$ be a finite subset of $\mathcal{F}_0$ such that the event $\{\mathcal{T} \preceq \tau_i^*\} := \{\exists t \in \mathcal{T},\, t \preceq \tau_i^*\}$ happens with probability at least $1-\varepsilon$. Let $\mathcal{T}(c)$ be the set of trees $t \in \mathcal{T}$ such that u_0 has colour c in t. Remember that we see (T_n, v_n) as a subtree of $\mathcal{U}_\infty^\bullet$. For all $t \in \mathcal{T}$, define the event $A_n(t) := \{t \preceq (T_n, v_n)\}$. Using Theorem 8, we have that for all $t \in \mathcal{T}$

$$\mathbb{P}\left(A_n(t)\right) \xrightarrow[n\to\infty]{} \mathbb{P}\left(t \preceq T_i^*\right). \tag{4}$$

The properties of τ_i^* and $t \in \mathcal{T}$ imply that $t \preceq T_i^*$ if and only if $t \preceq \tau_i^*$. Thus

$$\mathbb{P}\left(t \preceq T_i^*\right) = \mathbb{P}\left(t \preceq \tau_i^*\right).$$

Denote by E_n the event $\cup_{t\in\mathcal{T}} A_n(t)$. Notice that the event $E_n \cap \{v_n \text{ has colour } c \text{ in } T_n\}$ is equal to the event $\cup_{t\in\mathcal{T}(c)} A_n(t)$. It is a consequence of Lemma 9 and 10. Notice also that for t_1, t_2 distinct trees of $\mathcal{T}$, $A_n(t_1)$ and $A_n(t_2)$ are disjoint for all n. Consequently,

$$\mathbb{P}\left(\{v_n \text{ has colour } c \text{ in } T_n\} \cap E_n\right) = \sum_{t\in\mathcal{T}(c)} \mathbb{P}\left(A_n(t)\right) \xrightarrow[n\to\infty]{} \mathbb{P}\left(\mathcal{T}(c) \preceq \tau_i^*\right).$$

Since u_0 has colour c in τ_i^* if and only if u_0 has colour c in T_i^*,

$$\mathbb{P}\left(\mathcal{T}(c) \preceq \tau_i^*\right) = \mathbb{P}\left(u_0 \text{ has colour } c \text{ in } T_i^* \text{ and } \mathcal{T} \preceq \tau_i^*\right) \in [p_1(c)-\varepsilon, p_1(c)+\varepsilon].$$

Finally, using Theorem 8 again, one has

$$\mathbb{P}\left(E_n\right) \xrightarrow[n\to\infty]{} \mathbb{P}\left(\mathcal{T} \preceq \tau_i^*\right) \geq 1-\varepsilon.$$

The convergence $\mathbb{E}\left[n_c(T_n)/n\right] \to p_1(c)$ readily follows.

Convergence of the second moment in regime 1 and 2

The next step is to show convergence for the second moment in regime $i=1$ or $i=2$ with complete condensation. Let c' be another colour in {red, green, orange}. We will actually show that

$$\frac{1}{n^2}\mathbb{E}\left[n_c(T_n)n_{c'}(T_n)\right] \xrightarrow[n\to\infty]{} p_i(c)p_i(c').$$

We keep the notation of the previous part which shows the convergence of the first moment. For all $t \in \mathcal{T}$, let $A_n'(t) := \{t \preceq (T_n, v_n')\}$ and $E_n' := \cup_{t\in\mathcal{T}} A_n'(t)$. We have,

$$\mathbb{P}\left(\{v_n \text{ has colour } c \text{ in } T_n\} \cap E_n \cap E_n'\right) = \sum_{t\in\mathcal{T}(c)} \sum_{t'\in\mathcal{T}(c')} \mathbb{P}\left(A_n(t) \cap A_n'(t')\right). \tag{5}$$

Fix $t, t' \in \mathcal{T}$. Recall that the trees t, t' and T_n are rooted plane trees, thus we can consider their so-called Łukasiewicz walk. More precisely, let T be a rooted plane tree with n vertices and w be a vertex of T. Denote by $\ell(w,T)$ the rank of w in T for the lexicographic order. Equivalently, w is the $\ell(w,T)$-th vertex of T explored by the depth first search starting from the root of T. Let $w_1, \ldots, w_n$ be the vertices of T ordered according to the lexicographic order (so $\ell(w_i,T) = i$ for all i). The *Łukasiewicz walk* associated with T is the sequence $(s_k)_{1\leq k\leq n}$ such that $s_0 = 0$ and $s_k - s_{k-1} + 1$ is the out-degree of w_k for all $k \in \{1,\ldots,n\}$. An important property of the Łukasiewicz walk is that it uniquely encodes its tree, meaning that the tree T can be retrieved from $(s_k)_{1\leq k\leq n}$. Let $(S_k^{(n)})_{0\leq k\leq n}$ be the Łukasiewicz walk associated with the tree T_n. Let $X_1, \ldots, X_n, \ldots$ be i.i.d random variables such that $\mathbb{P}\left(X_1 = m\right) = \pi_{m+1}$ for all integer $m \geq -1$ and set $S_k := \sum_{i=1}^k X_i$ for all $k \geq 0$. It is well known that, the random walk $(S_k)_{0\le k\le n}$, starting at 0 and conditioned on reaching -1 for the first time at time n, has the same law as $(S_k^{(n)})_{0\le k\le n}$. Let $m := |t|$ be the number of vertices of t and $m' := |t'|$. Let $(s_k)_{0\le k\le m}$ and $(s'_k)_{0\le k\le m'}$ be, respectively, the Łukasiewicz walks associated with t and t' and denote by $x_k := s_{k+1} - s_k$ and $x'_k := s'_{k+1} - s'_k$ the associated steps. Write $k_0 := \ell(u_0, t)$, $k'_0 := \ell(u_0, t')$, $i_n := \ell(v_n, T_n)$ and $i'_n := \ell(v'_n, T_n)$. The indices i_n and i'_n are independent random elements of $\{1, \dots, n\}$ with uniform distribution. The event $A_n(t)$ happens if and only if the Łukasiewicz walk $(S_k^{(n)})_{0\le k\le n}$ coincides with $(s_k)_{0\le k\le m}$, up to a vertical shifting, on the interval $[\![i_n - k_0, i_n + m - k_0]\!]$. The same goes for $A'_n(t')$. More precisely

$$A_n(t) \cap A'_n(t') = \{X^{(n)}_{i+i_n-k_0} = x_i \ \forall i \in [\![1, m]\!] \text{ and } X^{(n)}_{i+i'_n-k'_0} = x'_i \ \forall i \in [\![1, m']\!]\}.$$

Applying the reverse Vervaat transform, we can change the initial excursion type conditioning into a bridge type conditioning (see e.g. [18, Sec. 6.1]). Namely

$$\begin{aligned}&\mathbb{P}\left(A_n(t) \cap A'_n(t')\right)\\ &\qquad = \mathbb{P}\left(X_{i+i_n-k_0} = x_i \ \forall i \in [\![1, m]\!] \text{ and } X_{i+i'_n-k'_0} = x'_i \ \forall i \in [\![1, m']\!] \,\middle|\, X_n = -1\right).\end{aligned}$$

Denote by D_n the event $\{[\![i_n - k_0 + 1, i_n - k_0 + m]\!] \cap [\![i'_n - k'_0 + 1, i'_n - k'_0 + m']\!] = \emptyset\}$. This event has a probability tending to 1 and one can see that

$$\begin{aligned}&\mathbb{P}\left(A_n(t) \cap A'_n(t') \cap D_n\right)\\ &\qquad = \mathbb{P}\left(X_i = x_i \ \forall i \in [\![1, m]\!] \text{ and } X_{i+m} = x'_i \ \forall i \in [\![1, m']\!] \,\middle|\, X_n = -1\right)\mathbb{P}\left(D_n\right).\end{aligned}$$

According to [11, Thm. 11.7] the steps $X_1, \dots, X_{m+m'}$ conditioned on $\{X_n = -1\}$ are asymptotically independent and the conditioning fades for large values of n, consequently

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}\left(A_n(t) \cap A'_n(t')\right) &= \mathbb{P}\left(X_i = x_i \ \forall i \in [\![1, m]\!]\right)\mathbb{P}\left(X_i = x'_i \ \forall i \in [\![1, m']\!]\right)\\ &= \lim_{n\to\infty} \mathbb{P}\left(A_n(t)\right)\mathbb{P}\left(A'_n(t')\right).\end{aligned}$$

Finally, using (4) and (5), we have that

$$\lim_{n\to\infty} \mathbb{P}\left(\{v_n \text{ has colour } c \text{ in } T_n\} \cap E_n \cap E'_n\right) = \mathbb{P}\left(\mathcal{T}(c) \preceq \tau_i^*\right)\mathbb{P}\left(\mathcal{T}(c) \preceq \tau_i^*\right).$$

and the result follows.

References

1. F. Araujo, J. Farinha, P. Domingues, G.C. Silaghi, and D. Kondo. A Maximum Independent Set Approach for Collusion Detection in Voting Pools. *Journal of Parallel and Distributed Computing*, 71(10):1356–1366, 2011.
2. C. Banderier, M. Kuba, and A. Panholzer. Analysis of three graph parameters for random trees. *Random Structures & Algorithms*, 35(1):42–69, 2009.
3. B. Bollobás. The independence ratio of regular graphs. *Proceedings of the American Mathematical Society*, 83(2):433–436, 1981.
4. S. Butenko, P. Pardalos, I. Sergienko, V. Shylo, and P. Stetsyuk. Finding maximum independent sets in graphs arising from coding theory. In *Proceedings of the 2002 ACM Symposium on Applied Computing*, SAC '02, pages 542–546, New York, NY, USA, 2002. Association for Computing Machinery.
5. F. Chapoton. On some varieties associated with trees. *Michigan Mathematical Journal*, 64(4):721–758, 2015.

6 S. Coulomb. Minimal vertex covers of random trees. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(06):P06007, June 2005.

7 S. Coulomb and M. Bauer. On vertex covers, matchings and random trees, 2004. `arXiv:math/0407456`.

8 L. Devroye, M.K. Goh, and R.Y. Zhao. The independence number of a Bienaymé-Galton-Watson tree and related parameters, 2021. `arXiv:2106.14389`.

9 M. Fuchs, C. Holmgren, D. Mitsche, and R. Neininger. A note on the independence number, domination number and related parameters of random binary search trees and random recursive trees. *Discrete Applied Mathematics*, 292:64–71, 2021.

10 M.R. Garey and D.S. Johnson. *Computers and intractability*, volume 174. freeman San Francisco, 1979.

11 S. Janson. Simply generated trees, conditioned Galton–Watson trees, random allocations and condensation. *Probability Surveys*, 9(none):103–252, 2012.

12 S. Janson. On the independence number of some random trees. *Electronic Communications in Probability*, 25(none):1–14, 2020.

13 C. Joo, X. Lin, J.Ryu, and N.B. Shroff. Distributed greedy approximation to maximum weighted independent set for scheduling with fading channels. In *Proceedings of the Fourteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, MobiHoc '13, pages 89–98, New York, NY, USA, 2013. Association for Computing Machinery.

14 I. Kortchemski. Limit theorems for conditioned non-generic Galton–Watson trees. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 51(2):489–511, 2015.

15 D. Lokshtanov, M. Vatshelle, and Y. Villanger. Independent set in p5-free graphs in polynomial time. *Proceedings of SODA*, pages 570–581, January 2014.

16 A. Meir and J. W. Moon. On the altitude of nodes in random trees. *Canadian Journal of Mathematics*, 30:997–1015, 1978.

17 G.J. Minty. On maximal independent sets of vertices in claw-free graphs. *Journal of Combinatorial Theory, Series B*, 28(3):284–304, 1980.

18 J Pitman. Combinatorial stochastic processes. *Lecture Notes in Mathematics*, 1875, 2006. Springer, New York.

19 J.M. Robson. Algorithms for maximum independent sets. *Journal of Algorithms*, 7(3):425–440, 1986.

20 B. Stufler. Local limits of large Galton–Watson trees rerooted at a random vertex. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 55(1):155–183, 2019.

21 M. Xiao and H. Nagamochi. Confining sets and avoiding bottleneck cases: A simple maximum independent set algorithm in degree-3 graphs. *Theor. Comput. Sci.*, 469:92–104, 2013.

22 M. Xiao and H. Nagamochi. Exact algorithms for maximum independent set. *Information and Computation*, 255:126–146, 2017.

23 J. Zito. The structure and maximum number of maximum independent sets in trees. *Journal of Graph Theory*, 15(2):207–221, 1991.