

# BOLTZMANN SAMPLERS, PÓLYA THEORY, AND CYCLE POINTING

MANUEL BODIRSKY, ÉRIC FUSY, MIHYUN KANG, AND STEFAN VIGERSKE

ABSTRACT. We introduce a general method to count unlabeled combinatorial structures and to efficiently generate them at random. The approach is based on pointing unlabeled structures in an “unbiased” way that a structure of size  $n$  gives rise to  $n$  pointed structures. We extend Pólya theory to the corresponding pointing operator, and present a random sampling framework based on both the principles of Boltzmann sampling and on Pólya operators. All previously known unlabeled construction principles for Boltzmann samplers are special cases of our new results. Our method is illustrated on several examples: in each case, we provide enumerative results and efficient random samplers. The approach applies to unlabeled families of plane and nonplane unrooted trees, and tree-like structures in general, but also to families of graphs (such as cacti graphs and outerplanar graphs) and families of planar maps.

*This is the extended and revised journal version of a conference paper with the title “An unbiased pointing operator for unlabeled structures, with applications to counting and sampling”, which appeared in the Proceedings of ACM-SIAM Symposium on Discrete Algorithms (SODA’07), 7-9 January 2007, New Orleans.*

## 1. INTRODUCTION

Pointing (or rooting) is an important tool to derive decompositions of combinatorial structures, with many applications in enumerative combinatorics. Such decompositions can for instance be used in polynomial-time algorithms that sample structures of a combinatorial class uniformly at random. For a class of *labeled* structures, pointing corresponds to taking the derivative of the (typically exponential) generating function of the class. In other words, each structure of size  $n$  gives rise to  $n$  pointed (or rooted) structures. Other important operations on classes of combinatorial structures are the disjoint union, the product, and the substitution operation – they correspond to addition, multiplication, and composition of the associated generating functions. Together with the usual basic classes of combinatorial structures (finite classes, the class of finite sets, the class of finite sequences, and the class of cycles), this collection of constructions is a powerful device to define a great variety of combinatorial families.

If a class of structures can be described by recursive specifications involving pointing, disjoint unions, products, substitutions, and the basic classes, then the techniques of analytic combinatorics can be applied to obtain enumerative results, to study statistical properties of random structures in the class, and to derive efficient random samplers. An expository account for this line of research is [10]. Among recent developments in the area of random sampling are *Boltzmann samplers* [9], which are an attractive alternative to the recursive method of sampling [25, 11]. Both approaches provide in a systematic way polynomial-time uniform random generators for decomposable combinatorial classes. The advantage of Boltzmann samplers over the recursive method of sampling is that Boltzmann samplers operate in linear time if a small relative tolerance is allowed for the size of the output, and that they have a small preprocessing cost, which makes it possible to sample very large structures.

A third general random sampling approach is based on Markov chains. This approach does not require recursive decompositions of structures, and is applicable on a wide range of combinatorial classes. However, Markov chain methods are mostly limited to *approximate* uniformity. Moreover, it is usually difficult to obtain bounds on the rate of convergence to the uniform distribution [20].

All the results in this paper concern classes of *unlabeled* combinatorial structures, i.e., the structures are considered up to isomorphism. In the case of the class of all graphs, the labeled and the unlabeled model do not differ much, which is due to the fact that almost all graphs do not have a non-trivial automorphism (see e.g., [5, Ch. 9.] and [17]). However, for many interesting classes of

combinatorial structures (and for most of the classes studied in this paper), the difference between the labeled and the unlabeled setting does matter.

The Markov chain approach can be adapted also to sample unlabeled structures approximately uniformly at random, based on the orbit counting lemma [19]. Again, the rate of convergence is usually very difficult to analyze, and frequently these Markov chains do not have a polynomial convergence rate [15]. For unlabeled structures, the approach faces additional difficulties: to computationally implement the transitions of the Markov chain, we have to be able to generate structures with a given symmetry uniformly at random (this is for instance a difficult task for planar graphs), and we have to be able to generate a symmetry of a given structure (for the class of all finite graphs, this task is computationally equivalent to the graph isomorphism problem).

To enumerate unlabeled structures, typically the *ordinary* generating functions are the appropriate tool. Disjoint unions and products for unlabeled structures then still correspond to addition and multiplication of the associated generating functions as for labeled cases. Boltzmann samplers for classes described by recursive specifications involving these operations have recently been developed in [13]. However, the *substitution* operation for unlabeled structures no longer corresponds to the composition of generating functions, due to the symmetries an unlabeled structure might have. This problem can be solved by Pólya theory, which uses the generalization of generating functions to *cycle index sums* to take care of potential symmetries. Pólya theory provides a computation rule for the cycle index sum associated to a substitution construction. The presence of symmetries leads to another problem with the *pointing* (or *rooting*) operator: the fundamental property that a structure of size  $n$  gives rise to  $n$  pointed structures does not hold. Indeed, if a structure of size  $n$  has a non-trivial automorphism, then it corresponds to less than  $n$  pointed structures (because pointing at two vertices in symmetric positions produces the same pointed structure). Thus, for unlabeled structures, the classical pointing operator does not correspond to the derivative of the ordinary generating function.

In this paper, we introduce an *unbiased* pointing operator for unlabeled structures. The operator is unbiased in the sense that a structure of size  $n$  gives rise to  $n$  pointed structures. It produces *cycle-pointed* structures, i.e., combinatorial structures  $A$  together with a marked cyclic sequence of atoms of  $A$  that is a cycle of an automorphism of  $A$ . Accordingly, we call our operator *cycle-pointing*. The idea is based on Parker's lemma in permutation group theory [7] (this will be discussed in Remark 3.1). We develop techniques to apply this new pointing operator to enumeration and random sampling of unlabeled combinatorial classes. The crucial point is that cycle-pointing is *unbiased*. As a consequence, performing both tasks of enumeration and uniform random sampling on a combinatorial class is equivalent to performing these tasks on the associated cycle-pointed class.

To understand how we use our operator, it is instructive to look at the class of *free trees*, i.e., unrooted and nonplane *trees* (equivalently, acyclic connected graphs). Building on the work of Cayley and Pólya, Otter [26] determined the exact and asymptotic number of free trees. To this end, he developed the by-now standard dissimilarity characteristic equation, which relates the number of free trees with the number of rooted nonplane trees; see [16]. The best-known method to sample free trees uniformly at random is due to Wilf [34], and uses the concept of the centroid of a tree. The method is an example of application of the recursive method of sampling and requires a pre-processing step where a table of quadratic size is computed.

Cycle-pointing provides a new way to count and sample free trees. Both tasks are carried out on cycle-pointed (nonplane) trees. The advantage of studying cycle-pointed trees is that the pointed cycle provides a starting point for a recursive decomposition. In the case of cycle-pointed trees, we can formulate such a decomposition using standard constructions such as disjoint union, product, and substitution (which requires to suitably adapt these constructions to the cycle-pointed framework). We want to stress that, despite some superficial similarities, this method for counting free trees is fundamentally different from the previously existing methods mentioned above, and it proves particularly fruitful in the context of random generation. Indeed, the *dissimilarity characteristic equation* [26] and the *dissymmetry theorem* [2] both lead to generating function equations involving *subtraction*. However, subtraction yields massive rejection when translated into a random generator for the class of structures (both for Boltzmann samplers and for the

recursive method). In contrast, the equations produced by the method based on cycle-pointing have only positive signs, and the existence of a Boltzmann sampler for (cycle-pointed) free trees, with no rejection involved, will follow directly from the general results derived in this paper. As usual, the Boltzmann samplers we obtain have a running time that is linear in the size of the structure generated, and have small pre-processing cost.

Similarly, we can decompose plane and nonplane trees, and more generally all sorts of tree-like structures. By the observation that the block decomposition of a graph has also a tree-like structure, we can apply the method to classes of graphs where the two-connected components can be explicitly enumerated. This leads to efficient Boltzmann samplers, for instance, for cacti graphs and outerplanar graphs, improving on the generators of [4]. Further, our strategy is not limited to only tree-like structures, but can also be applied to other classes of structures that allow for a recursive decomposition. To demonstrate this, we sketch how the method can be applied to count and sample certain classes of planar maps.

**Outline of the paper.** To formalize our general results on enumeration and sampling for classes of unlabeled combinatorial structures in full generality, we apply the concept of *combinatorial species*; we recall this concept in Section 2. In Section 3 we introduce *cycle-pointed species* and the cycle-pointing operator. Section 4 is devoted to applications of our cycle-pointing operator in enumeration. The technique in which we use the operator to obtain recursive decomposition strategies for unlabeled enumeration is very generally applicable; we illustrate it by the enumeration of (unrooted) non-plane and plane trees, cacti graphs, and maps. Finally, in Section 5, we present how to apply the concepts introduced in this paper to obtain highly efficient random sampling procedures for unlabeled structures. This is illustrated by applications for sampling of concrete and fundamental classes of unlabeled combinatorial structures; several of these concrete sampling results are either new or improve the state-of-the-art of sampling efficiency.

## 2. PRELIMINARIES

In this paper we work with classes of *combinatorial structures*, such as graphs, relational structures, functions, trees, plane trees, maps, words, terms, or permutations. There are many ways to define formally these objects; for example, rooted trees can be coded as special types of directed graphs, but also as terms. Which formal representation of a combinatorial structure is best usually depends on the application.

When we are interested in combinatorial enumeration, the differences in representation are not essential; in the example above, we have the same number of rooted trees, no matter how they are represented. We would thus like to have a formalism that is sufficiently abstract so that our results apply to broad classes of combinatorial structures. At the same time, we would like to have a formalism for classes of combinatorial structures that supports fundamental construction operations for combinatorial classes, such as formation of *disjoint unions*, *products*, and *substitution*, and allows to state general results about enumeration and random sampling.

The theory of combinatorial species is an elegant tool that fully satisfies the needs mentioned above. We give a brief introduction to species, and refer to Bergeron, Labelle, and Leroux [2] for a broader treatment of the topic. The fundamental and well-known classes of combinatorial structures that we treat in Section 4 provide ample illustration of the concepts we define in this section.

**2.1. Combinatorial Species.** We closely follow the presentation in [2]. A *species of structures* is a functor  $\mathcal{A}$  from finite sets  $U$  to finite sets  $\mathcal{A}[U]$ , together with a rule that produces for each bijection  $\sigma : U \rightarrow V$  a function from  $\mathcal{A}[U]$  to  $\mathcal{A}[V]$ . Slightly abusing notation, this function will also be denoted by  $\mathcal{A}[\sigma]$ . The functions  $\mathcal{A}[\sigma]$  must satisfy the following two (functorial) properties:

- for all bijections  $\sigma : U \rightarrow V$  and  $\tau : V \rightarrow W$ ,

$$\mathcal{A}[\tau \circ \sigma] = \mathcal{A}[\tau] \circ \mathcal{A}[\sigma] ;$$

- for the identity map  $\text{Id}_U : U \rightarrow U$ ,

$$\mathcal{A}[\text{Id}_U] = \text{Id}_{\mathcal{A}[U]} .$$

An element  $A \in \mathcal{A}[U]$  is called an  $\mathcal{A}$ -*structure on  $U$* . The elements of  $U$  are called the *atoms of  $\mathcal{A}[U]$* , and we refer to  $|U|$  as the *size of  $A$* . The function  $\mathcal{A}[\sigma]$  is called the *transport of  $\mathcal{A}$ -structures along  $\sigma$* . We write  $\sigma \cdot A$  for  $\mathcal{A}[\sigma](A)$ .

**Definition 2.1.** Let  $A_1 \in \mathcal{A}[U]$  and  $A_2 \in \mathcal{A}[V]$  be two  $\mathcal{A}$ -structures. A bijection  $\sigma : U \rightarrow V$  is called an *isomorphism from  $A_1$  to  $A_2$*  if  $A_2 = \sigma \cdot A_1 = \mathcal{A}[\sigma](A_1)$ . An isomorphism from  $A_1$  to  $A_1$  is called an *automorphism of  $A_1$* .

The advantage of species is that the rule  $\mathcal{A}$  that produces the structures  $\mathcal{A}[U]$  and the transport functions  $\mathcal{A}[\sigma]$  can be described in any way; the book [2] gives instructive examples where this description is by axiomatic systems, explicit constructions, algorithms, combinatorial operations, or functional equations.

**Definition 2.2.** A species  $\mathcal{A}$  is a *subspecies* of a species  $\mathcal{B}$  (and we write  $\mathcal{A} \subseteq \mathcal{B}$ ) if it satisfies the following two conditions:

- for any finite set  $U$ ,  $\mathcal{A}[U] \subseteq \mathcal{B}[U]$ ;
- for any  $\sigma : U \rightarrow V$  we have  $\mathcal{A}[\sigma] = \mathcal{B}[\sigma]|_{\mathcal{A}[U]}$ .

**2.2. Enumeration.** For all finite sets  $U$ , the number of  $\mathcal{A}$ -structures on  $U$  depends only on the number of elements of  $U$  (and not on the elements of  $U$ ). If  $\mathcal{A}$  is a species, we write  $a_n$  for  $|\mathcal{A}[\{1, \dots, n\}]|$ , the cardinality of the set of  $\mathcal{A}$ -structures on  $\{1, \dots, n\}$ . The series

$$\mathcal{A}(x) := \sum_{n \geq 0} \frac{1}{n!} a_n x^n \quad (1)$$

is called the *exponential generating series* of the species  $\mathcal{A}$ .

In this article we focus on *unlabeled* structures, i.e., we consider structures up to isomorphism. We may restrict ourselves to structures on sets of the form  $U = [1..n] := \{1, \dots, n\}$ , and write  $\mathcal{A}[n]$  for  $\mathcal{A}[\{1, \dots, n\}]$ . We define the equivalence relation  $\sim$  on  $\mathcal{A}[n]$  by setting  $A_1 \sim A_2$ , for  $A_1, A_2 \in \mathcal{A}[n]$ , if and only if there is an isomorphism  $\sigma : [1..n] \rightarrow [1..n]$  from  $A_1$  to  $A_2$ . We also say that  $A_1$  and  $A_2$  have the same *isomorphism type*, and the equivalence classes of  $\mathcal{A}$ -structures on  $[1..n]$  with respect to  $\sim$  are also called *unlabeled  $\mathcal{A}$ -structures of size  $n$* . The set of all those equivalence classes is denoted by  $\tilde{\mathcal{A}}_n$ , and we write  $\tilde{a}_n$  for its cardinality  $|\tilde{\mathcal{A}}_n|$ . The series

$$\tilde{\mathcal{A}}(x) := \sum_{n \geq 0} \tilde{a}_n x^n \quad (2)$$

is called the *ordinary generating series (OGS)* of  $\mathcal{A}$ . We use the classical notation  $[x^n]\tilde{\mathcal{A}}(x)$  to denote the  $n$ -th coefficient  $\tilde{a}_n$  in the power series  $\tilde{\mathcal{A}}(x)$ .

**2.3. Cycle index sums.** For a species  $\mathcal{A}$  and each  $n \geq 0$ , a *symmetry of  $\mathcal{A}$  of size  $n$*  is a pair  $(A, \sigma)$  where  $A$  is from  $\mathcal{A}[n]$  and  $\sigma$  is an automorphism of  $A$ . We call  $A$  the *underlying structure* of the symmetry  $(A, \sigma)$ . Notice that the automorphism  $\sigma$  can be the identity. We denote by  $\text{Sym}(\mathcal{A})$  the species defined by

$$\text{Sym}(\mathcal{A})[n] := \{(A, \sigma) \mid A \in \mathcal{A}[n] \text{ and } \sigma \text{ is an automorphism of } A\}.$$

The definition of the transport of  $\text{Sym}(\mathcal{A})$  is obvious (the definition of symmetries as an auxiliary species is standard; see Section 4.3 in [2]). The *weight-monomial* of a symmetry  $(A, \sigma)$  of size  $n$  is defined as

$$w_{(A, \sigma)} := \frac{1}{n!} \prod_{i=1}^n s_i^{c_i(\sigma)}, \quad (3)$$

where, for  $i \in [1..n]$ ,  $s_i$  is a formal variable and  $c_i(\sigma)$  is the number of cycles of  $\sigma$  of length  $i$ . For simplicity, in the following we will write  $c_i$  for  $c_i(\sigma)$  if the corresponding automorphism is clear from the context. The *cycle index sum* of  $\mathcal{A}$ , denoted by  $Z_{\mathcal{A}}(s_1, s_2, \dots)$ , or shortly  $Z_{\mathcal{A}}$ , is the formal power series defined as the sum of the weight-monomials over all the symmetries of  $\mathcal{A}$ ,

$$Z_{\mathcal{A}}(s_1, s_2, \dots) := \sum_{n \geq 0} \left( \sum_{(A, \sigma) \in \text{Sym}(\mathcal{A})[n]} w_{(A, \sigma)} \right). \quad (4)$$

Basic species	Notation	Cycle index sum
Empty species	$\mathbf{0}$	$Z_{\mathbf{0}} = 0$
Neutral species	$\mathbf{1}$	$Z_{\mathbf{1}} = 1$
Atomic species	$X$	$Z_X = s_1$
Sequences with $k$ atoms	$\text{SEQ}^{[k]}$	$Z_{\text{SEQ}^{[k]}} = s_1^k$
Sets with $k$ atoms	$\text{SET}^{[k]}$	$Z_{\text{SET}^{[k]}} = [x^k] \exp\left(\sum_{r \geq 1} \frac{1}{r} x^r s_r\right)$
Cycles with $k$ atoms	$\text{CYC}^{[k]}$	$Z_{\text{CYC}^{[k]}} = \frac{1}{k} \sum_{r k} \phi(r) s_r^{k/r}$
Species of sequences	SEQ	$Z_{\text{SEQ}} = \frac{1}{1 - s_1}$
Species of sets	SET	$Z_{\text{SET}} = \exp\left(\sum_{r \geq 1} \frac{1}{r} s_r\right)$
Species of cycles	CYC	$Z_{\text{CYC}} = 1 + \sum_{r \geq 1} \frac{\phi(r)}{r} \log\left(\frac{1}{1 - s_r}\right)$
Construction	Notation	Cycle index sum
Union of two species	$\mathcal{M} = \mathcal{A} + \mathcal{B}$	$Z_{\mathcal{M}} = Z_{\mathcal{A}} + Z_{\mathcal{B}}$
Product of two species	$\mathcal{M} = \mathcal{A} \cdot \mathcal{B}$	$Z_{\mathcal{M}} = Z_{\mathcal{A}} \cdot Z_{\mathcal{B}}$
Composition of two species	$\mathcal{M} = \mathcal{A} \circ \mathcal{B}$	$Z_{\mathcal{M}} = Z_{\mathcal{A}} \circ Z_{\mathcal{B}}$

FIGURE 1. The cycle index sums of basic species and of species composed by  $+$ ,  $\cdot$ , and  $\circ$ .

Cycle index sums for classes of combinatorial structures have been introduced by Pólya [28]. The following fact, which is based on Burnside's lemma, shows that cycle index sums are a refinement of ordinary generating series.

**Lemma 1** (Pólya). *Let  $\mathcal{A}$  be a species of structures. For  $n \geq 0$ , each unlabeled structure  $\tilde{A} \in \tilde{\mathcal{A}}_n$  gives rise to  $n!$  symmetries, i.e., there are  $n!$  symmetries  $(A, \sigma)$  such that  $A \in \tilde{A}$ . Hence,*

$$Z_{\mathcal{A}}(x, x^2, x^3, \dots) = \tilde{\mathcal{A}}(x). \quad (5)$$

In the proofs of combinatorial identities it will be convenient to identify species that are essentially the same.

**Definition 2.3.** Let  $\mathcal{A}$  and  $\mathcal{B}$  be two species. An *isomorphism* from  $\mathcal{A}$  to  $\mathcal{B}$  is a family of bijections  $\alpha_U : \mathcal{A}[U] \rightarrow \mathcal{B}[U]$  which satisfies the following *naturality condition*: for any bijection  $\sigma : U \rightarrow V$  between two finite sets  $U$  and  $V$ , and for any  $\mathcal{A}$ -structure  $A$  on  $U$ , one must have

$$\sigma \cdot \alpha_U(A) = \alpha_V(\sigma \cdot A).$$

It is known (see [2]) that when  $\mathcal{A}$  is isomorphic to  $\mathcal{B}$ , then  $\mathcal{A}(x) = \mathcal{B}(x)$ ,  $\tilde{\mathcal{A}}(x) = \tilde{\mathcal{B}}(x)$ , and  $Z_{\mathcal{A}}(x_1, x_2, \dots) = Z_{\mathcal{B}}(x_1, x_2, \dots)$ . Hence, for the purposes of combinatorial enumeration we can even identify isomorphic species, and as in [2] we write  $\mathcal{A} = \mathcal{B}$  when  $\mathcal{A}$  and  $\mathcal{B}$  are isomorphic species.

**2.4. Basic species and combinatorial constructions.** We now recall a collection of basic species and combinatorial constructions. We start with the description of some basic species, and then introduce the three fundamental constructions of disjoint union, product, and substitution.

**2.4.1. Basic Species.** The most elementary species are

- the *empty species*  $\mathbf{0}$  defined by  $\mathbf{0}[U] := \emptyset$  for all  $U$ ;
- the *neutral species*  $\mathbf{1}$  defined by  $\mathbf{1}[U] := \{\emptyset\}$  if  $|U| = 0$  and  $\mathbf{1}[U] := \emptyset$  otherwise;
- the *atomic species*  $X$  defined by  $X[U] := \{U\}$  if  $|U| = 1$  and  $X[U] := \emptyset$  otherwise;
- the *species of sets* SET defined by  $\text{SET}[U] := \{U\}$ ;
- the species SEQ of sequences defined by  $\text{SEQ}[U] := \{(U, <) \mid U \text{ is linearly ordered by } <\}$ ;

- the species  $\text{CYC}$  of oriented cycles (or cyclic permutations) defined by

$$\text{CYC}[U] := \{(U, \pi) \mid \pi \text{ is a cyclic permutation of } U\}.$$

In each case, the definition of the transport  $\mathcal{A}[\sigma]$  of  $\mathcal{A}$ -structures is obvious.

If  $\mathfrak{A}$  is a species, then  $\mathfrak{A}^{[n]}$  denotes the species defined by  $\mathfrak{A}^{[n]}[U] := \mathfrak{A}[U]$  if  $|U| = n$  and  $\mathfrak{A}[U] := \emptyset$  otherwise. The transport  $\mathfrak{A}^{[n]}[\sigma]$  for bijections  $\sigma : U \rightarrow V$  is defined by  $\mathfrak{A}^{[n]}[\sigma] := \mathfrak{A}[\sigma]$  if  $|U| = n$  and  $\mathfrak{A}^{[n]}[\sigma] := \text{Id}_{\mathfrak{A}[U]}$  otherwise. In particular,  $X$  is  $\text{SET}^{[1]}$ .

**2.4.2. Constructions.** We now describe the three fundamental constructions that we use to construct species from other species.

*Disjoint union.* For two species  $\mathcal{A}$  and  $\mathcal{B}$ , the species  $\mathcal{A} + \mathcal{B}$ , called the *disjoint union* (or sum) of  $\mathcal{A}$  and  $\mathcal{B}$ , is defined by  $(\mathcal{A} + \mathcal{B})[U] = (\{1\} \times \mathcal{A}[U]) \cup (\{2\} \times \mathcal{B}[U])$ . The transport along a bijection  $\sigma : U \rightarrow V$  is carried out by setting, for any  $(\mathcal{A} + \mathcal{B})$ -structure  $C$  on  $U$ ,

$$(\mathcal{A} + \mathcal{B})[\sigma](C) := \begin{cases} \mathcal{A}[\sigma](C) & \text{if } i = 1 \\ \mathcal{B}[\sigma](C) & \text{if } i = 2. \end{cases}$$

Note that when  $\mathcal{A}, \mathcal{A}', \mathcal{B}, \mathcal{B}'$  are species such that  $\mathcal{A} = \mathcal{A}'$  (i.e.,  $\mathcal{A}$  and  $\mathcal{A}'$  are isomorphic) and  $\mathcal{B} = \mathcal{B}'$ , then  $\mathcal{A} + \mathcal{B} = \mathcal{A}' + \mathcal{B}'$ .

The disjoint union of a countably infinite sequence  $\mathcal{A}_1, \mathcal{A}_2, \dots$  of species is defined analogously, setting  $(\sum_{i \geq 1} \mathcal{A}_i)[U] := \bigcup_{i \geq 1} (i, \mathcal{A}_i[U])$  (this species is well-defined provided the set on the right is finite for each finite set  $U$ ).

*Product.* The *cartesian* (often also called *partitional* or *dinary*) *product*  $\mathcal{A} \cdot \mathcal{B}$  of two species  $\mathcal{A}$  and  $\mathcal{B}$  is the species defined as follows. An  $(\mathcal{A} \cdot \mathcal{B})$ -structure on  $U$  is an ordered pair  $C = (A, B)$  where  $A$  is an  $\mathcal{A}$ -structure on a set  $U_1 \subseteq U$ ,  $B$  is a  $\mathcal{B}$ -structure on a set  $U_2 \subseteq U$ ,  $U_1 \cap U_2 = \emptyset$ , and  $U_1 \cup U_2 = U$ . The transport along a bijection  $\sigma : U \rightarrow V$  is carried out by setting, for each  $(\mathcal{A} \cdot \mathcal{B})$ -structure  $C = (A, B)$  on  $U$ ,

$$(\mathcal{A} \cdot \mathcal{B})[\sigma](C) := (\mathcal{A}[\sigma_1](A), \mathcal{B}[\sigma_2](B))$$

where  $\sigma_i$  is the restriction  $\sigma|_{U_i}$  of  $\sigma$  on  $U_i$  for  $i \in \{1, 2\}$ .

Again we note that when  $\mathcal{A}, \mathcal{A}', \mathcal{B}, \mathcal{B}'$  are species such that  $\mathcal{A} = \mathcal{A}'$  and  $\mathcal{B} = \mathcal{B}'$ , then  $\mathcal{A} \cdot \mathcal{B} = \mathcal{A}' \cdot \mathcal{B}'$ .

*Substitution.* Given two species  $\mathcal{A}$  and  $\mathcal{B}$  such that  $\mathcal{B}[\emptyset] = \emptyset$ , the *(partitional) composite* of  $\mathcal{B}$  in  $\mathcal{A}$ , denoted by  $\mathcal{A} \circ \mathcal{B}$ , is the species  $\mathcal{C}$  obtained as follows. A  $\mathcal{C}$ -structure on  $U$  is a triple  $(\pi, A, B)$  where

- $\pi$  is a partition of  $U$ ;
- $A$  is an  $\mathcal{A}$ -structure on the set of classes of  $\pi$ ;
- $B := (B_p)_{p \in \pi}$  where for each class  $p$  of  $\pi$ ,  $B_p$  is a  $\mathcal{B}$ -structure on  $p$ .

The transport along a bijection  $\sigma : U \rightarrow V$  is carried out by setting, for any  $\mathcal{C}$ -structure  $C = (\pi, A, (B_p)_{p \in \pi})$  on  $U$

$$\mathcal{C}[\sigma](C) := (\bar{\pi}, \bar{A}, (\bar{B}_{\bar{p}})_{\bar{p} \in \bar{\pi}})$$

where

- $\bar{\pi}$  is the partition of  $V$  obtained by transport of  $\pi$  along  $\sigma$ ;
- the structure  $\bar{A}$  is obtained from the structure  $A$  by  $\mathcal{A}$ -transport along the bijection  $\bar{\sigma} : \pi \rightarrow \bar{\pi}$  induced by  $\sigma$  on  $\pi$ ;
- for each  $\bar{p} = \sigma(p) \in \bar{\pi}$ , the structure  $\bar{B}_{\bar{p}}$  is obtained from the structure  $B_p$  by  $\mathcal{B}$ -transport along  $\sigma|_p$ .

We call the  $\mathcal{A}$ -structure  $A$  the *core* of  $(\pi, A, B)$ , and  $B_p$  in  $B = (B_p)_{p \in \pi}$  a *component* of  $(\pi, A, B)$ . Also for the substitution construction, we have that when  $\mathcal{A}, \mathcal{A}', \mathcal{B}, \mathcal{B}'$  are species such that  $\mathcal{A} = \mathcal{A}'$  and  $\mathcal{B} = \mathcal{B}'$ , then  $\mathcal{A} \circ \mathcal{B} = \mathcal{A}' \circ \mathcal{B}'$ .

Together with the basic species, these constructions provide an extremely powerful device for the description of combinatorial families. The substitution construction is particularly interesting,

as it allows us to express other combinatorial constructions, such as the formation of sequences, sets, and cycles of structures of a species  $\mathcal{A}$ , which are specified as  $\text{SEQ} \circ \mathcal{A}$ ,  $\text{SET} \circ \mathcal{A}$ , and  $\text{CYC} \circ \mathcal{A}$ , respectively.

As usual, and to avoid clumsy expressions with many brackets, we make the convention that in expressions involving several of the symbols  $+$ ,  $\cdot$ ,  $\circ$ , the symbol  $\circ$  binds stronger than the symbol  $\cdot$ , and the symbol  $\cdot$  binds stronger than the symbol  $+$ .

There are explicit rules to compute the cycle index sum for the basic species and for each construction.

**Definition 2.4.** Given two power series  $f := f(x_1, x_2, x_3, \dots)$  and  $g := g(x_1, x_2, x_3, \dots)$  such that  $g(0, 0, \dots) = 0$ , the *plethystic composition* of  $f$  and  $g$ , as defined in [2], is the power series

$$f \circ g := f(g_1, g_2, g_3, \dots), \quad \text{with } g_k = g(x_k, x_{2k}, x_{3k}, \dots). \quad (6)$$

In other words,  $f \circ g$  is the series  $f$  where each variable  $x_k$  is replaced by  $g(x_k, x_{2k}, x_{3k}, \dots)$ .

**Proposition 2** (Pólya, Bergeron et al. [2]). *For each of the basic species  $\{\mathbf{0}, \mathbf{1}, X, \text{SEQ}^{[k]}, \text{SET}^{[k]}, \text{CYC}^{[k]}, \text{SEQ}, \text{SET}, \text{CYC}\}$ , the associated cycle index sum has an explicit expression, as given in Figure 1. For each of the fundamental constructions  $\wedge \in \{+, \cdot, \circ\}$ , there is an explicit rule to compute the cycle index sum of the species  $\mathcal{A} \wedge \mathcal{B}$ , as given in Figure 1.*

*Remark 2.1.* The ordinary generating series of a species can be obtained from the cycle index sums for the species. For the sum and product constructions, we obtain

$$\mathcal{C} = \mathcal{A} + \mathcal{B} \quad \Rightarrow \quad \tilde{\mathcal{C}}(x) = \tilde{\mathcal{A}}(x) + \tilde{\mathcal{B}}(x), \quad (7)$$

$$\mathcal{C} = \mathcal{A} \cdot \mathcal{B} \quad \Rightarrow \quad \tilde{\mathcal{C}}(x) = \tilde{\mathcal{A}}(x) \cdot \tilde{\mathcal{B}}(x). \quad (8)$$

For the substitution construction, the computation rule is

$$\mathcal{C} = \mathcal{A} \circ \mathcal{B} \quad \Rightarrow \quad \tilde{\mathcal{C}}(x) = Z_{\mathcal{A}}(\tilde{\mathcal{B}}(x), \tilde{\mathcal{B}}(x^2), \tilde{\mathcal{B}}(x^3), \dots). \quad (9)$$

**2.5. Recursive Specifications.** It is possible to define species via recursive specifications that involve the fundamental constructions introduced above.

**Definition 2.5.** A (standard) *recursive specification* with variables  $x_1, \dots, x_m$  over the species  $\mathcal{A}_1, \dots, \mathcal{A}_\ell$  is a system  $\Psi$  of equations  $x_1 = e_1, \dots, x_m = e_m$  where each  $e_i$  is

- of the form  $a + b$  or  $a \cdot b$  with  $a, b \in \{x_1, \dots, x_m, \mathcal{A}_1, \dots, \mathcal{A}_\ell\}$ , or
- of the form  $a \circ b$  with  $a \in \{\mathcal{A}_1, \dots, \mathcal{A}_\ell\}$  and  $b \in \{x_1, \dots, x_m, \mathcal{A}_1, \dots, \mathcal{A}_\ell\}$ .

Under a certain condition, a recursive specification  $\Psi$  with variables  $x_1, \dots, x_m$  defines new species  $\mathcal{X}_1, \dots, \mathcal{X}_m$  as follows. We first define for each  $i \geq 0$  a vector of species  $\mathcal{X}_1^{(i)}, \dots, \mathcal{X}_m^{(i)}$ . For  $i = 1$ , we set  $(\mathcal{X}_1^{(1)}, \dots, \mathcal{X}_m^{(1)}) = (\mathbf{0}, \dots, \mathbf{0})$ . For  $i > 1$  and  $1 \leq j \leq m$ , the species  $\mathcal{X}_j^{(i)}$  is defined from  $e_j$  by substituting for all  $1 \leq k \leq m$  the occurrences of  $x_k$  in  $e_j$  by  $\mathcal{X}_k^{(i-1)}$ . The resulting expression only contains species and symbols  $+$ ,  $\cdot$ , and  $\circ$ , and hence evaluates to a species, unless  $e_j = a \circ b$  and  $b$  is substituted by a species that contains structures of size 0. If this case never occurs, i.e., if  $\mathcal{B}[\emptyset] = \emptyset$  whenever a species  $\mathcal{B}$  is substituted for  $b$  in an expression  $a \circ b$ , then we call  $\Psi$  *admissible*.

Note that the sequence  $\mathcal{X}_k^{(1)}, \mathcal{X}_k^{(2)}, \dots$  is *monotone*, that is,  $\mathcal{X}_k^{(i)} \subseteq \mathcal{X}_k^{(i+1)}$  for all  $1 \leq k \leq m$  and all  $i \geq 1$ ; this follows from the following basic fact (we omit the proof which is straightforward).

**Proposition 3.** *If  $\mathcal{A}, \mathcal{A}', \mathcal{B}, \mathcal{B}'$  are species, and  $\mathcal{A} \subseteq \mathcal{A}'$ ,  $\mathcal{B} \subseteq \mathcal{B}'$ , then  $\mathcal{A} + \mathcal{B} \subseteq \mathcal{A}' + \mathcal{B}'$ ,  $\mathcal{A} \cdot \mathcal{B} \subseteq \mathcal{A}' \cdot \mathcal{B}'$ , and  $\mathcal{A} \circ \mathcal{B} \subseteq \mathcal{A}' \circ \mathcal{B}'$ .*

Note that due to this proposition, it is easy to decide for a given recursive specification whether or not it is admissible (given also the information which of the species  $\mathcal{A}_1, \dots, \mathcal{A}_\ell$  contains structures of size 0).

**Definition 2.6.** Let  $\Psi$  be an admissible recursive specification with variables  $x_1, \dots, x_m$  over the species  $\mathcal{A}_1, \dots, \mathcal{A}_\ell$ , and let the species  $\mathcal{X}_j^{(i)}$  be as described above. If for each  $n$ , the set  $\bigcup_{i \geq 1} \mathcal{X}^{(i)}[n]$  is finite, then the species  $\mathcal{X}_1, \dots, \mathcal{X}_m$  specified by  $\Psi$  are defined as follows. We set  $\mathcal{X}_j[n] = \bigcup_{i \geq 1} \mathcal{X}_j^{(i)}[n]$ . By monotonicity, for each  $n$  we have that  $\mathcal{X}_j[n]$  equals  $\mathcal{X}_j^{(k)}[n]$  for some  $k$ , and so we can define the transport of  $\mathcal{X}_j$  by  $\mathcal{X}_j[\sigma] = \mathcal{X}_j^{(k)}[\sigma]$ .

**Definition 2.7.** Let  $\mathfrak{A}$  be a class of species. The class of species that is *decomposable over*  $\mathfrak{A}$  is the smallest class of species  $\mathfrak{B}$  that contains  $\mathfrak{A}$ , and that contains all species that can be defined by recursive specifications over species from  $\mathfrak{B}$ .

**2.6. Decomposition of symmetries.** In this subsection we provide a proof of Proposition 2. The proof relies on a precise description of the nature of the automorphisms for the basic species and for the species obtained from one of the constructions  $+$ ,  $\cdot$ , or  $\circ$ . Even though proofs and details can already be found in [2], we give our own presentation here since we build on this later on, in particular to define random generation rules (Section 5).

**2.6.1. Automorphisms of some basic species.** By convention, the neutral species  $\mathbf{1}$  has the cycle-index sum 1 (the structure of size 0 is assumed to be fixed by the “empty” automorphism, of weight 1).

The cycle index sum of  $X$  is  $s_1$ . The only  $X$ -structures are over  $U$  with  $|U| = 1$ , and all automorphisms of such structures consist of a single cycle that has weight  $s_1$ .

For the species SEQ, the only automorphisms are the identity, and the identity has weight  $s_1^k/k!$ . As there are  $k!$  sequences of length  $k$ , the cycle index sum of  $\text{SEQ}^{[k]}$  is  $s_1^k$ . Since  $\text{SEQ} = \sum_{k \geq 0} \text{SEQ}^{[k]}$ , the cycle index sum for SEQ is

$$Z_{\text{SEQ}} = \sum_{k \geq 0} s_1^k = \frac{1}{1 - s_1}. \quad (10)$$

For the species CYC, the automorphisms are exactly the ‘shifts’: for a cycle  $(v_1, \dots, v_k)$ , the *shift* of this cycle by  $m \in [0..k-1]$  maps  $v_i$  to  $v_{i+m}$  where the indices are modulo  $k$ . The automorphisms consist of  $k/r$  cycles of length  $r$ , where  $r$  is the order of  $m$  in  $\mathbb{Z}/(k\mathbb{Z})$ . For each divisor  $r$  of  $k$ , there are  $\phi(r)$  elements of order  $r$  in  $\mathbb{Z}/(k\mathbb{Z})$ , where  $\phi(\cdot)$  is the Euler totient function. Hence, for each cycle of length  $k$ , the sum of the weight-monomials over all the cycles of size  $r$  is  $\phi(r)s_r^{k/r}/k!$ , and the sum of the weight-monomials over all the symmetries is  $\sum_{r|k} \phi(r)s_r^{k/r}/k!$ . As there are  $(k-1)!$  cycles of length  $k$ , the cycle index sum of  $\text{CYC}^{[k]}$  is  $1/k \cdot \sum_{r|k} \phi(r)s_r^{k/r}$ . For the species  $\text{CYC} = \sum_{k \geq 1} \text{CYC}^{[k]}$  of all cycles, the sum of the weight-monomials over all symmetries of size  $r$ , which we denote by  $Z_{\text{CYC},r}$ , is thus

$$Z_{\text{CYC},r} = \sum_{k,r|k} \frac{\phi(r)}{k} s_r^{k/r} = \frac{\phi(r)}{r} \sum_{m \geq 1} \frac{1}{m} s_r^m = \frac{\phi(r)}{r} \log \left( \frac{1}{1 - s_r} \right). \quad (11)$$

Therefore, summing  $Z_{\text{CYC},r}$  over  $r \geq 1$ , one obtains the expression of  $Z_{\text{CYC}}$  given in Figure 1.

For the species SET, the automorphisms are all permutations. Hence, the cycle index sum for SET is simply the exponential generating series for all permutations, where each cycle of length  $i$  is marked by a variable  $s_i$ . In other words, when  $f(x)$  denotes the exponential generating series for permutations, and  $f(x; s_1, s_2, s_3, \dots)$  denotes the same generating series where each variable  $s_i$  marks the number of cycles of length  $i$ , then the rules for computing exponential generating series yield

$$f(x; s_1, s_2, s_3, \dots) = \exp \left( \sum_{i \geq 1} \frac{1}{i!} x^i (i-1)! s_i \right) = \exp \left( \sum_{i \geq 1} \frac{1}{i} x^i s_i \right). \quad (12)$$

Hence,  $f(1; s_1, s_2, s_3, \dots)$  is the cycle index sum for SET, and the cycle index sum for  $\text{SET}^{[k]}$  is  $[x^k]f(x; s_1, s_2, s_3, \dots)$ , which corresponds to a restriction to permutations of size  $k$ . Notice that  $Z_{\text{SET}^{[k]}}$  is always a polynomial; for instance, we have

$$Z_{\text{SET}^{[2]}} = \frac{1}{2}(s_1^2 + s_2), \quad Z_{\text{SET}^{[3]}} = \frac{1}{6}(s_1^3 + 3s_2s_1 + 2s_3), \quad Z_{\text{SET}^{[4]}} = \frac{1}{24}(s_1^4 + 6s_1^2s_2 + 8s_1s_3 + 3s_2^2 + 6s_4).$$

2.6.2. *Automorphisms related to the constructions.* We now describe precisely how a symmetry of a species  $\mathcal{C} = \mathcal{A} \wedge \mathcal{B}$ , with  $\wedge \in \{+, \cdot, \circ\}$ , is assembled from symmetries on the species  $\mathcal{A}$  and  $\mathcal{B}$ .

*Disjoint union.* Clearly, each symmetry of  $\mathcal{C} = \mathcal{A} + \mathcal{B}$  is either a symmetry of  $\mathcal{A}$  or of  $\mathcal{B}$  depending on whether the underlying structure is in  $\mathcal{A}$  or in  $\mathcal{B}$ . Consequently,  $\text{Sym}(\mathcal{A} + \mathcal{B})$  is the disjoint union of  $\text{Sym}(\mathcal{A})$  and  $\text{Sym}(\mathcal{B})$ . This directly yields the formula  $Z_{\mathcal{A}+\mathcal{B}} = Z_{\mathcal{A}} + Z_{\mathcal{B}}$ .

*Product.* Consider a product species  $\mathcal{C} = \mathcal{A} \cdot \mathcal{B}$ . Then there is the following bijective correspondence between symmetries of  $\mathcal{C}$  and ordered pairs of a symmetry of  $\mathcal{A}$  and a symmetry of  $\mathcal{B}$ : when  $((A, B), \sigma)$  is a symmetry of  $\mathcal{C}$  where  $U$  are the atoms of  $\mathcal{A}$  and  $V$  are the atoms of  $\mathcal{B}$ , then  $((A, B), \sigma)$  is in correspondence with the symmetry  $(A, \sigma|_U)$  of  $\mathcal{A}$  and the symmetry  $(B, \sigma|_V)$  of  $\mathcal{B}$  (where  $\sigma|_U$  and  $\sigma|_V$  denote the restriction of  $\sigma$  to  $U$  and  $V$ , respectively).

Hence,  $Z_{\mathcal{C}} = Z_{\mathcal{A}} \cdot Z_{\mathcal{B}}$ . (Indeed, for the species  $\mathcal{C}$  the cycle index sum acts like an exponential generating series for the species  $\text{Sym}(\mathcal{C})$  of symmetries when taking the refined weights  $s_1^{c_1} \cdots s_n^{c_n}/n!$  instead of  $x^n/n!$ .)

*Substitution.* For two species  $\mathcal{A}$  and  $\mathcal{B}$  with  $\mathcal{B}[\emptyset] = \emptyset$ , let  $\mathcal{C} = \mathcal{A} \circ \mathcal{B}$ . To understand the automorphisms of  $\mathcal{C}$ -structures, consider a  $\mathcal{C}$ -structure  $C = (\pi, A, (B_p)_{p \in \pi})$  over  $[1..n]$ . Let  $\sigma$  be an automorphism of  $C$ . It is clear that if two atoms  $v_1$  and  $v_2$  of  $C$  are in the same component, then  $\sigma(v_1)$  and  $\sigma(v_2)$  have to be on the same component as well, by definition of the transport for  $\mathcal{C}$ ; moreover,  $\sigma$  induces an automorphism  $\tau$  of  $A$ .

Consider an atom  $v$  from a component  $B_p$  of  $C$ , and let  $k$  be the length of the cycle of  $\tau$  containing  $v$ . Note that  $\sigma^k$  maps  $B_p$  to itself, so  $\sigma^k$  induces an automorphism on  $B_p$ , the resulting symmetry being denoted by  $(B_p, \sigma_p)$ . Consider the cycle  $c = (p_1, \dots, p_k)$  in  $\tau$  where  $p_1 = p$ . Observe that the symmetries  $(B_{p_i}, \sigma_{p_i})$ , for  $i \in [k]$ , can be seen as  $k$  copies of the same symmetry of  $\mathcal{B}$ , which we denote by  $(B_c, \sigma_c)$ . For each cycle  $d = (w_1, \dots, w_\ell)$  of  $\sigma_c$ , let  $(d_1, \dots, d_k)$  be the copies of  $d$  at  $(p_1, \dots, p_k)$ , respectively. Then one can merge  $d_1, \dots, d_k$  into a unique cycle of length  $\ell \cdot k$  using a specific operation which we call *composition of cycles*.

**Definition 2.8.** Let  $d = (v_1, \dots, v_\ell)$  be a cycle of atoms from  $[1..n]$ , with  $v_1$  the atom of  $d$  having the smallest label. Let  $d_1, \dots, d_k$  be a sequence of  $k$  copies of the cycle. Then the *composed cycle* of  $d_1, \dots, d_k$  is the cycle of atoms of length  $\ell k$  such that, for  $1 \leq i < k$  and  $1 \leq j \leq \ell$ , the successor of the atom  $v_j$  in  $d_i$  is the atom  $v_j$  in  $d_{i+1}$ ; and for  $1 \leq j \leq \ell$ , the successor of the atom  $v_j$  in  $d_k$  is the atom  $v_{(j+1) \bmod \ell}$  in  $d_1$ .

This definition correctly reflects how each cycle of  $\sigma$  is assembled from copies of cycles that are on isomorphic components. Indeed, walking  $k$  steps forward on the composed cycle corresponds to walking one step forward on one fixed cycle, which corresponds to the fact that the induced automorphism on each component  $B_{v_i}$  is the effect of  $\sigma$  iterated  $k$  times.

In each symmetry  $(A, \sigma)$  where  $A$  is an  $\mathcal{A}$ -structure over  $[1..n]$ , the automorphism  $\sigma$  induces a partition of  $[1..n]$  corresponding to the cycles of  $\sigma$ . We say that  $\sigma$  has *type*  $t = (c_1, \dots, c_n)$  when  $c_i$  is the number of cycles of length  $i$  in  $\sigma$ . Note that  $n = \sum_{i=1}^n i c_i$  (and such integer sequences  $(c_1, \dots, c_n)$  will also be called *partition sequences (of order  $n$ )*).

To compute the cycle index sum  $Z_{\mathcal{A} \circ \mathcal{B}}$ , we first choose a type  $t = (c_1, \dots, c_n)$  of a permutation of  $[1..n]$ . The number  $a(t)$  of symmetries  $(A, \sigma)$  of the species  $\mathcal{A}$  where  $\sigma$  is of type  $t$  equals  $n! \cdot [s_1^{c_1} s_2^{c_2} \cdots s_n^{c_n}](Z_{\mathcal{A}})$ . To obtain a symmetry  $(C, \rho)$  of  $\mathcal{A} \circ \mathcal{B}$  with  $C = (\pi, A, B)$  where  $\rho$  induces on  $A$  a permutation  $\sigma$  of type  $t$ , we choose a symmetry of  $\mathcal{A}$  of type  $t$ , and then choose  $c_i$  symmetries of  $\mathcal{B}$  for each  $i$ . The sum  $Z_{\mathcal{A} \circ \mathcal{B}}^{(t)}$  of the weight-monomials for all those symmetries is therefore

$$Z_{\mathcal{A} \circ \mathcal{B}}^{(t)} = \frac{a(t)}{n!} b_1^{c_1} b_2^{c_2} \cdots b_n^{c_n}, \quad \text{where } b_i := Z_{\mathcal{B}}(s_i, s_{2i}, s_{3i}, \dots). \quad (13)$$

Summing over all possible types  $t$  of permutations, one obtains

$$Z_{\mathcal{A} \circ \mathcal{B}} = Z_{\mathcal{A}}(b_1, b_2, b_3, \dots), \quad \text{where } b_i := Z_{\mathcal{B}}(s_i, s_{2i}, s_{3i}, \dots). \quad (14)$$

### 3. CYCLE-POINTED SPECIES

In this section we introduce cycle-pointed species, and our unbiased pointing operator.

**3.1. Cycle-Pointed Species.** Let  $\mathcal{A}$  be a species. Then the *cycle-pointed species* of  $\mathcal{A}$ , denoted by  $\mathcal{A}^\circ$ , is defined as follows. For a finite set  $U$ , the set  $\mathcal{A}^\circ[U]$  is defined to be the set of all pairs  $P = (A, c)$  where  $A \in \mathcal{A}[U]$  and  $c = (v_1, \dots, v_\ell)$  is a cycle of atoms of  $A$  such that there exists *at least one* automorphism of  $A$  having  $c$  as one of its cycles (i.e.,  $(v_1, \dots, v_\ell)$  is mapped to  $(v_2, \dots, v_\ell, v_1)$ ). The cycle  $c$  is called the *marked cycle* (or *pointed cycle*) of  $P$ , and  $A$  is called the *underlying structure* of  $P$ . Note that cycle-pointed species are in particular species. Thus, the theory from the previous section also applies to cycle-pointed species.

An automorphism  $\sigma$  of  $A$  having  $c$  as one of its cycles is called a *c-automorphism* of  $P$ , and the other cycles of  $\sigma$  are called *unmarked*. By definition, two cycle-pointed structures  $P$  and  $P'$  are isomorphic if there exists an isomorphism from the underlying structure of  $P$  to the underlying structure of  $P'$  that maps the marked cycle of  $P$  to the marked cycle of  $P'$  (i.e., each atom of the marked cycle of  $P$  is mapped to an atom of the marked cycle of  $P'$ , and the cyclic order is preserved).

**Definition 3.1.** A species  $\mathcal{P}$  is called *cycle-pointed* if  $\mathcal{P} \subseteq \mathcal{A}^\circ$ , for some species  $\mathcal{A}$ . For  $\ell \geq 1$ ,  $\mathcal{P}_{(\ell)}$  is the species that consists of those structures from  $\mathcal{P}$  whose marked cycle has length  $\ell$ .

We define  $\mathcal{A}^\circledast$  to be the subspecies of  $\mathcal{A}^\circ$  where in all structures the marked cycle has length greater than 1, i.e.,  $\mathcal{A}^\circledast = \sum_{\ell \geq 2} (\mathcal{A}^\circ)_{(\ell)}$ . (All cycle-pointed species that will be considered in the applications – except for maps – are either of the form  $\mathcal{A}^\circ$  or  $\mathcal{A}^\circledast$ .)

**Definition 3.2.** Let  $\mathcal{P}$  be a cycle-pointed species. We define  $\mathcal{P}^\dagger$  to be the species obtained from  $\mathcal{P}$  by removing the marked cycle from all  $\mathcal{P}$ -structures; that is, the set of  $\mathcal{P}^\dagger$ -structures on  $[1..n]$  is  $\{A \mid (A, c) \in \mathcal{P}_n \text{ for some cycle } c \text{ of atoms in } A\}$ .

Clearly, for any species  $\mathcal{A}$  we have that  $(\mathcal{A}^\circ)^\dagger = \mathcal{A}$ .

**3.2. Cycle Index Sums.** In order to develop Pólya theory for cycle-pointed species, we introduce the terminology of *c-symmetry* and *rooted c-symmetry*. Given a cycle-pointed species  $\mathcal{P}$ , a *c-symmetry* on  $\mathcal{P}$  is a pair  $(P, \sigma)$  where  $P = (A, c)$  is a cycle-pointed structure in  $\mathcal{P}$  and  $\sigma$  is a *c-automorphism* of  $A$ . A *rooted c-symmetry* is a triple  $(P, \sigma, v)$ , where  $(P, \sigma)$  is a *c-symmetry*, and  $v$  is one of the atoms of the marked cycle of  $P$ ; this atom  $v$  is called the *root* of the rooted *c-symmetry*. The species of rooted *c-symmetries* of  $\mathcal{P}$  is denoted by  $\text{RSym}(\mathcal{P})$ .

The *weight-monomial* of a rooted *c-symmetry* of size  $n$  is defined as

$$w_{(P, \sigma, v)} := \frac{1}{n!} t_\ell \prod_{i=1}^n s_i^{n_i(\sigma)}, \quad (15)$$

where  $t_\ell$  and the  $s_i$ 's are formal variables,  $\ell$  is the length of the marked cycle, and for  $i \in [n]$ ,  $n_i(\sigma)$  is the number of unmarked cycles of  $\sigma$  of length  $i$ , i.e.,  $n_i(\sigma) = c_i(\sigma)$  if  $i \neq \ell$  and  $n_\ell(\sigma) = c_\ell(\sigma) - 1$ . For simplicity, in the following we will write  $n_i$  for  $n_i(\sigma)$  if the corresponding automorphism is clear from the context. We define the *pointed cycle index sum*  $\bar{Z}_{\mathcal{P}}(s_1, t_1; s_2, t_2; \dots)$  of  $\mathcal{P}$ , denoted by  $\bar{Z}_{\mathcal{P}}$ , as the sum of the weight-monomials over all rooted *c-symmetries* of  $\mathcal{P}$ ,

$$\bar{Z}_{\mathcal{P}}(s_1, t_1; s_2, t_2; \dots) := \sum_{(P, \sigma, v) \in \text{RSym}(\mathcal{P})} w_{(P, \sigma, v)}. \quad (16)$$

The following lemma is the counterpart of Lemma 1 for cycle-pointed species; it shows that pointed cycle index sums refine ordinary generating functions. Recall that for a species  $\mathcal{P}$ , the set  $\tilde{\mathcal{P}}$  denotes the unlabeled  $\mathcal{P}$ -structures, see Section 2.2.

**Lemma 4.** *Let  $\mathcal{P}$  be a cycle-pointed species. For  $n \geq 0$ , each unlabeled structure  $\tilde{P} \in \tilde{\mathcal{P}}_n$  gives rise to exactly  $n!$  rooted *c-symmetries*, i.e., there are  $n!$  rooted *c-symmetries*  $(P, \sigma, v)$  such that  $P \in \tilde{P}$ . As a consequence,*

$$\tilde{\mathcal{P}}(x) = \bar{Z}_{\mathcal{P}}(x, x; x^2, x^2; \dots). \quad (17)$$

*Proof.* Lemma 1 implies that  $\tilde{\mathcal{P}}_n$  gives rise to  $n!$  symmetries. Now we establish a bijection between these symmetries and rooted *c-symmetries* from  $\tilde{\mathcal{P}}_n$ . Fix a *c-symmetry*  $(P_0, \sigma_0)$  where  $P_0$  is from  $\tilde{\mathcal{P}}_n$  (by definition of cycle-pointed species, such a *c-symmetry* exists). Now, let  $(P_1, \sigma_1)$  be a symmetry

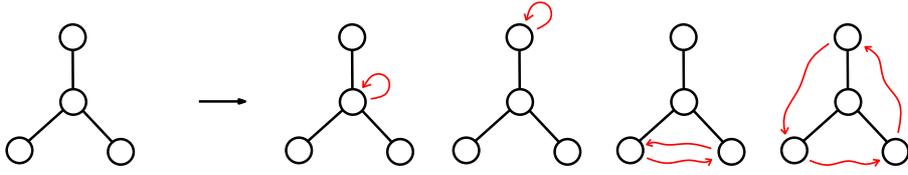


FIGURE 2. An unlabeled nonplane tree of size 4 yields 4 unlabeled cycle-pointed trees.

of  $\mathcal{P}$  where  $P_1$  is also from  $\tilde{P}_n$ . The marked cycle  $(v_1, \dots, v_\ell)$  of  $P_1$  is preserved by  $\sigma_1$ , so that all its elements are *shifted* by the same value  $r \in [1.. \ell]$  modulo  $\ell$ , i.e.,  $\sigma_1$  maps  $v_i$  to  $v_{(i+r) \bmod \ell}$ . Since both  $P_0$  and  $P_1$  are from  $\tilde{P}_n$ , there is an isomorphism  $\sigma$  from  $P_1$  to  $P_0$ . Moreover, the permutation  $\tau := \sigma^{-1}\sigma_0\sigma$  is a  $c$ -symmetry of  $P_1$ . Observe that the permutation  $\tau^{-r+1}\sigma_1$  is an automorphism of  $P_1$  that moves an atom of the marked cycle  $r$  steps forward (because of  $\sigma_1$ ) and then  $r-1$  steps backward (because of  $\tau^{-r+1}$ ). Hence,  $\tau^{-r+1}\sigma_1$  is a  $c$ -symmetry of  $P_1$ . The desired bijection maps  $(P_1, \sigma_1)$  to the rooted  $c$ -symmetry  $(P_1, \tau^{-r+1}\sigma_1, v)$  where  $v$  is the atom of the marked cycle having the  $r$ -th smallest label. This correspondence can be inverted easily, and hence we have found a bijection between the symmetries and the rooted  $c$ -symmetries for structures having  $\tilde{P}_n$  as unlabeled structure.  $\square$

Observe that a rooted  $c$ -symmetry of  $\mathcal{A}^\circ$  is obtained from a symmetry  $(A, \sigma)$  of  $\mathcal{A}$  by choosing an atom  $v$  of  $A$  and marking the cycle of  $\sigma$  containing  $v$ . Therefore, each symmetry  $(A, \sigma)$  of size  $n$  of the species  $\mathcal{A}$  yields  $n$  rooted  $c$ -symmetries on  $\mathcal{A}^\circ$ , and hence

$$\bar{Z}_{\mathcal{A}^\circ}(s_1, t_1; s_2, t_2; \dots)|_{\{t_1=s_1, t_2=s_2, \dots\}} = \frac{d}{dt} Z_{\mathcal{A}}(ts_1, t^2s_2, t^3s_3, \dots)|_{\{t=1\}}. \quad (18)$$

Further, a rooted  $c$ -symmetry of  $\mathcal{A}^\circ$  can equivalently be obtained by marking a cycle of atoms that corresponds to a cycle of  $\sigma$  and choosing an atom of the cycle as the root of the rooted  $c$ -symmetry. Together, these observations yield the equality

$$\bar{Z}_{\mathcal{A}^\circ(\ell)}(s_1, t_1; s_2, t_2; \dots)|_{\{t_1=s_1, t_2=s_2, \dots\}} = \ell t_\ell \frac{\partial}{\partial s_\ell} Z_{\mathcal{A}}(s_1, s_2, \dots) \quad \text{for all } \ell \geq 1. \quad (19)$$

For  $\ell = 1$ , which corresponds to structures of  $\mathcal{A}$  with a unique distinguished vertex (the *root*), we recover the well-known equation relating the cycle index sum of a species and of the associated *rooted* species; see [2, Sec. 1.4.] and [16].

As stated below and illustrated in Figure 2, pointing a cycle of symmetry instead of a single atom (as the classical pointing operator does) yields an unbiased pointing operator in the unlabeled setting.

**Theorem 5** (unbiased pointing). *Let  $\mathcal{A}$  be a species. Then, for  $n \geq 0$ , each unlabeled structure of  $\tilde{\mathcal{A}}_n$  gives rise to exactly  $n$  unlabeled structures in  $\tilde{\mathcal{A}}_n^\circ$ ; that is, for each  $A \in \mathcal{A}[n]$  there are exactly  $n$  non-isomorphic structures  $(A, c)$  in  $\mathcal{A}^\circ[n]$ . Hence, the ordinary generating series of  $\mathcal{A}^\circ$  satisfies*

$$\tilde{\mathcal{A}}^\circ(x) = x \frac{d}{dx} \tilde{\mathcal{A}}(x). \quad (20)$$

*Proof.* Given  $\tilde{A} \in \tilde{\mathcal{A}}_n$ , let  $S$  be the set of unlabeled pointed structures of  $\tilde{\mathcal{A}}^\circ$  whose underlying unpointed structure is  $\tilde{A}$ . The proof of the lemma reduces to proving that  $S$  has cardinality  $n$ . Let  $\text{Sym}(\tilde{A})$  be the set of symmetries for the structure  $\tilde{A}$ , and let  $\text{RSym}(S)$  be the set of rooted  $c$ -symmetries for structures from  $S$ . Lemma 1 shows that  $\tilde{A}$  has  $n!$  symmetries and Lemma 4 shows that each structure of  $\tilde{\mathcal{A}}_n^\circ$  has  $n!$  rooted  $c$ -symmetries. Hence  $|\text{Sym}(\tilde{A})| = n!$  and  $|\text{RSym}(S)| = |S|n!$ . In addition, we have seen that a symmetry  $(A, \sigma)$  of size  $n$  on  $\mathcal{A}$  has  $n$  rooted  $c$ -symmetries on  $\mathcal{A}^\circ$ . Hence,  $|\text{RSym}(S)| = n|\text{Sym}(\tilde{A})|$ . Thus, we obtain  $|S|n! = nn!$ , and therefore  $|S| = n$ .  $\square$

*Remark 3.1.* Theorem 5 is equivalent to a result known as Parker's lemma [7, Section 2.8]. For a subgroup  $G$  of the symmetric group  $\mathfrak{S}_n$ , say that a cycle  $c$  in  $g \in G$  is equivalent to a cycle  $c'$  in  $g' \in G$  if there exists  $h \in G$  that maps the elements of  $c$  to the elements of  $c'$  and preserves the

<p><b>Basic cycle-pointed species:</b> <math>\mathcal{A}^\circ, \mathcal{A}^\circledast</math>,  where <math>\mathcal{A} \in \{\mathbf{0}, \mathbf{1}, X, \text{SEQ}^{[k]}, \text{SET}^{[k]}, \text{CYC}^{[k]}, \text{SEQ}, \text{SET}, \text{CYC}\}</math></p> <p><b>Pointed cycle-index sum:</b></p> $\bar{Z}_{\mathcal{A}^\circ} = \sum_{\ell \geq 1} \ell t_\ell \frac{\partial}{\partial s_\ell} Z_{\mathcal{A}}$ $\bar{Z}_{\mathcal{A}^\circledast} = \sum_{\ell \geq 2} \ell t_\ell \frac{\partial}{\partial s_\ell} Z_{\mathcal{A}}$		
Construction	Notation	Pointed cycle index sum
<b>Pointed union</b>	$\mathcal{R} = \mathcal{P} + \mathcal{Q}$	$Z_{\mathcal{R}} = Z_{\mathcal{P}} + Z_{\mathcal{Q}}$
<b>Pointed product</b>	$\mathcal{R} = \mathcal{P} \star \mathcal{B}$	$Z_{\mathcal{R}} = \bar{Z}_{\mathcal{P}} \cdot Z_{\mathcal{B}}$
<b>Pointed substitution</b>	$\mathcal{R} = \mathcal{P} \odot \mathcal{B}$	$Z_{\mathcal{R}} = \bar{Z}_{\mathcal{P}} \odot Z_{\mathcal{B}}$

FIGURE 3. The rules to calculate the pointed cycle index sums of basic cycle-pointed species and of cycle-pointed species composed by  $+$ ,  $\star$ , and  $\odot$ .

cyclic order, i.e., for each element  $x \in c$ , the successor of  $x$  in  $c$  is mapped by  $h$  to the successor of  $h(x)$  in  $c'$ . Let  $a_k$  be the number of inequivalent cycles of length  $k$ . Then Parker's lemma states that  $\sum_{k=1}^n a_k = n$ .

If this lemma is applied to the automorphism group of a fixed structure  $A$  of size  $n$ , then  $a_k$  is the number of unlabeled cycle-pointed structures arising from  $A$  and such that the marked cycle has length  $k$ . So Parker's lemma states that there are  $n$  unlabeled cycle-pointed structures arising from  $A$ , i.e., it implies Theorem 5 (conversely, each permutation group is the automorphism group of a structure, so Parker's lemma can be deduced from Theorem 5).

*Remark 3.2.* The classical pointing operator, which selects a single atom in a structure, yields an equation similar to (20) for *exponential* generating series, which is useful for *labeled* enumeration. Given a species  $\mathcal{A}$ , let  $\mathcal{A}^\bullet$  be the species of structures from  $\mathcal{A}$  where an atom is distinguished. Then

$$\mathcal{A}^\bullet(x) = x \frac{d}{dx} \mathcal{A}(x). \quad (21)$$

An important contribution of this article is to define a pointing operator  $\mathcal{A} \mapsto \mathcal{A}^\circ$  that yields the same equation, Equation (20), in the *unlabeled* case.

### 3.3. Basic Cycle-pointed Species and Constructions.

**3.3.1. Basic cycle-pointed species.** Species of the form  $\mathcal{A}^\circ$  or  $\mathcal{A}^\circledast$  where  $\mathcal{A}$  is a basic species as introduced in Subsection 2.4 will be called *basic cycle-pointed species*. The derivation rule in Equation (19) allows us to compute the pointed cycle index sum of basic cycle-pointed species, as indicated in Figure 3 (upper part).

**3.3.2. Pointed constructions.** It is clear that the disjoint union of two cycle-pointed species as defined in Section 2.4.2 is again a cycle-pointed species. We now adapt the constructions of product and substitution to obtain a pointed product and a pointed substitution operation that produce cycle-pointed species.

*Pointed product.* Let  $\mathcal{A}, \mathcal{B}$  be species, and let  $\mathcal{P} \subseteq \mathcal{A}^\circ$  be a cycle-pointed species. Then the pointed product  $\mathcal{P} \star \mathcal{B}$  of  $\mathcal{P}$  and  $\mathcal{B}$  is the subspecies of  $(\mathcal{A} \cdot \mathcal{B})^\circ$  of all those structures  $((A, B), c)$  in  $(\mathcal{A} \cdot \mathcal{B})^\circ$  where the pointed cycle  $c$  is from  $\mathcal{A}$ , and  $(A, c) \in \mathcal{P}$ .

*Pointed substitution.* Let  $E = ((\pi, A, (B_p)_{p \in \pi}), c)$  be a structure in  $(\mathcal{A} \circ \mathcal{B})^\circ$ . The study of automorphisms of structures in  $\mathcal{A} \circ \mathcal{B}$  performed in Section 2.6.2 shows that

- $c$  is the cycle composed from cycles  $c_1, \dots, c_k$  on different components  $B_{p_1}, \dots, B_{p_k}$ ;
- the structure  $(A, c' = (p_1, \dots, p_k))$  is cycle-pointed.

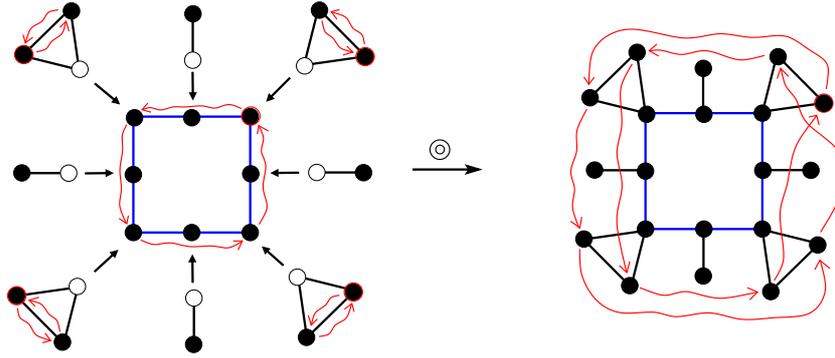


FIGURE 4. A cycle-pointed graph obtained from a substitution.

We call  $(A, c') \in \mathcal{A}^\circ$  the *core-structure* of  $E$ .

We can now define a substitution construction for cycle-pointed species. Let  $\mathcal{P} \subseteq \mathcal{A}^\circ$  and let  $\mathcal{B}$  be a species such that  $\mathcal{B}[\emptyset] = \emptyset$ . Then  $\mathcal{P} \odot \mathcal{B}$  is defined as the subspecies of structures from  $(\mathcal{A} \circ \mathcal{B})^\circ$  whose core-structure is in  $\mathcal{P}$ .

As in the previous section, we make the convention that in expressions involving several of the symbols  $+$ ,  $\cdot$ ,  $\star$ ,  $\circ$ ,  $\odot$ , the symbols  $\circ$ ,  $\odot$  bind stronger than the symbols  $\cdot$ ,  $\star$ , and the symbols  $\cdot$ ,  $\star$  bind stronger than the symbol  $+$ .

In a similar way to the labeled framework [2], our pointing operator behaves well with the three constructions  $+$ ,  $\cdot$ , and  $\circ$ :

**Proposition 6.** *The cycle-pointing operator obeys the following rules:*

$$(\mathcal{A} + \mathcal{B})^\circ = \mathcal{A}^\circ + \mathcal{B}^\circ, \quad (22)$$

$$(\mathcal{A} \cdot \mathcal{B})^\circ = \mathcal{A}^\circ \star \mathcal{B} + \mathcal{B}^\circ \star \mathcal{A}, \quad (23)$$

$$(\mathcal{A} \circ \mathcal{B})^\circ = \mathcal{A}^\circ \odot \mathcal{B}. \quad (24)$$

*Proof.* It is easy to see that  $(\mathcal{A} + \mathcal{B})^\circ$  is isomorphic to  $\mathcal{A}^\circ + \mathcal{B}^\circ$ . For the product, note that the pointed cycle  $c$  of a structures from  $(\mathcal{A} \cdot \mathcal{B})^\circ$  has to be entirely on  $A$  or entirely on  $B$ . The species that contains all structures  $((A, B), c)$  where  $c$  is on  $A$  is isomorphic to  $\mathcal{A}^\circ \star \mathcal{B}$ , and the species that contains all structures  $((A, B), c)$  where  $c$  is on  $B$  is isomorphic to  $\mathcal{B}^\circ \star \mathcal{A}$ . Hence,  $(\mathcal{A} \cdot \mathcal{B})^\circ = \mathcal{A}^\circ \star \mathcal{B} + \mathcal{B}^\circ \star \mathcal{A}$ . For the expression for the substitution operation, we in fact have not only isomorphism, but even equality of species. This is clear from the fact that *all* core-structures of structures from  $(\mathcal{A} \circ \mathcal{B})^\circ$  are from  $\mathcal{A}^\circ$ .  $\square$

As in the unpointed case, there are explicit rules to compute the pointed cycle index sums for each basic species and for each construction. To this end we need the following notion of composition for power series.

**Definition 3.3.** Let  $f$  and  $g$  be two power series of the form  $\bar{f} := f(x_1, y_1; x_2, y_2; \dots)$  and  $g := g(x_1, x_2, \dots)$  such that  $g(0, 0, \dots) = 0$ . Then the *pointed plethystic composition* of  $f$  with  $g$  is the power series

$$\bar{f} \odot g := f(g_1, \bar{h}_1; g_2, \bar{h}_2; \dots), \quad (25)$$

with  $g_k = g(x_k, x_{2k}, x_{3k}, \dots)$  and  $\bar{h}_k = \bar{h}(x_k, y_k; x_{2k}, y_{2k}; \dots)$  for  $\bar{h} := \sum_{\ell \geq 1} \ell t_\ell \frac{\partial}{\partial s_\ell} g$ .

The following proposition is the counterpart of Proposition 2 for cycle-pointed species.

**Proposition 7** (computation rules for pointed cycle index sums). *For each basic species  $\mathcal{A}$ , the pointed cycle index sum of the pointed species  $\mathcal{A}^\circ$  and  $\mathcal{A}^\circ$  is given by the explicit expression given in Figure 3 (upper part) in terms of the cycle index sum of  $\mathcal{A}$ . For each of the fundamental pointed constructions  $+$ ,  $\star$ , and  $\odot$  there is an explicit rule, given in Figure 3 (lower part), to compute the pointed cycle index sum of the resulting species.*

*Proof.* The statement is clear for the cycle-pointed atomic species and the pointed union. Let  $\mathcal{A}, \mathcal{B}$  be species, and let  $\mathcal{P} \subseteq \mathcal{A}^\circ$  be a cycle-pointed species. For the cycle-pointed product  $\mathcal{Q} = \mathcal{P} \star \mathcal{B}$  notice that, similarly as for a partitionial product for species, a rooted  $c$ -symmetry on  $\mathcal{Q}$  decomposes into a rooted  $c$ -symmetry on  $\mathcal{P}$  and a symmetry on  $\mathcal{B}$ , since the automorphism has to act separately on the two component structures. Therefore,  $\text{RSym}(\mathcal{Q})$  can be considered as a partitionial product of  $\text{RSym}(\mathcal{P})$  and  $\text{Sym}(\mathcal{B})$ , which yields  $\bar{Z}_{\mathcal{Q}} = \bar{Z}_{\mathcal{P}} \cdot Z_{\mathcal{B}}$ .

For the substitution construction, the proof is similar. Let  $\mathcal{P} \subseteq \mathcal{A}^\circ$  be a cycle-pointed species and let  $\mathcal{B}$  be a species such that  $\mathcal{B}[\emptyset] = \emptyset$ . Let  $\mathcal{Q} = \mathcal{P} \odot \mathcal{B} \subseteq (\mathcal{A} \circ \mathcal{B})^\circ$ . Consider a rooted  $c$ -symmetry  $(Q, \sigma, v)$ . As we have seen in Section 2.6, the core structure  $A$  is endowed with an induced automorphism  $\sigma'$ . In addition, the automorphism is naturally rooted at the atom  $v' \in A$  where the  $\mathcal{B}^\circ$ -component that contains the root  $v$  is substituted. We denote the cycle of  $\sigma'$  that contains  $v'$  by  $c'$ . Now we have that  $(A, c')$  is cycle-pointed and the automorphism  $\sigma'$  rooted at  $v'$  is a rooted  $c$ -symmetry on  $P = (A, c')$ . In addition, the components substituted at each atom of a cycle  $c = (u_1, \dots, u_k)$  of  $\sigma'$  are isomorphic copies of a same symmetry on  $\mathcal{B}$ . The components that are substituted at the atoms of the marked cycle  $c'$  are naturally rooted at the isomorphic representant of  $v$ . Finally, this decomposition is reversable: one can go back to the original composed symmetry using the composition of cycle operation.

To express these observations in an equation, we define the *type of the rooted  $c$ -symmetry*  $(P, \rho, v)$  with  $P = (A, c')$  to be the sequence  $(\ell; n_1, n_2, \dots, n_k)$  where  $\ell$  is the length of  $c'$ , and  $n_i$  is the number of unmarked cycles of length  $i$  in  $\rho$ . Note that the size of  $P$  is  $\ell + \sum_i i n_i$ . The number  $a(t)$  of rooted  $c$ -symmetries of type  $t = (\ell; n_1, n_2, \dots, n_n)$  on  $\mathcal{P}$  is  $n! [t^\ell s_1^{n_1} \dots s_n^{n_n}] \bar{Z}_{\mathcal{P}}$ . The *core type* of a rooted  $c$ -symmetry on  $\mathcal{P} \odot \mathcal{B}$  is defined as the type of the rooted  $c$ -symmetry induced on the core structure.

Let  $\bar{Z}_{\mathcal{R}}^{(t)}$  be the pointed cycle index sum of  $\mathcal{P}$  restricted to the rooted  $c$ -symmetries with core-type  $t = (\ell; n_1, n_2, \dots, n_n)$ . From the above discussion, we have

$$\bar{Z}_{\mathcal{R}}^{(t)} = \frac{a(t)}{n!} q_\ell b_1^{n_1} b_2^{n_2} \dots b_n^{n_n}, \text{ where } b_i := Z_{\mathcal{B}}(s_1, s_2, \dots), \text{ } q_\ell := \bar{Z}_{\mathcal{B}^\circ}(s_\ell, t_\ell; s_{2\ell}, t_{2\ell}; \dots). \quad (26)$$

Summing over all possible types of rooted  $c$ -symmetries  $t$ , we obtain

$$\bar{Z}_{\mathcal{R}} = \bar{Z}_{\mathcal{P}}(b_1, q_1; b_2, q_2; \dots), \text{ where } b_i := Z_{\mathcal{B}}(s_1, s_2, \dots), \text{ } q_\ell := \bar{Z}_{\mathcal{B}^\circ}(s_\ell, t_\ell; s_{2\ell}, t_{2\ell}; \dots). \quad \square$$

In the following we introduce recursive specifications that involve pointed constructions. Cycle-pointed specifications are like standard recursive specifications (Definition 2.5), but with two sorts of variables (where one is reserved for cycle-pointed species) and where we are allowed to use additionally the pointed constructions.

**Definition 3.4.** A *recursive cycle-pointed specification* with variables  $x_1, \dots, x_m, y_1, \dots, y_{m'}$  over the species  $\mathcal{A}_1, \dots, \mathcal{A}_\ell$  and over cycle-pointed species  $\mathcal{B}_1, \dots, \mathcal{B}_k$  is a system  $\Psi$  of equations  $x_1 = e_1, \dots, x_m = e_m, y_1 = f_1, \dots, y_{m'} = f_{m'}$  where each  $e_i$  is

- of the form  $a + b$  or  $a \cdot b$  with  $a, b \in \{x_1, \dots, x_m, \mathcal{A}_1, \dots, \mathcal{A}_\ell\}$ , or
- of the form  $a \circ b$  with  $a \in \{\mathcal{A}_1, \dots, \mathcal{A}_\ell\}$  and  $b \in \{x_1, \dots, x_m, \mathcal{A}_1, \dots, \mathcal{A}_\ell\}$ ,

and each  $f_i$  is

- of the form  $a + b$  with  $a, b \in \{y_1, \dots, y_{m'}, \mathcal{B}_1, \dots, \mathcal{B}_k\}$ , or
- of the form  $a \star b$  with  $a \in \{y_1, \dots, y_{m'}, \mathcal{B}_1, \dots, \mathcal{B}_k\}$  and  $b \in \{x_1, \dots, x_m, \mathcal{A}_1, \dots, \mathcal{A}_\ell\}$ , or
- of the form  $a \odot b$  with  $a \in \{\mathcal{B}_1, \dots, \mathcal{B}_k\}$  and  $b \in \{x_1, \dots, x_m, \mathcal{A}_1, \dots, \mathcal{A}_\ell\}$ .

To define the species  $\mathcal{X}_1, \dots, \mathcal{X}_m, \mathcal{Y}_1, \dots, \mathcal{Y}_{m'}$  that are given by a recursive cycle-pointed specification  $\Psi$  with variables  $x_1, \dots, x_m, y_1, \dots, y_{m'}$  over the species  $\mathcal{A}_1, \dots, \mathcal{A}_\ell, \mathcal{B}_1, \dots, \mathcal{B}_k$  where  $\mathcal{B}_1, \dots, \mathcal{B}_k$  are pointed, we again (as in Section 2.5) consider sequences of species  $\mathcal{X}_j^{(i)}$  and  $\mathcal{Y}_j^{(i)}$  for  $i \geq 1$ . For  $i = 1$ , we define  $\mathcal{X}_j^{(1)} = \mathbf{0}$  for all  $1 \leq j \leq m$ , and  $\mathcal{Y}_j^{(1)} = \mathbf{0}^\circ$  for all  $1 \leq j \leq m'$ . For  $i > 1$  the species  $\mathcal{X}_j^{(i)}$  and  $\mathcal{Y}_j^{(i)}$  are obtained by evaluating the corresponding expressions for  $x_j$  and  $y_j$ , respectively (as in Section 2.5). We say that  $\Psi$  is *admissible* if in expressions of the form  $a \circ b$  or  $a \odot b$  the species substituted for  $b$  never contain structures of size 0.

Note that also the new pointed constructions are monotone, so in case that for each  $n$  the sets  $\bigcup_{i \geq 1} \mathcal{X}^{(i)}[n]$  and  $\bigcup_{i \geq 1} \mathcal{Y}^{(i)}[n]$  are finite it is straightforward (and analogous to Definition 2.6) to define the species  $\mathcal{X}_1, \dots, \mathcal{X}_m, \mathcal{Y}_1, \dots, \mathcal{Y}_{m'}$  specified by admissible recursive specifications  $\Psi$  over  $\mathcal{A}_1, \dots, \mathcal{A}_\ell, \mathcal{B}_1, \dots, \mathcal{B}_k$ .

**Definition 3.5.** Let  $\mathfrak{A}$  be a class of species. The class of species that is *cycle-pointing decomposable over  $\mathfrak{A}$*  is the smallest class of species  $\mathfrak{B}$  that contains  $\mathfrak{A}$ , contains all species that can be specified by cycle-pointed recursive specifications over pointed and unpointed species from  $\mathfrak{B}$ , and contains all species obtained from species of the form  $\mathcal{A}^\circ$  in  $\mathfrak{B}$  by applying the unpointing operation.

Plenty of examples of species that are decomposable over simple basic species can be found in Section 4.

**Proposition 8.** *If a species  $\mathcal{A}$  is decomposable (in the sense of Definition 2.7), then the pointed species  $\mathcal{A}^\circ$  is cycle-pointing-decomposable.*

*Proof.* Follows directly from Proposition 6. □

*Remark 3.3.* The ordinary generating series inherit simple computation rules from the ones for pointed cycle index sums. As expected, for the sum and product constructions, one gets

$$\mathcal{R} = \mathcal{P} + \mathcal{Q} \quad \Rightarrow \quad \tilde{\mathcal{R}}(x) = \tilde{\mathcal{P}}(x) + \tilde{\mathcal{Q}}(x), \quad (27)$$

$$\mathcal{Q} = \mathcal{P} \star \mathcal{B} \quad \Rightarrow \quad \tilde{\mathcal{Q}}(x) = \tilde{\mathcal{P}}(x) \cdot \tilde{\mathcal{B}}(x). \quad (28)$$

For the substitution construction, the computation rule is:

$$\mathcal{Q} = \mathcal{P} \odot \mathcal{B} \quad \Rightarrow \quad \tilde{\mathcal{Q}}(x) = \bar{Z}_{\mathcal{P}}(\tilde{\mathcal{B}}(x), \tilde{\mathcal{B}}^\circ(x); \tilde{\mathcal{B}}(x^2), \tilde{\mathcal{B}}^\circ(x^2); \dots), \quad (29)$$

where  $\tilde{\mathcal{B}}^\circ(x) = x \frac{d}{dx} \tilde{\mathcal{B}}(x)$ . Hence, to compute the ordinary generating series of a decomposable cycle-pointed species, the only place where the cycle index sum or pointed cycle index sum is needed (as a refinement of ordinary generating series) is for the species that is the first argument of a substitution or pointed substitution construction.

*Remark 3.4.* As an exercise, the reader can check just by standard algebraic manipulations that the computation rules for cycle-index sums are consistent with Proposition 6. For instance, proving  $\bar{Z}_{(\mathcal{A} \circ \mathcal{B})^\circ} = \bar{Z}_{\mathcal{A}^\circ \circ \mathcal{B}}$  is equivalent (by the computation rules) to proving the equality  $\bar{Z}_{(\mathcal{A} \circ \mathcal{B})^\circ} = \bar{Z}_{\mathcal{A}^\circ} \odot \bar{Z}_{\mathcal{B}}$ , which reduces to checking the following identity on power series:

$$\Delta(f \circ g) = (\Delta f) \odot g, \quad (30)$$

where  $\Delta$  is the operator that associates to a power series  $f(x_1, x_2, x_3, \dots)$  the power series

$$\Delta f(x_1, y_1; x_2, y_2; \dots) := \sum_{\ell \geq 1} \ell y_\ell \frac{\partial}{\partial x_\ell} f(x_1, x_2, x_3, \dots).$$

Similarly, to prove  $\bar{Z}_{(\mathcal{A} + \mathcal{B})^\circ} = \bar{Z}_{\mathcal{A}^\circ + \mathcal{B}^\circ}$  and  $\bar{Z}_{(\mathcal{A} \cdot \mathcal{B})^\circ} = \bar{Z}_{\mathcal{A}^\circ \star \mathcal{B} + \mathcal{B}^\circ \star \mathcal{A}}$ , one has to check the identities  $\Delta(f + g) = \Delta f + \Delta g$ , and  $\Delta(f \cdot g) = \Delta f \cdot g + f \cdot \Delta g$ , respectively.

#### 4. APPLICATION TO ENUMERATION

In this section we demonstrate that cycle-pointing provides a new way of counting many classes of combinatorial structures in the unlabeled setting. Typically, species satisfying a “tree-like” decomposition are amenable to our method. This includes of course species of trees, but also species of graphs (provided that the species is closed under taking 2-connected components, and that the sub-species of 2-connected graphs is tractable), and species of planar maps.

The general scheme to enumerate unlabeled structures of a species  $\mathcal{A}$ , i.e., to obtain the coefficients  $|\widetilde{\mathcal{A}}_n|$ , is as follows. First, we observe that the task is equivalent to the task to enumerate unlabeled *cycle-pointed* structures from  $\mathcal{A}^\circ$ , because  $|\widetilde{\mathcal{A}}_n^\circ| = n|\widetilde{\mathcal{A}}_n|$ . Enumeration for  $\mathcal{A}^\circ$  turns out to be easier since the marked cycle usually provides a starting point for a recursive decomposition.

For a cycle-pointed species of trees (and more generally for species satisfying tree-like decompositions), the first step of the decomposition scheme is to distinguish whether the marked cycle has length 1 or greater than 1. The general equation is

$$\mathcal{A}^\circ = X^\circ \star \mathcal{A}' + \mathcal{A}^\circ, \quad (31)$$

where  $\mathcal{A}'$  is the *derived species* of  $\mathcal{A}$ , consisting of structures from  $\mathcal{A}$  where one atom is marked with a special label, say  $*$ , as defined in [2]<sup>1</sup>. Then each of the two species  $\mathcal{A}'$  and  $\mathcal{A}^\circ$  has to be decomposed. For derived structures (the species  $\mathcal{A}'$ ) we follow the classical root decomposition. For symmetric cycle-pointed structures (the species  $\mathcal{A}^\circ$ ) our decomposition strategy is different, and leads us to introduce the notion of *center of symmetry*.

**4.1. Trees.** We first illustrate our decomposition method for trees, which are defined as connected acyclic graphs (i.e., unless mentioned otherwise, trees are *unrooted*), and we start with the formal definition of the *center of symmetry*. Let  $T$  be a symmetric cycle-pointed tree. A path of  $T$  connecting two consecutive atoms of the marked cycle is called a *connecting path* (thus the number of connecting paths is the size of the marked cycle).

**Claim 9** (center of symmetry). *Given a symmetric cycle-pointed tree  $T$ , all connecting paths of  $T$  share the same middle  $v_c$ , called the central point for the marked cycle of  $T$ . The central point  $v_c$  is the middle of an edge  $e$  if these paths have odd length, and is a vertex  $v$  if these paths have even length. In the first (second) case, the edge  $e$  (the vertex  $v$ , resp.) is called the center of symmetry of  $T$ .*

*Proof.* We prove here that all connecting paths share the same middle. Let  $U$  be the subgraph of  $T$  formed by the union of all connecting paths. Observe that  $U$  is connected, so  $U$  is a subtree of  $T$ . In addition,  $U$  is globally fixed by any  $c$ -automorphism of  $T$  (indeed, the property of being on a connecting path is invariant under the action of a  $c$ -automorphism), and it contains the atoms of the marked cycle of  $T$ . Hence  $U$  is the underlying structure of a cycle-pointed tree  $(U, c)$ . Consider the classical center of  $U$ , obtained by pruning the leaves (at each step, all leaves are simultaneously deleted) until the resulting tree is reduced to an edge or a vertex [2]. The central point  $v_c$  of  $U$  is defined as follows: if the center of  $U$  is a vertex  $v$ , then  $v_c := v$ , if the center of  $U$  is an edge  $e$ , then  $v_c$  is the middle of  $e$ . Let  $\sigma$  be a  $c$ -automorphism of  $(U, c)$  and let  $\langle \sigma \rangle$  be the group of automorphisms generated by  $\sigma$ . It is well known that the central point is fixed by any automorphism on the tree, hence  $v_c$  is equidistant from all atoms of the marked cycle, as the group  $\langle \sigma \rangle$  acts transitively on the vertices of the marked cycle. In addition,  $v_c$  is on at least one connecting path of  $T$  (because  $U$  is the union of these connecting paths). Hence,  $v_c$  has to be on all connecting paths, as the group  $\langle \sigma \rangle$  acts transitively on the connecting paths. Thus,  $v_c$  has to be the middle of all connecting paths simultaneously.  $\square$

*Remark 4.1.* Notice that the center of symmetry might not coincide with the classical center of the tree, as shown in Figure 5. However, in the case of *plane trees*, the two notions of center coincide.

**4.1.1. Nonplane trees.** Let  $\mathcal{F}$  be the species of *free trees*, i.e., unrooted nonplane trees (equivalently, acyclic connected graphs), where the vertices are taken as atoms. Let  $\mathcal{F}'$  be the derived species of  $\mathcal{F}$  (also the species of derived nonplane trees). Rooted nonplane trees can be decomposed at the root. Since the root does not count as an atom and since the children of the root node are unordered, we classically have

$$\mathcal{F}' = \text{SET} \circ (X \cdot \mathcal{F}'). \quad (32)$$

In contrast, the decomposition of symmetric cycle-pointed trees does not start at atoms of the marked cycle, but at the center of symmetry, which is either an edge or a vertex (see Figure 5 for an illustration of the decomposition). In order to write down the decomposition, we introduce the

<sup>1</sup>Note that the derived species  $\mathcal{A}'$  is not cycle-pointed. However, it can be identified with  $\mathcal{A}'_{(1)}$ . Indeed, by adding a new label and a pointing loop on the marked atom, one obtains a bijective correspondence between  $\mathcal{A}'_{(1)}$  and  $X^\circ \star \mathcal{A}'$ .

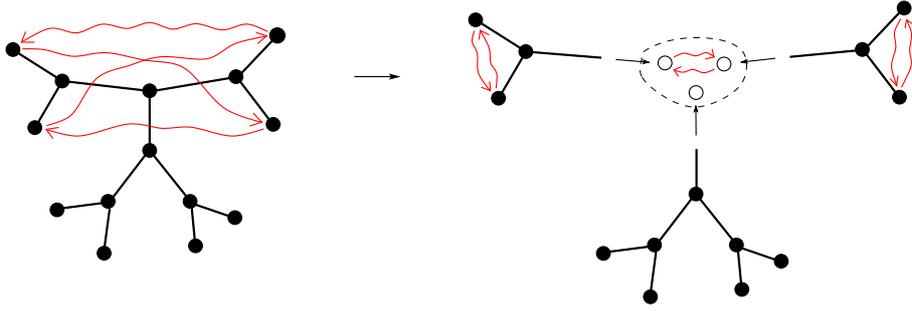


FIGURE 5. Decomposition of a nonplane tree at its center of symmetry (in case the center of symmetry is a vertex).

species  $\mathcal{L}$  consisting of a single one-edge graph. Note that  $\mathcal{L} \simeq \text{SET}^{[2]}$ , and that  $\mathcal{L}^\circledast$  consists of the link graph carrying a marked cycle of length 2 that exchanges the two extremities of the edge.

**Claim 10.** *The species  $\mathcal{F}^\circledast$  of symmetric cycle-pointed free trees satisfies*

$$\mathcal{F}^\circledast = \mathcal{L}^\circledast \circledcirc \mathcal{R} + X \cdot \text{SET}^\circledast \circledcirc \mathcal{R}, \quad (33)$$

where  $\mathcal{R} := X \cdot \mathcal{F}$  is the species of all pointed trees.

*Proof.* Consider a tree produced from the species  $\mathcal{L}^\circledast \circledcirc \mathcal{R} + X \cdot \text{SET}^\circledast \circledcirc \mathcal{R}$  (for an example of a tree produced from  $X \cdot \text{SET}^\circledast \circledcirc \mathcal{R}$ , see the transition between the right and the left drawing in Figure 5). Clearly, such a tree is free and cycle-pointed and it is symmetric because the marked cycle of the core-structure – an edge  $e$  in the first case, a cycle-pointed set attached to a vertex  $v$  in the second case – already has length greater than 1. Hence  $\mathcal{F}^\circledast \supseteq \mathcal{L}^\circledast \circledcirc \mathcal{R} + X \cdot \text{SET}^\circledast \circledcirc \mathcal{R}^\circ$ . Notice also that in the first (second) case,  $e$  ( $v$ , respectively) is the center of symmetry of the resulting tree. Indeed each connecting path connects vertices on two different subtrees attached at the center of symmetry, which, by symmetry, stands in the middle of such a path.

Conversely, for each symmetric cycle-pointed free tree  $T$ , we color blue its center of symmetry, which plays the role of a *core-structure* for  $T$ . Partition  $\mathcal{F}^\circledast$  as  $\mathcal{F}_v^\circledast + \mathcal{F}_e^\circledast$ , where  $\mathcal{F}_v^\circledast$  ( $\mathcal{F}_e^\circledast$ , respectively) gathers the trees in  $\mathcal{F}^\circledast$  whose center of symmetry is a vertex (an edge, respectively). Define also  $\mathcal{M}_v$  ( $\mathcal{M}_e$ ) as the species of free trees with a distinguished vertex (edge, resp.) that is colored blue. Clearly  $\mathcal{M}_v = X \cdot (\text{SET} \circ \mathcal{R})$  and  $\mathcal{M}_e = \mathcal{L} \circ \mathcal{R}$ . From Proposition 6, we obtain  $\mathcal{M}_v^\circledast = X^\circ \star (\text{SET} \circ \mathcal{R}) + X \cdot \text{SET}^\circ \circledcirc \mathcal{R}$  and  $\mathcal{M}_e^\circledast = \mathcal{L}^\circ \circledcirc \mathcal{R}$ . Observe that  $\mathcal{F}_v^\circledast$  ( $\mathcal{F}_e^\circledast$ ) contains the structures of  $\mathcal{M}_v^\circledast$  (of  $\mathcal{M}_e^\circledast$ ) where the blue vertex (edge, resp.) is the center of symmetry. It is clear that the structures of  $X^\circ \star (\text{SET} \circ \mathcal{R})$  have their marked cycle of length 1, so they are not in  $\mathcal{F}_v^\circledast$ . Concerning the structures of  $X \cdot \text{SET}_{(1)}^\circ \circledcirc \mathcal{R}$ , the atoms of the marked cycle are on a same subtree attached at the blue vertex, so that the blue vertex is not the center of symmetry. Hence the structures of  $X \cdot \text{SET}_{(1)}^\circ \circledcirc \mathcal{R}$  are not in  $\mathcal{F}_v^\circledast$ . Similarly, the structures of  $\mathcal{L}_{(1)}^\circ \circledcirc \mathcal{R}$  are not in  $\mathcal{F}_e^\circledast$ . Therefore we obtain the second inclusion  $\mathcal{F}^\circledast \subseteq \mathcal{L}^\circledast \circledcirc \mathcal{R} + X \cdot \text{SET}^\circledast \circledcirc \mathcal{R}$ .  $\square$

**Proposition 11** (decomposing and counting free trees). *The species  $\mathcal{F}^\circ$  of cycle-pointed free trees has the following cycle-pointed recursive specification over the species  $\text{SET}$ ,  $\mathcal{L}^\circledast$ ,  $X$ :*

$$\begin{cases} \mathcal{F}^\circ &= X^\circ \star \mathcal{F}' + \mathcal{F}^\circledast, \\ \mathcal{F}' &= \text{SET} \circ \mathcal{R}, \quad \mathcal{R} = X \cdot \mathcal{F}', \\ \mathcal{F}^\circledast &= \mathcal{L}^\circledast \circledcirc \mathcal{R} + X \cdot \text{SET}^\circledast \circledcirc \mathcal{R}, \\ \mathcal{R}^\circ &= X^\circ \star (\text{SET} \circ \mathcal{R}) + X \cdot \text{SET}^\circ \circledcirc \mathcal{R}, \end{cases} \quad (34)$$

The ordinary generating function  $f(x) := \widetilde{\mathcal{F}}(x)$  of free trees satisfies the equations

$$xf'(x) = r(x) + x^2r'(x^2) + \left( \sum_{\ell \geq 2} x^\ell r'(x^\ell) \right) r(x) \quad (35)$$

$$= xr'(x)(1 - r(x)) + x^2r'(x^2), \quad (36)$$

where  $r(x)$  is specified by  $r(x) = x \exp \left( \sum_{i \geq 1} \frac{1}{i} r(x^i) \right)$ .

*Proof.* The first three lines of the grammar are Equations (31), (32), and (33), respectively. The fourth line<sup>2</sup> is obtained from the second line (i.e.,  $\mathcal{R} = X \cdot (\text{SET} \circ \mathcal{R})$ ) using the derivation rules of Proposition 6.

Concerning the OGSs, let  $r(x) := \widetilde{\mathcal{R}}(x)$  be the OGS of the species  $\mathcal{R}$ . Note that  $\widetilde{\mathcal{F}}^\circ(x) = xf'(x)$  and  $\widetilde{\mathcal{R}}^\circ(x) = xr'(x)$  by Theorem 5. By the computation rules for OGSs (Remark 2.1 and Remark 3.3), the second line of the grammar, i.e.,  $\mathcal{R} = X \cdot (\text{SET} \circ \mathcal{R})$ , yields  $r(x) = x \exp(\sum_{i \geq 1} r(x^i)/i)$ ; and the third line of the grammar yields

$$\widetilde{\mathcal{F}}^\circ(x) = \overline{Z}_{\mathcal{L}^\circ}(r(x), xr'(x); r(x^2), x^2r'(x^2); \dots) + x \overline{Z}_{\text{SET}^\circ}(r(x), xr'(x); r(x^2), x^2r'(x^2); \dots).$$

Applying the derivation rule (19) to  $\mathcal{L} = \text{SET}^{[2]}$  and  $\text{SET}$ , we get the expressions  $\overline{Z}_{\mathcal{L}^\circ} = t_2$  and  $\overline{Z}_{\text{SET}^\circ} = (\sum_{\ell \geq 2} t_\ell) \cdot Z_{\text{SET}}$ . Hence  $\widetilde{\mathcal{F}}^\circ(x) = x^2r'(x^2) + (\sum_{\ell \geq 2} x^\ell r'(x^\ell)) \cdot r(x)$ . Finally, the first line of the grammar yields  $\widetilde{\mathcal{F}}^\circ(x) = r(x) + \widetilde{\mathcal{F}}^\circ(x)$ , which gives Expression (35) of  $xf'(x)$ . Using  $xr'(x) = r(x)(1 + \sum_{\ell \geq 1} x^\ell r'(x^\ell))$ , this expression simplifies to Expression (36) of  $xf'(x)$ .  $\square$

*Remark 4.2.* The expression  $xf'(x) = xr'(x)(1 - r(x)) + x^2r'(x^2)$  clearly agrees with Otter's formula [26]:

$$f(x) = r(x) - \frac{1}{2}(r^2(x) - r(x^2)), \quad (37)$$

which can be obtained either from Otter's dissimilarity equation or from the dissymmetry theorem [2]. The new result of our method is to yield an expression for  $xf'(x)$  – Equation (35) – that has only positive signs, as it reflects a positive decomposition grammar. This is crucial to obtain random generators without rejection in Section 5.

All the arguments we have used for free trees can be adapted to decompose and enumerate species  $\mathcal{F}_\Omega$  of trees where the degrees of the vertices lie in a finite integer set  $\Omega$  that contains 1. It is helpful to define the auxiliary species  $\mathcal{R}_\Omega$  that consists of trees from  $\mathcal{F}_\Omega$  rooted at a leaf that does not count as an atom. By decomposing trees at the root, we note that  $\mathcal{R}_\Omega$  has the recursive specification

$$\mathcal{R}_\Omega = X \cdot \text{SET}_{\Omega-1} \circ \mathcal{R}_\Omega, \quad \text{with } \text{SET}_{\Omega-1} := \cup_{k \in \Omega} \text{SET}^{[k-1]}.$$

The species  $\mathcal{R}_\Omega$  serves as elementary rooted species to express the pointed species arising from  $\mathcal{F}_\Omega$ .

**Proposition 12** (decomposing and counting degree-constrained trees). *For any finite set  $\Omega$  of positive integers containing 1, let  $\mathcal{F}_\Omega$  be the species of nonplane trees whose vertex degrees are in  $\Omega$ . Then the species  $\mathcal{F}_\Omega^\circ$  has the following cycle-pointed recursive specification, where  $\text{SET}_\Omega := \cup_{k \in \Omega} \text{SET}^{[k]}$  and  $\text{SET}_{\Omega-1} := \cup_{k \in \Omega} \text{SET}^{[k-1]}$ :*

$$\begin{cases} \mathcal{F}_\Omega^\circ &= X^\circ \star \mathcal{F}_\Omega' + \mathcal{F}_\Omega^\circ, \\ \mathcal{F}_\Omega' &= \text{SET}_\Omega \circ \mathcal{R}_\Omega, \quad \mathcal{R}_\Omega = X \cdot \text{SET}_{\Omega-1} \circ \mathcal{R}_\Omega, \\ \mathcal{F}_\Omega^\circ &= \mathcal{L}^\circ \circledast \mathcal{R}_\Omega + X \cdot \text{SET}_\Omega^\circ \circledast \mathcal{R}_\Omega, \\ \mathcal{R}_\Omega^\circ &= X^\circ \star \text{SET}_\Omega \circ \mathcal{R}_\Omega + X \cdot \text{SET}_\Omega^\circ \circledast \mathcal{R}_\Omega. \end{cases} \quad (38)$$

<sup>2</sup>The fourth line of Equation (34) is not needed for enumeration, but it is necessary to make the grammar completely recursive, and, as such, will be necessary for writing down a random generator in Section 5.

The ordinary generating function  $f_\Omega(x) := \widetilde{\mathcal{F}}_\Omega(x)$  satisfies the equation

$$xf'_\Omega(x) = xZ_{\text{SET}_\Omega}(r_\Omega(x), r_\Omega(x^2), \dots) + x^2r'_\Omega(x^2) + x\overline{Z}_{\text{SET}_\Omega^\circledast}(r_\Omega(x), xr'_\Omega(x); r_\Omega(x^2), x^2r'_\Omega(x^2); \dots), \quad (39)$$

where  $r_\Omega(x)$  is specified by  $r_\Omega(x) = x \cdot Z_{\text{SET}_{\Omega-1}}(r_\Omega(x), r_\Omega(x^2), r_\Omega(x^3), \dots)$ . The power series  $Z_{\text{SET}_{\Omega-1}}$ ,  $Z_{\text{SET}_\Omega}$  and  $\overline{Z}_{\text{SET}_\Omega^\circledast}$  appearing in the equation are polynomials that can be computed explicitly.

EXAMPLE 1. *Unrooted nonplane binary trees.* Trees whose vertex degrees are in  $\Omega := \{1, 3\}$  are called unrooted nonplane binary trees (note that rooting such a tree at a leaf, one obtains a rooted nonplane binary tree, i.e., each internal node has two unordered children). In that case, the elementary cycle index sums required in Equation (39) are

$$Z_{\text{SET}_{\Omega-1}} = 1 + \frac{1}{2}s_1^2 + \frac{1}{2}s_2, \quad Z_{\text{SET}_\Omega} = s_1 + \frac{1}{6}s_1^3 + \frac{1}{2}s_1s_2 + \frac{1}{3}s_3, \quad \overline{Z}_{\text{SET}_\Omega^\circledast} = t_2s_1 + t_3.$$

Let  $f(x)$  be the OGS of unrooted nonplane binary trees and  $r(x)$  the OGS of rooted nonplane binary trees (rooted at a leaf that does not count as an atom). Firstly, from the expression of  $Z_{\text{SET}_{\Omega-1}}$  one obtains

$$r(x) = x \cdot \left(1 + \frac{1}{2}r(x)^2 + \frac{1}{2}r(x^2)\right).$$

Then, Equation (39) yields

$$xf'(x) = x \cdot (r(x) + \frac{1}{6}r(x)^3 + \frac{1}{2}r(x)r(x^2) + \frac{1}{3}r(x^3)) + x^2r'(x^2) + x \cdot (x^2r'(x^2)r(x) + x^3r'(x^3)),$$

i.e.,

$$f'(x) = r(x) + \frac{1}{6}r(x)^3 + \frac{1}{2}r(x)r(x^2) + \frac{1}{3}r(x^3) + xr'(x^2) \cdot (1 + xr(x)) + x^3r'(x^3).$$

From this equation one can extract the counting coefficients of unrooted nonplane binary trees with respect to the number of vertices (after extracting first the coefficients of  $r(x)$ ):

$$xf'(x) = 2 \cdot 1 \cdot x^2 + 4 \cdot 1 \cdot x^4 + 6 \cdot 1 \cdot x^6 + 8 \cdot 1 \cdot x^8 + 10 \cdot 2 \cdot x^{10} + 12 \cdot 2 \cdot x^{12} + 14 \cdot 4 \cdot x^{14} + 16 \cdot 6 \cdot x^{16} + \dots$$

Hence the first counting coefficients with respect to the number of internal nodes (starting with 0 internal nodes) are 1, 1, 1, 1, 2, 2, 4, 6. Pushing further one gets 1, 1, 1, 1, 2, 2, 4, 6, 11, 18, 37, 66, 135, 265, 552, 1132, which coincides with Sequence A000672 in [32] (the number of trivalent trees with  $n$  nodes). ..  $\square$

4.1.2. *Plane trees.* A *plane tree* is a tree endowed with an explicit *embedding* in the plane. Hence, a plane tree is a tree where the cyclic order around each vertex matters. Let  $\mathcal{E}$  be the species of plane trees, where again the atoms are the vertices. As usual the strategy to count plane trees is to decompose  $\mathcal{E}^\circ$ , distinguishing whether the marked cycle has length 1 or larger than 1:

$$\mathcal{E}^\circ = X^\circ \star \mathcal{E}' + \mathcal{E}^\circledast. \quad (40)$$

The species  $\mathcal{E}'$  is decomposed with the help of another species of plane trees: denote by  $\mathcal{A}$  the species of plane trees rooted at a leaf which does not count as an atom. Decomposing  $\mathcal{A}$  at the root, we get

$$\mathcal{A} = X \cdot \text{SEQ} \circ \mathcal{A}. \quad (41)$$

Again the species  $\mathcal{A}$  serves as elementary rooted species to express species of pointed plane trees:

**Proposition 13** (decomposing and counting plane trees). *The species  $\mathcal{E}^\circ$  of cycle-pointed plane trees has the following cycle-pointed recursive specification.*

$$\left\{ \begin{array}{l} \mathcal{E}^\circ = X^\circ \star \mathcal{E}' + \mathcal{E}^\circledast, \\ \mathcal{E}' = \text{CYC} \circ \mathcal{A}, \quad \mathcal{A} = X \cdot \text{SEQ} \circ \mathcal{A}, \\ \mathcal{E}^\circledast = \mathcal{L}^\circledast \odot \mathcal{A} + X \cdot \text{CYC}^\circledast \odot \mathcal{A}, \\ \mathcal{A}^\circ = X^\circ \star \text{SEQ} \circ \mathcal{A} + X \cdot \text{SEQ}^\circ \odot \mathcal{A}, \end{array} \right. \quad (42)$$

The ordinary generating function  $\widetilde{\mathcal{E}}(x)$  of plane trees satisfies the equation:

$$e'(x) = 1 + \sum_{\ell \geq 1} \frac{\phi(\ell)}{\ell} \log \frac{1}{1 - a(x^\ell)} + xa'(x^2) + \sum_{\ell \geq 2} \phi(\ell) \frac{x^\ell a'(x^\ell)}{1 - a(x^\ell)}, \quad (43)$$

where  $a(x)$  is the series of Catalan numbers:  $a(x) = \frac{1}{2}(1 - \sqrt{1-4x}) = \sum_{n \geq 0} \frac{1}{n+1} \binom{2n}{n} x^{n+1}$ .

By coefficient extraction, one gets the following formula for the number  $e_n$  of plane trees with  $n+1$  vertices (entry A002995 in [32]):

$$e_n = \frac{1}{n+1} \left[ \frac{1}{2n} \binom{2n}{n} + \frac{n+1}{2n} \sum_{k|n, k \neq n} \phi(n/k) \binom{2k}{k} + \mathbf{1}_{\left\{ \frac{n \text{ odd}}{n=2n'+1} \right\}} \cdot \binom{2n'}{n'} \right]. \quad (44)$$

*Proof.* The grammar is obtained by arguments similar to those used to derive the grammar (34) for free trees. The only difference is that the cyclic order of the neighbors around each vertex matters, so a SET construction in the grammar for free trees typically has to be replaced by a CYC construction in the grammar for plane trees.  $\square$

All the arguments apply similarly for species of plane trees where the degrees of vertices are constrained. As a counterpart to Proposition 11, we obtain:

**Proposition 14** (decomposing and counting degree-constrained plane trees). *For any finite set  $\Omega$  of positive integers containing 1, let  $\mathcal{E}_\Omega$  be the species of free trees where the degrees of vertices are constrained to lie in  $\Omega$ . Then the cycle-pointed species  $\mathcal{E}_\Omega^\circ$  is decomposable, it satisfies the following decomposition grammar, where  $\text{CYC}_\Omega := \cup_{k \in \Omega} \text{CYC}^{[k]}$  and  $\text{SEQ}_{\Omega-1} := \cup_{k \in \Omega} \text{SEQ}^{[k-1]}$ :*

$$\begin{cases} \mathcal{E}_\Omega^\circ &= X^\circ \star \mathcal{E}_\Omega' + \mathcal{E}_\Omega^{\circledast}, \\ \mathcal{E}_\Omega' &= \text{CYC}_\Omega \circ \mathcal{A}_\Omega, \quad \mathcal{A}_\Omega = X \cdot \text{SEQ}_{\Omega-1} \circ \mathcal{A}_\Omega, \\ \mathcal{E}_\Omega^{\circledast} &= \mathcal{L}^{\circledast} \odot \mathcal{A}_\Omega + X \cdot \text{CYC}_\Omega^{\circledast} \odot \mathcal{A}_\Omega, \\ \mathcal{A}_\Omega^\circ &= X^\circ \star \text{SEQ}_{\Omega-1} \circ \mathcal{A}_\Omega + X \cdot \text{SEQ}_{\Omega-1}^\circ \odot \mathcal{A}_\Omega. \end{cases} \quad (45)$$

The ordinary generating function  $e_\Omega(x) := \widetilde{\mathcal{E}}_\Omega(x)$  satisfies the equation<sup>3</sup>:

$$e_\Omega'(x) = \sum_{\substack{k \in \Omega \\ r|k}} \frac{\phi(r)}{k} a_\Omega(x^r)^{k/r} + x a_\Omega'(x^2) + \sum_{\substack{k \in \Omega \\ r|k, r > 1}} \phi(r) x^r a_\Omega'(x^r) a_\Omega(x^r)^{k/r-1}, \quad (46)$$

where  $a_\Omega(x)$  is specified by  $a_\Omega(x) = x \sum_{k \in \Omega} a_\Omega(x)^{k-1}$ .

**EXAMPLE 2.** *d-regular plane trees.* For  $d \geq 3$ , a  $d$ -regular plane tree is a plane tree such that each internal node has degree  $d$ , which corresponds to the case  $\Omega = \{1, d\}$  in Proposition 14. It is easily shown that such a tree with  $n$  internal nodes has  $m = n(d-2) + 2$  leaves. Let  $\mathcal{E}_{[d]}$  be the species of  $d$ -regular plane trees, where the atoms are the *leaves* (it proves here more convenient to take leaves as atoms and to write the counting coefficients according to the number of internal vertices). Let  $\mathcal{A}_{[d]} := \mathcal{E}_{[d]}$  be the corresponding derived species, which satisfies  $\mathcal{A}_{[d]} = X + (\mathcal{A}_{[d]})^{d-1}$ . The decomposition for the cycle-pointed species  $\mathcal{E}_{[d]}^\circ$  is

$$\mathcal{E}_{[d]}^\circ = X^\circ \star \mathcal{A}_{[d]} + \mathcal{L}^{\circledast} \odot \mathcal{A}_{[d]} + (\text{CYC}^{[d]})^{\circledast} \odot \mathcal{A}_{[d]}.$$

Hence, the OGS  $e_{[d]}(x) := \widetilde{\mathcal{E}}_{[d]}(x)$  satisfies:

$$x e_{[d]}'(x) = x a_{[d]}(x) + x^2 a_{[d]}'(x^2) + \sum_{r|d, r > 1} \phi(r) x^r a_{[d]}'(x^r) a_{[d]}(x^r)^{d/r-1}, \quad \text{where } a_{[d]}(x) = x + a_{[d]}(x)^{d-1}.$$

The coefficients of each of the summand series (such as  $a_{[d]}'(x) a_{[d]}(x)^{k-1}$ ) have a closed formula, which can be found for instance using the univariate Lagrange inversion formula. From these formulas, we obtain the following expression for the number  $e_{n,[d]}$  of  $d$ -regular plane trees with  $n$  internal nodes:

$$e_{n,[d]} = \frac{1}{m} \left[ \frac{1}{m-1} \binom{n+m-2}{n} + \mathbf{1}_{\left\{ \frac{2|m}{(d-2)|(m-2)/2} \right\}} \cdot \binom{n'+m_2-1}{n'} + \sum_{\substack{r > 1, r|d, r|m \\ (d-2)|(m-d)/r}} \phi(r) \binom{n_r+m_r-1}{n_r} \right] \quad (47)$$

<sup>3</sup>To obtain this equation we use the formula  $Z_{\text{CYC}^{[k]}} = \frac{1}{k} \sum_{r|k} \phi(r) s_r^{k/r}$ .

where

$$m = n(d - 2) + 2, \quad n' = \frac{m - 2}{2(d - 2)} = \frac{n}{2}, \quad n_r = \frac{m - d}{r(d - 2)} = \frac{n - 1}{r}, \quad m_r = m/r.$$

One can extend this formula to any degree distribution on vertices, by adding variables marking the degree of each vertex and applying the multivariate Lagrange inversion formula. A general enumeration formula is given in [6] using the dissymmetry theorem. . . . .  $\square$

*Remark 4.3.* As the only automorphisms for plane trees are rotations, a simple application of Burnside’s lemma is enough to get (47). (An adaptation of Burnside’s lemma to unrooted plane graphs is given in [22].) In contrast, free trees require more involved counting techniques using cycle index sums. Currently, these techniques are Otter’s dissimilarity equation, the dissymmetry theorem (these two methods being closely related), and now cycle-pointing.

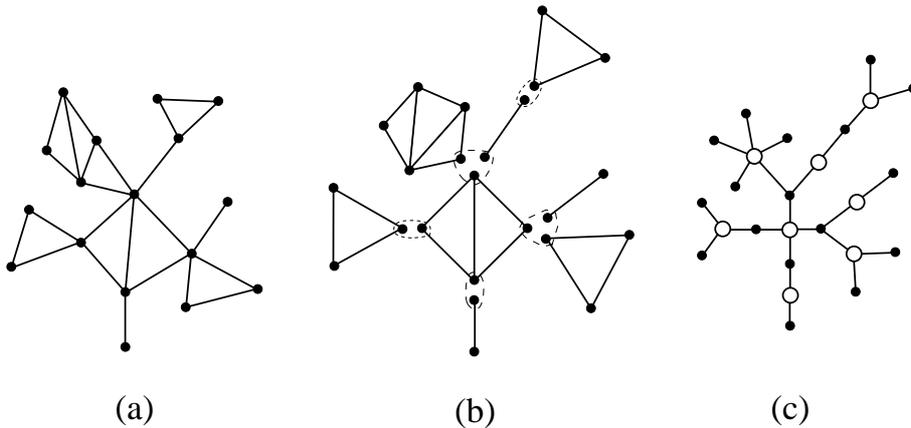


FIGURE 6. (a) A connected (outerplanar) graph, (b) its block-decomposition, (c) the associated Bv-tree.

**4.2. Graphs.** We extend here the decomposition principles which we have developed for trees to the more general case of a species of connected graphs, by taking advantage of a well-known “tree-like” decomposition of a connected graph into 2-connected components. (A 2-connected graph is a graph that has at least two vertices and has no separating vertex.) Given a connected graph  $G$ , a maximal 2-connected subgraph of  $G$  is called a *block* of  $G$ . The set of vertices of  $G$  is denoted  $\mathfrak{V}(G)$  and its set of blocks is denoted  $\mathfrak{B}(G)$ . The *Bv-tree* of  $G$  is the bicolored graph with vertex-set  $\mathfrak{V}(G) \cup \mathfrak{B}(G)$  and edges corresponding to the adjacencies between the blocks and the vertices of  $G$ , see Figure 6. It can be shown that the Bv-tree of  $G$  is indeed a tree, see [16, p.10] and [24] for details.

**Proposition 15.** *Let  $\mathcal{G}$  be a species of connected graphs that satisfy the following stability property: “a connected graph is in  $\mathcal{G}$  iff all its blocks are in  $\mathcal{G}$ ”. Let  $\mathcal{B}$  be the subspecies of graphs in  $\mathcal{G}$  that are 2-connected. Then  $\mathcal{G}$  admits a decomposition grammar from the species of 2-connected structures  $\mathcal{B}'$ ,  $\mathcal{B}^\circledast$ , and  $(\mathcal{B}')^\circ$ :*

$$\left\{ \begin{array}{l} \mathcal{G}^\circ = X \cdot \mathcal{G}' + \mathcal{G}^\circledast. \\ \mathcal{G}' = \text{SET} \circ \mathcal{K}, \quad \mathcal{K} = \mathcal{B}' \circ \mathcal{H}, \quad \mathcal{H} = X \cdot \mathcal{G}', \\ \mathcal{G}^\circledast = \mathcal{B}^\circledast \circledast \mathcal{H} + X \cdot \text{SET}^\circledast \circledast \mathcal{K}, \\ \mathcal{H}^\circ = X^\circ \star \text{SET} \circ \mathcal{K} + X \cdot \text{SET}^\circ \circledast \mathcal{K}, \quad \mathcal{K}^\circ = (\mathcal{B}')^\circ \circledast \mathcal{H}. \end{array} \right. \quad (48)$$

Hence, if the species of 2-connected structures  $\mathcal{B}^\circledast$  and  $\mathcal{B}'$  are decomposable (the latter implies that  $(\mathcal{B}')^\circ$  is decomposable), then the cycle-pointed species  $\mathcal{G}^\circ$  is decomposable as well. More generally, if  $Z_{\mathcal{B}'}$  and  $\overline{Z}_{\mathcal{B}^\circledast}$  are both solutions of an equation system involving the operations  $\{+, \star, \circ, \circledast\}$  and basic cycle-index sums, then  $Z_{\mathcal{G}^\circ}$  is also a solution of such an equation system.

*Proof.* The first line of the grammar is obtained as usual by distinguishing whether the marked cycle has length 1 or greater than 1. The second line easily follows from the block decomposition, as shown for instance in [14]. To wit, the marked vertex of a graph in  $\mathcal{G}'$  is incident to a collection of blocks, and a connected graph is possibly attached at each non-marked vertex of these blocks. Let us prove the third line in a similar way as for free trees (Claim 10). Consider a graph  $G$  in  $\mathcal{G}'$ , and let  $T$  be the Bv-tree of  $G$ . Clearly the Bv-tree of a graph has less structure than the graph itself, so any automorphism of  $G$  induces an automorphism on  $T$ . In particular  $T$  is a symmetric cycle-pointed tree, hence it has a center of symmetry that either corresponds to a block or to a vertex of  $G$ . The species of graphs in  $\mathcal{G}^{\otimes}$  whose center of symmetry in the Bv-tree is a vertex (a block) is denoted  $(\mathcal{G}^{\otimes})_{\text{V}}$  ( $(\mathcal{G}^{\otimes})_{\text{B}}$ , resp.). Let  $\mathcal{G}_{\text{V}}$  ( $\mathcal{G}_{\text{B}}$ ) be the species of graphs in  $\mathcal{G}$  with a marked vertex (block, resp.) that is colored blue. Then clearly  $\mathcal{G}_{\text{V}} = X \cdot \mathcal{G}'$  and  $\mathcal{G}_{\text{B}} = \mathcal{B} \circ (X \cdot \mathcal{G}')$ . Hence, following the notations introduced in the grammar,  $\mathcal{G}_{\text{V}} = \mathcal{H} = X \cdot \text{SET} \circ \mathcal{K}$  and  $\mathcal{G}_{\text{B}} = \mathcal{B} \circ \mathcal{H}$ . Note that the structures in  $(\mathcal{G}^{\otimes})_{\text{V}}$  ( $(\mathcal{G}^{\otimes})_{\text{B}}$ ) are the graphs in  $\mathcal{G}_{\text{V}}^{\otimes}$  ( $\mathcal{G}_{\text{B}}^{\otimes}$ ) such that the center of symmetry of the associated Bv-tree is the blue vertex (block, resp.). It is easy to check, in a similar way as for free trees, that this property holds only for the graphs of  $\mathcal{G}_{\text{V}}^{\otimes}$  that are in  $X \cdot \text{SET}^{\otimes} \odot \mathcal{K}$  and only for the graphs of  $\mathcal{G}_{\text{B}}^{\otimes}$  that are in  $\mathcal{B}^{\otimes} \odot \mathcal{H}$ . Finally, the 4th line, which is necessary to have only species of 2-connected structures as terminal species, is obtained from the second line by applying the derivation rules (Proposition 6).  $\square$

*Remark 4.4.* Trees are exactly connected graphs where each block is an edge. In other words, the species  $\mathcal{F}$  of free trees is the species  $\mathcal{G}$  of connected graphs formed from the species  $\mathcal{B} = \mathcal{L}$  (the one-element species that consists of the link graph). One easily checks that, in that case, the grammar (48) for  $\mathcal{G}$  is equivalent to the grammar (34) for free trees.

4.2.1. *Cacti graphs.* Cacti graphs form an important class of graphs that have several algorithmic applications. They consist of cycles attached together in a tree-like fashion; in other words, the species of cacti graphs arises from the species of 2-connected structures as  $\mathcal{B} = \mathcal{L} + \mathcal{P}$  where  $\mathcal{L}$  is the species of the link graph and  $\mathcal{P}$  is the species of polygons with at least 3 edges (i.e.,  $\mathcal{B}$  is the species of polygons where one allows the degenerated 2-sided polygon).

Thanks to the grammar (48), the unlabeled enumeration of connected cacti graphs reduces to the calculation of the cycle-index sums for the species of 2-connected structures  $\mathcal{B}'$  and  $\mathcal{B}^{\otimes}$  (the cycle-index sum  $\overline{Z}_{(\mathcal{B}')^{\circ}}$  is also required, but it can directly be deduced from  $Z_{\mathcal{B}'}$  by differentiation). Since the 2-connected cacti graphs are polygons, the possible automorphisms are from the dihedral group. In addition, the presence of a marked vertex (for  $\mathcal{B}'$ ) or cycle (for  $\mathcal{B}^{\otimes}$ ) restricts the symmetries. For instance, if a structure in  $\mathcal{B}'$  has a marked (unlabeled) vertex  $v$ , then the automorphisms have to fix  $v$ ; there are only two such symmetries for each polygon, the identity and the unique reflection whose axis passes by  $v$ . Accordingly, we have two terms in the expression of  $Z_{\mathcal{B}'}$  below, the first one for the identity, and the second one for reflections (where one distinguishes whether the polygon has odd or even length).

$$Z_{\mathcal{B}'} = \frac{1}{2} \frac{s_1}{1-s_1} + \frac{1}{2} \frac{s_1 + s_2}{1-s_2}$$

For  $\mathcal{B}^{\otimes}$ , all symmetries must be nontrivial and have to respect the marked cycle. These symmetries are of two types: rotation and reflection, which yields the two main terms in the expression of  $\overline{Z}_{\mathcal{B}^{\otimes}}$  below.

$$\overline{Z}_{\mathcal{B}^{\otimes}} = \sum_{r \geq 2} \frac{\phi(r)}{2} \frac{t_r}{1-s_r} + \frac{t_2(s_1^2 + 2s_1 + 1)}{2(1-s_2)^2}$$

The expressions for  $Z_{\mathcal{B}'}$  and  $\overline{Z}_{\mathcal{B}^{\otimes}}$  can be used to enumerate unlabeled cacti graphs. We just have to translate (using the computation rules in Remark 2.1 and Remark 3.3) the grammar (48) – applied to the species of cacti graphs – into an equation system satisfied by the corresponding ordinary generating functions.

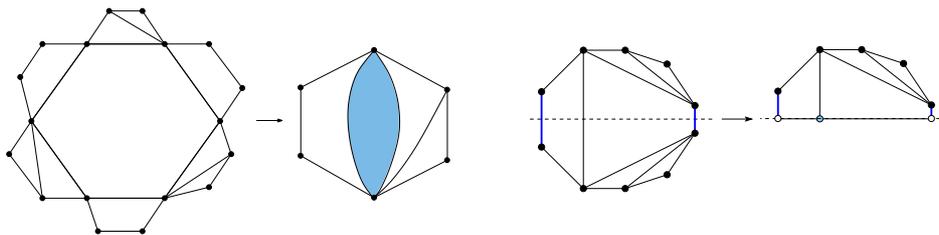


FIGURE 7. Left (Right, resp.): the quotient of a dissection under a rotation (reflection, resp.).

**Proposition 16** (enumeration of unlabeled cacti graphs). *The ordinary generating function  $c(x) = \sum_n c_n x^n$  of unlabeled cacti graphs counted with respect to the number of vertices satisfies*

$$\begin{aligned} xc'(x) &= H(x) + I(x), \\ H(x) &= x \exp\left(\sum_{r \geq 1} \frac{K(x^r)}{r}\right), \text{ with } K(x) = \frac{1}{2} \frac{H(x)}{1-H(x)} + \frac{1}{2} \frac{H(x) + H(x^2)}{1-H(x^2)}, \\ I(x) &= \sum_{r \geq 2} \frac{\phi(r)}{2} \frac{x^r H'(x^r)}{1-H(x^r)} + \frac{x^2 H'(x^2)(H(x)^2 + 2H(x) + 1)}{2(1-H(x^2))^2} + H(x) \sum_{r \geq 2} x^r K'(x^r), \end{aligned}$$

from which one can extract<sup>4</sup> the counting coefficients  $c_n$  (after firstly extracting the coefficients of  $H(x)$ ):

$$xc'(x) = 1 \cdot 1 \cdot x + 2 \cdot 1 \cdot x^2 + 3 \cdot 2 \cdot x^3 + 4 \cdot 4 \cdot x^4 + 5 \cdot 9 \cdot x^5 + 6 \cdot 23 \cdot x^6 + 7 \cdot 63 \cdot x^7 + \dots$$

**4.2.2. Outerplanar graphs.** Outerplanar graphs are graphs that can be drawn in the plane so that all vertices are incident to the outer face. They form a fundamental subspecies of the species of planar graphs, which already captures some difficulties of the species of all planar graphs; for example, the convergence rate of sampling procedures using the Markov Chain approach is not known. However, outerplanar graphs are easier to tackle with the decomposition approach. For enumeration, we use the well-known property that 2-connected outerplanar graphs, except for the one-edge graph, have a unique hamiltonian cycle. Hence, the species  $\mathcal{B}$  of 2-connected outerplanar graphs can be identified with the species of dissections of a polygon (allowing a degenerated 2-sided dissection). This time, to obtain the cycle index sums  $Z_{\mathcal{B}'}$  and  $\bar{Z}_{\mathcal{B} \oplus}$ , we have to count not polygons (as for cacti graphs) but *dissections* of a polygon under the action of the dihedral group. We only sketch the method here (the principles for counting such dissections are well known, going back to earlier articles of Read [30], see also [3] for more detailed calculations).

For each type of symmetry (rotation or reflection), one considers the “quotient dissection”, as shown in Figure 7. Notice that a dissection fixed by a rotation has either a central edge (only for the rotation of order two) or a central face. In case of a central edge  $e$ , it turns out to be more convenient to “double”  $e$ , so as to always have a central face before taking the quotient. Thus, the quotient dissection has a marked face (the quotient of the central face) that might have degree one (only for rotations of order at least three) or two (only for rotations of order at least two). Concerning quotient dissections under a reflection, there are two special vertices  $v_1$  and  $v_2$  on the boundary (the intersections of the original polygon with the reflection-axis), and there might be some other special vertices, all of degree three, on the boundary path from  $v_1$  to  $v_2$ ; see Figure 7.

The second ingredient is to take the dual of such quotient dissections in order to obtain plane trees, which are easier to decompose and to count. Notice that if the rotation is the identity rotation, then the associated plane tree is in the species  $\mathcal{F}$  of plane trees with no vertex of degree two. Notice also that each leaf of the tree corresponds to a vertex of the dissection; see Figure 8

<sup>4</sup>The calculations have been done with the help of the computer algebra system Maple.

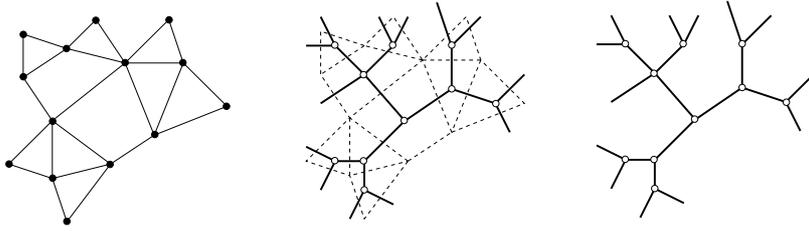


FIGURE 8. The dual tree of a dissection.

for an example. The generating function of  $\mathcal{F}$  with respect to the number of leaves satisfies

$$F(x) = x + \frac{F(x)^2}{1 - F(x)}. \quad (49)$$

To calculate  $Z_{\mathcal{B}'}$  and  $Z_{\mathcal{B}^\otimes}$ , one computes separately the contributions of rotations and reflections to  $Z_{\mathcal{B}'}$  and to  $Z_{\mathcal{B}^\otimes}$ . In each case, using duality, the contribution is easily expressed in terms of the series  $F(x)$ . All calculations done, one finds:

$$Z_{\mathcal{B}'} = \frac{1}{2}(F(s_1) + (s_1 + s_2)P(s_2)), \quad \overline{Z}_{\mathcal{B}^\otimes} = \frac{1}{2} \sum_{r \geq 2} \phi(r) t_r G(s_r) + \frac{1}{2} t_2 (1 + s_1^2 Q(s_2) + 2s_1 R(s_2) + S(s_2)),$$

where

$$G(x) = \frac{F'(x)}{1 - F(x)}, \quad P(x) = \frac{1}{x} \frac{F(x)}{1 - 2F(x)},$$

$$Q(x) = \frac{d}{dx} P(x), \quad R(x) = \frac{d}{dx} (xP(x)), \quad S(x) = \frac{d}{dx} (x^2 P(x)).$$

Similarly as for cacti graphs, the expressions for  $Z_{\mathcal{B}'}$  and  $\overline{Z}_{\mathcal{B}^\otimes}$  make it possible to enumerate unlabeled connected outerplanar graphs. Translating the grammar (48) into an equation system on the corresponding generating functions, we obtain the following.

**Proposition 17** (Enumeration of unlabeled connected outerplanar graphs). *The ordinary generating function  $o(x) = \sum_n o_n x^n$  of unlabeled connected outerplanar graphs counted with respect to the number of vertices satisfies the system:*

$$x o'(x) = H(x) + I(x),$$

$$H(x) = x \exp \left( \sum_{r \geq 1} \frac{K(x^r)}{r} \right), \quad \text{with } K(x) = \frac{1}{2} F(H(x)) + \frac{1}{2} (H(x) + H(x^2)) \cdot P(H(x^2)),$$

$$I(x) = \frac{1}{2} \sum_{r \geq 2} \phi(r) x^r H'(x^r) G(H(x^r)) + H(x) \sum_{r \geq 2} x^r K'(x^r)$$

$$+ \frac{1}{2} x^2 H'(x^2) (1 + H(x)^2 Q(H(x^2))) + 2H(x) R(H(x^2)) + S(H(x^2)),$$

where the series  $F, G, P, Q, R, S$  are defined above. One extracts from this system (extracting firstly the coefficients in  $F, Q, R, S$ , then in  $H$  and  $K$ , then in  $I$ ) the counting coefficients  $o_n$ :

$$x o'(x) = 1 \cdot \mathbf{1} \cdot x + 2 \cdot \mathbf{1} \cdot x^2 + 3 \cdot \mathbf{2} \cdot x^3 + 4 \cdot \mathbf{5} \cdot x^4 + 5 \cdot \mathbf{13} \cdot x^5 + 6 \cdot \mathbf{46} \cdot x^6 + 7 \cdot \mathbf{172} \cdot x^7 + \dots$$

**4.3. Maps.** A map is a planar graph embedded on a sphere up to isotopic deformation, i.e., it is a planar graph together with a cyclic order of the neighbors around each vertex. There is a huge literature on maps since the pioneering work of Tutte [33]. As we show next, the decomposition grammar (48) for maps is actually simpler than for graphs, and it allows us to enumerate (unrooted unlabeled) 2-connected maps in terms of not necessarily connected maps. To write down the grammar, it turns out to be more convenient to take half-edges as atoms instead of vertices. Denote by  $\mathcal{M}$  the species of maps – so  $\mathcal{R} := \mathcal{Z} \cdot \mathcal{M}'$  is the species of *rooted* maps (maps with a marked half-edge) – and by  $\mathcal{B}$  the species of 2-connected maps (the loop-map is considered as 2-connected).

**Proposition 18.** *The species of rooted maps and symmetric cycle-pointed maps (in each length  $\ell \geq 2$  of the marked cycle) have the following recursive specification over the corresponding species of rooted and cycle-pointed 2-connected maps.*

$$\begin{cases} \mathcal{R} = \mathcal{K} \cdot \text{SEQ} \circ \mathcal{K}, & \mathcal{K} = X \cdot \mathcal{B}' \circ \mathcal{H}, & \mathcal{H} = X \cdot (1 + \mathcal{R}), \\ \mathcal{M}_{(\ell)}^{\circ} = \mathcal{B}_{(\ell)}^{\circ} \odot \mathcal{H} + \text{CYC}_{(\ell)}^{\circ} \odot \mathcal{K} & \text{for } \ell \geq 2. \end{cases} \quad (50)$$

*Proof.* The arguments are similar to the proof for graphs (Proposition 15). The only difference is that one takes the embedding into account, hence corners (which are in one-to-one correspondence with half-edges for a given map) play for maps a similar role as vertices do for graphs, and a SET construction typically becomes a CYC construction here. Let us comment here on the decomposition for symmetric cycle-pointed maps (the one for rooted maps is well known, see [33]). One has

$$\mathcal{M}^{\circledast} = \mathcal{B}^{\circledast} \odot \mathcal{H} + \text{CYC}^{\circledast} \odot \mathcal{K}, \quad (51)$$

where the first (second) term takes account of the maps whose center of symmetry – for the associated block-decomposition tree – is a block (vertex, respectively). Further simplification is possible, since a rooted map has only the identity as automorphism. Hence, the species of rooted maps  $\mathcal{H}$  and  $\mathcal{K}$  satisfy  $\mathcal{H}^{\circ} = \mathcal{H}_{(1)}^{\circ}$  and  $\mathcal{K}^{\circ} = \mathcal{K}_{(1)}^{\circ}$ . Thus, Equation (51) can be “sliced” into a collection of equations, one for each length  $\ell \geq 2$  of the marked cycle.  $\square$

An important property of any map automorphism – as shown by Liskovets [22] – is that all its cycles have the same length  $\ell$ , which is also the order of the automorphism. Hence, the number of half-edges of a cycle-pointed map with a marked cycle of length  $\ell$  is divisible by  $\ell$ . For  $\ell \geq 1$ , denote by  $M_{(\ell)}(x)$  ( $B_{(\ell)}(y)$ ) the series counting unlabeled cycle-pointed maps (2-connected maps, respectively), according to the number of half-edges, divided by  $\ell$ . In particular,  $R(x) := M_{(1)}(x)$  and  $S(x) := B_{(1)}(x)$  are the series counting rooted maps and rooted 2-connected maps, respectively. We clearly have

$$\overline{Z}_{\mathcal{M}_{(\ell)}^{\circ}} = \frac{t_{\ell}}{s_{\ell}} M_{(\ell)}(s_{\ell}), \quad \overline{Z}_{\mathcal{B}_{(\ell)}^{\circ}} = \frac{t_{\ell}}{s_{\ell}} B_{(\ell)}(s_{\ell}).$$

Given this simplification, the grammar (50) is translated into the following system relating the series counting species of maps and species of 2-connected maps:

$$R(x) = \frac{K(x)}{1 - K(x)}, \quad H(x) = x(1 + R(x)), \quad K(x) = \frac{x}{H(x)} S(H(x)), \quad (52)$$

$$M_{(\ell)}(x) = \frac{xH'(x)}{H(x)} B_{(\ell)}(H(x)) + \phi(\ell) \frac{xK'(x)}{1 - K(x)} \quad \text{for } \ell \geq 2. \quad (53)$$

In the case of maps, the decomposition grammar is used in the other direction, i.e., one obtains the enumeration of (unrooted) 2-connected maps from maps. Indeed, unconstrained maps are easier to count, by a method of quotient [22] similar to the one we have used for counting dissections in Section 4.2.2.

Let us first review (from Tutte [33]) how one obtains an expression for the series  $S(y)$  counting rooted 2-connected maps from an expression for the series  $R(x)$  counting rooted maps. One starts from the following expression of  $R(x)$ :

$$R(x) = \frac{\beta(2 - 9\beta)}{(1 - 3\beta)^2}, \quad \text{with } \beta = \beta(x) \text{ specified by } \beta = x^2 + 3\beta^2. \quad (54)$$

Next, notice that the change of variable  $y = H(x) = x(1 + R(x))$  between rooted maps and rooted 2-connected maps is such that  $y^2$  is also rational in  $\beta$ :

$$y^2 = x^2(1 + R(x))^2 = \beta \frac{(1 - 4\beta)^2}{(1 - 3\beta)^3}.$$

Equivalently:

$$y^2 = \eta(1 - \eta)^2, \quad \text{where } \eta := \frac{\beta}{1 - 3\beta},$$

so the dependence between  $\eta$  and  $\beta$  is invertible:  $\beta = \eta/(1 + 3\eta)$ . Notice also from (52) that

$$S(y) = \frac{H(x)}{x}K(x) = (1 + R(x))K(x) = R(x) = \frac{\beta(2 - 9\beta)}{(1 - 3\beta)^2}.$$

Replacing  $\beta$  by  $\eta/(1 + 3\eta)$ , one gets

$$S(y) = \eta(2 - 3\eta), \quad \text{with } \eta = \eta(y) \text{ specified by } \eta = \frac{y^2}{(1 - \eta)^2},$$

which can also be written as

$$S(y) = \bar{S}(y^2), \quad \text{with } \bar{S}(y) = \bar{\eta}(2 - 3\bar{\eta}), \quad \text{and } \bar{\eta} := \bar{\eta}(y) \text{ specified by } \bar{\eta} = \frac{y}{(1 - \bar{\eta})^2}. \quad (55)$$

In a similar way, if two series  $f(x) = g(y)$  are related by the change of variables  $y = H(x)$  and if  $f(x)$  is rational in  $\beta(x)$ , then  $g(y)$  is rational in  $\eta(y)$  (replacing  $\beta$  by  $\eta/(1 + 3\eta)$ ). For instance, for  $\ell \geq 3$ , it has been shown by Liskovets using the quotient method (see [12] for the reformulation on series) that

$$M_{(\ell)}(x) = \phi(\ell) \frac{6\beta(x)}{1 - 6\beta(x)}.$$

Since  $H(x)/x = 1 + R(x)$  and  $K(x) = R(x)/(1 + R(x))$  are rational in  $\beta$ , as well as  $xH'(x)$  and  $xK'(x)$  (noticing that  $xd f/dx = 2x^2 d f/dx^2 = 2\beta(1 - 3\beta) \cdot (d f/d\beta)/(dx^2/d\beta)$ ), one finds from (53) a rational expression in  $\beta$  for the series  $B_{(\ell)}(H(x))$ . Replacing  $\beta$  by  $\eta/(1 + 3\eta)$  in that expression, one finally gets:

$$B_{(\ell)}(y) = \phi(\ell) \frac{2\eta(y)}{1 - 3\eta(y)} \quad \text{for } \ell \geq 3,$$

which can also be written as

$$B_{(\ell)}(y) = \phi(\ell)G(y^2) \quad \text{for } \ell \geq 3, \quad \text{with } G(y) = \frac{2\bar{\eta}(y)}{1 - 3\bar{\eta}(y)}. \quad (56)$$

In a similar way, starting from the expression (given in [12])

$$M_{(2)}(x) = -1 + \frac{1 - 2\beta(x)}{(1 - 6\beta(x))(1 - 3\beta(x))} + x \cdot \frac{2}{(1 - 6\beta(x))(1 - 3\beta(x))},$$

one obtains the following expression for  $B_{(2)}(y)$ :

$$B_{(2)}(y) = \frac{\eta(y)(3 - \eta(y))}{1 - 3\eta(y)} + y \frac{2}{1 - 3\eta(y)},$$

which can also be written as

$$B_{(2)}(y) = P(y^2) + yQ(y^2), \quad \text{with } P(y) = \frac{\bar{\eta}(y)(3 - \bar{\eta}(y))}{1 - 3\bar{\eta}(y)} \quad \text{and } Q(y) = \frac{2}{1 - 3\bar{\eta}(y)}. \quad (57)$$

**Proposition 19** (counting unrooted 2-connected maps, recover [23]). *The number  $t_n$  of (unrooted unlabeled) 2-connected maps with  $n$  edges satisfies:*

$$t_n = \frac{1}{2n} \left( s_n + u_n + \frac{1}{2} \sum_{k|n, k \neq n} \phi(n/k) \cdot (9k^2 - 9k + 1)s_k \right), \quad (58)$$

where  $s_n = \frac{2(3n - 3)!}{n!(2n - 1)!}$ ,  $u_n = \frac{n(n + 1)}{2} s_{(n+1)/2}$  if  $n$  is odd, and  $u_n = \frac{(3n - 4)n}{8} s_{n/2}$  if  $n$  is even.

*Proof.* Cycle-pointing ensures that the generating function  $\widetilde{\mathcal{B}}^\circ(y) = \sum_n 2nt_n y^{2n}$  satisfies

$$\widetilde{\mathcal{B}}^\circ(y) = S(y) + B_{(2)}(y^2) + \sum_{\ell \geq 3} B_{(\ell)}(y^\ell) = S(y) + \left( B_{(2)}(y^2) - G(y^4) \right) + \sum_{\ell \geq 2} \phi(\ell)G(y^{2\ell}).$$

Extracting the coefficient  $[y^{2n}]$  in this equation yields

$$2nt_n = s_n + u_n + \sum_{\ell \geq 2, \ell | n} \phi(\ell)v_{n/\ell} = s_n + u_n + \sum_{k|n, k \neq n} \phi(n/k)v_k, \quad (59)$$

where  $s_n = [y^n]\overline{S}(y)$ ,  $v_n = [y^n]G(y)$ ,  $u_n = [y^n](B_{(2)}(y) - G(y^2))$ , which is  $[y^{n/2}]P(y) - v_{n/2}$  if  $n$  is even and  $y^{(n-1)/2}Q(y)$  if  $n$  is odd. Notice that the series  $\overline{S}(y)$ ,  $G(y)$ ,  $P(y)$ , and  $Q(y)$  are rational in the simple series  $\overline{\eta}(y) = y/(1 - \overline{\eta}(y))^2$ . Hence the Lagrange inversion formula [2, Section 3.1] allows us to extract exact formulas for the coefficients  $s_n$ ,  $v_n$ , and  $u_n$ . Substituting these exact expressions in (59), one obtains the announced formula for  $t_n$ .  $\square$

The enumeration of unrooted 2-connected maps has first been done by Liskovets and Walsh [23] using the quotient method in a quite involved way. More recently, the counting formula has been recovered in [12] using a method of extraction at a center of symmetry on quadrangulations. What we do here is equivalent to [12], but the cycle-pointed framework allows us to write the equations on generating functions in a more systematic way.

**4.4. Asymptotic enumeration.** Cycle-pointing makes it possible to easily obtain an asymptotic estimate for the coefficients counting the number of unlabeled structures from a species, provided that the singular behavior of the OGS counting the associated species of *rooted* unlabeled structures is known.

We illustrate the method on free trees. Let  $R(x)$  be the OGS of rooted unlabeled nonplane trees, which is specified by  $R(x) = x \exp(\sum_{i \geq 1} R(x^i)/i)$ . It is well known that  $R(x)$  has a dominant singularity  $\rho < 1$  of the square-root type [10, VII.5]. That is, in the slit complex neighborhood  $D_\epsilon := \{x \mid x - \rho \notin \mathbb{R}_+ \text{ and } |x - \rho| < \epsilon\}$  we have the expansion

$$R(x) = 1 - aX + o(X), \quad \text{where } X = \sqrt{1 - x/\rho}, \quad (60)$$

which yields – using transfer theorems of analytic combinatorics [10, VI] – the asymptotic estimate

$$R_n \sim cn^{-3/2}\rho^{-n}, \quad \text{where } c = \frac{a}{2\sqrt{\pi}} \approx 0.43922, \quad \rho \approx 0.33832, \quad (61)$$

for the number of rooted unlabeled nonplane trees with  $n$  vertices.

To obtain a similar estimate for *free trees*, we consider the OGS  $P(x)$  of cycle-pointed nonplane trees and start from the expression of  $P(x)$  obtained in Proposition 11:

$$P(x) = x^2 R'(x^2) + \left(1 + \sum_{\ell \geq 2} x^\ell R'(x^\ell)\right) R(x).$$

Notice that, since  $\rho < 1$ , the series  $A := x^2 R'(x^2)$  and  $B := 1 + \sum_{\ell \geq 2} x^\ell R'(x^\ell)$  are analytic at  $x = \rho$ , and the value at  $x = \rho$  of  $B$  is the positive constant

$$b := 1 + \sum_{\ell \geq 2} \rho^\ell R'(\rho^\ell). \quad (62)$$

Therefore, from the singular expansion (60) of  $R(x)$ , we obtain

$$P(x) = P(\rho) - abX + o(X). \quad (63)$$

Let us simplify further the positive constant  $b = B(\rho)$ . First, by deriving the equation that specifies  $R(x)$ , one obtains  $xR'(x) = R(x)(1 + \sum_{i \geq 1} x^i R'(x^i))$ , which yields

$$B(x)R(x) = xR'(x)(1 - R(x)).$$

By deriving the singular expansion of  $R(x)$ , one obtains

$$R'(x) = \frac{a}{2\rho} X^{-1} + o(X^{-1}),$$

hence  $xR'(x)(1 - R(x))$  converges to  $a^2/2$  as  $x \rightarrow \rho$ , i.e.,  $b = a^2/2$ .

**Proposition 20** (asymptotic enumeration of free trees). *The number  $F_n$  of unlabeled free trees with  $n$  vertices satisfies*

$$F_n \sim (2\pi c^3)n^{-5/2}\rho^{-n}, \quad (64)$$

where  $c$  is the constant and  $\rho^{-1}$  is the growth ratio in the estimate (61) for rooted nonplane trees ( $R_n \sim cn^{-3/2}\rho^{-n}$ ).

*Proof.* From the singular expansion (63) of  $P(x)$  we obtain (again by the transfer theorems of singularity analysis)

$$x^n[P(x)] \sim \frac{ab}{2\sqrt{\pi}} n^{-3/2} \rho^{-n}.$$

Using  $c = a/(2\sqrt{\pi})$  and  $b = a^2/2$ , we have  $ab/(2\sqrt{\pi}) = 2\pi c^3$ . Finally, Theorem 5 (unbiased pointing) yields  $F_n = \frac{1}{n}[x^n]P(x)$ , so  $F_n \sim (2\pi c^3)n^{-5/2}\rho^{-n}$ .  $\square$

It is also possible to get the estimate of  $F_n$  from Otter's dissimilarity equation (or from the dissymmetry theorem). However, we find that cycle-pointing provides a more transparent explanation why the asymptotic estimate of the coefficients  $F_n$  counting unlabeled structures from an unrooted "tree-like" species  $\mathcal{F}$  is of the universal type  $cn^{-5/2}\rho^{-n}$ . The argument is very simple:

- The OGS  $P(x)$  of the cycle-pointed species  $\mathcal{F}^\circ$  is positively expressed in terms of the OGS of the rooted species, which has a square-root dominant singularity. Therefore,  $P(x)$  inherits the same singularity and singularity type (square-root).
- Transfer theorems of singularity analysis ensure that a square-root dominant singularity yields an asymptotic estimate in  $cn^{-3/2}\rho^{-n}$  for the coefficients  $[x^n]P(x)$ . Since  $F_n = \frac{1}{n}[x^n]P(x)$ , one gets  $F_n \sim cn^{-5/2}\rho^{-n}$ .

This strategy applies to all species of trees encountered in this section, as well as to cacti graphs and connected outerplanar graphs (in all cases one starts from the singular expansion of the OGS counting the corresponding *rooted* species).

## 5. APPLICATION TO RANDOM GENERATION

Recently, so-called Boltzmann samplers have been introduced by Duchon et al [9] as a general method to efficiently (typically in linear time) generate uniformly at random combinatorial structures that admit a decomposition. In contrast to the more costly *recursive method of sampling* [25], which is based on counting coefficients of the recursive decomposition, Boltzmann samplers are primarily based on generating functions. Until now Boltzmann samplers were developed in the labeled setting [9] and partially in the unlabeled setting [13].

In this section we provide a more complete method in the unlabeled setting. In order to deal with the substitution construction and the cycle-pointing operator (which are not covered in [13]), we have to describe samplers not solely based on ordinary generating functions, but on cycle index sums – also known as Pólya operators. Therefore we call these random generators *Pólya-Boltzmann samplers*.

With these refined samplers we are able to design in a systematic way (via specific generation rules) a Pólya-Boltzmann sampler for species that admit a recursive decomposition, thereby allowing in the decomposition all operators that have been described in this article. When specialized suitably, a Pólya-Boltzmann sampler reduces to an ordinary Boltzmann sampler, hence it provides a uniform random sampler for species of unlabeled structures. In particular, we obtain highly efficient random generators for the species in Section 4: for trees, cacti graphs, outerplanar graphs, etc.

**5.1. Ordinary Boltzmann Samplers.** Let  $\mathcal{A}$  be a species of structures, and let  $\tilde{\mathcal{A}}(x)$  be the ordinary generating series for  $\mathcal{A}$ . A real number  $x > 0$  is said to be *admissible* iff the sum defining  $\tilde{\mathcal{A}}(x)$  converges ( $x$  within the disk of convergence of the series). Given a fixed admissible value  $x > 0$ , an *ordinary Boltzmann sampler* for unlabeled structures from  $\mathcal{A}$  is a random generator  $\Gamma_{\tilde{\mathcal{A}}}(x)$  that draws each structure  $\gamma \in \tilde{\mathcal{A}}$  with probability

$$\mathbb{P}_x(\gamma) = \frac{x^{|\gamma|}}{\tilde{\mathcal{A}}(x)}. \quad (65)$$

Notice that this distribution has the fundamental property to be *uniform*, i.e., any two unlabeled structures of the species with the same size have the same probability.

5.1.1. *Automatic rules to design Boltzmann samplers.* As described in [9], there are simple rules to assemble Boltzmann samplers for the two classical constructions Sum and Product (Bern( $p$ ) stands for a Bernoulli law, returning “true” with probability  $p$  and “false” with probability  $1 - p$ ).

$$\begin{aligned} \mathcal{C} = \mathcal{A} + \mathcal{B}. \quad \Gamma\tilde{\mathcal{C}}(x): & \text{ if Bern}(\tilde{\mathcal{A}}(x)/\tilde{\mathcal{C}}(x)) \text{ return } \Gamma\tilde{\mathcal{A}}(x) \text{ else return } \Gamma\tilde{\mathcal{B}}(x) \\ \mathcal{C} = \mathcal{A} \cdot \mathcal{B}. \quad \Gamma\tilde{\mathcal{C}}(x): & \text{ return } (\Gamma\tilde{\mathcal{A}}(x), \Gamma\tilde{\mathcal{B}}(x)) \text{ \{independent calls\}} \end{aligned}$$

These rules can be used recursively. For instance, the species  $\mathcal{T}$  of rooted binary trees satisfies

$$\mathcal{T} = X + \mathcal{T} \cdot \mathcal{T}, \tag{66}$$

which translates to the following Boltzmann sampler:

$$\Gamma\tilde{\mathcal{T}}(x): \text{ if Bern}\left(\frac{x}{T(x)}\right) \text{ return leaf else return } (\Gamma\tilde{\mathcal{T}}(x), \text{node}, \Gamma\tilde{\mathcal{T}}(x)).$$

5.1.2. *The complexity model.* Typically, when  $\mathcal{F}$  has a recursive specification over the species  $\mathcal{A}_1, \dots, \mathcal{A}_l$ , then our sampling procedure  $\Gamma\tilde{\mathcal{F}}(x)$  will require that we can evaluate the ordinary generating functions for  $\mathcal{A}_1, \dots, \mathcal{A}_l$  at real values  $x$ . Indeed, for a species  $\mathcal{F}$  defined as  $\mathcal{A}_1 + \mathcal{A}_2$ , each Bernoulli choice requires to draw a uniform value in  $[0, 1]$  and compare it with a ratio of the form  $\tilde{\mathcal{A}}_1(x)/\tilde{\mathcal{F}}(x)$ . In the following we work with the complexity model where we assume that there exists an *oracle* that provides at unit cost the exact values of these generating functions at  $x$ , and that a random number in  $[0, 1]$  can be generated and compared with a fixed value such as  $\tilde{\mathcal{A}}(x)/\tilde{\mathcal{C}}(x)$  in constant time as well. We will refer to this complexity model as the *real-arithmetic complexity model* in the following.

The model is justified since in many applications we obtain expressions for the ordinary generating series that allow a rapid numeric evaluation of those series at given values, for example with the Newton method. Then, in practice, one works at a fixed precision, say  $N$  bits (typically  $N = 64$ , correspondingly roughly to 20 decimal digits).

Let us mention that the Boltzmann samplers for the constructions Multiset and Cycle—as given in [13] and recovered in a more general framework here—require typically the values of the generating functions not only at  $x$ , but at all powers  $x^i$ . Since combinatorial species often have exponential growth rate (which is the case for all examples presented here), the dominant singularity  $\rho$  satisfies  $\rho < 1$ , hence  $x \leq \rho < 1$ . Therefore the values  $\tilde{\mathcal{A}}(x^i)$  decrease exponentially fast with  $i$ . When working at fixed precision of  $N$  bits, one can thus discard the powers greater than  $k = N/\log_2(1/\rho)$  and assume that the oracle provides the evaluations of the generating functions at  $x, x^2, \dots, x^k$ . For a more detailed study and implementation of the evaluation procedures we refer to the recent article by Pivoteau, Salvy, and Soria [27].

**Proposition 21** (Duchon et al. [9]). *Let  $\mathcal{F}$  be a species that can be decomposed recursively from  $\{\mathbf{1}, X\}$  in terms of the constructions  $\{+, \cdot\}$  (this is meant analogous to but more restricted than Definition 2.7). Then one can obtain in a systematic way (from the recursive specification) an ordinary Boltzmann sampler  $\Gamma\tilde{\mathcal{F}}(x)$  for  $\mathcal{F}$ . In addition, in the real-arithmetic complexity model,  $\Gamma\tilde{\mathcal{F}}(x)$  operates in linear time in the size of the output.*

This result was recently extended in [13] to other constructions, such as the Multiset and Cycle constructions (and their counterpart with fixed number of components). In this section we extend this result to the substitution construction, and to the cycle-pointed constructions. It follows that any species that is cycle-pointed decomposable over species where we already have a Pólya-Boltzmann sampler also has a Pólya-Boltzmann sampler.

*Remark 5.1.* Note that in the general results on sampling we consider species *up to isomorphism*. Theoretically, this is a necessary assumption, since isomorphisms between species might in artificial examples be non-effective. However, in all the presented examples and applications in this article, the isomorphisms between species are straightforward and efficiently computable. The effectiveness of isomorphisms between the output species of the sampling procedures and the actual species are actually more related to the question how combinatorial structures are represented on a computer, and in particular they do not concern the complexity of the sampling task itself.

5.1.3. *Targeting Boltzmann samplers.* Boltzmann samplers often lead to very efficient *exact-size* and *approximate-size* random samplers. In order to draw unlabeled structures uniformly at random from a species  $\mathcal{A}$  at (in case of exact-size sampling) or around (in case of approximate-size sampling) a *target-size*  $n$ , one simply repeats calling the Boltzmann sampler  $\Gamma\tilde{\mathcal{A}}(x)$  – with a suitably chosen value of  $x$  – until the size of the output is  $n$  (exact-size sampling) or is in  $[n(1 - \epsilon), n(1 + \epsilon)]$  (approximate-size sampling), where  $\epsilon$  is a *tolerance-parameter* fixed by the user. It turns out that for a wide class of species, which covers the species encountered in Section 4 (trees, cacti graphs, outerplanar graphs), this method works very well, as proved in [9].

**Proposition 22** (Duchon et al [9]). *Let  $\mathcal{F}$  be a species such that asymptotically*

$$|\tilde{\mathcal{F}}_n| \sim c n^{-3/2} \rho^{-n}$$

for some positive constant  $c$ , and where  $\rho$  is the radius of convergence of  $\tilde{\mathcal{F}}(x)$ , assuming that  $\tilde{\mathcal{F}}(\rho)$  is convergent<sup>5</sup>. Also suppose that there is a Boltzmann sampler  $\Gamma\tilde{\mathcal{F}}(\rho)$  at  $x = \rho$  such that the cost of generating a structure is linearly bounded by the size of the structure all along the generation process. Then  $\Gamma\tilde{\mathcal{F}}(\rho)$  yields an exact-size (approximate-size, resp.) sampler for unlabeled structures from  $\mathcal{F}$  with expected complexity  $O(n^2)$  ( $O(n/\epsilon)$ , resp.), where  $n$  is a target-size and  $\epsilon$  is a tolerance-ratio.

The exact-size and approximate-size samplers are obtained by running  $\Gamma\tilde{\mathcal{F}}(\rho)$  until the size of the output is in the target domain  $\Omega_n$  (that is,  $\Omega_n = \{n\}$  for exact-size sampling, and  $\Omega_n = [n(1 - \epsilon), n(1 + \epsilon)]$  for approximate-size sampling). To obtain the stated complexity, it is necessary that the generation of too large structures is aborted as soon as the size of the generated object gets larger than  $\text{Max}(\Omega_n)$ .

5.2. **Pólya-Boltzmann Samplers for classical species.** Let  $\mathcal{A}$  be a species. Recall that a *symmetry* on a species  $\mathcal{A}$  is a pair  $(A, \sigma)$  where  $A \in \mathcal{A}$  and  $\sigma$  is an automorphism of  $A$ . A symmetry has a weight-monomial  $w_{(A, \sigma)}$ , as defined in (3); and the cycle index sum  $Z_{\mathcal{A}}(s_1, s_2, \dots)$  is the sum of the weight-monomials over all the symmetries on  $\mathcal{A}$ . Similarly as for the one-variable case, a vector  $(s_i)_{i \geq 1}$  of nonnegative real values is said to be *admissible* if the sum defining  $Z_{\mathcal{A}}(s_1, s_2, \dots)$  converges. Given an admissible vector  $(s_i)_{i \geq 1}$ , a *Pólya-Boltzmann sampler* is a procedure  $\Gamma Z_{\mathcal{A}}(s_1, s_2, \dots)$  that randomly samples symmetries on  $\mathcal{A}$  such that each symmetry  $(A, \sigma)$  is drawn with probability

$$\mathbb{P}_{(s_1, s_2, \dots)}(A, \sigma) = \frac{w_{(A, \sigma)}}{Z_{\mathcal{A}}(s_1, s_2, \dots)}, \quad (67)$$

where the weight-monomial  $w_{(A, \sigma)}$  is evaluated at  $(s_1, s_2, \dots)$ . This probability distribution is called the *Pólya-Boltzmann distribution* for  $\mathcal{A}$  at  $(s_i)_{i \geq 1}$ . The following simple lemma ensures that Pólya-Boltzmann samplers are a refinement of ordinary Boltzmann samplers, in the same way as cycle index sums are a refinement of ordinary generating functions.

**Lemma 23** (Pólya-Boltzmann samplers extend ordinary Boltzmann samplers). *Consider a species  $\mathcal{A}$  having a Pólya-Boltzmann sampler  $\Gamma Z_{\mathcal{A}}(s_i)_{i \geq 1}$ . Then, for any value  $x$  admissible for  $\tilde{\mathcal{A}}(x)$ , the sampler  $\Gamma Z_{\mathcal{A}}(x, x^2, x^3, \dots)$  is an ordinary Boltzmann sampler for unlabeled structures from  $\mathcal{A}$  at  $x$ .*

*Proof.* The generator  $\Gamma Z_{\mathcal{A}}(x, x^2, \dots)$  gives weight  $x^n / (n! Z_{\mathcal{A}}(x, x^2, \dots))$  to each symmetry of size  $n$ . Since  $Z_{\mathcal{A}}(x, x^2, \dots) = \tilde{\mathcal{A}}(x)$  by Lemma 1, this weight simplifies to  $x^n / (n! \tilde{\mathcal{A}}(x))$ . In addition, we have seen in Lemma 1 that each unlabeled structure  $\gamma \in \tilde{\mathcal{A}}_n$  gives rise to  $n!$  symmetries. Hence, each unlabeled structure of size  $n$  has weight  $x^n / \tilde{\mathcal{A}}(x)$  when calling  $\Gamma Z_{\mathcal{A}}(x, x^2, \dots)$ , i.e.,  $\Gamma Z_{\mathcal{A}}(x, x^2, \dots)$  is an ordinary Boltzmann sampler for unlabeled structures from  $\mathcal{A}$ .  $\square$

In the next two sections, we describe Pólya-Boltzmann samplers for unlabeled structures from basic species and for the constructions  $\{+, \cdot, \circ\}$ . Note that the output of a Pólya-Boltzmann sampler for a species  $\mathcal{A}$  consists of a species from  $\mathcal{A}[n]$  together with an automorphism on that structure. In all the random generators to be described (as well as in the procedures for cycle-pointed

<sup>5</sup>The asymptotic behaviour  $c\rho^{-n}n^{-3/2}$  is called *universal* [1], as it is widely encountered in combinatorics.

species), the resulting structure  $S$  is made well-labeled by applying a procedure `DISTRIBUTE LABELS` that substitutes  $[1, \dots, |S|]$  for the atoms of  $S$  uniformly at random (i.e., according to a permutation of size  $|S|$  taken uniformly at random).

We also assume to have generators for classical distributions:

- $\text{Geom}(p)$  returns an integer under the geometric law of parameter  $p \in [0, 1]$ :  $\Pr(k) = p^{k-1}(1-p)$ ;
- $\text{Pois}(\lambda)$  returns an integer under the Poisson law of parameter  $\lambda$ :  $\Pr(k) = e^{-\lambda}\lambda^k/k!$ ; and
- $\text{Loga}(\lambda)$  returns an integer under the distribution  $\Pr(k) = (\log(1/(1-\lambda)))^{-1}\lambda^k/k$ ; this distribution we call *Loga law of parameter  $\lambda$* .

The generators for those distributions (more generally, any generator for an explicit distribution on integers) can be easily obtained from the “inversion method” [8, §2.1] and [21, §4.1]).

**5.2.1. Pólya-Boltzmann samplers for basic species.** At first, let us describe Pólya-Boltzmann samplers for the basic species `SEQ`, `SET`, `CYC`, and their counterparts `SEQ[k]`, `SET[k]`, `CYC[k]`. In each case, the design of the sampler is guided by the expressions of the cycle index sums for the basic species as given in Figure 1.

**Proposition 24.** *The random generators shown in Figure 9 are Pólya-Boltzmann samplers for the corresponding species.*

*Proof.* For `SEQ`, the proof is easy. Since  $Z_{\text{SET}} = \sum_{k \geq 1} s_1^k$ , the probability of a sequence to have size  $k$  must be  $s_1^k/Z_{\text{SET}}$ , i.e., the size distribution is a geometric law of parameter  $s_1$ .

For `SET`, observe that the sum of weight-monomials over all symmetries of type  $(n_1, n_2, \dots, n_k)$  in  $Z_{\text{SET}}$  is

$$\frac{s_1^{n_1}}{n_1!} \frac{(s_2/2)^{n_2}}{n_2!} \dots \frac{(s_k/k)^{n_k}}{n_k!}.$$

Therefore, a Pólya-Boltzmann sampler has to draw a collection of cycles such that the number  $n_i$  of cycles of length  $i$  follows a Poisson law of parameter  $s_i/i$  for  $i \geq 1$ , and the  $n_i$ 's are independent. This is precisely what the algorithm  $\Gamma Z_{\text{SET}}$  in Figure 9 does, upon choosing a priori the size of the largest cycle to be drawn.

For `CYC`, the argument is similar. As we have seen in Section 2.6.1, the sum of the weight-monomials over all the symmetries of order  $r$  is

$$Z_{\text{CYC}}^{(r)} := \frac{\phi(r)}{r} \log \left( \frac{1}{1-s_r} \right).$$

Therefore the order of the automorphism has to be chosen with probability  $Z_{\text{CYC}}^{(r)}/Z_{\text{CYC}}$ . In addition, for each fixed order  $r$ , the probability of the cycle being  $r \times k$  has to be  $s_r^k/k/\log(1/(1-s_r))$ , i.e., the size (divided by  $r$ ) has to follow a Loga law of parameter  $s_r$ . Finally, for all automorphisms of size  $r$  and size  $r \times k$ , all possible ‘rotation angles’ (there are  $\phi(r)$  possibilities) have to be equiprobable. This is exactly what the generator  $\Gamma Z_{\text{CYC}}$  does.

The proof that the generators for the species `SEQ[k]`, `SET[k]`, and `CYC[k]` are Pólya-Boltzmann samplers follows similar arguments upon restricting to structures of size  $k$ .  $\square$

**5.2.2. Pólya-Boltzmann samplers for combinatorial constructions.** As shown in Figure 10, Pólya-Boltzmann samplers make it possible to have a simple sampling rule for each of the standard constructions; that is, sampling rules not only for sum and product, but also for the substitution construction.

**Proposition 25.** *Let  $\mathcal{C} = \mathcal{A} \wedge \mathcal{B}$ , with  $\wedge \in \{+, \cdot, \circ\}$ . When there are Pólya-Boltzmann samplers for  $\mathcal{A}$  and  $\mathcal{B}$ , then there is also a Pólya-Boltzmann sampler  $\Gamma Z_{\mathcal{C}}(s_1, s_2, \dots)$  for  $\mathcal{C}$  that can be constructed from the samplers for  $\mathcal{A}$  and  $\mathcal{B}$ , as given in Figure 10.*

*Proof.* For  $\mathcal{C} = \mathcal{A} + \mathcal{B}$ , the proof is easy. Note that  $\text{Sym}(\mathcal{A} + \mathcal{B}) = \text{Sym}(\mathcal{A}) + \text{Sym}(\mathcal{B})$ . As shown in [9] (for the standard weight  $x^n/n!$ ), a disjoint union yields a Bernoulli switch on Boltzmann samplers, with probability corresponding to the ratio of the series for  $\mathcal{A}$  divided by the series for  $\mathcal{C}$  (this argument works as well here, where we take the refined weight  $s_1^{n_1} s_2^{n_2} \dots s_k^{n_k}/n!$ ). Therefore the probability of the Bernoulli switch has to be  $Z_{\mathcal{A}}/Z_{\mathcal{C}}$ .

<p><b>(1) Sequence.</b></p> <p><b>Algorithm</b> <math>\Gamma Z_{\text{SEQ}}(s_1, s_2, \dots)</math>, with <math>s_1 &lt; 1</math>:  <math>k \leftarrow \text{Geom}(s_1)</math>;  <b>return</b> a sequence of <math>k</math> atoms (endowed with the identity-automorphism).</p>
<p><b>(2) Set.</b></p> <p>Define the probability distribution relative to <math>(s_i)_{i \geq 1}</math>:  <math display="block">\Pr(K \leq k) = \frac{1}{Z_{\text{SET}}(s_1, s_2, \dots)} \prod_{j \leq k} \exp\left(\frac{1}{j} s_j\right).</math> Let <math>\text{MAX\_INDEX}(s_1, s_2, \dots)</math> be a generator for this distribution.</p> <p><b>Algorithm</b> <math>\Gamma Z_{\text{SET}}(s_1, s_2, \dots)</math> :  <math>J \leftarrow \text{MAX\_INDEX}(s_1, s_2, \dots)</math>;  <b>for</b> <math>j</math> <b>from</b> 1 <b>to</b> <math>J - 1</math> <b>do</b> <math>k_j \leftarrow \text{Pois}\left(\frac{s_j}{j}\right)</math> <b>end for</b>  <math>k_J \leftarrow \text{Pois}_{\geq 1}\left(\frac{s_J}{J}\right)</math> {Poisson conditioned to output a strictly positive integer}  <b>return</b> a collection of cycles of atoms where there are <math>k_j</math> cycles in each length <math>j &gt; 0</math>.</p>
<p><b>(3) Cycle.</b></p> <p>Given <math>(s_i)_{i \geq 1}</math> such that <math>Z_{\text{CYC}}(s_1, s_2, \dots)</math> converges, consider the probability distribution  <math display="block">\Pr(R = r) = \frac{1}{Z_{\text{CYC}}(s_1, s_2, \dots)} \frac{\varphi(r)}{r} \log(1/(1 - s_r))</math> for <math>r \geq 1</math>.  Let <math>\text{REPLICORDER}(s_1, s_2, \dots)</math> be a generator of this distribution.</p> <p><b>Algorithm</b> <math>\Gamma Z_{\text{CYC}}(s_1, s_2, \dots)</math>  <math>r \leftarrow \text{REPLICORDER}(s_1, s_2, \dots)</math>;  <math>j \leftarrow \text{Loga}(s_r)</math>;  Draw an integer <math>b \in [1..r - 1]</math> that is relatively prime to <math>r</math> uniformly at random ;  <b>return</b> the cycle of length <math>j \times r</math> endowed with the automorphism:  “each atom is mapped to the atom that is <math>j \times b</math> units further on the cycle”.</p>
<p><b>(1') Sequence of size <math>k</math>.</b></p> <p><b>Algorithm</b> <math>\Gamma Z_{\text{SEQ}^{[k]}}(s_1, s_2, \dots)</math> :  <b>return</b> a sequence of <math>k</math> atoms (endowed with the identity-automorphism).</p>
<p><b>(2') Set of size <math>k</math>.</b></p> <p><b>Algorithm</b> <math>\Gamma Z_{\text{SET}^{[k]}}(s_1, s_2, \dots)</math> :  Draw a partition sequence <math>\pi</math> of order <math>k</math> such that <math>\Pr(\pi) = \frac{s_1^{n_1} s_2^{n_2} \dots s_k^{n_k} \cdot [s_1^{n_1} \dots s_k^{n_k}] Z_{\text{SET}^{[k]}}}{Z_{\text{SET}^{[k]}}(s_1, s_2, \dots)}</math>  <b>return</b> a collection of <math>n_1</math> cycles of length 1, <math>n_2</math> cycles of length 2, ..., <math>n_k</math> cycles of length <math>k</math>.</p>
<p><b>(3') Cycle of size <math>k</math>.</b></p> <p><b>Algorithm</b> <math>\Gamma Z_{\text{CYC}^{[k]}}(s_1, s_2, \dots)</math> :  Draw a divisor <math>r</math> of <math>k</math> with distribution <math>\Pr(r) = \frac{\phi(r) s_r^{k/r}}{k Z_{\text{CYC}^{[k]}}(s_1, s_2, \dots)}</math>  Draw an integer <math>b \in [1..r - 1]</math> that is relatively prime to <math>r</math> uniformly at random;  <b>return</b> the cycle of length <math>k</math> endowed with the automorphism:  “each atom is mapped to the atom that is <math>kb/r</math> units further on the cycle”.</p>

FIGURE 9. Pólya-Boltzmann samplers for basic species. In all these random samplers, the output structure is made well-labeled using the procedure `DISTRIBUTELABELS`.

<p><b>Sum:</b> <math>\mathcal{C} = \mathcal{A} + \mathcal{B}</math>.  Given <math>(s_i)_{i \geq 1}</math> such that <math>Z_a := Z_{\mathcal{A}}(s_1, s_2, \dots)</math> and <math>Z_b := Z_{\mathcal{B}}(s_1, s_2, \dots)</math> converge:</p> <p style="margin-left: 20px;"><b>Algorithm</b> <math>\Gamma Z_{\mathcal{C}}(s_1, s_2, \dots)</math> :</p> <p style="margin-left: 20px;"><b>if</b> <math>\text{Bern}(Z_a/(Z_a + Z_b))</math> <b>then return</b> <math>\Gamma Z_{\mathcal{A}}(s_1, s_2, \dots)</math></p> <p style="margin-left: 20px;"><b>else return</b> <math>\Gamma Z_{\mathcal{B}}(s_1, s_2, \dots)</math> <b>end if</b></p>
<p><b>Product:</b> <math>\mathcal{C} = \mathcal{A} \cdot \mathcal{B}</math>.  Given <math>(s_i)_{i \geq 1}</math> such that <math>Z_{\mathcal{A}}(s_1, s_2, \dots)</math> and <math>Z_{\mathcal{B}}(s_1, s_2, \dots)</math> converge:</p> <p style="margin-left: 20px;"><b>Algorithm</b> <math>\Gamma Z_{\mathcal{C}}(s_1, s_2, \dots)</math> :</p> <p style="margin-left: 20px;"><b>return</b> <math>(\Gamma Z_{\mathcal{A}}(s_1, s_2, \dots), \Gamma Z_{\mathcal{B}}(s_1, s_2, \dots))</math> {independent calls}</p>
<p><b>Substitution:</b> <math>\mathcal{C} = \mathcal{A} \circ \mathcal{B}</math>.  Given <math>(s_i)_{i \geq 1}</math> such that <math>b_i := Z_{\mathcal{B}}(s_i, s_{2i}, s_{3i}, \dots)</math> converges for each <math>i \geq 1</math>, and <math>Z_{\mathcal{A}}(b_1, b_2, b_3, \dots)</math> converges:</p> <p style="margin-left: 20px;"><b>Algorithm</b> <math>\Gamma Z_{\mathcal{C}}(s_1, s_2, \dots)</math> :</p> <p style="margin-left: 40px;">Compute <math>(A, \sigma_A) \leftarrow \Gamma Z_{\mathcal{A}}(b_1, b_2, b_3, \dots)</math>;</p> <p style="margin-left: 40px;"><b>for each</b> cycle <math>C = (u_1, \dots, u_k)</math> of <math>\sigma_A</math> (<math>u_1</math> has smallest label in <math>C</math>) <b>do</b></p> <p style="margin-left: 60px;">Compute <math>(B, \sigma_B) \leftarrow \Gamma Z_{\mathcal{B}}(s_k, s_{2k}, \dots)</math>;</p> <p style="margin-left: 60px;">Replace each atom of <math>C</math> by a copy of <math>B</math>;</p> <p style="margin-left: 40px;"><b>for each</b> cycle <math>D</math> of <math>\sigma_B</math> <b>do</b></p> <p style="margin-left: 60px;">Let <math>E_D</math> be the cycle composed of the copies of <math>D</math> at <math>u_1, \dots, u_k</math></p> <p style="margin-left: 40px;"><b>end for</b></p> <p style="margin-left: 20px;"><b>end for</b></p> <p style="margin-left: 20px;"><b>return</b> the resulting structure and the automorphism consisting of the composed cycles <math>E_D</math></p>

FIGURE 10. The rules to specify a Pólya-Boltzmann sampler for a species that has a recursive specification. In all these random samplers, the finally returned structure is made well-labeled using the procedure DISTRIBUTE LABELS.

For product,  $\mathcal{C} = \mathcal{A} \cdot \mathcal{B}$ ,  $\text{Sym}(\mathcal{C})$  is like a partitionial product of  $\text{Sym}(\mathcal{A})$  and  $\text{Sym}(\mathcal{B})$ . Therefore, a Boltzmann sampler classically consists of two independent calls to Boltzmann samplers, as shown in [9] (again, for the standard weight  $x^n/n!$ ). All the arguments work the same way for the refined weight  $s_1^{n_1} s_2^{n_2} \dots s_k^{n_k}/n!$ . Therefore, one has to call independently a Pólya-Boltzmann sampler for  $\mathcal{A}$  and a Pólya-Boltzmann sampler for  $\mathcal{B}$ .

For substitution,  $\mathcal{C} = \mathcal{A} \circ \mathcal{B}$ , recall (from Section 2.6.2, Equation (13)) that for each partition sequence  $\pi$ , the sum of the weight-monomials over all the symmetries on  $\mathcal{C}$  of type  $\pi = (n_1, \dots, n_n)$  satisfies the expression

$$Z_{\mathcal{C}}^{(\pi)} = \frac{a_{\pi}}{n!} b_1^{n_1} b_2^{n_2} \dots b_k^{n_k}, \quad \text{where } a_{\pi} = n! [s_1^{n_1} \dots s_n^{n_n}] Z_{\mathcal{A}}(s_1, s_2, \dots) \text{ and } b_i = Z_{\mathcal{B}}(s_i, s_{2i}, \dots).$$

Hence, a Pólya-Boltzmann sampler for  $\mathcal{C}$  must draw the core structure following the Pólya-Boltzmann distribution for  $\mathcal{A}$  with parameters  $(b_1, b_2, \dots)$ . In addition, as discussed in Section 2.6.2, once the type  $\pi$  of the core symmetry is fixed, the structures substituted at the cycles of the core-automorphism form a partitionial product of the form

$$\text{Sym}(\mathcal{B})^{n_1} \cdot \{\text{Sym}(\mathcal{B}) \text{ duplicated}\}^{n_2} \cdot (\dots) \cdot \{\text{Sym}(\mathcal{B}) \text{ replicated } k \text{ times}\}^{n_k}.$$

Recall that a partitionial product yields independent Boltzmann samplers. Hence, once the core-automorphism  $(A, \sigma_A)$  is drawn, the symmetries in  $\mathcal{B}$  that are substituted at each cycle of  $\sigma_A$  must be independent calls of a Pólya-Boltzmann sampler for  $\mathcal{B}$ , and the parameters of the sampler must

be  $(s_i, s_{2i}, \dots)$  if the replication order is  $i$  ( $i$  is also the length of the cycle of the core structure where the substitution occurs). This is precisely what the generator  $\Gamma Z_e$  does.  $\square$

5.2.3. *Pólya-Boltzmann samplers for decomposable species.* The random generation rules shown in Figure 9 for basic species and Figure 10 for general constructions can be combined to design a Pólya-Boltzmann sampler for species that are decomposable over those basic species.

**Definition 5.1.** A Pólya-Boltzmann sampler  $\Gamma Z_{\mathcal{A}}(s_1, s_2, \dots)$  is called *linear* if, in the real-arithmetic complexity model described above, the number of computation steps for generating a structure from  $\mathcal{A}$  is linearly bounded by the size of the structures all along the generation process.

**Theorem 26.** *Let  $\mathcal{A}$  be a species that is decomposable over species  $\mathcal{A}_1, \dots, \mathcal{A}_l$  (Definition 2.7), each having a linear Pólya-Boltzmann sampler. Then also the species  $\mathcal{A}$  has a linear Pólya-Boltzmann sampler  $\Gamma Z_{\mathcal{A}}(s_1, s_2, \dots)$ .*

*Proof.* Given Proposition 25, the Pólya-Boltzmann sampler for  $\mathcal{A}$  is straightforward by ‘unrolling’ the recursive specifications that show that  $\mathcal{A}$  is decomposable; so we are left with the task to show that we also have a *linear* Pólya-Boltzmann sampler for  $\mathcal{A}$ .

Our argument is based on the following concept. The *decomposition tree* for structures  $A$  from  $\mathcal{A}$  is defined as follows. The leaves of this tree are structures from the species  $\mathcal{A}_1, \dots, \mathcal{A}_l$ . The other vertices of this tree are structures from the species  $\mathcal{X}_j^{(i)}$  that appear in the definition of recursive specifications (Definition 2.6). Suppose that  $\mathcal{A}$  is specified by variable  $x_s$  in the recursive specification, and that  $n$  is the size of  $A$ . The root of the decomposition tree is  $A$  itself, which is a structure from  $\mathcal{X}_s^{(n)}$ . To define the children of a vertex  $A'$  from  $\mathcal{X}_j^{(i)}$  in the decomposition tree, consider the species  $\mathcal{X}_k^{(i-1)}$  for which  $x_k$  appears in the expression  $e_j$  in the recursive specification. The children are the structures of those species from which  $A$  has been constructed. The *size* of a decomposition tree is the number of its internal (i.e., non-leaf) vertices.

Let  $A$  be a structure from  $\mathcal{A}$ , and let  $A_1, \dots, A_t$  be the leafs of the decomposition tree for  $A$ . Note that the cost of the sampling procedure for producing  $A$  is linearly bounded by the size of the decomposition tree for  $A$  plus the total cost for sampling  $A_1, \dots, A_t$ . The size of  $A$  is distributed over all structures  $A_1, \dots, A_t$  from terminal classes, consequently  $|A| = |A_1| + \dots + |A_t|$ . Since the samplers for the terminal classes are linear, we conclude that the total cost for sampling  $A_1, \dots, A_t$  is linear in  $|A|$ .

We finally show that the size of the decomposition tree for  $A$  is linear in the size  $n$  of  $A$ . More specifically, when  $m$  is the number of variables in the recursive specification of  $\mathcal{A}$ , then the size of the decomposition tree is bounded by  $2mn$ . Otherwise, there would be a branch in the decomposition tree that contains two vertices  $B_1$  from  $\mathcal{X}_j^{(n_1)}$  and  $B_2$  from  $\mathcal{X}_j^{(n_2)}$  (corresponding to the same variable  $x_j$ ) such that  $B_1$  and  $B_2$  have the same size  $n$ . But then, by looping through the corresponding cycle in the decomposition grammar, we can produce infinitely many distinct structures of size  $n$  from the recursive specification, contradicting the fact that  $\mathcal{X}_j[n]$  is finite. The details of this argument are easy and left to the reader.  $\square$

Consequently, the rooted species we have encountered (i.e., with a single marked atom) in Section 4 can be endowed with efficient random samplers.

**Proposition 27.** *In the real-arithmetic complexity model (oracle assumption), the following species admit an exact-size sampler of expected complexity  $O(n^2)$  ( $n$  being the target-size) and an approximate-size sampler of expected complexity  $O(n/\epsilon)$  ( $\epsilon$  being the tolerance-ratio) for the uniform distribution on unlabeled structures: rooted nonplane (plane, resp.) trees, rooted nonplane (plane, resp.) trees whose node degrees are constrained to lie in a finite integer set  $\Omega$ , rooted cacti graphs, rooted connected outerplanar graphs.*

*Proof.* All the rooted tree species stated here are decomposable over the discussed basic species, as we have seen in Section 4. (For instance, the species  $\mathcal{F}'$  of rooted nonplane trees satisfies  $\mathcal{F}' = \text{SET} \circ (X \cdot \mathcal{F}')$ .) Hence, by Theorem 26, the species of rooted trees can be endowed with a linear Pólya-Boltzmann sampler. A species  $\mathcal{G}'$  of rooted connected graphs – provided that it is

closed under taking 2-connected components – is decomposable in terms of the subspecies  $\mathcal{B}'$  of rooted 2-connected graphs:

$$\mathcal{G}' = \text{SET} \circ (\mathcal{B}' \circ (X \cdot \mathcal{G}')).$$

As we have explained in Sections 4.2.1 and 4.2.2, there is a decomposition strategy for rooted 2-connected cacti graphs (rooted polygons) and rooted 2-connected outerplanar graphs (rooted dissections of a polygon). The linear Pólya-Boltzmann sampler  $\Gamma Z_{\mathcal{B}'}$  yields in turn a linear Pólya-Boltzmann sampler for  $\mathcal{G}'$  (using the specification of  $\mathcal{G}'$  in terms of  $\mathcal{B}'$  stated above). Hence, each of the rooted species stated above has a linear Pólya-Boltzmann sampler, which becomes a linear ordinary Boltzmann sampler when specialized to  $s_i = x^i$ .

Moreover, all these rooted species obey the universal asymptotic form  $c \rho^{-n} n^{-3/2}$ , as shown in [26] for trees, in [31] for cacti graphs, and in [3] for outerplanar graphs. Hence, by Proposition 22, a Boltzmann sampler run at the dominant singularity yields an exact-size (approximate-size, respectively) sampler with expected complexity  $O(n^2)$  ( $O(n/\epsilon)$ , respectively).  $\square$

### 5.3. Pólya-Boltzmann samplers for cycle-pointed species.

5.3.1. *Definition.* Given a cycle-pointed species  $\mathcal{P}$ , a vector  $(s_i, t_i)_{i \geq 1}$  of nonnegative real values is said to be *admissible* if the sum of weight-monomials defining  $Z_{\mathcal{P}}$  converges when evaluated at this vector. Given a fixed admissible vector  $(s_i, t_i)_{i \geq 1}$ , a *Pólya-Boltzmann sampler* is a procedure  $\Gamma \overline{Z}_{\mathcal{P}}(s_i, t_i)_{i \geq 1}$  that generates a rooted  $c$ -symmetry on  $\mathcal{P}$  at random such that each rooted  $c$ -symmetry  $(P, \sigma, v)$  of  $\mathcal{R}(\mathcal{P})$  is drawn with probability

$$\mathbb{P}(P, \sigma, v) = \frac{w_{(P, \sigma, v)}}{\overline{Z}_{\mathcal{P}}((s_i, t_i)_{i \geq 1})}, \quad (68)$$

with  $w_{(P, \sigma, v)}$  as defined in (15). This probability distribution is called the Pólya-Boltzmann distribution for  $\mathcal{P}$  at  $(s_i, t_i)_{i \geq 1}$ . Similarly as for classical species, the procedure of calling  $\Gamma Z_{\mathcal{P}}(x^i, x^i)_{i \geq 1}$  (where  $x$  is admissible for the ordinary generating function  $\hat{\mathcal{P}}(x)$ ) and then returning the underlying unlabeled structure yields an ordinary Boltzmann sampler  $\Gamma \mathcal{P}(x)$ . The following sampling rules make it possible to systematically assemble Pólya-Boltzmann samplers for cycle-pointed species.

#### 5.3.2. Pólya-Boltzmann samplers for basic cycle-pointed species.

**Proposition 28.** *The random generators shown in Figure 11 are Pólya-Boltzmann samplers for the corresponding basic cycle-pointed species.*

*Proof.* The arguments are very similar to the ones in the proof of Proposition 24. Observe that  $\text{SEQ}^\circ = (X^\circ \star \text{SEQ}) \star \text{SEQ}$ . The marked atom (the atom bearing the marked cycle, which has length 1 here) must be preceded by a sequence of  $k_1$  atoms and followed by a sequence of  $k_2$  atoms such that  $k_1$  and  $k_2$  follow independently a geometric law of parameter  $s_1$ .

Next, we have  $\text{SET}^\circ_{(\ell)} = \mathcal{C}_{(\ell)} \star \text{SET}$  where  $\mathcal{C}_{(\ell)}$  is the cycle-pointed species of cycles of  $\ell$  atoms (the cycle being marked), which explains the samplers  $\Gamma \overline{Z}_{\text{SET}^\circ}$  and  $\Gamma \overline{Z}_{\text{SET}^\circ}$  in Figure 11.

A cycle-pointed structure in  $\text{CYC}_\ell^\circ$  consists of  $\ell$  isomorphic copies (attached cyclically in a chain) of an object in  $X^\circ \star \text{SEQ}$ . Additionally, one needs to specify the shift of the automorphism; if the cycle has length  $n\ell$ , the possible shifts are  $n \cdot i$  where  $i \in [1.. \ell]$  is relatively prime to  $\ell$ , hence there are  $\phi(\ell)$  possibilities for the shift.

The proofs for the samplers with  $k$  components follow similar arguments.  $\square$

#### 5.3.3. Pólya-Boltzmann samplers for cycle-pointed constructions.

**Proposition 29.** *Let  $\mathcal{R}$  be a species with a recursive specification over other species having a Pólya-Boltzmann sampler. Then the random sampler  $\Gamma \overline{Z}_{\mathcal{R}}(s_1, t_1; s_2, t_2; \dots)$ , as given in Figure 12, is a Pólya-Boltzmann sampler for  $\mathcal{R}$ .*

*Proof.* The arguments are very similar to the ones in the proof of Proposition 25. For the cycle-pointed sum,  $\mathcal{R} = \mathcal{P} + \mathcal{Q}$ , we have  $\text{RSym}(\mathcal{R}) = \text{RSym}(\mathcal{P}) + \text{RSym}(\mathcal{Q})$ . Therefore the Pólya-Boltzmann sampler has to be a Bernoulli switch with probability  $\overline{Z}_{\mathcal{P}}/\overline{Z}_{\mathcal{R}}$  followed by a call to the Pólya-Boltzmann sampler of either  $\mathcal{P}$  or  $\mathcal{Q}$  (depending on the Bernoulli output to be “true” or

<p><b>(1) Cycle-pointed sequence.</b></p> <p><b>Algorithm</b> <math>\Gamma \overline{Z}_{\text{SEQ}^\circ}(s_1, s_2, \dots)</math> :</p> <p style="margin-left: 20px;"><math>k_1 \leftarrow \text{Geom}(s_1); k_2 \leftarrow \text{Geom}(s_1); \{\text{indep. calls}\}</math></p> <p><b>return</b> a sequence of <math>k_1 + k_2 + 1</math> atoms (endowed with the identity-automorphism) where the atom at position <math>k_1 + 1</math> is marked.</p>
<p><b>(2) Cycle-pointed (symmetric cycle-pointed, resp.) set.</b></p> <p>Given <math>(s_i, t_i)_{i \geq 1}</math> such that <math>\sum_{i \geq 1} t_i</math> converges, define the distribution:</p> $\Pr(K) = \frac{t_K}{\sum_{i \geq 1} t_i} \text{ for } K \geq 1 \quad (\Pr(K) = \frac{t_K}{\sum_{i \geq 2} t_i} \text{ for } K \geq 2, \text{ resp.}).$ <p>Let <math>\text{ROOT\_CYCLE\_SIZE}(t_1, t_2, \dots)</math> (<math>\text{ROOT\_CYCLE\_SIZE}_{\geq 2}(t_1, t_2, \dots)</math>, resp.) be a generator for this distribution.</p> <p><b>Algorithm</b> <math>\Gamma \overline{Z}_{\text{SET}^\circ}(s_1, t_1; s_2, t_2; \dots)</math> (<math>\Gamma \overline{Z}_{\text{SET}^\circ}(s_1, t_1; s_2, t_2; \dots)</math>, resp.) :</p> <p style="margin-left: 20px;"><math>K \leftarrow \text{ROOT\_CYCLE\_SIZE}(t_1, t_2, \dots)</math> (<math>K \leftarrow \text{ROOT\_CYCLE\_SIZE}_{\geq 2}(t_1, t_2, \dots)</math>, resp.);</p> <p style="margin-left: 20px;"><math>\gamma \leftarrow \Gamma Z_{\text{SET}}(s_1, s_2, \dots)</math>;</p> <p style="margin-left: 20px;">Add to <math>\gamma</math> a marked cycle of length <math>K</math>;</p> <p><b>return</b> <math>\gamma</math>.</p>
<p><b>(3) Cycle-pointed (symmetric cycle-pointed, resp.) cycle.</b></p> <p>Given <math>(s_i, t_i)_{i \geq 1}</math> such that <math>Z := \overline{Z}_{\text{CYC}^\circ}(s_1, t_1; s_2, t_2; \dots)</math> (<math>Z := \overline{Z}_{\text{CYC}^\circ}(s_1, t_1; s_2, t_2; \dots)</math>, resp.) converges, consider the probability distribution</p> $\Pr(R = r) = \frac{1}{Z} \varphi(r) \frac{t_r}{1 - s_r} \text{ for } r \geq 1 \quad (r \geq 2, \text{ resp.}).$ <p>Let <math>\text{REPLICORDER}(s_1, t_1; s_2, t_2; \dots)</math> be a generator of this distribution.</p> <p><b>Algorithm</b> <math>\Gamma \overline{Z}_{\text{CYC}^\circ}(s_1, t_1; s_2, t_2; \dots)</math> (<math>\Gamma \overline{Z}_{\text{CYC}^\circ}(s_1, t_1; s_2, t_2; \dots)</math>, resp.) :</p> <p style="margin-left: 20px;"><math>r \leftarrow \text{REPLICORDER}(s_1, t_1; s_2, t_2; \dots)</math>;</p> <p style="margin-left: 20px;"><math>j \leftarrow 1 + \text{Geom}(s_r)</math>;</p> <p style="margin-left: 20px;">Draw an integer <math>b \in [1..r - 1]</math> that is relatively prime to <math>r</math> uniformly at random ;</p> <p><b>return</b> the cycle of length <math>j \times r</math> with a marked atom, and endowed with the automorphism: “each atom is mapped to the atom that is <math>j \times b</math> units further on the cycle” (the marked cycle is the automorphism-cycle containing the marked atom).</p>
<p><b>(1') Cycle-pointed sequence of size <math>k</math>; denote <math>\mathcal{E} := \text{Seq}^{[k]}</math>.</b></p> <p><b>Algorithm</b> <math>\Gamma \overline{Z}_{\mathcal{E}^\circ}(s_1, t_1; s_2, t_2; \dots)</math> :</p> <p><b>return</b> a sequence of <math>k</math> atoms (with the identity-automorphism) where one atom taken u.a.r. is marked</p>
<p><b>(2') Cycle-pointed (symmetric cycle-pointed, resp.) set of size <math>k</math>; denote <math>\mathcal{S} := \text{Set}^{[k]}</math>.</b></p> <p>A marked (marked symmetric, resp.) partition sequence of order <math>k</math> is a sequence <math>\overline{\pi} = (\ell, n_1, n_2, \dots, n_k)</math> such that <math>\ell \geq 1</math> (<math>\ell \geq 2</math>, resp.) and <math>\ell + \sum_i i n_i = k</math> (one block is marked).</p> <p>The corresponding coefficient <math>[t^\ell s_1^{n_1} \dots s_k^{n_k}]</math> is denoted <math>\text{coef}_{\overline{\pi}}(\overline{Z}_{\mathcal{S}^\circ})</math> (<math>\text{coef}_{\overline{\pi}}(\overline{Z}_{\mathcal{S}^\circ})</math>, resp.)</p> <p><b>Algorithm</b> <math>\Gamma \overline{Z}_{\mathcal{S}^\circ}(s_1, t_1; s_2, t_2; \dots)</math> (<math>\Gamma \overline{Z}_{\mathcal{S}^\circ}(s_1, t_1; s_2, t_2; \dots)</math>, resp.) :</p> <p style="margin-left: 20px;">Draw a partition-sequence <math>\overline{\pi}</math> of order <math>k</math> with probability:</p> $\Pr(\overline{\pi}) = \frac{t^\ell s_1^{n_1} s_2^{n_2} \dots s_k^{n_k} \cdot \text{coef}_{\overline{\pi}}(\overline{Z}_{\mathcal{S}^\circ})}{Z_{\mathcal{S}^\circ}(s_1, t_1; s_2, t_2; \dots)} \quad \left( \Pr(\overline{\pi}) = \frac{t^\ell s_1^{n_1} s_2^{n_2} \dots s_k^{n_k} \cdot \text{coef}_{\overline{\pi}}(\overline{Z}_{\mathcal{S}^\circ})}{Z_{\mathcal{S}^\circ}(s_1, t_1; s_2, t_2; \dots)}, \text{ resp.} \right)$ <p><b>return</b> a collection of <math>n_1</math> cycles of length 1, <math>n_2</math> cycles of length 2, <math>\dots</math>, <math>n_k</math> cycles of length <math>k</math>, together with a marked cycle of length <math>\ell</math> (with a marked atom on it taken u.a.r.).</p>
<p><b>(3') Cycle-pointed (symmetric cycle-pointed, resp.) cycle of size <math>k</math>; denote <math>\mathcal{C} := \text{Cyc}^{[k]}</math>.</b></p> <p><b>Algorithm</b> <math>\Gamma \overline{Z}_{\mathcal{C}^\circ}(s_1, t_1; s_2, t_2; \dots)</math> (<math>\Gamma \overline{Z}_{\mathcal{C}^\circ}(s_1, t_1; s_2, t_2; \dots)</math>, resp.) :</p> <p style="margin-left: 20px;">Draw a divisor <math>r</math> of <math>k</math> (divisor <math>r \geq 2</math> of <math>k</math>) with distribution</p> $\Pr(r) = \frac{\phi(r) t_r s_r^{-1+k/r}}{Z_{\mathcal{C}^\circ}(s_1, t_1; s_2, t_2; \dots)} \quad \left( \Pr(r) = \frac{\phi(r) t_r s_r^{-1+k/r}}{Z_{\mathcal{C}^\circ}(s_1, t_1; s_2, t_2; \dots)}, \text{ resp.} \right)$ <p style="margin-left: 20px;">Draw an integer <math>b \in [1..r - 1]</math> that is relatively prime to <math>r</math> uniformly at random;</p> <p><b>return</b> the cycle of length <math>k</math> with a marked atom and endowed with the automorphism: “each atom is mapped to the atom that is <math>kb/r</math> units further on the cycle”.</p>

FIGURE 11. Pólya-Boltzmann sampler for basic cycle-pointed species. In all these random samplers, the finally returned structure is made well-labeled using the procedure `DISTRIBUTELABELS`.

<p><b>Cycle-pointed sum:</b> <math>\mathcal{R} = \mathcal{P} + \mathcal{Q}</math>.  Given <math>(s_i, t_i)_{i \geq 1}</math> such that <math>\overline{Z}_p := \overline{Z}_{\mathcal{P}}(s_1, t_1; s_2, t_2; \dots)</math> and <math>\overline{Z}_q := \overline{Z}_{\mathcal{Q}}(s_1, t_1; s_2, t_2; \dots)</math> converge:</p> <p style="margin-left: 20px;"><b>Algorithm</b> <math>\Gamma \overline{Z}_{\mathcal{R}}(s_1, t_1; s_2, t_2; \dots)</math> :</p> <p style="margin-left: 40px;"><b>if</b> <math>\text{Bern}(\overline{Z}_p / (\overline{Z}_p + \overline{Z}_q))</math> <b>then return</b> <math>\Gamma \overline{Z}_{\mathcal{P}}(s_1, t_1; s_2, t_2; \dots)</math></p> <p style="margin-left: 40px;"><b>else return</b> <math>\Gamma \overline{Z}_{\mathcal{Q}}(s_1, t_1; s_2, t_2; \dots)</math> <b>end if</b></p>
<p><b>Cycle-pointed product:</b> <math>\mathcal{R} = \mathcal{P} \star \mathcal{B}</math> (analog for <math>\mathcal{R} = \mathcal{B} \star \mathcal{P}</math>)  Given <math>(s_i, t_i)_{i \geq 1}</math> such that <math>Z_{\mathcal{B}}(s_1, s_2, \dots)</math> and <math>\overline{Z}_{\mathcal{P}}(s_1, t_1; s_2, t_2; \dots)</math> converge:</p> <p style="margin-left: 20px;"><b>Algorithm</b> <math>\Gamma \overline{Z}_{\mathcal{R}}(s_1, t_1; s_2, t_2; \dots)</math> :</p> <p style="margin-left: 20px;"><b>return</b> <math>(\Gamma \overline{Z}_{\mathcal{P}}(s_1, t_1; s_2, t_2; \dots), \Gamma Z_{\mathcal{B}}(s_1, s_2, \dots))</math> {independent calls}</p>
<p><b>Cycle-pointed substitution:</b> <math>\mathcal{R} = \mathcal{P} \odot \mathcal{B}</math>.  Given <math>(s_i, t_i)_{i \geq 1}</math> such that <math>b_i := Z_{\mathcal{B}}(s_i, s_{2i}, s_{3i}, \dots)</math> and <math>q_i := \overline{Z}_{\mathcal{B}^\circ}(s_i, t_i; s_{2i}, t_{2i}; \dots)</math> converge for each <math>i \geq 1</math>, and such that <math>\overline{Z}_{\mathcal{P}}(b_1, q_1; b_2, q_2; \dots)</math> converges:</p> <p style="margin-left: 20px;"><b>Algorithm</b> <math>\Gamma \overline{Z}_{\mathcal{R}}(s_1, t_1; s_2, t_2; \dots)</math> :</p> <p style="margin-left: 40px;">Compute <math>(P, \sigma_P, v) \leftarrow \Gamma \overline{Z}_{\mathcal{P}}(b_1, q_1; b_2, q_2; \dots)</math>;</p> <p style="margin-left: 40px;"><b>for each</b> unmarked cycle <math>C = (u_1, \dots, u_k)</math> of <math>\sigma_P</math> (<math>u_1</math> has smallest label in <math>C</math>)</p> <p style="margin-left: 60px;">Compute <math>(B, \sigma_B) \leftarrow \Gamma Z_{\mathcal{B}}(s_k, s_{2k}, \dots)</math>;</p> <p style="margin-left: 60px;">Replace each atom of <math>C</math> by a copy of <math>B</math>;</p> <p style="margin-left: 40px;"><b>for each</b> cycle <math>D</math> of <math>\sigma_B</math> <b>do</b></p> <p style="margin-left: 60px;">Let <math>E</math> be the cycle composed from the copies of <math>D</math> at <math>u_1, \dots, u_k</math>;</p> <p style="margin-left: 40px;"><b>end for</b></p> <p style="margin-left: 40px;"><b>end for</b></p> <p style="margin-left: 40px;">Let <math>F = (v_1, \dots, v_\ell)</math> be the marked cycle of <math>P</math> (<math>v_1</math> has smallest label in <math>F</math>);</p> <p style="margin-left: 40px;">Compute <math>(Q, \sigma_Q, q) \leftarrow \Gamma \overline{Z}_{\mathcal{B}^\circ}(s_\ell, t_\ell; s_{2\ell}, t_{2\ell}; \dots)</math>;</p> <p style="margin-left: 40px;">Replace each atom of <math>F</math> by a copy of <math>Q</math>;</p> <p style="margin-left: 40px;"><b>for each</b> cycle <math>G</math> of <math>\sigma_Q</math> <b>do</b></p> <p style="margin-left: 60px;">Let <math>H</math> be the cycle composed from the copies of <math>G</math> at <math>v_1, \dots, v_\ell</math>;</p> <p style="margin-left: 40px;"><b>end for</b></p> <p style="margin-left: 40px;">In the resulting structure <math>R</math>, mark the cycle composed from the copies of the marked cycle of <math>Q</math>;</p> <p style="margin-left: 20px;"><b>return</b> <math>(R, \sigma_R, r)</math>, where <math>\sigma_R</math> is the automorphism consisting of the cycles <math>E</math> and the cycles <math>H</math>, and where <math>r</math> is the atom <math>q</math> in the copy of the marked cycle of <math>Q</math> substituted at <math>v</math>.</p>

FIGURE 12. The rules to specify a Pólya-Boltzmann sampler for a cycle-pointed species assembled from other species using the cycle-pointed constructions  $\{+, \star, \odot\}$ . In all these random samplers, the output structure is made well-labeled using the procedure `DISTRIBUTE LABELS`.

“false”). (We work here with the refined weight  $t^\ell s_1^{n_1} s_2^{n_2} \dots s_k^{n_k} / n!$  instead of the standard weight  $x^n / n!$ , but the arguments given in [9] work the same way with these refined weights.)

For the cycle-pointed product,  $\mathcal{R} = \mathcal{P} \star \mathcal{B}$ ,  $\text{RSym}(\mathcal{R})$  is like a partitional product of  $\text{RSym}(\mathcal{P})$  and  $\text{Sym}(\mathcal{B})$ . Therefore, a Boltzmann sampler for  $\mathcal{R}$  consists of two independent calls to Pólya-Boltzmann samplers for  $\mathcal{P}$  and for  $\mathcal{B}$ . (Again the only difference here with [9] is that we consider refined weights:  $t^\ell s_1^{n_1} s_2^{n_2} \dots s_k^{n_k} / n!$  for  $\mathcal{P}$ , and  $s_1^{n_1} s_2^{n_2} \dots s_k^{n_k} / n!$  for  $\mathcal{B}$ .) The arguments are the same for  $\mathcal{B} \star \mathcal{P}$ .

For cycle-pointed substitution,  $\mathcal{R} = \mathcal{P} \odot \mathcal{B}$ , recall (from Equation (26) in the proof of Proposition 7) that for each marked integer partition  $\overline{\pi}$ , the sum of the weight-monomials over all the

symmetries on  $\mathcal{R}$  whose core has type  $\bar{\pi} = (\ell; n_1, n_2, \dots, n_k)$  satisfies the expression

$$\bar{Z}_{\mathcal{R}}^{(\bar{\pi})} = \frac{a_{\bar{\pi}}}{n!} q_{\ell} b_1^{n_1} b_2^{n_2} \dots b_k^{n_k}, \quad \text{where } b_i = Z_{\mathcal{B}}(s_i, s_{2i}, \dots), \quad q_{\ell} = \bar{Z}_{\mathcal{B}^{\circ}}(s_{\ell}, t_{\ell}; s_{2\ell}, t_{2\ell}; \dots).$$

Hence, a Pólya-Boltzmann sampler for  $\mathcal{R}$  must draw the core structure following the Pólya-Boltzmann distribution for  $\mathcal{P}$  with parameters  $(b_1, q_1; b_2, q_2; \dots)$ . In addition, once the type  $\bar{\pi}$  of the core symmetry is fixed, the structures substituted at the cycles of the core automorphism form a partitionial product of the form

$$\{\text{RSym}(\mathcal{B}^{\circ}) \text{ replicated } \ell \text{ times}\} \star \text{Sym}(\mathcal{B})^{n_1} \star \{\text{Sym}(\mathcal{B}) \text{ duplicated}\}^{n_2} \star \dots \star \{\text{Sym}(\mathcal{B}) \text{ replicated } k \text{ times}\}^{n_k}.$$

Recall that a partitionial product yields independent Boltzmann samplers. Hence, once the core symmetry  $(P, \sigma)$  is drawn, the symmetries in  $\mathcal{B}$  that are substituted at each cycle of  $\sigma_{\mathcal{A}}$  must be independent calls of a Pólya-Boltzmann sampler for  $\mathcal{B}$ , except for the marked cycle where we have to call a Pólya-Boltzmann sampler for  $\mathcal{B}^{\circ}$ . In addition, for an unmarked (marked, resp.) cycle, the parameters of  $\Gamma Z_{\mathcal{B}}$  (of  $\Gamma \bar{Z}_{\mathcal{B}^{\circ}}$ , resp.) must be  $(s_i, s_{2i}, \dots)$  ( $(s_i, t_i; s_{2i}, t_{2i}; \dots)$ , resp.) if the cycle has length  $i$ , as indicated by the expression of  $\bar{Z}_{\mathcal{R}}^{(\bar{\pi})}$  given above. This is precisely what the generator  $\Gamma \bar{Z}_{\mathcal{R}}$  does.  $\square$

**5.3.4. Pólya-Boltzmann samplers for decomposable cycle-pointed species.** Similarly as for decomposable species, the random generation rules shown in Figure 11 (basic cycle-pointed species) and Figure 12 (cycle-pointed constructions) can be combined to design a Pólya-Boltzmann sampler for any species with a *cycle-pointed* recursive decomposition over basic species. We assume here again that an oracle provides the required evaluations of cycle-index sums and pointed cycle-index sums for the species appearing in the decomposition; and that the cost of drawing  $k$  under a specific integer distribution (such as `REPLICORDER` in  $\Gamma Z_{\text{CYC}}$ ) has linear cost in  $k$ .

**Theorem 30.** *Any species  $\mathcal{P}$  with a cycle-pointed recursive specification (Definition 3.4) over species  $\mathcal{A}_1, \dots, \mathcal{A}_l$  having a linear Pólya-Boltzmann sampler can be endowed with a linear Pólya-Boltzmann sampler  $\Gamma Z_{\mathcal{P}}(s_1, t_1; s_2, t_2; \dots)$ .*

*Proof.* Analogous to the proof of Theorem 26.  $\square$

Consequently, the *unrooted species* we have encountered in Section 4 can be endowed with efficient random samplers.

**Proposition 31.** *In the real-arithmetic complexity model (oracle assumption), the following unlabeled species admit an exact-size sampler and an approximate-size sampler of expected complexities  $O(n^2)$  and  $O(n/\epsilon)$  ( $n$  being the target-size,  $\epsilon$  the tolerance-ratio): unrooted nonplane (plane, resp.) trees, unrooted nonplane (plane, resp.) trees whose node degrees are constrained to lie in a finite integer set  $\Omega$ , unrooted cacti graphs, unrooted connected outerplanar graphs.*

*Proof.* The crucial point is that cycle-pointing is unbiased, hence finding an exact-size (approximate-size, resp.) sampler for a species  $\mathcal{A}$  is equivalent to finding one for the cycle-pointed species  $\mathcal{A}^{\circ}$ .

For each of the unrooted tree species listed above, we have shown in Section 4 that the corresponding cycle-pointed species is decomposable. If  $\mathcal{G}$  is a species of connected graphs (closed under taking 2-connected components), the grammar (48) given in Proposition 15 ensures that the cycle-pointed species  $\mathcal{G}^{\circ}$  is decomposed in terms of the 2-connected graph species  $\mathcal{B}^{\circ}$ ,  $\mathcal{B}'$ , and  $(\mathcal{B}')^{\circ}$ . For cacti graphs and outerplanar graphs, there is a decomposition strategy for the 2-connected subspecies (polygons for cacti graphs, dissections of a polygon for outerplanar graphs), which easily yields linear Pólya-Boltzmann samplers for the species  $\mathcal{B}^{\circ}$ ,  $\mathcal{B}'$ , and  $(\mathcal{B}')^{\circ}$ . Since  $\mathcal{G}^{\circ}$  is specified over these three species, there is also a linear Pólya-Boltzmann sampler for  $\mathcal{G}^{\circ}$ .

Hence, for each unrooted species  $\mathcal{A}$  stated above, there is a linear Pólya-Boltzmann sampler for  $\mathcal{A}^{\circ}$ , which becomes a linear ordinary Boltzmann sampler when specializing to  $(s_i = x^i, t_i = x^i)$ . Moreover, the counting coefficients  $|\widetilde{\mathcal{A}}_n|$  obey the asymptotic form  $c \rho^{-n} n^{-5/2}$ , as shown in [26] for trees, [31] for cacti graphs, and [3] for outerplanar graphs. Therefore the coefficients  $|\widetilde{\mathcal{A}}_n^{\circ}|$  obey the asymptotic form  $c \rho^{-n} n^{-3/2}$ .

Hence, by Proposition 22, running the Boltzmann sampler  $\Gamma\overline{Z}_{\mathcal{A}^\circ}$  at  $(s_i = \rho^i, t_i = \rho^i)$  – with  $\rho$  the dominant singularity of  $\tilde{\mathcal{A}}(x)$  – yields an exact-size (approximate-size, resp.) sampler for unlabeled structures from  $\mathcal{A}$  with expected complexity  $O(n^2)$  ( $O(n/\epsilon)$ , respectively).  $\square$

#### 5.4. Simplifications.

5.4.1. *Regarding the labels.* Notice that the last step of all the random generation rules – as written in Figure 9, 10, 11, and 12 – is a call to a procedure that distributes distinct labels uniformly at random on the atoms of the output structure. A Pólya-Boltzmann sampler for a decomposable species calls each of these random generation rules a certain number of times until the structure is completely build. It is however not necessary to call the label-distribution procedure at each step; one can just wait until the complete structure is build to distribute the labels. (If one is only interested in the underlying unlabeled structure of the final output, one can simply forget about any label-distribution procedure.)

5.4.2. *Regarding the cycles in automorphisms.* We want to remark that one can simplify the random generation rules by forgetting about the cyclic order of the atoms in each part of the symmetries (automorphisms), i.e., storing only the partition of the atoms induced by the automorphism. The composition of cycles then corresponds to merging the corresponding parts of the partition; and the line “draw an integer that is relatively prime to  $r$ ” is deleted in the Pólya-Boltzmann samplers for cycle-pointed species. These simplified random samplers output the *profile* of a symmetry drawn under the Pólya-Boltzmann distribution (by profile we mean that the cyclic order of the atoms in each part is forgotten). This is actually enough if one is just interested in the underlying unlabeled structure, which is mostly the case in practice, since Pólya-Boltzmann samplers are used as a tool to design ordinary Boltzmann samplers.

5.4.3. *Specialization to ordinary Boltzmann samplers.* Let  $\mathcal{F}$  be a species with a recursive decomposition (cycle-pointed or not) over finitely many other species. Remark 2.1 and Remark 3.3 ensure that, to compute  $\tilde{\mathcal{F}}(x)$ , the only species in the recursive specification for which it is necessary to know the cycle-index sum are the core species of the substitution operations in the recursive specification. For all the other species that appear in the recursive specification, it is enough to compute just the ordinary generating function.

A similar remark holds for random generation. Namely, if one wants an ordinary Boltzmann sampler  $\Gamma\tilde{\mathcal{F}}(x)$ , it suffices to have Pólya-Boltzmann samplers for the species that appear as core species of substitution operations in the recursive specification. For all the other species, an ordinary Boltzmann sampler is enough.

5.5. **Examples.** To illustrate how our samplers (and the simplifications discussed above) operate in practice, we give here two examples: the species  $\mathcal{R}$  of rooted nonplane trees, already treated in [13], and the species  $\mathcal{F}$  of unrooted nonplane trees, also called free trees; here, one has to consider the associated cycle-pointed species, which is new.

In order to obtain uniform random samplers for unlabeled structures from  $\mathcal{R}$ , we design an ordinary Boltzmann sampler for  $\mathcal{R}$  via Pólya-Boltzmann samplers. Recall that  $\mathcal{R}$  is specified by  $\mathcal{R} = X \cdot \text{SET}(\mathcal{R})$ . Hence, according to Section 5.4.3, we need a Pólya-Boltzmann sampler for SET in order to recursively specify an ordinary Boltzmann sampler  $\Gamma\tilde{\mathcal{R}}(x)$ . Actually, according to the discussion in Section 5.4.2, we just need the profile of the automorphism computed by  $\Gamma Z_{\text{SET}}$ . Given these simplifications and the definition of  $\Gamma Z_{\text{SET}}$  (Figure 9), we obtain the ordinary Boltzmann sampler  $\Gamma\tilde{\mathcal{R}}(x)$  shown in Figure 13.

Let us now discuss the species of free trees  $\mathcal{F}$ . We want to sample unlabeled structures from  $\mathcal{F}$  uniformly at random. Since cycle-pointing is unbiased, it suffices to sample unlabeled structures from the cycle-pointed species  $\mathcal{F}^\circ$ . From the decomposition grammar for  $\mathcal{F}^\circ$  given in Proposition 11, we can design an ordinary Boltzmann sampler via Pólya-Boltzmann samplers. According to Section 5.4.3, the only species for which we need Pólya-Boltzmann samplers (as refinements of ordinary Boltzmann samplers) are the species SET, SET $^\circ$ , and SET $^\circ$ . Again, only the profile of the automorphisms returned by these samplers is necessary. Given the definition of  $\Gamma Z_{\text{SET}}$ ,  $\Gamma\overline{Z}_{\text{SET}^\circ}$ ,

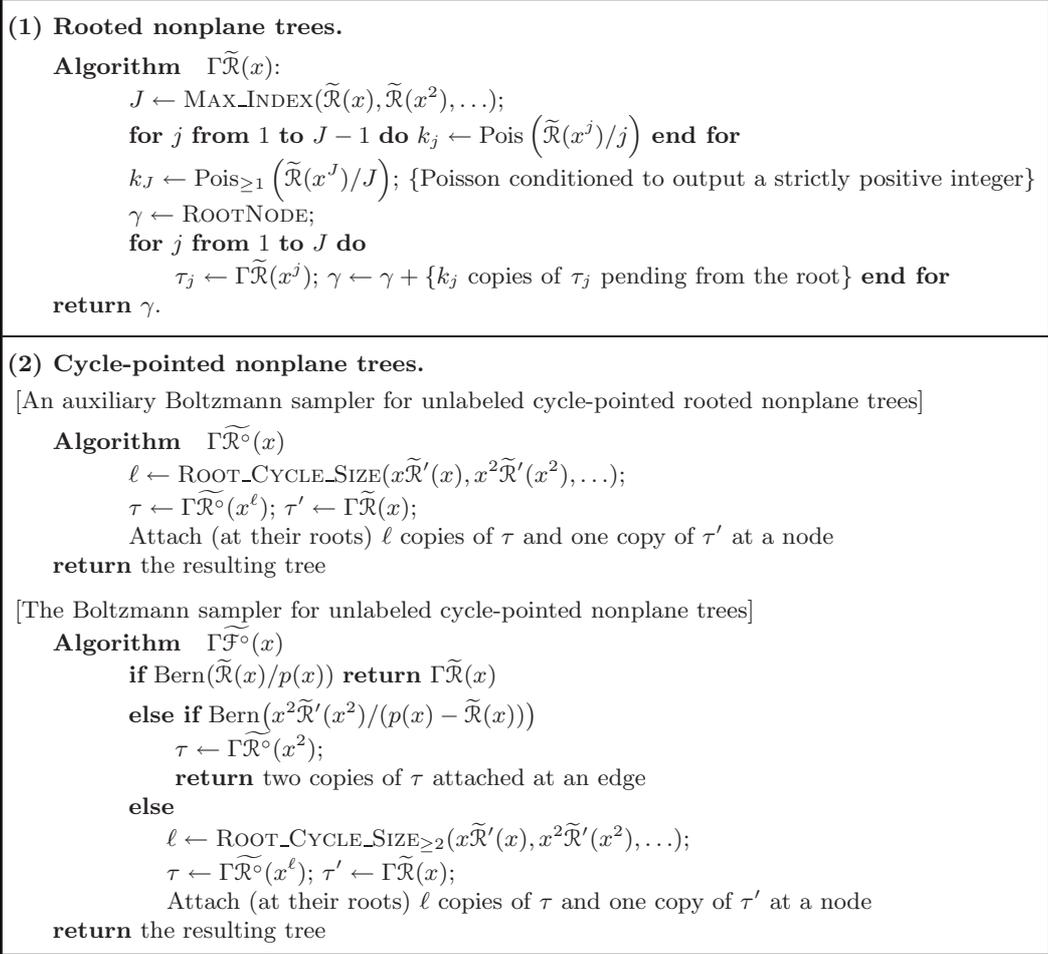


FIGURE 13. Ordinary Boltzmann samplers for rooted nonplane trees and cycle-pointed nonplane trees. Taking  $x = \rho$  the singularity of  $\widetilde{\mathcal{R}}(x)$ , we repeatedly call  $\Gamma\widetilde{\mathcal{R}}(x)$  ( $\Gamma\widetilde{\mathcal{F}}^\circ(x)$ , respectively) until we reach the target size; this yields an exact-size sampler of expected complexity  $O(n^2)$  and an approximate-size sampler of complexity  $O(n/\epsilon)$  for rooted nonplane trees (for free trees, respectively).

and  $\Gamma\overline{\mathcal{Z}}_{\text{SET}^\circ}$  (Figure 9), we obtain the ordinary Boltzmann sampler  $\Gamma\widetilde{\mathcal{F}}^\circ$  for unlabeled structures shown in Figure 13.

Note that in these Boltzmann samplers, the only evaluations required are those of the series  $p(x)$ , and of  $r(x^i)$  and  $x^i r'(x^i)$  for any  $i \geq 1$ . As discussed in Section 5.1.2, in practice we evaluate these series up to the power  $i = N/\log_2(1/\rho)$ , where  $N$  is the precision (number of bits).

*Acknowledgements.* Omid Amini, Olivier Bodini, Philippe Flajolet, and Pierre Leroux are greatly thanked for fruitful discussions and suggestions. Further, we thank two anonymous referees for their helpful comments.

#### REFERENCES

- [1] J. P. Bell, S. N. Burris, and K. A. Yeats. Counting rooted trees: The universal law  $t(n) \sim c\rho^{-n}n^{-3/2}$ . *Electronic Journal of Combinatorics* Volume 13(1) R63, 2006.
- [2] F. Bergeron, G. Labelle, and P. Leroux. *Combinatorial Species and Tree-like Structures*. Cambridge University Press, 1997.

- [3] M. Bodirsky, É. Fusy, M. Kang, and S. Vigerske. Enumeration of unlabeled outerplanar graphs. *Electronic Journal of Combinatorics*, 14, R66, 2007.
- [4] M. Bodirsky and M. Kang. Generating outerplanar graphs uniformly at random. *Comb. Prob. and Computing*, 15:333–343, 2006.
- [5] B. Bollobás. *Random Graph*. Cambridge University Press, 2nd ed., 2001.
- [6] M. Bona, M. Bousquet, G. Labelle, and P. Leroux. Enumeration of  $m$ -ary cacti, *Advances in Applied Mathematics*, 24:22–56, 2000.
- [7] P. Cameron. *Permutation groups*. Cambridge University Press, 1999.
- [8] L. Devroye. *Non-uniform Random Variate Generation*. Springer, 1986.
- [9] P. Duchon, P. Flajolet, G. Louchard, and G. Schaeffer. Boltzmann Samplers for the Random Generation of Combinatorial Structures. *Combinatorics, Probability and Computing*, 13(4-5):577–625, 2004.
- [10] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.
- [11] P. Flajolet, P. Zimmerman, and B. van Cutsem. A calculus for the random generation of labelled combinatorial structures. *Theoretical Computer Science*, 132(1-2):1–35, 1994.
- [12] É. Fusy. Counting unrooted maps using tree-decomposition. *Séminaire Lotharingien de Combinatoire*, B54A1 (2007), 44 pp.
- [13] É. Fusy, P. Flajolet, and C. Pivoteau. Boltzmann sampling of unlabelled structures. In *Proceedings of the Fourth Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, 2007.
- [14] O. Giménez and M. Noy. Asymptotic enumeration and limit laws of planar graphs *J. Amer. Math. Soc.* 22, 309-329, 2009.
- [15] L. A. Goldberg, M. Jerrum. The “Burnside Process” Converges Slowly, *Combinatorics, Probability and Computing*, 11(1):21-34, 2002.
- [16] F. Harary and E. M. Palmer. *Graphical Enumeration*. Academic Press, New York and London, 1973.
- [17] P. Hell and J. Nešetřil. *Graphs and Homomorphisms*. Oxford University Press, 2004.
- [18] J. A. Howell, T. F. Smith, and M. S. Waterman. Computation of generating functions for biological molecules. *SIAM J. Appl. Math.*, 39:119–133, 1980.
- [19] M. Jerrum. Uniform sampling modulo a group of symmetries using Markov chain simulation. Technical Report ECS-LFCS-94-288, Univ. Edinburgh, 1994.
- [20] M. Jerrum and A. Sinclair. The Markov Chain Monte Carlo Method: An Approach to Approximate Counting and Integration. Approximation Algorithms (D. Hochbaum), PWS Publishing Company, Boston, 1997.
- [21] D. E. Knuth. *The Art of Computer Programming*, 3rd edn, Vol. 2, *Seminumerical Algorithms*, Addison-Wesley, 1998.
- [22] V. Liskovets. A census of non-isomorphic planar maps. In *Coll. Math. Soc. J. Bolyai, Proc. Conf. Algebr. Meth. in Graph Th.*, volume 25:2, pages 479–494, 1981.
- [23] V. Liskovets and T. Walsh. The enumeration of non-isomorphic 2-connected planar maps. *Canad. J. Math.*, volume 35:3, pages 417–435, 1983.
- [24] B. Mohar and C. Thomassen. *Graphs on surfaces*. Johns Hopkins University Press, 2001.
- [25] A. Nijenhuis and H. Wilf. *Combinatorial algorithms*. Academic Press Inc., 1979.
- [26] Otter. The number of trees. *Annals of Math.*, 49:583–599, 1948.
- [27] C. Pivoteau, B. Salvy and M. Soria. Boltzmann Oracles for Combinatorial Systems. *Proceedings of the Fifth Colloquium on Mathematics and Computer Science: Algorithms, Trees, Combinatorics and Probabilities*, 2008.
- [28] G. Pólya. *Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen*. Acta Mathematica 68 (1): 145–254, 1937.
- [29] G. Pólya and R. C. Read. *Combinatorial Enumeration of Groups, Graphs and Chemical Compounds*. Springer Verlag, 1987.
- [30] R. C. Read. On general dissections of a polygon. *Aequationes Mathematicae*, 18, University of Waterloo, pages 370–388, 1978.
- [31] B. Shoilekova. Unlabelled enumeration of cacti graphs. *Manuscript*, 2007.
- [32] N. Sloane. The On-Line Encyclopedia of Integer Sequences. <http://www.research.att.com/~njas/sequences/A000672>, 2003.
- [33] W. T. Tutte. A census of planar maps. *Canad. J. Math.*, 15:249–271, 1963.
- [34] H. S. Wilf. The uniform selection of free trees. *J. Algorithms*, 2(2):204–207, 1981.

MANUEL BODIRSKY: LIX (CNRS UMR 7161), ÉCOLE POLYTECHNIQUE, 91128 PALAISEAU, FRANCE.

ÉRIC FUSY: LIX (CNRS UMR 7161), ÉCOLE POLYTECHNIQUE, 91128 PALAISEAU, FRANCE.

MIHYUN KANG: INSTITUT FÜR MATHEMATIK, TECHNISCHE UNIVERSITÄT BERLIN, STRASSE DES 17. JUNI 136, D-10623 BERLIN, GERMANY.

STEFAN VIGERSKE: INSTITUT FÜR MATHEMATIK, HUMBOLDT-UNIVERSITÄT ZU BERLIN, UNTER DEN LINDEN 6, 10099 BERLIN, GERMANY.