UNICELLULAR MAPS VS HYPERBOLIC SURFACES IN LARGE GENUS: SIMPLE CLOSED CURVES

SVANTE JANSON AND BAPTISTE LOUF

Abstract. We study uniformly random maps with a single face, genus g, and size n, as $n, g \to \infty$ with $g = o(n)$, in continuation of several previous works on the geometric properties of "high genus maps". We calculate the number of short simple cycles, and we show convergence of their lengths (after a well-chosen rescaling of the graph distance) to a Poisson process, which happens to be exactly the same as the limit law obtained by Mirzakhani and Petri (2019) when they studied simple closed geodesics on random hyperbolic surfaces under the Weil–Petersson measure as $g \to \infty$.

This leads us to conjecture that these two models are somehow "the same" in the limit, which would allow to translate problems on hyperbolic surfaces in terms of random trees, thanks to a powerful bijection of Chapuy, Féray and Fusy (2013).

1. Introduction

1.1. Combinatorial maps.

1.1. Combinatorial maps. Maps are defined as gluings of polygons forming a (compact, connected, oriented) surface. They have been studied extensively in the past 60 years, especially in the case of planar maps, i.e., maps of the sphere. They were first approached from the combinatorial point of view, both enumeratively, starting with [34], and bijectively, starting with [32].

More recently, relying on previous combinatorial results, geometric properties of large random maps have been studied. More precisely, one can study the geometry of random maps picked uniformly in certain classes, as their size tends to infinity. In the case of planar maps, this culminated in the identification of two types of "limits" (for two well defined topologies on the set of planar maps): the local limit (the *UIPT*[1] [2]) and the scaling limit (the *Brownian map* [18; 24]).

All these works have been extended to maps with a fixed genus $g > 0$ [3; 9; 4].

1.2. High genus maps. Very recently, another regime has been studied: *high genus maps* are defined as (sequences of) maps whose genus grow linearly in the size of the map. They have a negative average discrete curvature, and can therefore be considered as a discrete model of hyperbolic geometry. Their geometric properties have been studied, first on unicellular maps

Date: 23 November, 2021; revised 10 December, 2021.
Supported by the Knut and Alice Wallenberg Foundation.

[1] In the case of triangulations, i.e., maps made out of triangles.

[1; 31; 22; 15] (i.e., maps with one face), and shortly after on more general models of maps [6; 7; 21].

1.3. **Our results.** While all these works focuses on the regime where g grows linearly in n, we are here interested in the slightly different regime where $g \to \infty$ but $g = o(n)$. We will study the distribution of lengths of simple cycles in unicellular maps (which we studied in the "linear genus regime" in a previous work [15]). The main interest here is that, with the right rescaling of the graph distance, our result matches exactly a result of Mirzakhani and Petri [26] on random hyperbolic surfaces, which leads us to conjecture that these random hyperbolic surfaces can in some sense be approximated by unicellular maps (see Section 1.4 for more details).

Let $\mathbf{U}_{n,g}$ be a uniform unicellular map of genus g and size n, and set

$$L_n := \sqrt{\frac{n}{12g}}, \tag{1.1}$$

which will turn out to be the typical order of the size of the smallest cycles.

Theorem 1.1. *Suppose that $n \to \infty$ and that $g = g_n \to \infty$ with $g = o(n)$. Let $\{\zeta_i\}$ be the set of simple cycles in $\mathbf{U}_{n,g}$, and consider the (multi)set of their lengths $Z_i := |\zeta_i|$, scaled as $\Xi_n := \bigl\{Z_i/L_n\bigr\} = \bigl\{(12g/n)^{1/2}Z_i\bigr\}$. Then the random set Ξ_n, regarded as a point process on $[0,\infty)$, converges in distribution to a Poisson process on $[0,\infty)$ with intensity $(\cosh t - 1)/t$.*

For a background on point processes, see e.g. [16, Chapter 12 and 16] or [17]. The convergence to a Poisson process in Theorem 1.1 can be expressed in several, equivalent forms. One equivalent version is the following, stated similarly to the main result of Mirzakhani and Petri [26].

Theorem 1.2. *Let $\mathcal{C}_n^{x,y}$ be the number of simple cycles of $\mathbf{U}_{n,g}$ whose length belongs to $[xL_n, yL_n]$. Then, for every finite set of disjoint intervals $[x_1, y_1]$, $[x_2, y_2]$,...,$[x_k, y_k]$, the random variables $\mathcal{C}_n^{x_i,y_i}$ converge in distribution, as $n \to \infty$, to independent Poisson variables with parameters $\lambda(x_i, y_i)$ where*

$$\lambda(x,y) = \int_x^y \frac{\cosh t - 1}{t}\,\mathrm{d}t. \tag{1.2}$$

For comparison, we state the theorem by Mirzakhani and Petri [26].[2] (See Section 1.4 and the references there for definitions.)

Theorem 1.3. *[Mirzakhani–Petri [26]] Let $\widehat{\mathcal{C}}_g^{x,y}$ be the number of simple closed geodesics in the random hyperbolic surface $\mathbf{S}_g$ whose lengths belong to $[x,y]$. Then, for every finite set of disjoint intervals $[x_1,y_1], [x_2,y_2], \dots,$ $[x_k,y_k]$, the random variables $\widehat{\mathcal{C}}_g^{x_i,y_i}$ converge jointly in distribution, as $g \to \infty$, to independent Poisson variables with parameters $\lambda(x_i,y_i)$ where $\lambda(x,y)$ is given by (1.2).*

[2]The theorem in [26] is stated for primitive closed geodesics, but it follows from the proof there that whp every primitive closed geodesic with length $\leq C$ is simple, and thus the same result holds for simple closed geodesics. The same holds in our Theorem 1.2, see Remark 5.9.

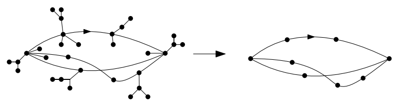

FIGURE 1. A unicellular map and its 2-core.

Another equivalent version of Theorem 1.1 is that if we order the cycles according to increasing length, so that $Z_1 \leq Z_2 \leq \dots$, and extend the sequence $(Z_1, Z_2, \dots)$ to an infinite one by adding a tail $\infty, \infty, \dots$, then the resulting sequence, after rescaling as above, converges to the sequence of points in the (inhomogeneous) Poisson process defined above, in the usual product topology on $[0, \infty)^\infty$. (See, e.g., [14, Lemma 4].) In particular, this yields the following corollary, cf. [26, Theorem 5.1].

Corollary 1.4. *Let $Z_1^{(n)}$ be the length of the shortest cycle in $\mathbf{U}_{n,g}$. Then,*

$$Z_1^{(n)}/L_n \xrightarrow{\mathrm{d}} Z, \tag{1.3}$$

where Z is a random variable with the distribution function

$$\mathbb{P}(Z \leq z) = 1 - \exp\Bigl(-\int_0^z \frac{\cosh t - 1}{t}\, \mathrm{d}t\Bigr), \qquad z \geq 0. \tag{1.4}$$

In unicellular maps, all simple cycles are non-contractible; hence $Z_1^{(n)}$ is the law of the *systole* (the size of the smallest non-contractible cycle) of $\mathbf{U}_{n,g}$.

1.4. **A conjecture.** For $g \geq 2$, there is a natural way of defining a random hyperbolic surface $\mathbf{S}_g$ of genus g: the *Weil–Petersson probability measure* (we refer to [26] and references therein for more details). It is natural to try to understand the geometric behaviour of these random hyperbolic surfaces as $g \to \infty$, and this has been done rather extensively in the recent years [13; 25; 26; 27; 29; 33; 36; 30; 28; 37; 12; 19].

The similarity between the geometric behaviour of maps and hyperbolic surfaces had been noticed before, but in the precise regime considered in this paper, the "numerical evidence" provided by Theorems 1.2 and 1.3 leads us to conjecture that these two models are somehow "the same" as $g \to \infty$. We do not know yet what the exact formulation should be, but two features seem to be necessary for this conjecture to be true: first, one should remove the "fractal part" of the map, and consider the 2-core of $\mathbf{U}_{n_g,g}$ instead (see Fig. 1). It is also important to consider 2-core($\mathbf{U}_{n_g,g}$) as a hyperbolic polygon whose sides were glued, and not just an embedded graph (see Section 2.4 for a proper definition of unicellular maps as gluings of a polygon). We can now conjecture the following.

Conjecture 1.5. Let n_g be such that $g = o(n_g)$ as $g \to \infty$. Then, $\mathbf{U}_{n_g,g}$, with distances rescaled by the factor $L_{n_g}^{-1}$, and $\mathbf{S}_g$ can be coupled such that

$$d_{\mathrm{GH}}\left(\text{2-core}(\mathbf{U}_{n_g,g}), \mathbf{S}_g\right) \xrightarrow[g\to\infty]{} 0 \tag{1.5}$$

in probability, where d_{GH} is the Gromov–Hausdorff distance between metric spaces.

There may be small adjustments to make to this conjecture: for instance, maybe a different notion of distance than Gromov–Hausdorff is needed, and on the other hand, we can hope for something stronger that accounts for the topology, e.g. separating curves. Furthermore, there might be a slightly simpler equivalent combinatorial model (for instance, cubic unicellular maps with random edge lengths). We leave these questions open for now.

We believe the conjecture it is an interesting question for two reasons. First, it would reinforce the "universality" principle in two-dimensional geometry (i.e., different models behave the same). And what's more, if this conjecture is true, any geometric property that we prove on our model of unicellular maps would hold for hyperbolic surfaces in large genus. But unicellular maps are easier to work with especially because they are in bijection with a certain model of trees [8], see Section 2.5 below. Therefore, Conjecture 1.5, if true, would allow us to transfer any geometric problem on hyperbolic surfaces onto a problem on *random trees*, which are very well understood.

There are several open questions remaining for random hyperbolic surfaces, and the conjecture above (whether true or not) suggests studying the corresponding problems on $\mathbf{U}_{n_g,g}$. Perhaps the most natural such question is about the diameter:

Open problem 1.6. What is the diameter of $\mathbf{U}_{n_g,g}$ (rescaled by $L_{n_g}^{-1}$ as above) ?

For the diameter of random hyperbolic surfaces, a simple area argument gives a deterministic lower bound of $(1+o(1))\log g$, while, so far, the best upper bound is $(4+o(1))\log g$ whp, which can be derived from an inequality linking the diameter to the spectral gap (see [23], combined with [37; 20]); it is natural to try and find the 'right constant" in front of $\log g$.

Several spectral properties of $\mathbf{S}_g$ are still open problems, and might be more tractable on $\mathbf{U}_{n_g,g}$ (and the associated model of random trees):

Open problem 1.7. Study the spectral gap, the Cheeger constant and Laplacian eigenfunctions of $\mathbf{U}_{n_g,g}$.

1.5. **Structure of the paper.** We will end this section with an index of notations, and we will give some definitions in Section 2. In Section 3, we prove some results about C-permutations, and in Section 4, we study the number of occurrences of paths in uniformly random trees. We use these results in Section 5 to calculate the law of cycles in unicellular maps.

Acknowledgements. We are grateful to Bram Petri and Stephan Wagner for enlightening discussions. We also thank Thomas Budzinski, Nicolas Curien and Yunhui Wu for comments on the first version of this paper.

Index of notations. (Not including some that are only used locally.)

- $g = g_n$: the genus of the map. (We assume $1 \ll g_n \ll n$.)
- $\mathbf{U}_{n,g}$: a uniformly random unicellular map of genus g and size n.
- $\mathbf{T} = \mathbf{T}_n$: a uniformly random tree of size n.
- $(T_n)_{n\geq 1}$: a deterministic sequence of trees.
- $L_n := \sqrt{\frac{n}{12g}}$: the scaling factor for the graph distance.

- M: a large integer. (Usually fixed.)
- $\mathfrak{S}^{\mathsf{C}}_{n,m}$: the set of C-permutations on n elements and m cycles.
- $\boldsymbol{\sigma}$: a uniformly random element in $\mathfrak{S}^{\mathsf{C}}_{n+1,n+1-2g}$. (Depends thus implicitly on n.)
- T: a fixed rooted tree.
- $\mathbf{t}$: a fixed rooted tree.
- $N_{\mathbf{t}}(T)$: the number of occurrences of $\mathbf{t}$ in T.
- $P_i(T)$: the number of paths of length $\ell \in [\frac{i}{M}L_n, \frac{i+1}{M}L_n)$ in T.
- $\mathbf{m}$: a finite sequence of non-negative integers.
- $\mathbf{P}$: a list of pairwise (vertex) disjoint paths.
- $s(\mathbf{P})$: the number of paths in $\mathbf{P}$.
- $\ell(\mathbf{P})$: the total length of the paths in $\mathbf{P}$.
- $\mathfrak{P}(T)$, $\mathfrak{P}^{[\mathbf{m}]}(T)$, $\mathfrak{P}^{[m]}_k(T)$ $\mathfrak{C}^{[m]}_k(T,\sigma)$, $\mathfrak{C}^{[a,b)}_k(T,\sigma)$, $\widetilde{\mathfrak{C}}^{[m]}_k(T,\sigma)$: sets of lists of disjoint paths in T.
- $P^{[\mathbf{m}]}(T)$, $P^{[m]}_k(T)$, $C^{[m]}_k(T,\sigma)$, $C^{[a,b)}_k(T,\sigma)$, $\widetilde{C}^{[m]}_k(T,\sigma)$: cardinalities of these sets.
- $\kappa^{(\mathbf{m})}$: the constant (2.6).

2. Definitions and notations

2.1. **Parameters.** We will discretize our problem in order to be able to reason on a finite number of quantities. For most of the proof, we will fix a (large) integer $M > 0$. Only in Section 5.4 we will let $M \to \infty$, which eventually will yield our final results. For notational convenience, we will usually omit n and M from the notation when there is no risk of confusion, but it should be remembered that most variables introduced below depend on both n and M.

Recall that L_n was defined in (1.1). Note that, by our assumptions on $g = g_n$, we have $L_n \to \infty$ and $L_n = o(n^{1/2})$. We define also

$$L^\bullet := (\log g)L_n. \tag{2.1}$$

The exact definition is not important; we will only use the properties $L_n \ll L^\bullet \ll n^{1/2}$ as $n \to \infty$.

2.2. **Paths, cycles and trees.** By a path p, we mean a simple path, i.e., a list of $\ell + 1$ distinct vertices $v_0, \dots, v_\ell$ and $\ell \geq 1$ edges $v_{i-1}v_i$, where ℓ is the *length* or *size* of the path, denoted $|p|$. (Note that we require $|p| > 0$.) All our paths are *oriented*, i.e., they have a start $\mathrm{start}(p) = v_0$ and an end $\mathrm{end}(p) = v_\ell$, which together are the *endpoints* $\mathrm{Ext}(p) := \{\mathrm{start}(p), \mathrm{end}(p)\}$.

Similarly, a cycle means a simple cycle, i.e., a set of ℓ distinct vertices $v_1, \dots, v_\ell$ and $\ell \geq 2$ edges v_iv_{i+1} (where $v_{\ell+1}$ is interpreted as v_1), where ℓ is the *length* or *size* of the cycle. Our cycles are unoriented, and they do not have any designated starting point; thus the vertices $v_1, \dots, v_\ell$ can be ordered in 2ℓ different ways yielding the same cycle.

Our trees will be plane trees, i.e., trees embedded in the plane (up to obvious isomorphism). The size $|\mathbf{t}|$ of a tree $\mathbf{t}$ is its number of edges. At each vertex v of $\mathbf{t}$, the gaps between two adjacent edges are called *corners*; thus, there are d corners at a vertex of degree d, and hence in total $2|\mathbf{t}|$ corners in a tree $\mathbf{t}$.

Our trees are usually rooted; the root of a tree is a corner. (This is equivalent to the slightly different definition of rooted plane trees in e.g. [10, Section 1.1.2].)

We emphasize that the size of a path, cycle or tree is its number of edges.

Let T be a rooted tree. For any tree $\mathbf{t}$, let $N_{\mathbf{t}}(T)$ be the number of occurrences of $\mathbf{t}$ in T. Furthermore, let

$$P_i(T) := \sum_{\mathbf{t}} N_{\mathbf{t}}(T), \qquad i \geq 0, \tag{2.2}$$

where the sum spans over all paths of size belonging to $[\frac{i}{M}L_n, \frac{i+1}{M}L_n)$. (See Section 2.1 for the (implicit) parameter M and L_n.)

We denote by $\mathcal{T}_n$ the set of rooted plane trees of size n, and by $\mathbf{T} = \mathbf{T}_n$ a uniformly random element of $\mathcal{T}_n$.

2.3. **Lists of paths.** Given a rooted plane tree T, let $\mathfrak{P}(T)$ be the set of all lists $\mathbf{P} = (p_1, \dots, p_k)$ of pairwise vertex disjoint paths in T, of arbitrary length $k \geq 1$. For a list $\mathbf{P} = (p_1, \dots, p_k) \in \mathfrak{P}(T)$, let $s(\mathbf{P}) := k$, the number of paths in the list, and $\ell(\mathbf{P}) := \sum_1^k |p_i|$, their total length. Also, let $\operatorname{Ext}(\mathbf{P}) := \bigcup_i \operatorname{Ext}(p_i) = \{\operatorname{start}(p_i), \operatorname{end}(p_i) : i = 1, \dots, k\}$, the set of endpoints of the paths in $\mathbf{P}$; note that $|\operatorname{Ext}(\mathbf{P})| = 2s(\mathbf{P})$ since the paths are disjoint.

Furthermore, let $\mathcal{M}$ be the set of all (non-empty) finite sequences of non-negative integers. If $\mathbf{m} = (m_1, \dots, m_k) \in \mathcal{M}$, we write $|\mathbf{m}| = m_1 + m_2 + \dots + m_k$ and $s(\mathbf{m}) = k \geq 1$. We define

$$\mathfrak{P}^{[\mathbf{m}]}(T) := \Big\{\mathbf{P} = (p_1, \dots, p_k) \in \mathfrak{P}(T) : |p_i| \in \Big[\frac{m_i}{M}L_n, \frac{m_i+1}{M}L_n\Big) \, \forall i\Big\}, \tag{2.3}$$

$$\mathfrak{P}_k^{[m]}(T) := \bigcup_{\substack{|\mathbf{m}|=m \\ s(\mathbf{m})=k}} \mathfrak{P}^{[\mathbf{m}]}(T). \tag{2.4}$$

We let $P^{[\mathbf{m}]}(T) := |\mathfrak{P}^{[\mathbf{m}]}(T)|$ and $P_k^{[m]}(T) := |\mathfrak{P}_k^{[m]}(T)|$ be the cardinalities of these sets of lists.

Note that it follows from the definition (2.3) that if $\mathbf{P} \in \mathfrak{P}^{[\mathbf{m}]}(T)$, then

$$\frac{|\mathbf{m}|}{M}L_n \leq \ell(\mathbf{P}) \leq \frac{|\mathbf{m}| + s(\mathbf{m})}{M}L_n. \tag{2.5}$$

Define also, for $\mathbf{m} = (m_1, \dots, m_k) \in \mathcal{M}$ as above,

$$\kappa^{(\mathbf{m})} = \prod_{i=1}^{s(\mathbf{m})} (2m_i + 1). \tag{2.6}$$

2.4. **Unicellular maps.** A unicellular map of *size* n is a $2n$-gon whose sides were glued two by two to form a (compact, connected, oriented) surface. The *genus* of the map is the genus of the surface created by the gluings (its number of handles). After the gluing, the sides of the polygon become the *edges* of the map, and the vertices of the polygon become the *vertices* of the map. Note that the number of edges equals the size n. By Euler's formula, a unicellular map of genus g and size n has $n+1-2g$ vertices. As for trees, the gaps between two adjacent (half-)edges around a vertex are called *corners*, and there are $2n$ corners in a unicellular map of size n. The underlying graph of a unicellular map is the graph obtained from this map by only remembering its edges and vertices. (In general, this is a multigraph.)

We consider in this paper only *rooted* unicellular maps, where a corner is marked as the *root*. The underlying graph is then a rooted graph.

A rooted unicellular map of genus 0 is the same as a rooted plane tree.

We denote by $\mathbf{U}_{n,g}$ a uniformly random unicellular map of size n and genus g.

2.5. **C-permutations and C-decorated trees.** A *C-permutation* is a permutation whose cycles are of odd length. Let $\mathfrak{S}_n^{\mathsf{C}}$ be the set of C-permutations of length n, and $\mathfrak{S}_{n,m}^{\mathsf{C}}$ the subset of permutations in $\mathfrak{S}_n^{\mathsf{C}}$ with exactly m cycles. (This is empty unless $n \equiv m \pmod 2$; we assume tacitly in the sequel that we only consider cases with $\mathfrak{S}_{n,m}^{\mathsf{C}} \neq \emptyset$.) Note that our definition of a C-permutation differs from the one given in [8], where each cycle carries an additional sign. Here we do not include the signs as they will not play a role in our proofs.

A *C-decorated tree* of size n and genus g is a pair $(T, \sigma) \in \mathcal{T}_n \times \mathfrak{S}_{n+1,n+1-2g}^{\mathsf{C}}$ where σ is seen as a C-permutation of the vertices of T (given an arbitrary labeling of the vertices of T, for example the one given by a depth first search with left to right child ordering). The underlying graph of (T, σ) is the graph obtained by merging the vertices of T that belong to the same cycle in σ. If $v, v' \in T$, we write $v \sim v'$ if v and v' belong to the same cycle in σ.

Theorem 2.1 ([8], Theorem 5)**.** *Unicellular maps of size n and genus g are in 2^{2g} to 1 correspondence with C-decorated trees of size n and genus g. This correspondence preserves the underlying graph.*

Therefore, with this correspondence, it is sufficient to study C-decorated trees.

2.6. **Further notation.** We let $(n)_r$ denote the descending factorial $n(n-1)\cdots(n-r+1)$.

For a real number x, let $(x)_+ := x \vee 0 := \max\{x, 0\}$.

$\mathrm{Poi}(\lambda)$ denotes the Poisson distribution with parameter λ.

Convergence in distribution and in probability are denoted $\xrightarrow{\mathrm{d}}$ and $\xrightarrow{\mathrm{p}}$, respectively.

whp means with probability $1 - o(1)$ as $n \to \infty$.

Unspecified limits are as $n \to \infty$.

3. Cycles in C-permutations

In this section, we give several lemmas on cycles in random C-permutations; the only results that are used outside this section are Lemmas 3.4 and 3.5.

We will use ν as an index denoting cycle lengths in a C-permutation. Recall that only cycles of odd lengths are allowed; thus it is tacitly understood that ν ranges over the odd natural numbers (or a subset of them if indicated). (The same applies to ν_i and μ.)

Let n and g be given, and let $m := n - 2g$. Let $\boldsymbol{\sigma} = \boldsymbol{\sigma}_{n,m}$ be a uniformly random element of $\mathfrak{S}^{\mathsf{C}}_{n,m}$, and let $N_\nu = N_{\nu;n,m}$ be its number of cycles of length ν.

Assume that $\mathbf{x} = (x_1, x_3, \dots)$ is a sequence of non-negative integers, with only finitely many $x_\nu \neq 0$. Let $\mathcal{N}(\mathbf{x})$ be the number of C-permutations with exactly x_ν cycles of size ν for every $\nu \geq 1$. Recall that these permutations belong to $\mathfrak{S}^{\mathsf{C}}_{n,m}$ if and only if

$$\sum_{\nu\geq 1} x_\nu = m = n - 2g, \qquad \sum_{\nu\geq 1} \nu x_\nu = n, \tag{3.1}$$

which imply

$$x_3 = \tfrac{1}{2}\Big(n - (n-2g) - \sum_{\nu\geq 5}(\nu-1)x_\nu\Big) = g - \sum_{\nu\geq 5}\frac{\nu-1}{2}x_\nu, \tag{3.2}$$

$$x_1 = n - 3x_3 - \sum_{\nu\geq 5}\nu x_\nu = n - 3g + \sum_{\nu\geq 5}\frac{\nu-3}{2}x_\nu. \tag{3.3}$$

Fix n and m and let $\mathcal{X}_{n,m}$ be the set of all non-negative integer sequences $\mathbf{x} = (x_1, x_3, \dots)$ that satisfy (3.1), and thus (3.2)–(3.3). If $\mathbf{x} \in \mathcal{X}_{n,m}$, it is easily shown that

$$\mathcal{N}(\mathbf{x}) = \frac{n!}{\prod_{\nu\geq 1} x_\nu!\, \nu^{x_\nu}}. \tag{3.4}$$

For $\mathbf{x} \in \mathcal{X}_{n,m}$, let also

$$p(\mathbf{x}) := \mathbb{P}\big(N_{\nu;n,m} = x_\nu,\ \forall \nu\big) = \frac{\mathcal{N}(\mathbf{x})}{|\mathfrak{S}^{\mathsf{C}}_{m,n}|}. \tag{3.5}$$

Lemma 3.1. *Suppose that $g < n/3$. Let $\mathbf{x} \in \mathcal{X}_{n,m}$, let $\mu = 2k+1 \geq 5$ be odd and assume $x_\mu > 0$. Let $\mathbf{x}'$ be given by*

$$x'_\nu = \begin{cases} x_1 - \frac{\mu-3}{2}, & \nu = 1,\\ x_3 + \frac{\mu-1}{2}, & \nu = 3,\\ x_\nu - \delta_{\nu,\mu}, & \nu \geq 5. \end{cases} \tag{3.6}$$

Then $\mathbf{x}' \in \mathcal{X}_{n,m}$ and

$$x_\mu p(\mathbf{x}) \leq \frac{(3g)^k}{\mu(n-3g)^{k-1}} p(\mathbf{x}') = \frac{(3g)^{(\mu-1)/2}}{\mu(n-3g)^{(\mu-3)/2}} p(\mathbf{x}'). \tag{3.7}$$

Proof. Note that x'_1 and x'_3 are defined such that (3.2)–(3.3) hold for $\mathbf{x}'$. Furthermore, $x'_\mu = x_\mu - 1 \geq 0$ by assumption, and thus also, by (3.3) and the assumption $g < n/3$,

$$x'_1 = n - 3g + \sum_{\nu\geq 5}\frac{\nu-3}{2}x'_\nu \geq n - 3g \geq 0 \tag{3.8}$$

Hence, $x'_\nu \geq 0$ for all ν, and thus $\mathbf{x}' \in \mathcal{X}_{n,m}$.

Note that $x_1 = x'_1 + k - 1$ and $x_3 = x'_3 - k$, and also that $x'_1 \geq n - 3g$ by (3.8) and $x'_3 \leq g$ by (3.2). Hence, (3.5) and (3.4) yield

$$\frac{p(\mathbf{x})}{p(\mathbf{x}')} = \frac{\mathcal{N}(\mathbf{x})}{\mathcal{N}(\mathbf{x}')} = \frac{\prod_{\nu\geq 1} x'_\nu!\, \nu^{x'_\nu}}{\prod_{\nu\geq 1} x_\nu!\, \nu^{x_\nu}}$$

$$= \frac{x_1'!\, x_3'!\, 3^{x_3'}(x_\mu - 1)!\, \mu^{x_\mu - 1}}{(x_1' + k - 1)!\, (x_3' - k)!\, 3^{x_3' - k} x_\mu!\, \mu^{x_\mu}}$$
$$\le \frac{(x_3')^k 3^k}{(x_1')^{k-1} x_\mu \mu} \le \frac{g^k 3^k}{(n - 3g)^{k-1} x_\mu \mu}. \tag{3.9}$$

The result (3.7) follows. □

We can now easily estimate the mean of $N_{\nu;n,m}$ as well as higher (mixed) factorial moments.

Lemma 3.2. *Suppose that $g < n/3$. Then, for every $\nu \ge 3$,*

$$\mathbb{E}\, N_{\nu;n,m} \le \lambda_\nu := \frac{(3g)^{(\nu-1)/2}}{\nu(n-3g)^{(\nu-3)/2}}. \tag{3.10}$$

More generally, for any sequence $(\alpha_\nu)_3^\infty$ of non-negative integers (with only finitely many non-zero),

$$\mathbb{E}\Bigl[\prod_{\nu\ge3}(N_{\nu;n,m})_{\alpha_\nu}\Bigr] \le \prod_{\nu\ge3}\lambda_\nu^{\alpha_\nu}. \tag{3.11}$$

Proof. Let $\mu = 2k+1 \ge 5$. For $\mathbf{x} \in \mathcal{X}_{n,m}$ with $x_\mu \ge 1$, define $\mathbf{x}'$ as in (3.6) and note that with λ_μ defined in (3.10), (3.7) says

$$x_\mu p(\mathbf{x}) \le \lambda_\mu p(\mathbf{x}'). \tag{3.12}$$

Hence, Lemma 3.1 implies, noting that the map $\mathbf{x} \mapsto \mathbf{x}'$ is injective,

$$\mathbb{E}\, N_{\nu;n,m} = \sum_{\mathbf{x}\in\mathcal{X}_{n,m}} x_\mu p(\mathbf{x}) = \sum_{\mathbf{x}\in\mathcal{X}_{n,m},\, x_\mu\ge1} x_\mu p(\mathbf{x})$$
$$\le \sum_{\mathbf{x}\in\mathcal{X}_{n,m},\, x_\mu\ge1} \lambda_\mu p(\mathbf{x}') \le \lambda_\mu \sum_{\mathbf{x}'\in\mathcal{X}_{n,m}} p(\mathbf{x}') = \lambda_\mu. \tag{3.13}$$

This shows (3.10) for $\nu = \mu \ge 5$. For $\nu = 3$, we simply note that by (3.2),

$$N_{3;n,m} \le g = \lambda_3. \tag{3.14}$$

We prove (3.11) similarly. Suppose first that $\alpha_3 = 0$, and let $\boldsymbol{\alpha} = (0, 0, \alpha_5, \alpha_7, \dots)$. If $\mathbf{x} \in \mathcal{X}_{n,m}$ and $x_\nu \ge \alpha_\nu$ for all ν, let $\overline{\mathbf{x}-\boldsymbol{\alpha}}$ denote the element in $\mathcal{X}_{n,m}$ with coordinates $x_\nu - \alpha_\nu$ for $\nu \ge 5$ (and for $\nu = 1, 3$, given from these by (3.2)–(3.3)). Then repeated use of (3.12) yields

$$p(\mathbf{x})\prod_{\nu\ge5}(x_\nu)_{\alpha_\nu} \le \prod_{\nu\ge5}\lambda_\nu^{\alpha_\nu}\cdot p\bigl(\overline{\mathbf{x}-\boldsymbol{\alpha}}\bigr). \tag{3.15}$$

Hence,

$$\mathbb{E}\Bigl[\prod_{\nu\ge5}(N_{\nu;n,m})_{\alpha_\nu}\Bigr] = \sum_{\mathbf{x}\ge\boldsymbol{\alpha}} p(\mathbf{x})\prod_{\nu\ge5}(x_\nu)_{\alpha_\nu} \le \sum_{\mathbf{x}\ge\boldsymbol{\alpha}}\prod_{\nu\ge5}\lambda_\nu^{\alpha_\nu}\cdot p\bigl(\overline{\mathbf{x}-\boldsymbol{\alpha}}\bigr)$$
$$\le \prod_{\nu\ge5}\lambda_\nu^{\alpha_\nu}. \tag{3.16}$$

This proves (3.11) when $\alpha_3 = 0$. The general case follows by this and the deterministic bound (3.14). □

The estimate in Lemma 3.2 shows that in our range $g = o(n)$, cycles of length 5 or more are few, and we will see in results and proofs below that they are insignificant for our purposes. The estimate in Lemma 3.2 seems to be rather sharp for all ν that are not very large, but we will not study this further. We give only a matching lower bound in the case $\nu = 3$.

Lemma 3.3. *If $g < n/6$, then*

$$g - \frac{5g^2}{n-6g} \le \mathbb{E}\, N_{3;n,m} \le g. \tag{3.17}$$

In particular, as $n \to \infty$ with $g = o(n)$, $\mathbb{E}\, N_{3;n,m} \sim g$.

Proof. By (3.2) and Lemma 3.2,

$$\begin{aligned}
\mathbb{E}\bigl(g - N_{3;n,m}\bigr) &= \mathbb{E} \sum_{\nu\ge5} \frac{\nu-1}{2} N_{\nu;n,m} \le \sum_{\nu\ge5} \frac{\nu-1}{2}\lambda_\nu \\
&= \sum_{k\ge2} \frac{(3g)^k}{2(n-3g)^{k-1}} = \frac{(3g)^2}{2(n-3g)}\Bigl(1-\frac{3g}{n-3g}\Bigr)^{-1} \\
&\le \frac{5g^2}{n-6g}.
\end{aligned} \tag{3.18}$$

The lower bound in (3.17) follows. The upper bound follows trivially from the deterministic bound (3.14). ☐

Lemma 3.4. *Let $\mathcal{E}^{(r)}_{n,g}$ be the event where, in $\boldsymbol{\sigma}$, $2i-1$ and $2i$ belong to the same cycle for all $1 \le i \le r$, and these r cycles are distinct, and let $\mathbb{P}^{(r)}_{n,g} = \mathbb{P}(\mathcal{E}^{(r)}_{n,g})$. If $g \le n/7$, then for all $r \le n/2$,*

$$\mathbb{P}^{(r)}_{n,g} \le \frac{1}{(n)_{2r}}\Lambda^r \le \Bigl(\frac{C_0 g}{n^2}\Bigr)^r, \tag{3.19}$$

where $C_0 < 2200$ is an absolute constant and

$$\Lambda = 6g\Bigl(\frac{1-3g/n}{1-6g/n}\Bigr)^2. \tag{3.20}$$

Moreover, as $n \to \infty$ with $g \to \infty$ and $g = o(n)$, $\Lambda \sim 6g$ and, for any fixed $r \ge 1$,

$$\mathbb{P}^{(r)}_{n,g} \sim \Bigl(\frac{6g}{n^2}\Bigr)^r. \tag{3.21}$$

By symmetry, the probability is the same if we replace the nodes $1, \dots, 2r$ by any other $2r$ fixed nodes in $[n]$.

Proof. Let τ be a uniformly random permutation of $[n]$ independent of $\boldsymbol{\sigma}$. By the invariance just mentioned, the probability is the same if we instead consider the nodes $\tau(1), \dots, \tau(2r)$, which will be convenient in the proof.

For a sequence $\nu_1, \dots, \nu_r$, let $\mathcal{E}(\nu_1, \dots, \nu_r)$ be the event that there are distinct cycles $\mathsf{C}_1, \dots, \mathsf{C}_r$ in $\boldsymbol{\sigma}$ such that $|\mathsf{C}_i| = \nu_i$ and $\tau(2i-1), \tau(2i) \in \mathsf{C}_i$ for every $i \le r$. We may assume $\nu_i \ge 3$ for every i, since otherwise the event is impossible.

We first compute the conditional probability $\mathbb{P}\bigl(\mathcal{E}(\nu_1, \dots, \nu_r) \mid \boldsymbol{\sigma}\bigr)$. Let $\alpha_\nu := |\{i : \nu_i = \nu\}|$, the number of the cycles C_i that are required to have length ν. Given $\boldsymbol{\sigma}$, there are $N_{\nu;n,m}$ cycles of length ν, and thus

$\prod_\nu (N_{\nu;n,m})_{\alpha_\nu}$ ways to choose the cycles $\mathsf{C}_1, \dots, \mathsf{C}_r$. Given these cycles, the probability that $\tau(2i-1), \tau(2i) \in \mathsf{C}_i$ for all i is

$$\frac{1}{(n)_{2r}} \prod_{i=1}^{r} \bigl(\nu_i(\nu_i - 1)\bigr) = \frac{1}{(n)_{2r}} \prod_{\nu} \bigl(\nu(\nu-1)\bigr)^{\alpha_\nu}. \tag{3.22}$$

Hence,

$$\mathbb{P}\bigl(\mathcal{E}(\nu_1, \dots, \nu_r) \mid \boldsymbol{\sigma}\bigr) = \prod_{\nu} (N_{\nu;n,m})_{\alpha_\nu} \cdot \frac{1}{(n)_{2r}} \prod_{\nu} \bigl(\nu(\nu-1)\bigr)^{\alpha_\nu}. \tag{3.23}$$

Define, for $\nu \geq 3$,

$$\widehat{\lambda}_\nu := \nu(\nu-1)\lambda_\nu = (\nu - 1)\frac{(3g)^{(\nu-1)/2}}{(n-3g)^{(\nu-3)/2}}. \tag{3.24}$$

Taking the expectation in (3.23), we obtain by Lemma 3.2,

$$\begin{aligned}\mathbb{P}\bigl(\mathcal{E}(\nu_1, \dots, \nu_r)\bigr) &= \mathbb{E}\,\mathbb{P}\bigl(\mathcal{E}(\nu_1, \dots, \nu_r) \mid \boldsymbol{\sigma}\bigr) \\ &= \frac{1}{(n)_{2r}} \prod_{\nu} \bigl(\nu(\nu-1)\bigr)^{\alpha_\nu} \cdot \mathbb{E} \prod_{\nu} (N_{\nu;n,m})_{\alpha_\nu} \\ &\leq \frac{1}{(n)_{2r}} \prod_{\nu} \bigl(\nu(\nu-1)\bigr)^{\alpha_\nu} \cdot \prod_{\nu} \lambda_\nu^{\alpha_\nu} \\ &= \frac{1}{(n)_{2r}} \prod_{\nu} \widehat{\lambda}_\nu^{\alpha_\nu} = \frac{1}{(n)_{2r}} \prod_{i=1}^{r} \widehat{\lambda}_{\nu_i}. \end{aligned} \tag{3.25}$$

Summing over all $\nu_1, \dots, \nu_r$ yields the result

$$\mathbb{P}_{n,g}^{(r)} = \sum_{\nu_1,\dots,\nu_r} \mathbb{P}\bigl(\mathcal{E}(\nu_1, \dots, \nu_r)\bigr) \leq \frac{1}{(n)_{2r}} \sum_{\nu_1,\dots,\nu_r} \prod_{i=1}^{r} \widehat{\lambda}_{\nu_i} = \frac{1}{(n)_{2r}} \Bigl(\sum_{\nu \geq 3} \widehat{\lambda}_\nu\Bigr)^r. \tag{3.26}$$

We have

$$\begin{aligned}\sum_{\nu \geq 3} \widehat{\lambda}_\nu &= \sum_{\nu \geq 3} (\nu-1)\frac{(3g)^{(\nu-1)/2}}{(n-3g)^{(\nu-3)/2}} = \sum_{k \geq 1} 2k \frac{(3g)^k}{(n-3g)^{k-1}} \\ &= 6g\Bigl(1 - \frac{3g}{n-3g}\Bigr)^{-2} = 6g\Bigl(\frac{1-3g/n}{1-6g/n}\Bigr)^2 = \Lambda, \end{aligned} \tag{3.27}$$

as defined in (3.20). Hence, (3.26) proves the first inequality in (3.19).

Moreover, using Stirling's formula,

$$(n)_{2r}^{1/2r} \geq (n!)^{1/n} \geq n/e \tag{3.28}$$

and if $g/n \leq 1/7$, then $\Lambda \leq 6g(1-6/7)^{-2} = 294g$. Hence, the second inequality in (3.19) holds with $C_0 = 294e^2 < 2200$.

For a fixed r, we have $\Lambda \sim 6g$ since $g/n \to 0$, and thus (3.19) yields the (implicit) upper bound in (3.21). For a matching lower bound, we consider only the case $\nu_1 = \dots = \nu_r = 3$. We have, by again taking the expectation in (3.23),

$$\mathbb{P}_{n,g}^{(r)} \geq \mathbb{P}\bigl(\mathcal{E}(3, \dots, 3)\bigr) = \frac{1}{(n)_{2r}} 6^r\, \mathbb{E}(N_{3;n,m})_r. \tag{3.29}$$

Furthermore, by Jensen's inequality and Lemma 3.3, for any fixed r,

$$\begin{aligned}\mathbb{E}(N_{3;n,m})_r \geq \mathbb{E}(N_{3;n,m}-r)_+^r \geq (\mathbb{E}\,N_{3;n,m}-r)_+^r = \bigl(g+O(g^2/n)+O(1)\bigr)^r \\ \sim g^r. \qquad (3.30)\end{aligned}$$

The (implicit) lower bound in (3.21) follows from (3.29) and (3.30). □

In Lemma 3.4, each of the cycles that contain one of the distinguished points $1,\dots,2k$ contains exactly two of them. We will also use an estimate of the probability that some cycle contains more than two of the distinguished points.

Lemma 3.5. *Assume $g \leq n/7$. For every fixed $k \geq 1$ and $r \leq k/2$, the probability that $1,\dots,k$ belong to exactly r different cycles in $\boldsymbol{\sigma}$, with at least two of these points in each cycle, is, for some constants $C = C(k)$,*

$$\leq Cg^r/(n)_k \leq Cg^r/n^k. \qquad (3.31)$$

(For $r = k/2$, this is a just weaker version of Lemma 3.4.)

Proof. We must have $n \geq k$, and thus the two bounds in (3.31) are equivalent (with different C).

We argue similarly to the proof of Lemma 3.4. We use again randomization, and let τ be a random permutation of $[n]$ independent of $\boldsymbol{\sigma}$. For sequences $\nu_1,\dots,\nu_r$ and $\beta_1,\dots,\beta_r$, let $\mathcal{E}(\nu_1,\dots,\nu_r;\beta_1,\dots\beta_r)$ be the event that there are distinct cycles $\mathsf{C}_1,\dots,\mathsf{C}_r$ in $\boldsymbol{\sigma}$ such that $|\mathsf{C}_i| = \nu_i$ and exactly β_i of the points $\tau(1),\dots,\tau(k)$ belong to C_i for every $i \leq r$. We assume $\beta_i \geq 2$ and $\sum_i \beta_i = k$, and we also assume $\nu_i \geq 3$ for every i, since otherwise the event is impossible.

As in the proof of Lemma 3.4, let $\alpha_\nu := |\{i : \nu_i = \nu\}|$. Then, again, given $\boldsymbol{\sigma}$, there are $\prod_\nu (N_{\nu;n,m})_{\alpha_\nu}$ ways to choose the cycles $\mathsf{C}_1,\dots,\mathsf{C}_r$. Given these cycles, the conditional probability that β_i of $\tau(1),\dots,\tau(k)$ belong to C_i, $i = 1,\dots,r$, is

$$\leq C\frac{1}{(n)_k}\prod_{i=1}^r (\nu_i)_{\beta_i}. \qquad (3.32)$$

(In this proof, C denotes constants that may depend on k but not on other variables. In (3.32) we may take $C = k!$.) Hence, using Lemma 3.2,

$$\begin{aligned}\mathbb{P}\bigl[\mathcal{E}(\nu_1,\dots,\nu_r;\beta_1,\dots\beta_r)\bigr] &\leq \frac{C}{(n)_k}\prod_{i=1}^r(\nu_i)_{\beta_i}\,\mathbb{E}\prod_\nu (N_{\nu;n,m})_{\alpha_\nu} \\ &\leq \frac{C}{(n)_k}\prod_{i=1}^r(\nu_i)_{\beta_i}\cdot\prod_\nu \lambda_\nu^{\alpha_\nu} = \frac{C}{(n)_k}\prod_{i=1}^r(\nu_i)_{\beta_i}\lambda_{\nu_i}. \qquad (3.33)\end{aligned}$$

We sum (3.33) first over all $\nu_i,\dots,\nu_r$, and note that by (3.10) and the assumption $g \leq n/7$, for every fixed $\beta \geq 2$,

$$\sum_{\nu\geq 3}(\nu)_\beta\lambda_\nu \leq C(\beta)g. \qquad (3.34)$$

Since we only consider $\beta_i \leq k$, we thus obtain from (3.33),

$$\sum_{\nu_1,\dots,\nu_r} \mathbb{P}\big[\mathcal{E}(\nu_1,\dots,\nu_r;\beta_1,\dots\beta_r)\big] \leq \frac{C}{(n)_k}\prod_{i=1}^{r}\sum_{\nu_i\geq 3}(\nu_i)_{\beta_i}\lambda_{\nu_i} \leq \frac{C}{(n)_k}g^r. \quad (3.35)$$

The result follows by summing over the $O(1)$ allowed $(\beta_1,\dots,\beta_r)$. □

4. Counting paths in trees

As in Section 3, many lemmas here are local; only Lemmas 4.11, 4.12 and 4.13 are used outside this section.

4.1. **Generating-functionology.** We start by introducing some generating functions. First, the generating function of rooted plane trees enumerated by edges:

$$B(z) := \frac{1-\sqrt{1-4z}}{2z} = \sum_{n=0}^{\infty} Cat_n z^n, \quad (4.1)$$

where $Cat_n := \frac{1}{n+1}\binom{2n}{n}$ is the nth Catalan number. We also introduce

$$T(z) := zB(z) = \frac{1-\sqrt{1-4z}}{2}, \quad (4.2)$$

which satisfies

$$T(z) = \frac{z}{1-T(z)}. \quad (4.3)$$

We will also use doubly rooted plane trees. These trees have two roots, labelled first and second root. Both roots are corners of T; the roots may be the same corner, but that case we also distinguish between two different orderings of the roots. A rooted plane tree with n edges has $2n$ corners, and thus a second root may be added in $2n+1$ different places (including 2 places in the corner of the first root). Therefore, the generating function of doubly rooted plane trees, enumerated by edges, is

$$A(z) := \Big(2z\frac{\partial}{\partial z}+1\Big)B(z). \quad (4.4)$$

We recall the Lagrange–Bürmann formula that will be useful to us in this section.

Theorem 4.1 (Lagrange–Bürmann formula ([11], Theorem A.2))**.** *Let F and ϕ be power series satisfying*

$$F(z) = z\phi(F(z)), \quad (4.5)$$

then, for any (analytic) function f, we have

$$[z^n]f(F(z)) = \frac{1}{n}[z^{n-1}]\big(\phi(z)^n f'(z)\big), \qquad n \geq 1. \quad (4.6)$$

We will use this formula to estimate the number of occurrences of a certain pattern in $\mathbf{T}_n$.

Lemma 4.2. *Let $\mathbf{t}$ be a rooted plane tree of size ℓ. Then the number of rooted plane trees of size n with a marked rooted subtree isomorphic to $\mathbf{t}$ is*

$$T_{n,\ell} = 2n[z^{n-\ell}]B(z)^{2\ell} = 2\ell\binom{2n}{n-\ell}, \qquad n \geq \ell, \quad (4.7)$$

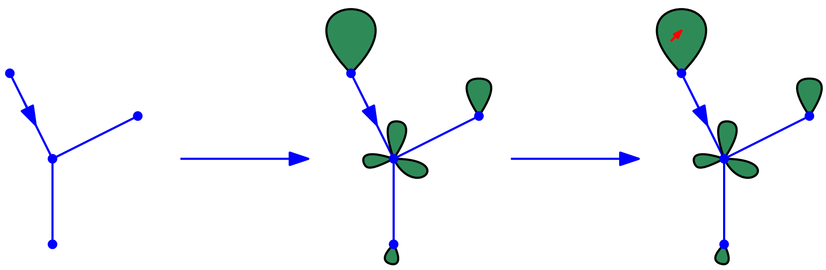

FIGURE 2. Building a tree T with a marked copy of $\mathbf{t}$. Here $\ell = 3$, $\mathbf{t}$ is in blue, the 2ℓ trees are in green, and the root of T is in red.

and we have

$$\mathbb{E}\, N_{\mathbf{t}}(\mathbf{T}_n) = \frac{T_{n,\ell}}{Cat_n} = (1 + o(1))2\ell n \tag{4.8}$$

where the $o(1)$ *is uniform over* $\ell \in [1, L^\bullet]$*.*

Proof. We want to enumerate the number of trees of size n with a marked subtree isomorphic to $\mathbf{t}$. One can build such a tree in a bijective way (see Figure 2). Starting from a copy of $\mathbf{t}$, to each corner c of $\mathbf{t}$, pick a rooted tree and glue its root corner to c. The only constraint is that the total size of all the 2ℓ trees that we graft must be $n - \ell$. One obtains an unrooted tree T of size n with a marked copy of $\mathbf{t}$. Finally, one just needs to pick one of its $2n$ corners as the root; note that the resulting pairs $(T, \mathbf{t})$ will be distinct, since T has no automorphisms (preserving order and $\mathbf{t}$). Hence the number of rooted such trees is

$$T_{n,\ell} = 2n[z^{n-\ell}]B(z)^{2\ell}. \tag{4.9}$$

Using the Lagrange–Bürmann formula (4.6) and (4.3), we get

$$\begin{aligned}
[z^{n-\ell}]B(z)^{2\ell} &= [z^{n+\ell}]T(z)^{2\ell} \\
&= \frac{1}{n+\ell}[z^{n+\ell-1}]\Big(\frac{1}{(1-z)^{n+\ell}}2\ell z^{2\ell-1}\Big) \\
&= \frac{2\ell}{n+\ell}[z^{n-\ell}]\frac{1}{(1-z)^{n+\ell}} \\
&= \frac{2\ell}{n+\ell}\binom{2n-1}{n-\ell} = \frac{2\ell}{2n}\binom{2n}{n-\ell}.
\end{aligned} \tag{4.10}$$

This, together with (4.9), shows (4.7).

The Stirling formula gives us

$$\binom{2n}{n-\ell} \sim \binom{2n}{n} = (n+1)\,Cat_n \tag{4.11}$$

uniformly for $|\ell| \le L^\bullet$ (because $L^\bullet = o(n^{1/2})$); hence (4.7) implies

$$\frac{T_{n,\ell}}{Cat_n} \sim 2\ell n \tag{4.12}$$

uniformly in $\ell \in [1, L^\bullet]$, which shows (4.8). ∎

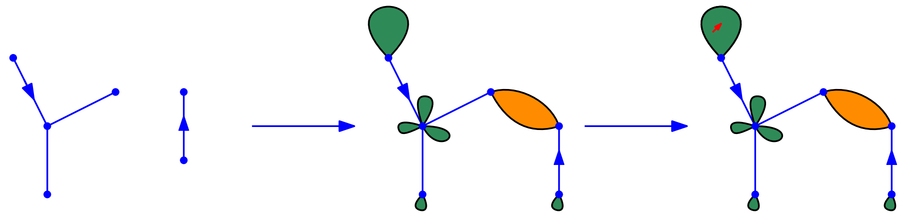

FIGURE 3. Building a tree T with a marked pair of trees. Here $\ell_1 = 3$ and $\ell_2 = 1$, $\mathbf{t}_1$ and $\mathbf{t}_2$ are in blue, the doubly rooted tree is in orange, the $2\ell - 2$ trees are in green, and the root of T is in red.

With the same method, we can also get an estimate for the number of pairs of patterns.

Lemma 4.3. *Let $\mathbf{t}_1$ and $\mathbf{t}_2$ be two rooted plane trees of sizes ℓ_1 and ℓ_2, and let $\ell = \ell_1 + \ell_2$. Then the number of rooted plane trees of size n with a marked pair of non intersecting rooted subtrees that are isomorphic to $\mathbf{t}_1$ and $\mathbf{t}_2$, respectively, is*

$$T_{n,\ell_1,\ell_2} \leq 8n\ell_1\ell_2[z^{n-\ell}]\big(A(z)B(z)^{2\ell-2}\big) = 4\ell_1\ell_2(n+\ell)\binom{2n}{n+\ell}. \tag{4.13}$$

and we have

$$\frac{T_{n,\ell_1,\ell_2}}{Cat_n} \leq (1+o(1))(2\ell_1 n)(2\ell_2 n) \tag{4.14}$$

where the $o(1)$ is uniform over $\ell_1, \ell_2 \in [1, L^\bullet]$.

Remark 4.4. It can be shown that the inequality in (4.14) is actually an equality, but we do not need this here. □

Proof. This is similar to the proof of Lemma 4.2. The decomposition now is the following (see Figure 3): start from a copy of $\mathbf{t}_1$ and a copy of $\mathbf{t}_2$, choose one corner on each and graft a doubly rooted tree $\tilde{T}$, identifying its first root (second root) with the root corner of $\mathbf{t}_1$ ($\mathbf{t}_2$). Then graft rooted trees to each of the $2\ell - 2$ remaining corners of $\mathbf{t}_1$ and $\mathbf{t}_2$ (as done in the proof of Lemma 4.2), to obtain an unrooted tree T of size n, and pick one of its $2n$ corners as the root. This way, we can build all rooted trees with non intersecting copies of $\mathbf{t}_1$ and $\mathbf{t}_2$, plus some cases where they intersect (namely, when $\mathbf{t}_1$ and $\mathbf{t}_2$ are grafted at a same vertex of $\tilde{T}$).

This yields

$$T_{n,\ell_1,\ell_2} \leq (2\ell_1)(2\ell_2)[z^{n-\ell}]\big(A(z)B(z)^{2\ell-2}\big) \cdot (2n), \tag{4.15}$$

showing the first part of (4.13). We then use (4.4) and, again, the Lagrange–Bürmann formula, and obtain

$$\begin{aligned}
[z^{n-\ell}]\big(A(z)B(z)^{2\ell-2}\big) &= 2[z^{n-\ell-1}]\big(B'(z)B(z)^{2\ell-2}\big) + [z^{n-\ell}]B(z)^{2\ell-1} \\
&= 2[z^{n-\ell-1}]\frac{(B^{2\ell-1})'}{2\ell-1} + [z^{n-\ell}]B(z)^{2\ell-1} \\
&= \left(2\frac{n-\ell}{2\ell-1} + 1\right)[z^{n-\ell}]B(z)^{2\ell-1}
\end{aligned}$$

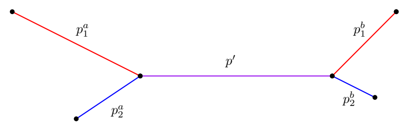

FIGURE 4. The union of two paths

$$= \left(2\frac{n-\ell}{2\ell-1}+1\right)\frac{2\ell-1}{n+\ell-1}\binom{2n-2}{n-\ell}$$
$$= \frac{2n-1}{n+\ell-1}\binom{2n-2}{n-\ell} = \frac{n+\ell}{2n}\binom{2n}{n-\ell}, \tag{4.16}$$

and (4.13) follows. Using (4.11) again, we obtain

$$\frac{T_{n,\ell_1,\ell_2}}{Cat_n} \le (1+o(1))(2n\ell_1)(2n\ell_2), \tag{4.17}$$

uniformly for $\ell_1, \ell_2 \le L^\bullet$. □

4.2. **Unions of paths.**

Definition 4.5. The set $\mathrm{Bic}(\ell_1, \ell_2)$ is the set of unrooted trees $\mathbf{t}$ with edges either blue, red, or bicolored such that $\mathbf{t}$ is the union of a blue path of length ℓ_1 and a red path of length ℓ_2.

Lemma 4.6. *There are at most* $16(\ell_1+1)(\ell_2+1)(\min(\ell_1,\ell_2)+1)$ *trees in* $\mathrm{Bic}(\ell_1, \ell_2)$.

Proof. In this proof, we make an exception, and allow paths to have length 0. We describe a procedure to build a tree in $\mathrm{Bic}(\ell_1, \ell_2)$ (see Figure 4).

(1) Create a bicolored path p' of length $0 \le \ell' \le \min(\ell_1, \ell_2)$ ($\min(\ell_1,\ell_2)+1$ possibilities).
(2) Create two red paths p_1^a and p_1^b of total length $\ell_1 - \ell'$ ($\le \ell_1 + 1$ possibilities).
(3) Create two blue paths p_2^a and p_2^b of total length $\ell_2 - \ell'$ ($\le \ell_2 + 1$ possibilities).
(4) Attach p_1^a and p_2^a to $\mathrm{start}(p')$ (2 possibilities).
(5) Attach p_1^b and p_2^b to $\mathrm{end}(p')$ (2 possibilities).
(6) Orient the blue and red paths ($2 \times 2 = 4$ possibilities).

This procedure is surjective; hence this proves what we wanted. □

Now, for $i, j \le M$, we define

$$\mathcal{B}(i,j) = \bigsqcup \mathrm{Bic}(\ell_1, \ell_2), \tag{4.18}$$

where the union is over all $\ell_1, \ell_2 \in [\frac{i}{M}L_n, \frac{i+1}{M}L_n) \times [\frac{j}{M}L_n, \frac{j+1}{M}L_n)$.

Lemma 4.7. *For every tree* T *and* $\mathbf{m} \in \mathcal{M}$, *we have*

$$1 \ge \frac{P^{[\mathbf{m}]}(T)}{\prod_{i=1}^{s(\mathbf{m})} P_{m_i}(T)} \ge 1 - \sum_{i,j} \frac{\sum_{\mathbf{t}\in\mathcal{B}(m_i,m_j)} N_{\mathbf{t}}(T)}{P_{m_i}(T)P_{m_j}(T)} \tag{4.19}$$

Proof. The proof is direct from the inclusion–exclusion principle: indeed, $\prod_{i=1}^{s(\mathbf{m})} P_{m_i}(T)$ counts lists of $s(\mathbf{m})$ paths with the right sizes, without any constraint of non-intersection between these paths. Additionally,

$$\sum_{i,j} \sum_{\mathbf{t}\in\mathcal{B}(m_i,m_j)} N_{\mathbf{t}}(T) \prod_{\substack{k=1 \\ k\neq i,j}}^{s(\mathbf{m})} P_{m_k}(T) \tag{4.20}$$

(over)counts such lists where two of the paths intersect. □

Definition 4.8. A *path tree* is a tree T together with a list of $q \geq 1$ paths $p_1, p_2, \dots, p_q$ such that $T = \bigcup_{i=1}^{q} p_i$, and for every $i > 1$, there exists $j < i$ such that $\mathrm{Ext}(p_i) \cap \mathrm{Ext}(p_j) \neq \emptyset$. (For convenience, we denote the path tree simply by T.) For a path tree T, we write $\mathrm{Ext}(T) := \bigcup_{i=1}^{q} \mathrm{Ext}(p_i)$.

Let $\mathcal{P}_{q,w}(\ell)$ be the set of path trees T with q paths p_i such that $\left|\mathrm{Ext}(T)\right| = w$, and $|p_i| \leq \ell$ for every path p_i.

Lemma 4.9. (i) *If $T \in \mathcal{P}_{q,w}(\ell)$, then $\max_{v\in V(T)} \deg(v) \leq q+1$.*
(ii) *For every q and w, there exists a constant $C_{q,w}$ such that $|\mathcal{P}_{q,w}(\ell)| \leq C_{q,w}\ell^{2w-3}$.*

Proof. We will prove this by induction. Both parts are verified for $q = 1$ because $\mathcal{P}_{1,w}$ is empty unless $w = 2$, and $\mathcal{P}_{1,2}(\ell)$ is just the set of paths of length $\leq \ell$, so $|\mathcal{P}_{1,2}(\ell)| = \ell$.

Now assume $q \geq 2$, and let $T = \bigcup_{i=1}^{q} p_i \in \mathcal{P}_{q,w}(\ell)$. Then $T' := \bigcup_{i=1}^{q-1} p_i$ is also a path tree, with $T' \in \mathcal{P}_{q-1,w'}(\ell)$ for some $w' \in \{w-1, w\}$.

Starting from T', one can reconstruct T by adding a path p_q. If $w' = w$, then both endpoints of p_q have to be in $\mathrm{Ext}(T')$, which yields $\leq w^2$ choices.

If $w' = w-1$, then p_q must have one endpoint v in $\mathrm{Ext}(T')$, but its other endpoint may be either in $T' \setminus \mathrm{Ext}(T')$ or outside T'; in the latter case, let v' by last point in $p_q \cup T'$ (starting from v). We may then reconstruct T from T' as follows (with some overcounting, since not all choices below are allowed):

(1) Choose a vertex $v \in \mathrm{Ext}(T')$ ($w' \leq w$ possibilities).
(2) Choose a vertex $v' \in V(T')$ ($\leq |V(T')| \leq q\ell$ possibilities).
(3) Either stop at v' and let $v^* := v'$, or attach a path of length $\leq \ell$ to one of the corners of v', and let v^* be the other endpoint of this path ($\leq \ell q$ possibilities, by (i)).
(4) Declare the path from v to v^* as p_q, and give it an orientation (2 choices).

It is clear that with this procedure, the vertex degrees increase by at most 1 and thus (i) holds for T, since it holds for T' by the induction hypothesis.

Furthermore, since the procedure is surjective, it follows that

$$|\mathcal{P}_{q,w}(\ell)| \leq w^2|\mathcal{P}_{q-1,w}(\ell)| + 2wq\ell\big(1+q\ell\big)|\mathcal{P}_{q-1,w-1}(\ell)|, \tag{4.21}$$

and (ii) follows by induction. □

4.3. **Patterns in uniform trees.**

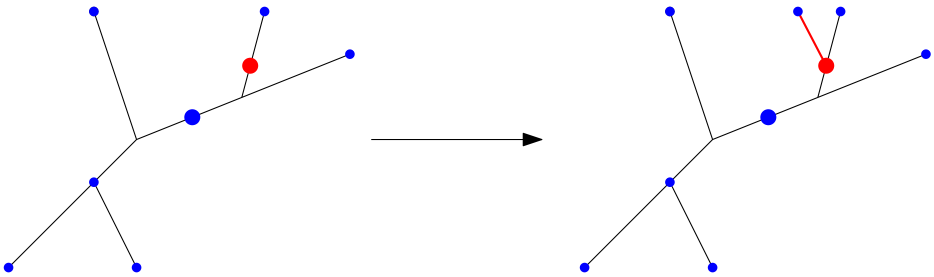

FIGURE 5. Building a path tree recursively. Here, $w' = 7$ and $w = 8$. The vertices of $\mathrm{Ext}(T')$ are blue, v is the large blue dot, and v' is the large red dot. The path between v' and v^* is in red.

Lemma 4.10. *Fix $i, j \geq 0$. Then*

$$\mathbb{E}\left[\sum_{\mathbf{t}\in\mathcal{B}(i,j)} N_{\mathbf{t}}(\mathbf{T})\right] = O\big(nL_n^6\big) = o\big(n^2L_n^4\big). \tag{4.22}$$

Proof. A tree $\mathbf{t}$ in $\mathcal{B}(i,j)$ will have size at most $\frac{i+j+2}{M}L_n = O(L_n)$. Hence, by (4.18) and Lemma 4.6, the cardinality of $\mathcal{B}(i,j)$ is bounded by $\big(\frac{L_n}{M}+1\big)^2 \times O\big(L_n^3\big) = O\big(L_n^5\big)$. Since $L_n = o(L^\bullet)$, we may also use (4.8). Consequently,

$$\begin{aligned}
\mathbb{E}\left[\sum_{\mathbf{t}\in\mathcal{B}(i,j)} N_{\mathbf{t}}(\mathbf{T})\right] &= \sum_{\mathbf{t}\in\mathcal{B}(i,j)} \mathbb{E}\, N_{\mathbf{t}}(\mathbf{T}) \\
&= \sum_{\mathbf{t}\in\mathcal{B}(i,j)} (1+o(1))2n|\mathbf{t}| \\
&= |\mathcal{B}(i,j)| \cdot O(nL_n) \\
&= O(nL_n^6),
\end{aligned} \tag{4.23}$$

and we conclude by recalling that $L_n = o(\sqrt{n})$. □

Here is a similar lemma that uses Lemma 4.9 and the notation there.

Lemma 4.11. *For every fixed q, w, and c, we have*

$$\mathbb{E}\left[\sum_{\mathbf{t}\in\mathcal{P}_{q,w}(cL_n)} N_{\mathbf{t}}(\mathbf{T})\right] = o\big(nL_n^{2w-2}\log g\big). \tag{4.24}$$

Proof. Since $\mathbf{t} \in \mathcal{P}_{q,w}(cL_n)$ implies $|\mathbf{t}| \leq qcL_n = o(L^\bullet)$, we have, by (4.8) and Lemma 4.9,

$$\begin{aligned}
\mathbb{E}\left[\sum_{\mathbf{t}\in\mathcal{P}_{q,w}(cL_n)} N_{\mathbf{t}}(\mathbf{T})\right] &= \sum_{\mathbf{t}\in\mathcal{P}_{q,w}(cL_n)} \mathbb{E}\, N_{\mathbf{t}}(\mathbf{T}) \\
&= \sum_{\mathbf{t}\in\mathcal{P}_{q,w}(cL_n)} (1+o(1))2n|\mathbf{t}| \\
&\leq (1+o(1))|\mathcal{P}_{q,w}(cL_n)|\, 2nqcL_n
\end{aligned}$$

$$= O(nL_n^{2w-2}), \tag{4.25}$$

which implies (4.24). □

Lemma 4.12. *For every fixed $i \geq 0$,*

$$\frac{P_i(\mathbf{T})}{nL_n^2} \xrightarrow{\text{p}} \frac{2i+1}{M^2}. \tag{4.26}$$

Proof. We will prove a slightly stronger result, namely L^2 convergence.

Let $J_i := [\frac{i}{M}L_n, \frac{i+1}{M}L_n)$. Using (2.2) and (4.8), we have

$$\mathbb{E}\big[P_i(\mathbf{T})\big] = \sum_{\ell\in J_i} \frac{T_{n,\ell}}{Cat_n} = \big(1+o(1)\big)\sum_{\ell\in J_i} 2\ell n \sim nL_n^2\frac{2i+1}{M^2}. \tag{4.27}$$

Now, let us count pairs of paths in $\mathbf{T}$, in order to estimate $\mathbb{E}\big[P_i(\mathbf{T})^2\big]$. There are two cases:

(1) Two disjoint paths. Such pairs are enumerated by Lemma 4.3, which implies that the total expectation of the number of such pairs is

$$\sum_{\ell_1,\ell_2\in J_i} \frac{T_{n,\ell_1,\ell_2}}{Cat_n} \leq \big(1+o(1)\big) \sum_{\ell_1,\ell_2\in J_i} (2\ell_1 n)(2\ell_2 n), \tag{4.28}$$

which by comparison with (4.27) is $\sim \big(\mathbb{E}\,P_i(\mathbf{T})\big)^2$.

(2) Two paths that intersect in at least one vertex. Their union is then a bicolored tree, and by Lemma 4.10, the expectation of the number of such pairs is $o(n^2L_n^4)$, which by (4.27) is $o\big(\mathbb{E}\,P_i(\mathbf{T})\big)^2$.

In summary, this establishes that

$$\mathbb{E}\big[P_i(\mathbf{T})^2\big] \sim \big(\mathbb{E}\,P_i(\mathbf{T})\big)^2, \tag{4.29}$$

which implies $P_i(\mathbf{T})/\mathbb{E}\,P_i(\mathbf{T}) \to 1$ in L^2. By (4.27), this yields (4.26) in L^2, and thus in probability. □

Lemma 4.13. *For every fixed $\mathbf{m} \in \mathcal{M}$,*

$$\frac{P^{[\mathbf{m}]}(\mathbf{T})}{\prod_{i=1}^{s(\mathbf{m})} P_{m_i}(\mathbf{T})} \xrightarrow{\text{p}} 1. \tag{4.30}$$

Proof. For fixed i and j, we have by Lemma 4.10 and the Markov inequality

$$\frac{1}{n^2L_n^4}\sum_{\mathbf{t}\in\mathcal{B}(i,j)} N_{\mathbf{t}}(\mathbf{T}) \xrightarrow{\text{p}} 0. \tag{4.31}$$

Hence by Lemma 4.12, for every fixed i and j,

$$\frac{\sum_{\mathbf{t}\in\mathcal{B}(i,j)} N_{\mathbf{t}}(\mathbf{T})}{P_i(\mathbf{T})P_j(\mathbf{T})} \xrightarrow{\text{p}} 0, \tag{4.32}$$

and we conclude by using the inequalities (4.19). □

5. Cycles in C-decorated trees

5.1. **Definitions.** If (T,σ) is a C-decorated tree, then any simple cycle of length ℓ in its underlying graph can be decomposed into a list $\mathbf{P} = (p_1, p_2, \dots, p_k)$ of non-intersecting simple paths in T such that

(C1) $\sum_{i=1}^k |p_i| = \ell$;
(C2) $\mathrm{end}(p_i) \sim \mathrm{start}(p_{(i+1 \mod k)})$ for all i;
(C3) for every other pair of vertices $v, v' \in (p_1, p_2, \dots, p_k)$, we have $v \not\sim v'$.

This decomposition is unique up to cyclically reordering the p_i, or reversing them all and their order, or a combination of both. Conversely, every list satisfying (C1)–(C3) yields a simple cycle in the underlying graph.

For two lists $\mathbf{P}, \mathbf{P}' \in \mathfrak{P}(T)$, we write $\mathbf{P} \equiv \mathbf{P}'$ if and only if $\mathbf{P}'$ can be obtained from $\mathbf{P}$ by cyclically reordering its paths, or reversing them all and their order, or a combination of both. Note that $\mathbf{P} \equiv \mathbf{P}'$ entails $s(\mathbf{P}) = s(\mathbf{P}')$ and $\ell(\mathbf{P}) = \ell(\mathbf{P}')$, and that each list $\mathbf{P}$ is in an equivalence class $[\mathbf{P}]$ with exactly $2s(\mathbf{P})$ elements. Let $\widehat{\mathfrak{P}}(T)$ be a subset of $\mathfrak{P}(T)$ obtained by selecting exactly one element from each equivalence class in $\mathfrak{P}(T)$.

Given also a C-permutation σ of the vertex set of T, so that (T,σ) is a C-decorated tree, let $\mathfrak{C}(T,\sigma)$ be the set of lists $\mathbf{P} \in \widehat{\mathfrak{P}}(T)$ that satisfy (C2)–(C3) above. There is thus a 1–1 correspondence between $\mathfrak{C}(T,\sigma)$ and the set of simple cycles in the underlying graph. Let also, recalling (2.4),

$$\mathfrak{C}_k^{[m]}(T,\sigma) := \mathfrak{C}(T,\sigma) \cap \mathfrak{P}_k^{[m]}(T), \tag{5.1}$$

$$\mathfrak{C}_k^{[a,b)}(T,\sigma) := \bigcup_{a \le m < b} \mathfrak{C}_k^{[m]}(T,\sigma), \tag{5.2}$$

and denote the cardinalities of these sets by $C_k^{[m]}(T,\sigma) := |\mathfrak{C}_k^{[m]}(T,\sigma)|$ and $C_k^{[a,b)}(T,\sigma) := |\mathfrak{C}_k^{[a,b)}(T,\sigma)|$.

Furthermore, let $\widetilde{\mathfrak{C}}_k^{[m]}(T,\sigma)$ be the set of lists $\mathbf{P} \in \mathfrak{P}_k^{[m]}(T) \cap \widehat{\mathfrak{P}}(T)$ that satisfy (C2). Thus $\widetilde{\mathfrak{C}}_k^{[m]}(T,\sigma) \supseteq \mathfrak{C}_k^{[m]}(T,\sigma)$. Let further $\widetilde{C}_k^{[m]}(T,\sigma) := |\widetilde{\mathfrak{C}}_k^{[m]}(T,\sigma)|$, and note that $\widetilde{C}_k^{[m]}(T,\sigma) \ge C_k^{[m]}(T,\sigma)$.

We ultimately want to work with random trees, but it will be easier to work with deterministic sequences of trees at first. We say that a sequence (T_n) is a *good sequence of trees* if for all n, T_n is a tree of size n and that the following properties hold: for every fixed $M \ge 1$ and every fixed $\mathbf{m} \in \mathcal{M}$,

$$P^{[\mathbf{m}]}(T_n) \sim \left(\frac{nL_n^2}{M^2}\right)^{s(\mathbf{m})} \kappa^{(\mathbf{m})} = \left(\frac{n^2}{12M^2 g}\right)^{s(\mathbf{m})} \kappa^{(\mathbf{m})}, \tag{5.3}$$

and for each fixed q, w, and c, with $\mathcal{P}_{q,w}(\ell)$ defined in Definition 4.8,

$$\sum_{\mathbf{t} \in \mathcal{P}_{q,w}(cL_n)} N_{\mathbf{t}}(T_n) = O\big(nL_n^{2w-2} \log g\big). \tag{5.4}$$

We will see in Lemma 5.8 that we may assume that the sequence $\mathbf{T}_n$ of random trees is good.

5.2. **Expectation.**

Lemma 5.1. *Let (T_n) be a good sequence of trees. Then, for every $M \geq 1$, $m \geq 0$ and $k \geq 1$, as $n \to \infty$,*

$$\mathbb{E}\big[C_k^{[m]}(T_n, \boldsymbol{\sigma})\big] \to \sum_{|\mathbf{m}|=m, s(\mathbf{m})=k} \frac{\kappa^{(\mathbf{m})}}{2k} \left(\frac{1}{2M^2}\right)^k =: \Lambda_k(m). \tag{5.5}$$

Furthermore,

$$\mathbb{E}\big[\widetilde{C}_k^{[m]}(T_n, \boldsymbol{\sigma}) - C_k^{[m]}(T_n, \boldsymbol{\sigma})\big] \to 0. \tag{5.6}$$

Proof. We start by estimating $\mathbb{E}\,\widetilde{C}_k^{[m]}(T_n, \boldsymbol{\sigma})$. For each given list $\mathbf{P} = (p_1, \dots, p_k) \in \widehat{\mathfrak{P}}(T_n)$, let $\tilde{\pi}(\mathbf{P})$ be the probability that (C2) holds. Then, by definitions and symmetry,

$$\mathbb{E}\big[\widetilde{C}_k^{[m]}(T_n, \boldsymbol{\sigma})\big] = \sum_{\mathbf{P} \in \mathfrak{P}_k^{[m]}(T_n) \cap \widehat{\mathfrak{P}}(T_n)} \tilde{\pi}(\mathbf{P}) = \frac{1}{2k} \sum_{\mathbf{P} \in \mathfrak{P}_k^{[m]}(T_n)} \tilde{\pi}(\mathbf{P}). \tag{5.7}$$

To find $\tilde{\pi}(\mathbf{P})$, note that we may relabel the $2k$ endpoints in $\operatorname{Ext}(\mathbf{P})$ as $1, \dots, 2k$ in an order such that (C2) becomes $2i-1 \sim 2i$ for $i = 1, \dots, k$, i.e., that $2i-1$ and $2i$ belong to the same cycle in $\boldsymbol{\sigma}$. There are two cases: either these k cycles are distinct, or at least two of them coincide. The first event is $\mathcal{E}_{n,g}^{(k)}$ in Lemma 3.4, and that lemma shows that its probability is

$$\mathbb{P}_{n,g}^{(k)} = \mathbb{P}\big(\mathcal{E}_{n,g}^{(k)}\big) \sim \left(\frac{6g}{n^2}\right)^k. \tag{5.8}$$

The second event means that $1, \dots, 2k$ belong to at most $k-1$ different cycles of $\boldsymbol{\sigma}$. Hence, Lemma 3.5 shows that the probability of this event is

$$\sum_{r=1}^{k-1} O\left(\frac{g^r}{n^{2k}}\right) = o\Big(\Big(\frac{g}{n^2}\Big)^k\Big). \tag{5.9}$$

Summing (5.8) and (5.9), we see that

$$\tilde{\pi}(\mathbf{P}) \sim \left(\frac{6g}{n^2}\right)^k, \tag{5.10}$$

uniformly for all $\mathbf{P}$ with $s(\mathbf{P}) = k$.

We develop the sum in (5.7) using (5.10) and (5.3), noting that if $\mathbf{P} \in \mathfrak{P}^{[\mathbf{m}]}$, then $s(\mathbf{P}) = s(\mathbf{m})$; we thus obtain

$$\begin{aligned}
\mathbb{E}\big[\widetilde{C}_k^{[m]}(T_n, \boldsymbol{\sigma})\big] &= \frac{1}{2k} \sum_{|\mathbf{m}|=m, s(\mathbf{m})=k} P^{[\mathbf{m}]}(T_n)(1+o(1)) \left(\frac{6g}{n^2}\right)^k \\
&= \frac{1}{2k} \left(\frac{1}{2M^2}\right)^k \sum_{|\mathbf{m}|=m, s(\mathbf{m})=k} \kappa^{(\mathbf{m})} + o(1).
\end{aligned} \tag{5.11}$$

Next, we consider the difference $\widetilde{C}_k^{[m]}(T_n, \boldsymbol{\sigma}) - C_k^{[m]}(T_n, \boldsymbol{\sigma})$. A given list $\mathbf{P} = (p_1, \dots, p_k) \in \widehat{\mathfrak{P}}(T_n)$ belongs to $\widetilde{\mathfrak{C}}_k^{[m]}(T_n, \boldsymbol{\sigma}) \setminus \mathfrak{C}_k^{[m]}(T_n, \boldsymbol{\sigma})$ if it satisfies (C2) but not (C3).

This means that one of the following holds:

- The $2k$ endpoints belong to at most $k-1$ cycles, which happens with probability $o\big((g/n^2)^k\big)$ by (5.9).

- There exists a vertex v in $\mathbf{P} \setminus \mathrm{Ext}(\mathbf{P})$ such that v belongs to the same cycle as $\mathrm{end}(p_i)$ and $\mathrm{start}(p_{i+1})$ for some i. Hence, we have $2k+1$ points belonging to k cycles, which by Lemma 3.5, happens with probability $O\left(g^k/n^{2k+1}\right)$ for a given v. There are $O(L_n) = O\left(\sqrt{n/g}\right)$ vertices in $\mathbf{P}$; hence by a union bound, the probability of this event is
$$O\left(\frac{g^k}{n^{2k+1}}\right) \times O\left(\sqrt{\frac{n}{g}}\right) = o\left(\left(\frac{g}{n^2}\right)^k\right). \tag{5.12}$$
- There exist two vertices v, v' in $\mathbf{P}\setminus\mathrm{Ext}(\mathbf{P})$ such that v and v' belongs to the same cycle. Hence, we have $2k+2$ points belonging to $k+1$ cycles, which by Lemma 3.5, happens with probability $O\left(g^{k+1}/n^{2k+2}\right)$ for a given pair v, v'. There are $O\big(L_n^2\big) = O\left(n/g\right)$ pairs of vertices in $\mathbf{P}$; hence by a union bound, the probability of this event is
$$O\left(\frac{g^{k+1}}{n^{2k+2}}\right) \times O\left(\frac{n}{g}\right) = o\left(\left(\frac{g}{n^2}\right)^k\right). \tag{5.13}$$

Hence, letting $\pi(\mathbf{P})$ be the probability that both (C2) and (C3) hold, we have
$$0 \le \tilde{\pi}(\mathbf{P}) - \pi(\mathbf{P}) = o\left(\left(\frac{g}{n^2}\right)^k\right), \tag{5.14}$$
uniformly for all $\mathbf{P}$ with $s(\mathbf{P}) = k$. Consequently, arguing as in (5.11), but more crudely, using again (5.3),
$$\mathbb{E}\big[\widetilde{C}_k^{[m]}(T_n,\boldsymbol{\sigma}) - C_k^{[m]}(T_n,\boldsymbol{\sigma})\big] = \sum_{|\mathbf{m}|=m, s(\mathbf{m})=k} \frac{P^{[\mathbf{m}]}(T_n)}{2k} o\Big(\Big(\frac{g}{n^2}\Big)^k\Big) = o(1). \tag{5.15}$$
The proof is completed by (5.11) and (5.15). ∎

5.3. **Higher moments.**

Lemma 5.2. *Let (T_n) be a good sequence of trees, let $(m_1,k_1),\dots,(m_q,k_q)$ be distinct pairs of integers in $\mathbb{Z}_{\ge 0}\times\mathbb{Z}_+$, and let $r_1, r_2, \dots, r_q$ be fixed positive integers, for some $q \ge 1$. Then*
$$\mathbb{E}\Big[\prod_{i=1}^q \Big(\widetilde{C}_{k_i}^{[m_i]}(T_n,\boldsymbol{\sigma})\Big)_{r_i}\Big] = \prod_{i=1}^q \Big(\mathbb{E}\big[\widetilde{C}_{k_i}^{[m_i]}(T_n,\boldsymbol{\sigma})\big]\Big)^{r_i} + o(1) = \prod_{i=1}^q \Lambda_{k_i}(m_i)^{r_i} + o(1). \tag{5.16}$$

The proof is somewhat lengthy, but the main idea is to show that, given a fixed number of distinct cycles in $(T_n, \boldsymbol{\sigma})$ of lengths $O(L_n)$, they are pairwise disjoint whp.

Proof. We argue similarly as in the special case $q = 1$ and $r_1 = 1$ in Lemma 5.1, and write this expectation as
$$\mathbb{E}\prod_{i=1}^q \Big(\widetilde{C}_{k_i}^{[m_i]}(T_n,\boldsymbol{\sigma})\Big)_{r_i} = \widehat{\sum_{(\mathbf{P}(i,j))_{ij}}} \tilde{\pi}((\mathbf{P}(i,j))_{ij}), \tag{5.17}$$

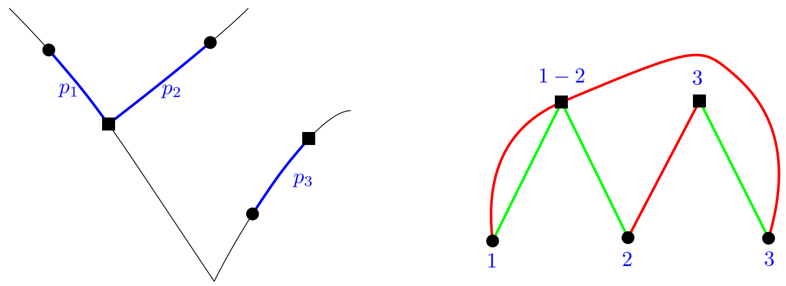

FIGURE 6. Three paths in a tree (left) and their associated graph H (right). The paths are in blue, the start of a path is represented as a square, and its end as a dot.

where we sum over all sequences of distinct lists $(\mathbf{P}(i,j))_{1\le i\le q,\ 1\le j\le r_i}$ such that $\mathbf{P}(i,j) \in \mathfrak{P}_{k_i}^{[m_i]}(T_n) \cap \widehat{\mathfrak{P}}(T_n)$, and $\tilde{\pi}((\mathbf{P}(i,j))_{ij})$ is the probability that every $\mathbf{P}(i,j) \in \widetilde{\mathfrak{C}}_{k_i}^{[m_i]}(T_n, \boldsymbol{\sigma})$. Recalling the definition of $\widehat{\mathfrak{P}}$, we can rewrite (5.17) as

$$\mathbb{E}\prod_{i=1}^{q}\big(\widetilde{C}_{k_i}^{[m_i]}(T_n,\boldsymbol{\sigma})\big)_{r_i} = \sum_{(\mathbf{P}(i,j))_{ij}}^{*} \frac{\tilde{\pi}((\mathbf{P}(i,j))_{ij})}{\prod_{i,j}(2k_i)}, \tag{5.18}$$

where we now sum over all sequences of lists $(\mathbf{P}(i,j))_{ij}$ such that $\mathbf{P}(i,j) \in \mathfrak{P}_{k_i}^{[m_i]}(T_n)$ and no two $\mathbf{P}(i,j)$ are equivalent (for $\equiv$).

For each such sequence $(\mathbf{P}(i,j))_{ij}$, define a graph H with vertex set $V(H) := \bigcup_{i,j}\operatorname{Ext}(\mathbf{P}(i,j))$, the set of endpoints of all participating paths, and edges of two colours as follows: For each list $\mathbf{P}(i,j) = (p_\nu)_1^k$, add for each ν a *green* edge between $\operatorname{start}(p_\nu)$ and $\operatorname{end}(p_\nu)$, and a *red* edge between $\operatorname{end}(p_\nu)$ and $\operatorname{start}(p_{\nu+1})$ (see Figure 6 for an example). (We use here and below the convention $p_{k+1} := p_1$.) Hence, by the definitions above, every $\mathbf{P}(i,j) \in \widetilde{\mathfrak{C}}_{k_i}^{[m_i]}(T_n, \boldsymbol{\sigma})$ if and only if each red edge in H joins two vertices in the same cycle of $\boldsymbol{\sigma}$. Thus, $\tilde{\pi}((\mathbf{P}(i,j))_{ij})$ in (5.18) is the probability of this event.

For each graph H constructed in this way, let H_G be the subgraph consisting of all green edges, and say that a connected component of H_G is a *green component* of H. Define a *red component* in the same way, and let $\zeta_G(H)$ and $\zeta_R(H)$ be the numbers of green and red components, respectively.

Let $M_H = M_H(n)$ be the number of terms in (5.18) with a given graph H. (For some fixed q, $m_1, \dots, m_q$, $k_1, \dots, k_q$, and $r_1, \dots, r_q$.) We estimate M_H as follows. Each green component of H with v vertices corresponds to some set of paths $p_{i,j,\nu}$ such that their union is a connected subtree $\mathbf{t}$ of T_n. All these paths have lengths $O(L_n)$. Furthermore, we can arrange these paths in some order such that for each path after the first, the corresponding green edge in H has an endpoint in common with a previous path, and thus $\mathbf{t}$ is a path tree in $\mathcal{P}_{u,v}(L_n)$ for some $u \le \sum_i r_i k_i$. Hence, the assumption (5.4) implies that there are $O\big(nL_n^{2v-2}\log g\big)$ possible choices for the paths $p_{i,j,\nu}$ corresponding to this green component. Consequently, taking the product over all green components of H, we have

$$M_H = O\big((\log g)^{\zeta_G(H)} L_n^{2v(H)-2\zeta_G(H)} n^{\zeta_G(H)}\big) \\ = O\big((\log g)^{\zeta_G(H)} g^{-v(H)+\zeta_G(H)} n^{v(H)}\big). \tag{5.19}$$

Moreover, we have seen that $\tilde{\pi}((\mathbf{P}(i,j))_{ij})$ in (5.18) is the probability that each red component lies in a single cycle of $\boldsymbol{\sigma}$. This entails that the $v(H)$ vertices in H lie in at most $\zeta_R(H)$ different cycles of $\boldsymbol{\sigma}$, with at least two of the vertices in each cycle, and thus Lemma 3.5 shows that

$$\tilde{\pi}((\mathbf{P}(i,j))_{ij}) = O\big(g^{\zeta_R(H)} n^{-v(H)}\big). \tag{5.20}$$

Consequently, the total contribution to (5.18) for all sequences of lists yielding a given H is, by (5.19) and (5.20),

$$O\big((\log g)^{\zeta_G(H)} g^{\zeta_G(H)+\zeta_R(H)-v(H)}\big). \tag{5.21}$$

Since each green or red component has size at least 2, it follows that $v(H) \geq 2\zeta_G(H)$ and $v(H) \geq 2\zeta_R(H)$, and thus $\zeta_G(H) + \zeta_R(H) \leq v(H)$. If we here have strict inequality, then, since $g \to \infty$, (5.21) shows that the contribution is $o(1)$ and may be ignored. (There is only a finite number of possible H to consider.)

Hence, it suffices to consider the case $\zeta_G(H) = \zeta_R(H) = v(H)/2$. This implies that all green or red components have size 2, and thus are isolated edges. It follows that if two different lists $\mathbf{P}(i_1, j_1)$ and $\mathbf{P}(i_2, j_2)$ contain two paths p_{i_1,j_1,ν_1} and p_{i_2,j_2,ν_2} that have a common endpoint, then these paths have to coincide (up to orientation). Furthermore, if they coincide, and have, say, the same orientation so $\mathrm{end}(p_{i_1,j_1,\nu_1}) = \mathrm{end}(p_{i_2,j_2,\nu_2})$, then the red edges from that vertex have to coincide, so $\mathrm{start}(p_{i_1,j_1,\nu_1+1}) = \mathrm{start}(p_{i_2,j_2,\nu_2+1})$. It follows easily that the two lists $\mathbf{P}(i_1, j_1)$ and $\mathbf{P}(i_2, j_2)$ are equivalent in the sense $\mathbf{P}(i_1, j_1) \equiv \mathbf{P}(i_2, j_2)$ defined above. However, we have excluded this possibility, and this contradiction shows that all paths $p_{i,j,\nu}$ in the lists have disjoint sets of endpoints $\mathrm{Ext}(p_{i,j,\nu})$.

Let $w := \sum_i r_i k_i$ be the total number of paths in the lists in $(\mathbf{P}(i,j))_{ij}$. We have proved that in (5.18), the contribution of the terms where the paths in $(\mathbf{P}(i,j))_{ij}$ do not have $2w$ distinct endpoints is $o(1)$. Hence we may now consider the case where these endpoints are distinct. The calculation is very similar to the one performed in [15], therefore we will omit some details. First, the total number of sequences of lists $(\mathbf{P}(i,j))_{ij}$ such that $\mathbf{P}(i,j) \in \mathfrak{P}_{k_i}^{[m_i]}(T_n)$ and no two $\mathbf{P}(i,j)$ are equivalent is

$$\prod_{i=1}^{q} \prod_{j=1}^{r_i} \big(P_{k_i}^{[m_i]}(T_n) + O(1)\big) \sim \prod_{i=1}^{q} P_{k_i}^{[m_i]}(T_n)^{r_i}, \tag{5.22}$$

which by (5.3) is

$$\prod_{i=1}^{q} \Theta\Big(\Big(\frac{n^2}{g}\Big)^{r_i k_i}\Big) = \Theta\Big(\Big(\frac{n^2}{g}\Big)^{w}\Big). \tag{5.23}$$

If the endpoints of the lists are not distinct, then the construction above yields a graph H with $v(H) \leq 2w - 1$. For each such graph H, the number of such sequences of lists is by (5.19), recalling $\zeta_G(H) \le v(H)/2$,

$$M_H = O\big((\log g)^{v(H)/2} g^{-v(H)/2} n^{v(H)}\big) = O\Big((\log g)^{v(H)/2}\Big(\frac{n}{g^{1/2}}\Big)^{2w-1}\Big)$$
$$= o\Big(\Big(\frac{n}{g^{1/2}}\Big)^{2w}\Big) = o\Big(\Big(\frac{n^2}{g}\Big)^{w}\Big). \tag{5.24}$$

Hence, comparing with (5.23), we see that the number of sequences of lists where the endpoints are not distinct is a fraction $o(1)$ of the total number. In other words, the number of sequences of lists that have $2w$ distinct endpoints is $1 - o(1)$ times the total number in (5.22). For each such sequence of lists $(\mathbf{P}(i,j))_{ij}$, we have

$$\tilde{\pi}((\mathbf{P}(i,j)_{ij}) \sim \Big(\frac{6g}{n^2}\Big)^{w} = \prod_{i=1}^{q}\Big(\frac{6g}{n^2}\Big)^{r_i k_i} \tag{5.25}$$

by Lemmas 3.4 and 3.5 (for the case that some cycle in $\boldsymbol{\sigma}$ covers more than one pair of endpoints). Consequently, (5.18), (5.22) and (5.25) yield

$$\mathbb{E}\prod_{i=1}^{q}\big(\widetilde{C}_{k_i}^{[m_i]}(T_n,\boldsymbol{\sigma})\big)_{r_i} = \big(1+o(1)\big)\prod_{i=1}^{q}\bigg(\frac{P_{k_i}^{[m_i]}(T_n)}{2k_i}\bigg)^{r_i}\Big(\frac{6g}{n^2}\Big)^{r_i k_i} + o(1), \tag{5.26}$$

and (5.16) follows, recalling (5.11) and (5.5)–(5.6), ☐

Lemma 5.3. *Let (T_n) be a good sequence of trees. Then, for every $m \ge 0$ and $k \ge 1$,*

$$C_k^{[m]}(T_n,\boldsymbol{\sigma}) \xrightarrow{\mathrm{d}} \operatorname{Poi}\big(\Lambda_k(m)\big) \tag{5.27}$$

as $n \to \infty$. Moreover, this holds jointly for any (finite) number of pairs (m,k), with the limit Poisson variables being independent.

Proof. Lemma 5.2 implies by the method of moments that $\widetilde{C}_{k_i}^{[m_i]}(T_n,\boldsymbol{\sigma}) \xrightarrow{\mathrm{d}} \operatorname{Poi}\big(\Lambda_{k_i}(m_i)\big)$ jointly, with independent limits, for any set of pairs (m_i,k_i). Furthermore, (5.6) implies that whp $C_{k_i}^{[m_i]}(T_n,\boldsymbol{\sigma}) = \widetilde{C}_{k_i}^{[m_i]}(T_n,\boldsymbol{\sigma})$ for each (m_i,k_i), and thus $C_{k_i}^{[m_i]}(T_n,\boldsymbol{\sigma})$ converge to the same limits. ☐

5.4. **Letting $M \to \infty$.** We have so far kept M fixed. Now it is time to let $M \to \infty$. We therefore add M to the notations when necessary.

Recall $\Lambda_k(m) = \Lambda_k(m;M)$ defined in (5.5). We define also, for integers a and b with $0 \le a \le b < \infty$,

$$\Lambda_k[a,b;M] := \sum_{m=a}^{b-1}\Lambda_k(m;M). \tag{5.28}$$

We begin by finding the asymptotics of this as $M \to \infty$.

Lemma 5.4. *Let $a(M)$ and $b(M)$ be integers depending on M such that $a(M) \le b(M)$ and, as $M \to \infty$,*

$$\frac{a(M)}{M} \to x \tag{5.29}$$

and

$$\frac{b(M)}{M} \to y. \tag{5.30}$$

Then, as $M \to \infty$, for every fixed k,

$$\Lambda_k[a(M), b(M); M] \to \lambda_k^{x,y} := \frac{y^{2k} - x^{2k}}{(2k)(2k)!} = \int_x^y \frac{t^{2k-1}}{(2k)!} \,\mathrm{d}t. \tag{5.31}$$

Proof. We first note that on the one hand, we have, if $\ell \geq k$,

$$\begin{aligned}
\sum_{\substack{|\mathbf{m}|=\ell \\ s(\mathbf{m})=k}} \kappa^{(\mathbf{m})} &\geq \sum_{\substack{|\mathbf{m}|=\ell \\ s(\mathbf{m})=k}} \prod_{i=1}^k 2m_i \\
&= 2^k [z^\ell] \left(\frac{z}{(1-z)^2} \right)^k \\
&= 2^k \binom{\ell+k-1}{\ell-k} \\
&\geq 2^k \frac{(\ell-k)^{2k-1}}{(2k-1)!}
\end{aligned} \tag{5.32}$$

and, similarly,

$$\begin{aligned}
\sum_{\substack{|\mathbf{m}|=\ell \\ s(\mathbf{m})=k}} \kappa^{(\mathbf{m})} &\leq \sum_{\substack{|\mathbf{m}|=\ell \\ s(\mathbf{m})=k}} \prod_{i=1}^k (2m_i+2) \\
&\leq 2^k \sum_{\substack{|\mathbf{m}|=\ell+k \\ s(\mathbf{m})=k}} \prod_{i=1}^k m_i \\
&= 2^k \binom{\ell+2k-1}{\ell} \\
&\leq 2^k \frac{(\ell+2k)^{2k-1}}{(2k-1)!}.
\end{aligned} \tag{5.33}$$

Combining (5.32) and (5.33), we obtain

$$\sum_{\substack{|\mathbf{m}|=\ell \\ s(\mathbf{m})=k}} \kappa^{(\mathbf{m})} = \frac{2^k}{(2k-1)!} \ell^{2k-1} + O\bigl(1 + \ell^{2k-2}\bigr), \tag{5.34}$$

uniformly in $\ell \geq 0$. Hence, since $b(M) = O(M)$ by (5.30),

$$\begin{aligned}
\sum_{\substack{a(M) \leq |\mathbf{m}| < b(M) \\ s(\mathbf{m})=k}} \kappa^{(\mathbf{m})} &= \sum_{\ell=a(M)}^{b(M)} \frac{2^k}{(2k-1)!} \ell^{2k-1} + O\bigl(M^{2k-1}\bigr) \\
&= \frac{2^k}{(2k-1)!} \frac{b(M)^{2k} - a(M)^{2k}}{2k} + O\bigl(M^{2k-1}\bigr)
\end{aligned} \tag{5.35}$$

Consequently, (5.5) and the assumptions (5.29) and (5.30) yield

$$\Lambda_k[a(M), b(M); M] = \sum_{m=a(M)}^{b(M)-1} \Lambda_k(m; M)$$

$$= \frac{1}{2k}\left(\frac{1}{2M^2}\right)^k \Big(\frac{2^k}{(2k-1)!}\frac{b(M)^{2k}-a(M)^{2k}}{2k} + O\big(M^{2k-1}\big)\Big)$$
$$\to \frac{y^{2k}-x^{2k}}{(2k)(2k)!}. \tag{5.36}$$

which completes the proof. □

As we have seen above, a cycle C in the underlying graph of $(T_n, \boldsymbol{\sigma})$ is given by a list $\mathbf{P}$ of paths satisfying (C1)–(C3). Define $s(\mathsf{C}) := s(\mathbf{P})$, the number of paths in the list.

Lemma 5.5. *Let (T_n) be a good sequence of trees, and let $\mathfrak{C}_n$ be the set of simple cycles in the underlying graph of $(T_n, \boldsymbol{\sigma})$. Further, for $k \geq 1$, let $\mathfrak{C}_n^{(k)} := \{\mathsf{C} \in \mathfrak{C}_n : s(\mathsf{C}) = k\}$, and consider the (multi)set of their lengths $\Xi_n^{(k)} := \big\{|\mathsf{C}|/L_n : \mathsf{C} \in \mathfrak{C}_n^{(k)}\big\}$, with L_n given by (1.1). Then the random set $\Xi_n^{(k)}$, regarded as a point process on $[0,\infty)$, converges in distribution to a Poisson process on $[0,\infty)$ with intensity $t^{2k-1}/(2k)!$. Moreover, this holds jointly for any finite number of k.*

Proof. Let $\mathcal{C}_{kn}^{x,y}$ be the number of elements of $\Xi_n^{(k)} \cap [x, y)$, i.e., the number of simple cycles C in the underlying graph of $(T_n, \boldsymbol{\sigma})$ such that $s(\mathsf{C}) = k$ and $xL_n \leq |\mathsf{C}| < yL_n$. The conclusion is equivalent to, with $\lambda_k^{x,y}$ defined in (5.31),

$$\mathcal{C}_{kn}^{x,y} \xrightarrow{\mathrm{d}} \mathrm{Poi}\big(\lambda_k^{x,y}\big), \qquad \text{as } n \to \infty, \tag{5.37}$$

for every interval $[x, y)$ with $0 \leq x < y < \infty$, with joint convergence (to independent limits) for any finite set of disjoint such intervals; moreover, this is to hold jointly for several k. We show that this follows from the similar statement in Lemma 5.3 by letting $M \to \infty$. For notational convenience, we consider only a single k and a single interval $[x, y)$; the general case follows in the same way.

Let

$$a^+(M) = \big(\lfloor xM \rfloor - k\big) \vee 0, \tag{5.38}$$
$$a^-(M) = \lceil xM \rceil, \tag{5.39}$$
$$b^-(M) = \big(\lfloor yM \rfloor - k\big) \vee 0, \tag{5.40}$$
$$b^+(M) = \lceil yM \rceil. \tag{5.41}$$

Consider only M that are so large that $b^-(M) > a^-(M)$. Then, it follows from (2.5) that

$$C_{kMn}^- := C_k^{[a^-(M), b^-(M); M]} \leq \mathcal{C}_{kn}^{x,y} \leq C_{kMn}^+ := C_k^{[a^+(M), b^+(M); M]}. \tag{5.42}$$

By Lemma 5.3, for every fixed M, as $n \to \infty$,

$$C_{kMn}^- = \sum_{m=a^-(M)}^{b^-(M)-1} C_k^{[m]} \xrightarrow{\mathrm{d}} Z_{kM} := \mathrm{Poi}\big(\Lambda_k[a^-(M), b^-(M); M]\big). \tag{5.43}$$

Moreover, by Lemma 5.4, as $M \to \infty$

$$Z_{kM} \xrightarrow{\mathrm{d}} \mathrm{Poi}\big(\lambda_k(x, y)\big). \tag{5.44}$$

Similarly, for every fixed M, as $n \to \infty$,

$$\begin{aligned} C^+_{kMn} - C^-_{kMn} &= C_k^{[a^+(M),a^-(M);M]} + C_k^{[b^-(M),b^+(M);M]} \\ &\xrightarrow{\mathrm{d}} W_M := \operatorname{Poi}\bigl(\widehat{\lambda}(M)\bigr), \end{aligned} \tag{5.45}$$

where

$$\widehat{\lambda}(M) := \Lambda_k[a^+(M), a^-(M); M] + \Lambda_k[b^-(M), b^+(M); M], \tag{5.46}$$

and thus, by Lemma 5.4 again,

$$\widehat{\lambda}(M) \to 0 \qquad \text{as } M \to \infty. \tag{5.47}$$

Consequently, using (5.42), (5.45)

$$\limsup_{n\to\infty} \mathbb{P}\bigl[\mathcal{C}^{x,y}_{kn} \neq C^-_{kMn}\bigr] \le \limsup_{n\to\infty} \mathbb{P}\bigl[C^+_{kMn} \neq C^-_{kMn}\bigr] = \mathbb{P}\bigl(W_M > 0\bigr), \tag{5.48}$$

and thus, by (5.47),

$$\limsup_{M\to\infty} \limsup_{n\to\infty} \mathbb{P}\bigl[\mathcal{C}^{x,y}_{kn} \neq C^-_{kMn}\bigr] \le \lim_{M\to\infty} \mathbb{P}\bigl(W_M > 0\bigr) = 0. \tag{5.49}$$

Finally, (5.43), (5.44) and (5.49) imply (5.37) by [5, Theorem 4.2]. ☐

5.5. **Finishing the proof for good trees.** Let $\mathcal{C}^{x,y}_{kn} = \mathcal{C}^{x,y}_{kn}(T_n, \boldsymbol{\sigma})$ be as in the proof of Lemma 5.5, i.e., the number of simple cycles C in the underlying graph of $(T_n, \boldsymbol{\sigma})$ such that $s(\mathsf{C}) = k$ and $|\mathsf{C}|/L_n \in [x, y)$.

Lemma 5.6. *Let (T_n) be a good sequence of trees. Then, for every $y < \infty$, there exist K and N such that if $n > N$ and $k > K$, then*

$$\mathbb{E}\, \mathcal{C}^{0,y}_{kn} < 2^{-k}. \tag{5.50}$$

Proof. We may assume $y \ge 1$. Fix $M = \lceil 20000y \rceil$. By (5.42),

$$\mathcal{C}^{0,y}_{kn} \le C_k^{[0,\lceil yM \rceil;M]}(T_n, \boldsymbol{\sigma}) = \sum_{m<yM} C_k^{[m]}(T_n, \boldsymbol{\sigma}). \tag{5.51}$$

We have, similarly as in (5.7),

$$\mathbb{E}\bigl[C_k^{[m]}(T_n, \boldsymbol{\sigma})\bigr] = \frac{1}{2k} \sum_{\mathbf{P}\in\mathfrak{P}_k^{[m]}(T_n)} \pi(\mathbf{P}). \tag{5.52}$$

Furthermore, arguing as in the proof of Lemma 5.1, if (C2) and (C3) hold for some $\mathbf{P}$ with $s(\mathbf{P}) = k$, then $\mathcal{E}^{(k)}_{n,g}$ holds (up to a relabelling), and thus by Lemma 3.4,

$$\pi(\mathbf{P}) \le \mathbb{P}\bigl(\mathcal{E}^{(k)}_{n,g}\bigr) \le \Bigl(\frac{C_0 g}{n^2}\Bigr)^k. \tag{5.53}$$

(We may assume that $g/n < 1/7$ for $n > N$.) Consequently, (5.52) yields

$$\begin{aligned} \mathbb{E}\bigl[C_k^{[m]}(T_n, \boldsymbol{\sigma})\bigr] &\le \bigl|\mathfrak{P}_k^{[m]}(T_n)\bigr| \Bigl(\frac{C_0 g}{n^2}\Bigr)^k = P_k^{[m]}(T_n) \Bigl(\frac{C_0 g}{n^2}\Bigr)^k \\ &= \Bigl(\frac{C_0 g}{n^2}\Bigr)^k \sum_{|\mathbf{m}|=m, s(\mathbf{m})=k} P^{[\mathbf{m}]}(T_n). \end{aligned} \tag{5.54}$$

As a special case of (5.3) (with $s(\mathbf{m}) = 1$), we have for every fixed m,

$$P_m(T_n) \sim \frac{n^2}{12M^2 g}(2m+1). \tag{5.55}$$

Hence, there exists N such that

$$P_m(T_n) \le \frac{n^2}{M^2 g}(m+1) \tag{5.56}$$

for every $0 \le m < yM$ and $n > N$.

Consider only $n > N$. Then, (5.56) implies that for every $\mathbf{m} = (m_1, \dots, m_k)$ with $|\mathbf{m}| < yM$ and $s(\mathbf{m}) = k$,

$$P^{[\mathbf{m}]}(T_n) \le \prod_{i=1}^{k} P_{m_i}(T_n) \le \left(\frac{n^2}{M^2 g}\right)^k \prod_{i=1}^{k} (m_i + 1). \tag{5.57}$$

Hence, by the arithmetic-geometric inequality, if also $k \ge K := yM$,

$$P^{[\mathbf{m}]}(T_n) \le \left(\frac{n^2}{M^2 g}\right)^k \left(\frac{|\mathbf{m}| + k}{k}\right)^k \le \left(\frac{2n^2}{M^2 g}\right)^k. \tag{5.58}$$

Consequently, for $k \ge K$ and $n > N$, (5.51) and (5.54) yield

$$\mathbb{E}\, \mathcal{C}^{0,y}_{kn} \le \left(\frac{C_0 g}{n^2}\right)^k \sum_{|\mathbf{m}| < yM, s(\mathbf{m}) = k} P^{[\mathbf{m}]}(T_n). \tag{5.59}$$

There are less than $(yM+1)^k$ lists $\mathbf{m}$ with $|\mathbf{m}| < yM$ and $s(\mathbf{m}) = k$. Hence, (5.59) and (5.58) yield

$$\mathbb{E}\, \mathcal{C}^{0,y}_{kn} \le \left(\frac{C_0 g}{n^2}\right)^k (yM+1)^k \left(\frac{2n^2}{M^2 g}\right)^k \le \left(\frac{4yC_0}{M}\right)^k, \tag{5.60}$$

and the result (5.50) follows by our choice of M. □

Proposition 5.7. *Let (T_n) be a good sequence of trees, and let as in Lemma 5.5 $\mathfrak{C}_n$ be the set of simple cycles in the underlying graph of $(T_n, \boldsymbol{\sigma})$. Consider the (multi)set of the cycle lengths $\Xi_n := \{|\mathsf{C}|/L_n : \mathsf{C} \in \mathfrak{C}_n\}$, with L_n given by (1.1). Then the random set Ξ_n, regarded as a point process on $[0, \infty)$, converges in distribution to a Poisson process Ξ on $[0, \infty)$ with intensity $(\cosh(t) - 1)/t$.*

Proof. Let, recalling (5.31),

$$\lambda^{x,y} := \sum_{k=1}^{\infty} \lambda^{x,y}_k = \int_x^y \sum_{k=1}^{\infty} \frac{t^{2k-1}}{(2k)!} \,\mathrm{d}t = \int_x^y \frac{\cosh t - 1}{t} \,\mathrm{d}t. \tag{5.61}$$

Then, the conclusion is equivalent to

$$\mathcal{C}^{x,y}_n \xrightarrow{\mathrm{d}} \mathrm{Poi}\big(\lambda^{x,y}\big), \qquad \text{as } n \to \infty, \tag{5.62}$$

for every interval $[x, y)$ with $0 \le x < y < \infty$, with joint convergence (to independent limits) for any finite set of disjoint such intervals. As in the proof of Lemma 5.5, for notational convenience, we consider only a single interval $[x, y)$; the general case follows in the same way.

We have $\mathcal{C}^{x,y}_n = \sum_{k=1}^{\infty} \mathcal{C}^{x,y}_{kn}$. Define, for $K \ge 1$,

$$\mathcal{C}^{x,y}_{\le K.n} := \sum_{k=1}^{K} \mathcal{C}^{x,y}_{kn}. \tag{5.63}$$

For each fixed K, Lemma 5.5 implies that, as $n \to \infty$,

$$\mathcal{C}^{x,y}_{\le K,n} \xrightarrow{\text{d}} \operatorname{Poi}\bigl(\lambda^{x,y}_{\le K}\bigr), \tag{5.64}$$

with

$$\lambda^{x,y}_{\le K} := \sum_{k=1}^{K} \lambda^{x,y}_{k}. \tag{5.65}$$

Hence, as $K \to \infty$, $\lambda^{x,y}_{\le K} \to \lambda^{x,y}$, and thus

$$\operatorname{Poi}\bigl(\lambda^{x,y}_{\le K}\bigr) \xrightarrow{\text{d}} \operatorname{Poi}\bigl(\lambda^{x,y}\bigr). \tag{5.66}$$

Moreover, by Lemma 5.6, for large enough K,

$$\limsup_{n\to\infty} \mathbb{P}(\mathcal{C}^{x,y}_{\le K,n} \neq \mathcal{C}^{x,y}_{n}) \le \limsup_{n\to\infty} \sum_{k>K} \mathbb{P}\bigl[\mathcal{C}^{0,y}_{kn} > 0\bigr] \le \sum_{k>K} 2^{-k} = 2^{-K}, \tag{5.67}$$

which tends to 0 as $K \to \infty$. Hence, by [5, Theorem 4.2] again, (5.62) follows. □

5.6. **Finishing the proof for random trees and maps.**

Lemma 5.8. *For every fixed M and $\mathbf{m} \in \mathcal{M}$,*

$$\left(\frac{M^2}{nL_n^2}\right)^{s(\mathbf{m})} P^{[\mathbf{m}]}(\mathbf{T}_n) \xrightarrow{\text{p}} \kappa^{(\mathbf{m})}, \tag{5.68}$$

and for every fixed q, w and c,

$$\frac{1}{nL_n^{2w-2}\log g} \sum_{\mathbf{t}\in\mathcal{P}_{q,w}(cL_n)} N_{\mathbf{t}}(\mathbf{T}_n) \xrightarrow{\text{p}} 0. \tag{5.69}$$

Moreover, we may assume that the random trees $\mathbf{T}_n$ are coupled such that the convergences (5.68) *and* (5.69) *hold a.s., and thus* $(\mathbf{T}_n)$ *is good a.s.*

Proof. First, (5.68) is a consequence of Lemmas 4.12 and Lemma 4.13, and (5.69) follows from Lemma 4.11.

Consider the infinite vector $\mathbf{Y}$ of the left-hand sides of (5.68) and (5.69), for all M, $\mathbf{m}$, q, w and integers c. Then (5.68)–(5.69) say that $\mathbf{Y}$ converges in probability to some non-random vector $\mathbf{y}$, in the product space $\mathbb{R}^\infty$. By the Skorohod coupling theorem [16, Theorem 4.30], we may couple the random trees $\mathbf{T}_n$ such that $\mathbf{Y} \to \mathbf{y}$ a.s., and the result follows. □

Proof of Theorems 1.1 and 1.2. By Lemma 5.8, we may without loss of generality assume that the sequence of random graphs $\mathbf{T}_n$ is good a.s. Hence, if we condition on $(\mathbf{T}_n)$, we may apply Proposition 5.7. Consequently, the conclusion of Proposition 5.7 holds also for the sequence of random trees $\mathbf{T}_n$. Moreover, the underlying graph $G(\mathbf{T}_n, \boldsymbol{\sigma})$ has the same distribution as the random unicellular map $\mathbf{U}_{n,g}$, and thus the result holds also for it. □

5.7. **Further remarks.**

Remark 5.9. We have, for definiteness, considered only simple cycles in this paper. However, it follows from the proofs above, in particular the proof of Lemma 5.2, that whp all simple cycles in $\mathbf{U}_{n,g}$ with length $\leq C$ are disjoint, and thus every primitive cycle of length $\leq C$ is simple. (Recall that a primitive cycle may intersect itself, but it may not consist of another cycle repeated several times.) We omit the details. □

Remark 5.10. It follows from the proofs above that the convergence in Proposition 5.7 holds jointly with the convergence for each fixed k in Lemma 5.5. An alternative interpretation of this is that if $\overline{\mathbb{N}} := \{1, 2, \dots, \infty\}$ is the one-point compactification of $\mathbb{N}$, then the (multi)set of points $\widehat{\Xi}_n := \{(|\mathsf{C}|/L_n, s(\mathsf{C})) : \mathsf{C} \in \mathfrak{C}_n\}$, regarded as a point process in $\mathcal{S} := [0, \infty) \times \overline{\mathbb{N}}$, converges to a certain Poisson process $\widehat{\Xi}$ on $\mathcal{S}$. (Cf. e.g. [14, Section 4] for the importance of using $\overline{\mathbb{N}}$ instead of $\mathbb{N}$ here.) □

This joint convergence to Poisson processes implies, for example, by standard arguments the following.

Corollary 5.11. *Let C_1 be the shortest cycle in the underlying graph of $(\mathbf{T}_n, \boldsymbol{\sigma})$. (This is whp unique by Theorem 1.1.) Then $s(\mathsf{C}_1)$ has a limiting distribution, as $n \to \infty$, given by*

$$\mathbb{P}\big(s(\mathsf{C}_1) = k\big) \to p_k := \int_0^\infty \frac{z^{2k-1}}{(2k)!} \exp\Big(-\int_0^z \frac{\cosh t - 1}{t} \,\mathrm{d}t\Big) \,\mathrm{d}z. \tag{5.70}$$

Numerically we have $p_1 \doteq 0.792$, $p_2 \doteq 0.177$, $p_3 \doteq 0.028$, $p_4 \doteq 0.003$.

However, as far as we know, $s(\mathsf{C})$ has no natural interpretation for cycles in the unicellular map.

References

[1] O. Angel, G. Chapuy, N. Curien, and G. Ray. The local limit of unicellular maps in high genus. *Electron. Commun. Probab.*, 18(86):1–8, 2013.

[2] O. Angel and O. Schramm. Uniform infinite planar triangulations. *Comm. Math. Phys.*, 241(2-3):191–213, 2003.

[3] E. A. Bender and E. Canfield. The asymptotic number of rooted maps on a surface. *J. Combin. Theory Ser. A*, 43(2):244–257, 1986.

[4] J. Bettinelli. Geodesics in Brownian surfaces (Brownian maps). *Ann. Inst. Henri Poincaré Probab. Stat.*, 52(2):612–646, 2016.

[5] P. Billingsley. *Convergence of Probability Measures.* Wiley, New York, 1968.

[6] T. Budzinski and B. Louf. Local limits of uniform triangulations in high genus. *Invent. Math.*, 223(1):1–47, 2021.

[7] T. Budzinski and B. Louf. Local limits of bipartite maps with prescribed face degrees in high genus. *Ann. Probab.*, to appear. `arXiv`:2012.05813.

[8] G. Chapuy, V. Féray, and E. Fusy. A simple model of trees for unicellular maps. *J. Combin. Theory Ser. A*, 120(8), 2064–2092, 2013.

[9] G. Chapuy, M. Marcus, and G. Schaeffer. A bijection for rooted maps on orientable surfaces. *SIAM J. Discrete Math.*, 23(3):1587–1611, 2009.

[10] M. Drmota. *Random Trees. An interplay between Combinatorics and Probability*. Springer, Vienna, 2009.
[11] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge Univ. Press, Cambridge, 2009.
[12] C. Gilmore, E. Le Masson, T. Sahlsten and J. Thomas. Short geodesic loops and L^p norms of eigenfunctions on large genus random surfaces. *Geom. Funct. Anal.*, online, 2021.
[13] L. Guth, H. Parlier, and R. Young. Pants decompositions of random surfaces. *Geom. Funct. Anal.*, 21(5):1069–1090, 2011.
[14] S. Janson, Cycles and unicyclic components in random graphs. *Combin. Probab. Comput.* 12 (2003), 27–52.
[15] S. Janson and B. Louf. Short cycles in high genus unicellular maps. *Ann. Inst. Henri Poincaré Probab. Stat.*, to appear. arXiv:2103.02549.
[16] O. Kallenberg, *Foundations of Modern Probability.* 2nd ed., Springer-Verlag, New York, 2002.
[17] O. Kallenberg, *Random Measures, Theory and Applications*. Springer, Cham, Switzerland, 2017.
[18] J.-F. Le Gall. Uniqueness and universality of the Brownian map. *Ann. Probab.*, 41:2880–2960, 2013.
[19] M. Le Masson and T. Sahlsten. Quantum ergodicity for Eisenstein series on hyperbolic surfaces of large genus. Preprint, 2020. arXiv:2006.14935.
[20] M. Lipnowski and A. Wright. Towards optimal spectral gaps in large genus. Preprint, 2021. arXiv:2103.07496.
[21] B. Louf. Planarity and non-separating cycles in uniform high genus quadrangulations. *Probab. Theory Relat. Fields*, online, 2021.
[22] B. Louf. Large expanders in high genus unicellular maps. Preprint, 2021. arXiv:2102.11680.
[23] M. Magee. Letter to Bram Petri. 2020. https://www.maths.dur.ac.uk/users/michael.r.magee/diameter.pdf
[24] G. Miermont. The Brownian map is the scaling limit of uniform random plane quadrangulations. *Acta Math.*, 210(2):319–401, 2013.
[25] M. Mirzakhani. Growth of Weil–Petersson volumes and random hyperbolic surfaces of large genus. *J. Differential Geom.*, 94(2):267–300, 2013.
[26] M. Mirzakhani and B. Petri. Lengths of closed geodesics on random surfaces of large genus. *Comment. Math. Helv.*, 94(4):869–889, 2019.
[27] L. Monk. Benjamini–Schramm convergence and spectrum of random hyperbolic surfaces of high genus. Preprint, 2020. arXiv:2002.00869.
[28] L. Monk and J. Thomas. The tangle-free hypothesis on random hyperbolic surfaces. *Int. Math. Res. Not.*, online, 2021.
[29] X. Nie, Y. Wu, and Y. Xue. Large genus asymptotics for lengths of separating closed geodesics on random surfaces. Preprint, 2020. arXiv:2009.07538.
[30] H. Parlier, Y. Wu, and Y. Xue. The simple separating systole for hyperbolic surfaces of large genus. *J. Inst. Math. Jussieu*, online, 2021.
[31] G. Ray. Large unicellular maps in high genus. *Ann. Inst. H. Poincaré Probab. Statist.*, 51(4):1432–1456, 2015.

[32] G. Schaeffer. *Conjugaison d'arbres et cartes combinatoires aléatoires.* Thèse de doctorat, Université Bordeaux I, 1998.

[33] J. Thomas. Delocalisation of eigenfunctions on large genus random surfaces. Preprint, 2020. arXiv:2002.01403

[34] W. T. Tutte. A census of planar maps. *Canad. J. Math.*, 15:249–271, 1963.

[35] S. Wolpert. On the Weil–Petersson geometry of the moduli space of curves. *Amer. J. Math.*, 107(4):969–997, 1985.

[36] A. Wright. A tour through Mirzakhani's work on moduli spaces of Riemann surfaces. *Bull. Amer. Math. Soc. (N.S.)* 57(3):359–408, 2020.

[37] Y. Wu and Y. Xue. Random hyperbolic surfaces of large genus have first eigenvalues greater than $\frac{3}{16} - \varepsilon$. Preprint, 2021. arXiv:2102.05581

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

Email address: svante.janson@math.uu.se

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

Email address: baptiste.louf@math.uu.se