Combinatorics, Probability and Computing (2020), **29**, pp. 293–309
doi:10.1017/S0963548319000336

CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Unlabelled Gibbs partitions

Benedikt Stufler

Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, CH-8057, Zürich, Switzerland
Email: benedikt.stufler@math.uzh.ch

(Received 3 October 2016; revised 30 August 2019; first published online 4 November 2019)

Abstract
We study random composite structures considered up to symmetry that are sampled according to weights on the inner and outer structures. This model may be viewed as an unlabelled version of Gibbs partitions and encompasses multisets of weighted combinatorial objects. We describe a general setting characterized by the formation of a giant component. The collection of small fragments is shown to converge in total variation toward a limit object following a Pólya–Boltzmann distribution.

2010 MSC Codes: Primary 60C05, 05A18; Secondary 60B10

1. Introduction

The study of the evolution of shapes of random ensembles, as the total size becomes large, has a long history, and connections to a variety of fields such as statistical mechanics, representation theory and combinatorics are well known. A sketch of the history of limit shapes may be found in the work by Erlihson and Granovsky [10] on Gibbs partitions in the expansive case, and we refer the reader to this informative summary and references given therein for an adequate treatment of the historical development.

The term 'Gibbs partitions' was coined by Pitman [18] in his comprehensive survey on combinatorial stochastic processes. It describes a model of random partitions of sets, where the collection of classes as well as each individual partition class are endowed with a weighted structure. For example, in a simply generated random plane forest, each component is endowed with a tree structure carrying a non-negative weight, and the collection of components carries a linear order. Likewise, Gibbs partitions also encompass various types of random graphs whose vertex sets are partitioned by their connected components.

Many structures such as classes of graphs may also be viewed up to symmetry. The symmetric group acts in a canonical way on the collection of composite structures over a fixed set, and its orbits are called unlabelled objects. Sampling such an isomorphism class with probability proportional to its weight is the natural unlabelled version of the Gibbs partition model. This encompasses as a special case the important model of random multisets, which has been studied by Bell, Bender, Cameron and Richmond [3], and which is also encompassed in the setting by Arratia, Barbour and Tavaré [1] and Barbour and Granovsky [2]. The important example of forests of unlabelled trees has been considered by Mutafchiev [17]. General unlabelled Gibbs partitions, however, appear not to have received any attention in the literature so far. This is possibly due to the fact that this model of random ensembles is quite involved, as the symmetries of both the inner and outer structures influence its behaviour. This makes it particularly hard to arrive at general results that characterize the asymptotic behaviour for a wide range of species of structures. Nevertheless, it is natural to consider combinatorial objects up to symmetry, and to ask whether similar regimes such as, for example, the expansive case [10] or the convergent case [20] may also be found in the unlabelled setting.

For this reason, the present work aims to make a first step in this direction, with the hope that this may incite further research. We study a general setting characterized by the formation of a giant component with a stochastically bounded remainder. This phenomenon may, for example, be observed for uniformly sampled unordered forests of unlabelled trees as the total number of vertices tends to infinity, regardless of whether we consider trees that are rooted or unrooted, ordered or unordered. The small fragments are shown to converge in total variation towards a limit object following a Pólya–Boltzmann distribution, a term coined by Bodirsky, Fusy, Kang and Vigerske [5], who generalized and further developed the theory of Boltzmann samplers initiated in [7, 11]. Rather than taking a pure generating function viewpoint, our approach is to use the methods from [5] to reduce each problem to probabilistic questions. This allows us to prove our results in great generality and economically make use of available results for heavy-tailed and subexponential probability distributions [6, 8, 9, 13].

The present work is also the logical continuation of [20], where a gelation phenomenon was observed for labelled Gibbs partitions. The Pólya–Boltzmann sampler framework of [5] allows us to pursue a similar overall strategy as in [20], but our proofs are more involved and technical, as we have to consider objects up to symmetry.

The motivation of this particular line of research stems from the study of random graphs from restricted classes. McDiarmid [15, 16] showed that the small fragments of a random graph from a minor-closed addable class converge toward a Boltzmann Poisson random graph. In this work, McDiarmid poses the question whether similar behaviour may be observed for unlabelled graphs. As was shown in [20], an approach via Gibbs partitions and conditioned Galton–Watson trees is possible in the labelled setting. Hence it is natural to ask whether a similar strategy also works in the unlabelled setting. The present work provides a first piece of the puzzle, and we hope to pursue this question further in future work.

Plan of the paper

In Section 2 we fix notations and recall some background related to Gibbs partitions, combinatorial species, Pólya–Boltzmann distributions and subexponential sequences. Section 3 presents our main results for unlabelled Gibbs partitions. In Section 4 we collect all proofs.

2. Preliminaries

2.1 Notation

We use the notation

$$\mathbb{N} = \{1, 2, \ldots\}, \quad \mathbb{N}_0 = \{0\} \cup \mathbb{N}, \quad [n] = \{1, 2, \ldots, n\}, \quad n \in \mathbb{N}_0,$$

and let $\mathbb{R}_{>0}$ and $\mathbb{R}_{\geqslant 0}$ denote the sets of positive and non-negative real numbers, respectively. Throughout, we assume that all considered random variables are defined on a common probability space $(\Omega, \mathscr{F}, \mathbb{P})$. All unspecified limits are taken as n becomes large, possibly along an infinite subset of $\mathbb{N}$.

A function $h : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ is termed *slowly varying* if, for any fixed $t > 0$,

$$\lim_{x \to \infty} \frac{h(tx)}{h(x)} = 1.$$

For any power series $f(z) = \sum_n f_n z^n$, we let $[z^n]f(z) = f_n$ denote the coefficient of z^n. A sequence of $\mathbb{R}$-valued random variables $(X_n)_{n \geqslant 1}$ is *stochastically bounded* if, for each $\varepsilon > 0$, there is a constant $M > 0$ with

$$\limsup_{n\to\infty} \mathbb{P}(|X_n| \geqslant M) \leqslant \varepsilon.$$

The *total variation distance* between two random variables X and Y with values in a countable state space S is defined by

$$d_{\mathrm{TV}}(X, Y) = \sup_{\mathcal{E}\subset S} |\mathbb{P}(X \in \mathcal{E}) - \mathbb{P}(Y \in \mathcal{E})|.$$

2.2 Weighted combinatorial species and cycle index sums

The present section recalls the necessary species-theory following Joyal [14]. A *species of combinatorial structures* $\mathcal{F}^\omega$ with non-negative weights is a functor that produces for each finite set U a finite set $\mathcal{F}[U]$ of $\mathcal{F}$*-structures* and a map

$$\omega_U : \mathcal{F}[U] \to \mathbb{R}_{\geqslant 0}.$$

We will often write $\omega(F)$ instead of $\omega_U(F)$ for the weight of a structure $F \in \mathcal{F}[U]$. For the special case $U = [k]$ we write $\mathcal{F}[k]$ instead of $\mathcal{F}[[k]]$. If no weighting is specified explicitly, we assume that any structure receives weight 1. We refer to the set U as the set of *labels* or *atoms* of the structure. For any $\mathcal{F}$-object $F \in \mathcal{F}[U]$, we let $|F| := |U| \in \mathbb{N}_0$ denote its *size*. The species $\mathcal{F}$ is further required to produce for each bijection $\sigma : U \to V$ a corresponding bijection

$$\mathcal{F}[\sigma] : \mathcal{F}[U] \to \mathcal{F}[V]$$

that preserves the ω-weights. In other words, the following diagram must commute:

$$\begin{tikzcd} \mathcal{F}[U] \arrow[r, "\mathcal{F}[\sigma]"] \arrow[dr, "\omega_U"] & \mathcal{F}[V] \arrow[d, "\omega_V"] \\ & \mathbb{R}_{\geqslant 0} \end{tikzcd}$$

Species are also subject to the usual functoriality requirements: the identity map id_U on U gets mapped to the identity map $\mathcal{F}[\mathrm{id}_U] = \mathrm{id}_{\mathcal{F}[U]}$ on the set $\mathcal{F}[U]$. For any bijections $\sigma : U \to V$ and $\tau : V \to W$, the following diagram commutes:

$$\begin{tikzcd} \mathcal{F}[U] \arrow[r, "\mathcal{F}[\sigma]"] \arrow[dr, "\mathcal{F}[\tau\sigma]"] & \mathcal{F}[V] \arrow[d, "\mathcal{F}[\tau]"] \\ & \mathcal{F}[W] \end{tikzcd}$$

We further assume that $\mathcal{F}[U] \cap \mathcal{F}[V] = \emptyset$ whenever $U \neq V$. This is not much of a restriction, as we may always replace $\mathcal{F}[U]$ by $\{U\} \times \mathcal{F}[U]$ for all sets U, to make sure that it is satisfied.

Two weighted species $\mathcal{F}^\omega$ and $\mathcal{H}^\gamma$ are said to be *structurally equivalent* or *isomorphic*, denoted by $\mathcal{F}^\omega \simeq \mathcal{H}^\gamma$, if there is a family of weight-preserving bijections $(\alpha_U : \mathcal{F}[U] \to \mathcal{H}[U])_U$ with U ranging over all finite sets, such that the following diagram commutes for each bijection $\sigma : U \to V$ of finite sets:

$$\begin{tikzcd} \mathcal{F}[U] \arrow[r, "\mathcal{F}[\sigma]"] \arrow[d, "\alpha_U"] & \mathcal{F}[V] \arrow[d, "\alpha_V"] \\ \mathcal{H}[U] \arrow[r, "\mathcal{H}[\sigma]"] & \mathcal{H}[V] \end{tikzcd} \tag{2.1}$$

For any finite set U, the symmetric group $\mathscr{S}_U$ acts on the set $\mathcal{F}[U]$ via

$$\sigma.F = \mathcal{F}[\sigma](F)$$

https://doi.org/10.1017/S0963548319000336 Published online by Cambridge University Press

for all $F \in \mathcal{F}[U]$ and $\sigma \in \mathscr{S}_U$. A bijection σ with $\sigma.F = F$ is termed an *automorphism* of F. We let $\tilde{\mathcal{F}}[U]$ denote the orbits of this group action. All $\mathcal{F}$-objects of an orbit $\tilde{F}$ have the same size and same ω-weight, which we denote by $|\tilde{F}|$ and $\omega(\tilde{F})$. It will be convenient to use the notation

$$\mathscr{U}(\mathcal{F}) = \bigcup_{k \geqslant 0} \mathscr{U}_k(\mathcal{F}) \quad \text{with} \quad \mathscr{U}_k(\mathcal{F}) = \tilde{\mathcal{F}}[\{1, \ldots, k\}].$$

Formally, an *unlabelled* $\mathcal{F}$-object is defined as an isomorphism class of $\mathcal{F}$-objects. We may also identify the unlabelled objects of a given size n with the orbits of the action of the symmetric group on any n-sized set. In particular, the collection of unlabelled $\mathcal{F}$-objects may be identified with the set $\mathscr{U}(\mathcal{F})$. By abuse of notation, we treat unlabelled objects as if they were regular $\mathcal{F}$-objects. The power series

$$\tilde{\mathcal{F}}^{\omega}(z) = \sum_{\tilde{F} \in \mathscr{U}(\mathcal{F})} \omega(\tilde{F}) z^{|\tilde{F}|}$$

is the *ordinary generating series* of the species. Note that we really need condition (2.1) to ensure that isomorphic species have the same ordinary generating series.

To any species $\mathcal{F}$ we may assign the corresponding species $\mathrm{Sym}(\mathcal{F})$ of $\mathcal{F}$-*symmetries* such that

$$\mathrm{Sym}(\mathcal{F})[U] = \{(F, \sigma) \mid F \in \mathcal{F}[U], \sigma \in \mathscr{S}_U, \sigma.F = F\}.$$

In other words, a symmetry is a pair of an $\mathcal{F}$-object and an automorphism. The transport along a bijection $\gamma : U \to V$ is given by

$$\mathrm{Sym}(\mathcal{F})[\gamma](F, \sigma) = (\mathcal{F}[\gamma](F), \gamma\sigma\gamma^{-1}).$$

For any permutation σ we let σ_i denote its number of i-cycles. The *cycle index sum* of a species $\mathcal{F}$ is defined as the formal power series

$$Z_{\mathcal{F}^{\omega}}(z_1, z_2, \ldots) = \sum_{k \geqslant 0} \sum_{(F,\sigma) \in \mathrm{Sym}(\mathcal{F})[k]} \frac{\omega(F)}{k!} z_1^{\sigma_1} \cdots z_k^{\sigma_k}$$

in countably infinitely many indeterminates $z_1, z_2, \ldots$. The following standard result is given, for example, by Bergeron, Labelle and Leroux [4, Chapter 2.3] and shows how the ordinary generating series and the cycle index sum of a species are related.

Lemma 2.1. *For any finite set U and any unlabelled $\mathcal{F}$-object $\tilde{F} \in \tilde{\mathcal{F}}[U]$ there are precisely $|U|!$ many symmetries $(F, \sigma) \in \mathrm{Sym}(\mathcal{F})[U]$ such that F belongs to the orbit $\tilde{F}$. Consequently,*

$$\tilde{\mathcal{F}}^{\omega}(z) = Z_{\mathcal{F}^{\omega}}(z, z^2, z^3, \ldots).$$

We illustrate the concepts of this section with an example.

Example 2.2. *The species* CYC *of cycles associates to any finite set U the subset* $\mathrm{CYC}[U] \subset \mathscr{S}_U$ *of cyclic permutations with length $k := |U|$. Hence* $\mathrm{CYC}[U]$ *has $(k-1)!$ elements for $k \geqslant 1$. By convention,* $\mathrm{CYC}[\emptyset]$ *contains a single element, the trivial bijection from the empty set to itself. The transport along a bijection $\gamma : U \to V$ is given by*

$$\mathrm{CYC}[\gamma] : \mathrm{CYC}[U] \to \mathrm{CYC}[V], \qquad \tau \mapsto \gamma\tau\gamma^{-1}.$$

Conjugating a cycle $\tau = (u_1, \ldots, u_k) \in \mathrm{CYC}[U]$ by a permutation $\sigma \in \mathscr{S}_U$ yields the cycle $(\sigma(u_1), \ldots, \sigma(u_k))$. In particular, $\sigma\tau\sigma^{-1} = \tau$ holds if and only if σ is a power of τ. Hence

$$\mathrm{Sym}(\mathrm{CYC})[U] = \{(\tau, \tau^i) \mid \tau \in \mathrm{CYC}[U], 0 \leqslant i < |U|\}.$$

The power τ^i is the product of k/r disjoint cycles with length r, where r is the order of the coset $\bar{i}$ in the cyclic group $\mathbb{Z}/k\mathbb{Z}$. This entails that $r \mid k$, that is, r is a divisor of k. There are $\varphi(r)$ elements with order r in $\mathbb{Z}/k\mathbb{Z}$, with φ denoting Euler's totient function. This leads to the well-known equality

$$Z_{\mathrm{CYC}} = 1 + \sum_{k \geqslant 1} \sum_{r \mid k} \frac{\varphi(r)}{k} z_r^{k/r} = 1 + \sum_{r \geqslant 1} \frac{\varphi(r)}{r} \log\left(\frac{1}{1 - z_r}\right).$$

2.3 Constructions on species

There are many ways to form species of structures by combining other species. Most prominently, composite structures are formed by partitioning a set and endowing both the partition classes and the collection of all classes with additional weighted structures. Derived structures are regular structures over a set of labels together with a distinguished $*$-placeholder that does not count as a regular atom. We recall the details following the classical literature by Joyal [14] and Bergeron, Labelle and Leroux [4].

2.3.1 Composite structures

Let $\mathcal{F}^\omega$ and $\mathcal{G}^\nu$ be two combinatorial species with non-negative weights. We assume that $\mathcal{G}^\nu[\emptyset] = \emptyset$. The *composition* $\mathcal{F}^\omega \circ \mathcal{G}^\nu = (\mathcal{F} \circ \mathcal{G})^\mu$ is a weighted species that describes partitions of finite sets, where each partition class is endowed with a $\mathcal{G}$-structure, and the collection of partition classes carries an $\mathcal{F}$-structure. That is, for each finite set U,

$$(\mathcal{F} \circ \mathcal{G})[U] = \bigcup_{\pi} \mathcal{F}[\pi] \times \prod_{Q \in \pi} \mathcal{G}[Q],$$

with the index π ranging over all unordered partitions of U with non-empty partition classes. In other words, π is a set of non-empty subsets of U such that $U = \bigcup_{Q \in \pi} Q$ and $Q \cap Q' = \emptyset$ for all $Q, Q' \in \pi$ with $Q \neq Q'$. The weight of a composite structure $(F, (G_Q)_{Q \in \pi})$ is given by

$$\mu(F, (G_Q)_{Q \in \pi}) = \omega(F) \prod_{Q \in \pi} \nu(G_Q).$$

For any bijection $\sigma : U \to V$, the corresponding transport function

$$(\mathcal{F} \circ \mathcal{G})[\sigma] : (\mathcal{F} \circ \mathcal{G})[U] \to (\mathcal{F} \circ \mathcal{G})[V]$$

is given as follows. For each element $(F, (G_Q)_{Q \in \pi}) \in (\mathcal{F} \circ \mathcal{G})[U]$, we let $\bar{\pi} = \{\sigma(Q) \mid Q \in \pi\}$ denote the corresponding partition of V and set

$$\bar{\sigma} : \pi \to \bar{\pi}, Q \mapsto \sigma(Q).$$

For each $Q \in \pi$ we let

$$\sigma|_Q : Q \to \sigma(Q), x \mapsto \sigma(x)$$

denote the restriction of σ to the class Q. We set

$$(\mathcal{F} \circ \mathcal{G})[\sigma](F, (G_Q)_{Q \in \pi}) = (\mathcal{F}[\bar{\sigma}](F), (\mathcal{G}[\sigma|_{\sigma^{-1}(P)}](G_{\sigma^{-1}(P)}))_{P \in \bar{\pi}}).$$

The cycle index sum of the composition is given by

$$Z_{\mathcal{F}^\omega \circ \mathcal{G}^\nu}(z_1, z_2, \ldots) = Z_{\mathcal{F}^\omega}(Z_{\mathcal{G}^\nu}(z_1, z_2, \ldots), Z_{\mathcal{G}^{\nu^2}}(z_2, z_4, \ldots), Z_{\mathcal{G}^{\nu^3}}(z_3, z_6, \ldots), \ldots).$$

Here we let ν^i denote the weighting that assigns to each $\mathcal{G}$-object G the weight $\nu(G)^i$.

Example 2.3. *The species* SET *given by* $\mathrm{SET}[U] = \{U\}$ *for all* U *has cycle index sum given by*

$$Z_{\mathrm{SET}}(z_1, z_2, \ldots) = \exp\left(\sum_{i=1}^{\infty} \frac{z_i}{i}\right).$$

https://doi.org/10.1017/S0963548319000336 Published online by Cambridge University Press

Hence, for any weighted species $\mathcal{G}^\nu$ the generating series for multisets of unlabelled $\mathcal{G}$-objects is given by

$$\exp\left(\sum_{i=1}^{\infty} \frac{\tilde{\mathcal{G}}^{\nu^i}(z^i)}{i}\right).$$

By Example 2.2 it follows that the generating series for cyclically ordered collections of unlabelled $\mathcal{G}$-objects is given by

$$\sum_{r\geqslant 1} \frac{\varphi(r)}{r} \log\left(\frac{1}{1-\tilde{\mathcal{G}}^{\nu^r}(z^r)}\right).$$

2.3.2 Derived structures

Let $\mathcal{F}^\omega$ be a weighted species. The *derived* species $(\mathcal{F}')^\omega$ is defined as follows. For each set U we let $*_U$ denote a placeholder object not contained in U. For example, we could set $*_U := U$, as no set is allowed to be an element of itself. By abuse of notation, we will usually drop the index and just refer to it as the $*$-placeholder atom. We set

$$\mathcal{F}'[U] = \mathcal{F}[U \cup \{*_U\}].$$

The weight of an element $F' \in \mathcal{F}'[U]$ is its ω-weight as an $\mathcal{F}$-structure. Any bijection $\sigma : U \to V$ may canonically be extended to a bijection

$$\sigma' : U \cup \{*_U\} \to V \cup \{*_V\},$$

and we set

$$\mathcal{F}'[\sigma] = \mathcal{F}[\sigma'].$$

Thus, an $\mathcal{F}'$-object with size n is an $\mathcal{F}$-object with size $n+1$, since we do not count the $*$-placeholder.

Note that an $\mathcal{F}'$-symmetry over the set U corresponds to an $\mathcal{F}$-symmetry over $U \cup \{*_U\}$ that fixes the $*$-object. The cycle index sum of $(\mathcal{F}')^\omega$ is given by the formal derivative

$$Z_{(\mathcal{F}')^\omega}(z_1, z_2, \ldots) = \frac{\mathrm{d}}{\mathrm{d}z_1} Z_{\mathcal{F}^\omega}(z_1, z_2, \ldots).$$

Example 2.4. *The species* SEQ *of ordered sequences has cycle index sum given by*

$$Z_{\mathrm{SEQ}} = \frac{1}{1-z_1}.$$

Objects of the derived species SEQ$'$ *correspond to pairs of ordered sequences that are separated by the $*$-placeholder. Hence*

$$Z_{\mathrm{SEQ}'} = \frac{1}{(1-z_1)^2}.$$

Deriving a species restricts the number of symmetries. For example, objects of the derived species CYC$'$ *correspond to a $*$-placeholder followed by a linearly ordered list, yielding*

$$Z_{\mathrm{CYC}'} = \frac{1}{1-z_1}.$$

2.4 Pólya–Boltzmann distributions for composite structures

Given a weighted species $\mathcal{F}^\omega$ and a parameter $y > 0$ with $0 < \tilde{\mathcal{F}}^\omega(y) < \infty$, we may consider the corresponding Boltzmann probability measure

$$\mathbb{P}_{\tilde{\mathcal{F}}^\omega, y}(\tilde{F}) = \tilde{\mathcal{F}}^\omega(y)^{-1} y^{|\tilde{F}|} \omega(\tilde{F}), \quad \tilde{F} \in \mathscr{U}(\mathcal{F}).$$

https://doi.org/10.1017/S0963548319000336 Published online by Cambridge University Press

Likewise, given parameters $y_1, y_2, \ldots, \geqslant 0$ with

$$0 < Z_{\mathcal{F}^\omega}(y_1, y_2, \ldots) < \infty,$$

we may consider the Pólya–Boltzmann distribution

$$\mathbb{P}_{Z_{\mathcal{F}^\omega},(y_j)_j}(F, \sigma) = Z_{\mathcal{F}^\omega}(y_1, y_2, \ldots)^{-1} \frac{\omega(F)}{k!} y_1^{\sigma_1} \cdots y_k^{\sigma_k}$$

for

$$(F, \sigma) \in \bigcup_{k \geqslant 0} \mathrm{Sym}(\mathcal{F})[k].$$

Note that if we condition a $\mathbb{P}_{\tilde{\mathcal{F}}^\omega, y}$-distributed random variable on having a fixed size n, then the result gets drawn from $\mathscr{U}_n(\mathcal{F})$ with probability proportional to its ω-weight. In a way, this is analogous to the fact that simply generated trees (with analytic weights) may be viewed as Galton–Watson trees conditioned on having a fixed number of vertices, and the viewpoint is equally useful in this context.

Lemma 2.1 implies the useful fact that the orbit of the $\mathcal{F}$-object of a $\mathbb{P}_{Z_{\mathcal{F}^\omega},(y,y^2,\ldots)}$-distributed symmetry follows a $\mathbb{P}_{\tilde{\mathcal{F}}^\omega, y}$-distribution. This provides a systematic way for sampling Boltzmann distributed structures, as the cycle index sums for constructions on species admit explicit expressions with concrete combinatorial interpretations. For composite structures in particular, the following result is given in Bodirsky, Fusy, Kang and Vigerske [5, Proposition 25] for species without weights, and the generalization to the weighted setting is straightforward.

Lemma 2.5. *Let $\mathcal{F}^\omega$ and $\mathcal{G}^\nu$ be weighted species with $\mathcal{G}[\emptyset] = \emptyset$. Let $y > 0$ be a parameter with*

$$\widetilde{\mathcal{F}^\omega \circ \mathcal{G}^\nu}(y) = Z_{\mathcal{F}^\omega}(\tilde{\mathcal{G}}^\nu(y), \tilde{\mathcal{G}}^{\nu^2}(y^2), \tilde{\mathcal{G}}^{\nu^3}(y^3), \ldots) \in]0, \infty[.$$

Then the following procedure terminates with an unlabelled $(\mathcal{F}^\omega \circ \mathcal{G}^\nu)$-object that follows a $\mathbb{P}_{\widetilde{\mathcal{F}^\omega \circ \mathcal{G}^\nu}, y}$-distribution.

(1) *Let (F, σ) be a $\mathbb{P}_{Z_{\mathcal{F}^\omega},(\tilde{\mathcal{G}}^\nu(y), \tilde{\mathcal{G}}^{\nu^2}(y^2),\ldots)}$-distributed $\mathcal{F}$-symmetry.*
(2) *For each cycle τ of σ let $|\tau|$ denote its length and draw independently a $\mathcal{G}$-object G_τ according to a $\mathbb{P}_{\tilde{\mathcal{G}}^{\nu^{|\tau|}}, y^{|\tau|}}$-distribution.*
(3) *Construct an $\mathcal{F} \circ \mathcal{G}$-object by assigning for each cycle τ and each atom v of τ an identical copy of G_τ to v.*

2.5 Subexponential sequences

Subexponential sequences correspond up to tilting and rescaling to subexponential densities of random variables with values in a lattice, and may be put into the more general context of subexponential distributions [6, 8, 13].

Definition. Let $d \geqslant 1$ be an integer. A power series $g(z) = \sum_{n=0}^\infty g_n z^n$ with non-negative coefficients and radius of convergence $\rho > 0$ belongs to the class $\mathscr{S}_d$ if $g_n = 0$ whenever n is not divisible by d, and

$$\frac{g_n}{g_{n+d}} \sim \rho^d, \quad \frac{1}{g_n} \sum_{i+j=n} g_i g_j \sim 2g(\rho) < \infty \tag{2.2}$$

as $n \equiv 0 \mod d$ becomes large.

We are going to make use of the following basic properties of subexponential sequences.

https://doi.org/10.1017/S0963548319000336 Published online by Cambridge University Press

Lemma 2.6 **([13, Theorems 4.8, 4.11, 4.30], [9]).** *Let $g(z)$ belong to $\mathscr{S}_d$ with radius of convergence ρ.*

(1) *For each $\varepsilon > 0$ there is a $c(\varepsilon) > 0$ and an $n_0 > 0$ such that, for all $n \geqslant n_0$ with $n \equiv 0 \mod d$, and each $k \geqslant 0$,*
$$[z^n]g(z)^k \leqslant c(\varepsilon)(g(\rho)+\varepsilon)^k[z^n]g(z).$$

(2) *If $f(z)$ is a non-constant power series with non-negative coefficients that is analytic at ρ, then $f(g(z))$ belongs to $\mathscr{S}_d$ and*
$$[z^n]f(g(z)) \sim f'(g(\rho))[z^n]g(z), \quad n \to \infty, \quad n \equiv 0 \mod d.$$

(3) *If $a_n = h(n)n^{-\beta}\rho^{-n}$ for some constants $\rho > 0$, $\beta > 1$ and a slowly varying function h, then the series $\sum_{n \in d\mathbb{N}} a_n z^n$ belongs to the class $\mathscr{S}_d$.*

The following criterion will prove to be useful as well.

Lemma 2.7 **([13, Theorem 4.9]).** *Let $f(z)$ belong to $\mathscr{S}_1$ with radius of convergence ρ, and $g_1(z), g_2(z)$ be power series with non-negative coefficients. If*
$$\frac{[z^n]g_1(z)}{[z^n]f(z)} \to c_1 \quad \textit{and} \quad \frac{[z^n]g_2(z)}{[z^n]f(z)} \to c_2$$
as $n \to \infty$ with $c_1, c_2 \geqslant 0$, then
$$\frac{[z^n]g_1(z)g_2(z)}{[z^n]f(z)} \to c_1 g_2(\rho) + c_2 g_1(\rho).$$
If additionally $c_1 g_2(\rho) + c_2 g_1(\rho) > 0$, then $g_1(z)g_2(z)$ belongs to $\mathscr{S}_1$.

3. Unlabelled Gibbs partitions

Let $\mathcal{F}^\omega$ and $\mathcal{G}^\nu$ be weighted combinatorial species with $\mathcal{G}[\emptyset] = \emptyset$, so that the weighted composition
$$(\mathcal{F} \circ \mathcal{G})^\mu = \mathcal{F}^\omega \circ \mathcal{G}^\nu$$
is well-defined. Throughout we assume $[z^k]\tilde{\mathcal{F}}^\omega(z) > 0$ for at least one $k \geqslant 1$ and that $\widetilde{\mathcal{F}^\omega \circ \mathcal{G}^\nu}(z)$ is not a polynomial. For each integer $n \geqslant 0$ with
$$[z^n]\widetilde{\mathcal{F}^\omega \circ \mathcal{G}^\nu}(z) > 0,$$
we may sample a random composite structure
$$\mathsf{S}_n = (\mathsf{F}_n, (\mathsf{G}_Q)_{Q \in \pi_n})$$
from the set $\mathscr{U}_n(\mathcal{F} \circ \mathcal{G})$ with probability proportional to its μ-weight.

We are going to study the asymptotic behaviour of the remainder R_n when deleting 'the' largest component from S_n. More specifically, we pick an arbitrary representative of S_n and construct R_n as follows. We make a uniform choice of a component $Q_0 \in \pi_n$ having maximal size, and let F'_n denote the $\mathcal{F}'$-object obtained from the $\mathcal{F}$-object F_n by relabelling the Q_0 atom of F_n to a $*$-placeholder.

Thus
$$\mathsf{F}'_n = \mathcal{F}[\gamma](\mathsf{F}_n) \in \mathcal{F}'[\pi_n \setminus \{Q_0\}]$$
for the bijection $\gamma : \pi_n \to (\pi_n \setminus \{Q_0\}) \cup \{*\}$ with $\gamma(Q_0) = *$ and $\gamma(Q) = Q$ for $Q \neq Q_0$. This yields an unlabelled $\mathcal{F}' \circ \mathcal{G}$-object
$$\mathsf{R}_n := (\mathsf{F}'_n, (\mathsf{G}_Q)_{Q \in \pi_n \setminus \{Q_0\}}) \in \mathscr{U}(\mathcal{F}' \circ \mathcal{G}).$$

https://doi.org/10.1017/S0963548319000336 Published online by Cambridge University Press

We let ρ denote the radius of convergence of the ordinary generating series $\tilde{\mathcal{G}}^\nu(z)$ and suppose throughout that

$$Z_{\mathcal{F}^\omega}(\tilde{\mathcal{G}}^\nu(\rho)+\varepsilon, \tilde{\mathcal{G}}^{\nu^2}((\rho+\varepsilon)^2), \tilde{\mathcal{G}}^{\nu^3}((\rho+\varepsilon)^3), \ldots) < \infty \tag{3.1}$$

for some $\varepsilon > 0$.

For example, in the special case of multisets we have $\mathcal{F}^\omega = \text{SET}$ and each $\mathcal{G}$-object receives weight 1. It follows from the expression for Z_{SET} in Example 2.3 that condition (3.1) is satisfied in this case if $\rho < 1$ and $\tilde{\mathcal{G}}(\rho) < \infty$. A classical example where condition (3.1) is not satisfied are multisets of positive integers, where $\mathcal{G}$ is given by the species $\text{SEQ}_{\geqslant 1}$of ordered non-empty sequences.

Let R be a random unlabelled $\mathcal{F}' \circ \mathcal{G}$-element that follows a Boltzmann distribution

$$\mathbb{P}(\mathsf{R} = R) = \frac{\mu(R)\rho^{|R|}}{\widetilde{(\mathcal{F}')^\omega \circ \mathcal{G}^\nu}(\rho)}, \qquad R \in \mathscr{U}(\mathcal{F}' \circ \mathcal{G}).$$

Theorem 3.1. *If the series $\tilde{\mathcal{G}}^\nu(z)$ belongs to the class $\mathscr{S}_d$, then*

$$d_{\text{TV}}(\mathsf{R}_n, \mathsf{R}) \to 0$$

as $n \to \infty$ with $n \equiv 0 \mod d$.

The main challenge for verifying Theorem 3.1 is that we consider objects up to symmetry. Lemma 2.5 provides a way of sampling S_n as a conditioned Boltzmann-distributed composite structure consisting of an $\mathcal{F}$-symmetry with identical $\mathcal{G}$-objects dangling from each cycle. The key idea will be that the largest $\mathcal{G}$-object is likely to correspond to a fixpoint of the symmetry. A similar congelation phenomenon was observed for random labelled composite structures sampled from $(\mathcal{F}^\omega \circ \mathcal{G}^\nu)[n]$ with probability proportional to their weight [20, Theorem 3.1]. Our overall strategy is similar, but treating unlabelled structures is more involved.

Theorem 3.1 is relevant for the structure of connected components in certain models of random graphs such as uniform random unlabelled series-parallel graphs or uniform random unlabelled outerplanar graphs. We refer the reader to the subsequent paper [19] for details.

We require the following enumerative result for the proof of our main theorem.

Lemma 3.2. *Let ρ denote the radius of convergence of the series $\tilde{\mathcal{G}}^\nu(z)$. If $\tilde{\mathcal{G}}^\nu(z)$ belongs to the class $\mathscr{S}_d$, then*

$$[z^n]\widetilde{\mathcal{F}^\omega \circ \mathcal{G}^\nu}(z) \sim \widetilde{(\mathcal{F}')^\omega \circ \mathcal{G}^\nu}(\rho)[z^n]\tilde{\mathcal{G}}^\nu(z),$$

as $n \to \infty$ with $n \equiv 0 \mod d$. Here

$$\widetilde{(\mathcal{F}')^\omega \circ \mathcal{G}^\nu}(\rho) = \left(\frac{\mathrm{d}}{\mathrm{d}z_1} Z_{\mathcal{F}^\omega}\right)(\tilde{\mathcal{G}}^\nu(\rho), \tilde{\mathcal{G}}^{\nu^2}(\rho^2), \ldots).$$

If $\tilde{\mathcal{G}}^\nu(z)$ is amenable to singularity analysis, then Lemma 3.2 may also be verified using analytic methods [12]. But we make no assumptions at all about the singularities of $\tilde{\mathcal{G}}^\nu(z)$ on the circle $|z| = \rho$. We only require that this series belongs to the class $\mathscr{S}_d$, which is far more general.

Clearly Theorem 3.1 also implies distributional convergence of the number of components, which has been studied in [3] for the case of weighted multisets where $\mathcal{F}^\omega = \text{SET}$.

Corollary 3.3. *Suppose that the series $\tilde{\mathcal{G}}^\nu(z)$ belongs to the class $\mathscr{S}_d$. Let $c(\cdot)$ denote the number of components in a composite structure. Then $c(\mathsf{S}_n)$ converges towards $1 + c(\mathsf{R})$ in total variation.*

https://doi.org/10.1017/S0963548319000336 Published online by Cambridge University Press

If we condition R_n on having a fixed size $k < n/2$, then the $\mathcal{G}$-object of the largest component gets drawn with probability proportional to its weight from $\mathscr{U}_{n-k}$. And clearly, with probability tending to 1, R has size less than $n/2$. Hence we may rephrase Theorem 3.1 as follows.

Corollary 3.4. *Suppose that the series $\tilde{\mathcal{G}}^\nu(z)$ belongs to the class $\mathscr{S}_d$. If R has size less than n, let $\hat{\mathsf{S}}_n$ denote the random unlabelled $\mathcal{F} \circ \mathcal{G}$-object constructed by drawing a $\mathcal{G}$-object $\mathsf{G}_{n-|\mathsf{R}|}$ from $\mathscr{U}_{n-|\mathsf{R}|}$ with probability proportional to its weight, and attaching it to R. If $\mathsf{R} \geqslant n$, set $\hat{\mathsf{S}}_n$ to some placeholder value $\diamond$. Then*

$$d_{\mathrm{TV}}(\mathsf{S}_n, \hat{\mathsf{S}}_n) \to 0,$$

as $n \to \infty$ with $n \equiv 0 \mod d$.

4. Proofs

Before starting with the proofs of our main results, we make an elementary observation.

Lemma 4.1. *Let $\mathcal{F}^\omega$ and $\mathcal{G}^\nu$ be weighted species with $\mathcal{G}^\nu[\emptyset] = \emptyset$, and let (S, σ) be a random symmetry that follows a $\mathbb{P}_{Z_{\mathcal{F}^\omega \circ \mathcal{G}^\nu}, (\rho^j)_j}$-distribution for some $\rho > 0$. The composite structure of S is of the form $(\mathsf{F}, (\mathsf{G}_Q)_{Q \in \pi})$ with π a partition of a finite set, F an $\mathcal{F}$-structure on π, and G_Q a $\mathcal{G}$-structure on Q for each $Q \in \pi$. As σ is an automorphism, it follows that*

$$\bar{\sigma} : \pi \to \pi, Q \mapsto \sigma(Q)$$

is a well-defined permutation of the collection π of partition classes. For each $i \geqslant 1$, let X_i denote the number of cycles of length i in the induced permutation $\bar{\sigma}$, $Y_i = iX_i$ the total number of atoms contained in cycles of length i, and Z_i the sum of sizes of all $\mathcal{G}$-objects corresponding to atoms of $\bar{\sigma}$ that are contained in cycles of length i. Then

$$\mathbb{E}\left[\prod_{i \geqslant 1} x_i^{X_i} y_i^{Y_i} z_i^{Z_i}\right] = \frac{Z_{\mathcal{F}^\omega}(x_1 y_1 \tilde{\mathcal{G}}^\nu(z_1 \rho), x_2 y_2^2 \tilde{\mathcal{G}}^{\nu^2}((z_2\rho)^2), x_3 y_3^3 \tilde{\mathcal{G}}^{\nu^3}((z_3\rho)^3), \ldots)}{\widetilde{\mathcal{F}^\omega \circ \mathcal{G}^\nu}(\rho)}.$$

Lemma 4.1 is a minor extension of the proof of the well-known enumerative formula

$$\widetilde{\mathcal{F}^\omega \circ \mathcal{G}^\nu}(z) = Z_{\mathcal{F}^\omega}(\tilde{\mathcal{G}}^\nu(z), \tilde{\mathcal{G}}^{\nu^2}(z^2), \ldots)$$

given, for example, in [14, Theorem 3 and Section 6] or [4, Proposition 11 of Section 2.3]. Instead of using a single formal variable z in the proof for counting the total size, all involved counting series may be replaced by versions with additional formal variables $(x_i, y_i, z_i)_{i \geqslant 1}$, that keep track of the required fine-grained statistics. We do not aim to go through the details. Roughly speaking, the idea behind it is that symmetries of composite $\mathcal{F} \circ \mathcal{G}$-structures correspond, up to a certain relabelling and cycle composition process, to an $\mathcal{F}$-symmetry, where each cycle τ with length $|\tau|$ gets endowed with $|\tau|$ identical copies of a $\mathcal{G}$-symmetry. Thus, in the sum

$$Z_{\mathcal{F}^\omega}(x_1 y_1 \tilde{\mathcal{G}}^\nu(z_1 z), x_2 y_2^2 \tilde{\mathcal{G}}^{\nu^2}((z_2 z)^2), x_3 y_3^3 \tilde{\mathcal{G}}^{\nu^3}((z_3 z)^3), \ldots),$$

the variable z keeps track of the total size, the x_i of the number of cycles of length i in the symmetry and consequently the y_i of the total mass of these cycles. The powers $(z_i z)^i$ are due to the fact that each $\mathcal{G}$-object assigned to a cycle with length i gets counted i times due to the identical copies corresponding to each atom of the cycle.

Proof of Lemma 3.2. Throughout, we let n denote an integer that is divisible by d. We assumed that $\mathcal{F}^\omega$ and $\mathcal{G}^\nu$ are weighted species such that the ordinary generating function $\tilde{\mathcal{G}}^\nu(z)$ belongs to $\mathscr{S}_d$. We further assumed by inequality (3.1) that

$$Z_{\mathcal{F}^\omega}(\tilde{\mathcal{G}}^{\nu}(\rho)+\varepsilon, \tilde{\mathcal{G}}^{\nu^2}((\rho+\varepsilon)^2), \tilde{\mathcal{G}}^{\nu^3}((\rho+\varepsilon)^3), \dots) < \infty \tag{4.1}$$

for some $\varepsilon > 0$, with ρ denoting the radius of convergence of the series $\tilde{\mathcal{G}}^{\nu}(z)$.

We start by constructing a $\mathbb{P}_{\widetilde{\mathcal{F}^\omega \circ \mathcal{G}^\nu},\rho}$-distributed composite structure according to Lemma 2.5. Let (F,σ) follow a $\mathbb{P}_{Z_{\mathcal{F}^\omega},(\tilde{\mathcal{G}}^{\nu}(\rho),\tilde{\mathcal{G}}^{\nu^2}(\rho^2),\dots)}$-distribution. For each cycle τ of σ, let $|\tau|$ denote its length and draw a $\mathcal{G}$-object G_τ according to a $\mathbb{P}_{\tilde{\mathcal{G}}^{\nu^{|\tau|}},\rho^{|\tau|}}$-distribution. We construct the $\mathcal{F}\circ\mathcal{G}$-object S by assigning, for each cycle τ and each atom v of τ, an identical copy of G_τ to v. Thus S corresponds to $(\mathsf{F},(\mathsf{G}_v)_v)$.

Let f denote the number of fixpoints of the permutation σ, and $\mathsf{G}_1,\dots \mathsf{G}_f$ the corresponding $\mathcal{G}$-structures. We set $g_i = |\mathsf{G}_i|$ for all i. Let H denote the structure obtained from S by deleting all $\mathcal{G}$-objects that correspond to fixpoints of σ, and let h denote the total size of its remaining $\mathcal{G}$-objects. Thus

$$|\mathsf{S}| = \sum_{i=1}^{f} g_i + h. \tag{4.2}$$

The $(g_i)_i$ are independent, but f and h may very well depend on each other. By Lemma 4.1, their joint probability generating function is given by

$$\mathbb{E}[y^f w^h] = \frac{Z_{\mathcal{F}^\omega}(y\tilde{\mathcal{G}}^{\nu}(\rho), \tilde{\mathcal{G}}^{\nu^2}(w^2\rho^2), \tilde{\mathcal{G}}^{\nu^3}(w^3\rho^3),\dots)}{\widetilde{\mathcal{F}^\omega \circ \mathcal{G}^\nu}(\rho)}. \tag{4.3}$$

Hence the assumption (4.1) states precisely that the vector (f,h) has finite exponential moments. We are going to show that

$$\mathbb{P}(|\mathsf{S}|=n) \sim \mathbb{E}[f]\mathbb{P}(g=n), \tag{4.4}$$

where g denotes the size of a $\mathbb{P}_{\tilde{\mathcal{G}}^{\nu},\rho}$-distributed random $\mathcal{G}$-object. Since equation (4.3) implies

$$\mathbb{E}[f] = \frac{(\mathrm{d}/\mathrm{d}z_1)Z_{\mathcal{F}^\omega}(\tilde{\mathcal{G}}^{\nu}(\rho), \tilde{\mathcal{G}}^{\nu^2}(\rho^2), \tilde{\mathcal{G}}^{\nu^3}(\rho^3),\dots)\tilde{\mathcal{G}}^{\nu}(\rho)}{\widetilde{\mathcal{F}^\omega \circ \mathcal{G}^\nu}(\rho)},$$

it is clear that equation (4.4) is equivalent to

$$[z^n]\widetilde{\mathcal{F}^\omega \circ \mathcal{G}^\nu}(z) \sim \widetilde{(\mathcal{F}')^\omega \circ \mathcal{G}^\nu}(\rho)[z^n]\tilde{\mathcal{G}}^{\nu}(z), \quad n\to\infty, \quad n \equiv 0 \mod d.$$

We have thus successfully reduced the task of asymptotically determining the coefficients of $\widetilde{\mathcal{F}^\omega \circ \mathcal{G}^\nu}(z)$ to the probabilistic task of verifying (4.4), and we may apply available results for subexponential probability distributions. Equation (4.2) implies that

$$\mathbb{P}(|\mathsf{S}|=n) = \mathbb{P}\left(\sum_{i=1}^{f} g_i + h = n\right) = \sum_{k\geqslant 0}\mathbb{P}(f=k)\mathbb{P}\left(\sum_{i=1}^{k} g_i + h = n \,\middle|\, f=k\right). \tag{4.5}$$

Let g denote a random variable that is distributed like the size of a $\mathbb{P}_{\tilde{\mathcal{G}}^{\nu},y}$-distributed random $\mathcal{G}$-object. Given $f=k$, the $(g_i)_{1\leqslant i\leqslant k}$ are independent and identically distributed copies of g. Lemma 2.6 implies that, for each fixed k,

$$\mathbb{P}\left(\sum_{i=1}^{k} g_i = n \,\middle|\, f=k\right) = \mathbb{P}\left(\sum_{i=1}^{k} g_i = n\right) \sim k\mathbb{P}(g=n).$$

https://doi.org/10.1017/S0963548319000336 Published online by Cambridge University Press

As the vector (f, h) has finite exponential moments, it also holds that the conditioned version $(h \mid f = k)$ has finite exponential moments. It follows from Lemma 2.7 that

$$\mathbb{P}\left(\sum_{i=1}^{k} g_i + h = n \,\middle|\, f = k\right) \sim k\mathbb{P}(g = n)$$

and hence

$$\mathbb{P}\left(\sum_{i=1}^{k} g_i + h = n, f = k\right) \sim \mathbb{P}(f = k)k\mathbb{P}(g = n).$$

Consequently, if we can find a summable sequence $(C_k)_{k\geqslant 0}$ such that

$$\mathbb{P}\left(\sum_{i=1}^{k} g_i + h = n, f = k\right) \leqslant C_k\mathbb{P}(g = n), \tag{4.6}$$

for all k, then it follows by dominated convergence that

$$\mathbb{P}(|\mathsf{S}| = n) = \sum_{k\geqslant 0} \mathbb{P}(f = k)\mathbb{P}\left(\sum_{i=1}^{k} g_i + h = n \,\middle|\, f = k\right) \sim \mathbb{E}[f]\mathbb{P}(g = n).$$

Thus, in order to show (4.4) it remains to establish inequality (4.6). By Lemma 2.6 for each $\varepsilon > 0$ there is an integer $x_0 = x_0(\varepsilon) > 0$ and a constant $c(\varepsilon) > 0$ such that, for all integers $x \geqslant x_0$ and each $k \geqslant 0$, it holds that

$$\mathbb{P}\left(\sum_{i=1}^{k} g_i = x\right) \leqslant c(\varepsilon)(1+\varepsilon)^k\mathbb{P}(g = x). \tag{4.7}$$

Clearly we have

$$\begin{aligned}\mathbb{P}&\left(\sum_{i=1}^{k} g_i + h = n, f = k\right)\\ &= \mathbb{P}\left(\sum_{i=1}^{k} g_i + h = n, h > n - x_0, f = k\right) + \mathbb{P}\left(\sum_{i=1}^{k} g_i + h = n, h \leqslant n - x_0, f = k\right).\end{aligned} \tag{4.8}$$

Since h has finite exponential moments, there are constants $C, c > 0$ such that, for all n,

$$\mathbb{P}\left(\sum_{i=1}^{k} g_i + h = n, h > n - x_0, f = k\right) \leqslant \mathbb{P}(h > n - x_0) \leqslant C\exp(-cn).$$

We know that g is heavy-tailed because it belongs to $\mathscr{S}_d$. Hence it follows that

$$\mathbb{P}\left(\sum_{i=1}^{k} g_i + h = n, h > n - x_0, f = k\right) = o(\mathbb{P}(g = n)) \tag{4.9}$$

uniformly for all $k \geqslant 0$ as n becomes large. As for the other summand in (4.8), it holds that

$$\mathbb{P}\left(\sum_{i=1}^{k} g_i + h = n, h \leqslant n - x_0, f = k\right) = \sum_{\ell=0}^{n-x_0} \mathbb{P}(h = \ell, f = k)\mathbb{P}\left(\sum_{i=1}^{k} g_i = n - \ell\right).$$

https://doi.org/10.1017/S0963548319000336 Published online by Cambridge University Press

Since $n-k \geqslant x_0$, it follows from inequality (4.7) that, for all $\varepsilon > 0$,

$$\sum_{\ell=0}^{n-x_0} \mathbb{P}(h=\ell, f=k)\mathbb{P}\left(\sum_{i=1}^{k} g_i = n-\ell\right) \leqslant c(\varepsilon) \sum_{\ell=0}^{n} \mathbb{P}(h=\ell, f=k)(1+\varepsilon)^k \mathbb{P}(g=n-\ell). \quad (4.10)$$

As the vector (f, h) has finite exponential moments, there is a $\delta > 0$ such that

$$\mathbb{E}[(1+\delta)^f(1+\delta)^w] < \infty. \quad (4.11)$$

Since $\varepsilon > 0$ was arbitrary, we may choose it small enough such that $0 < \varepsilon < \delta$. Thus

$$\frac{1+\varepsilon}{1+\delta} < 1$$

and

$$\begin{aligned} &\mathbb{P}\left(\sum_{i=1}^{k} g_i + h = n, h \leqslant n - x_0, f = k\right) \\ &\qquad \leqslant c(\varepsilon)\left(\frac{1+\varepsilon}{1+\delta}\right)^k \sum_{\ell=0}^{n} \mathbb{P}(h=\ell, f=k)(1+\delta)^k \mathbb{P}(g=n-\ell) \\ &\qquad \leqslant c(\varepsilon)\left(\frac{1+\varepsilon}{1+\delta}\right)^k \sum_{\ell=0}^{n} p_\ell \mathbb{P}(g=n-\ell), \end{aligned} \quad (4.12)$$

with

$$p_\ell = \sum_{k \geqslant 0} \mathbb{P}(h=\ell, f=k)(1+\delta)^k,$$

satisfying

$$\sum_{\ell \geqslant 0} p_\ell (1+\delta)^\ell < \infty$$

by inequality (4.11). Hence we may apply Lemma 2.7 to obtain

$$\sum_{\ell=0}^{n} p_\ell \mathbb{P}(g=n-\ell) \sim \mathbb{P}(g=n).$$

So equation (4.9) and inequality (4.12) imply that, for all $k \geqslant 0$,

$$\mathbb{P}\left(\sum_{i=1}^{k} g_i + h = n, f = k\right) \leqslant C_k \mathbb{P}(g=n)$$

for a summable sequence $(C_k)_{k \geqslant 0}$. This verifies inequality (4.6) and hence (4.4) follows by dominated convergence. □

Proof of Theorem 3.1. We use the same notation as in the proof of Lemma 3.2, that is, we let S denote a random $\mathbb{P}_{\widetilde{\mathcal{F}^\omega \circ \mathcal{G}^\nu}, \rho}$-distributed composite structure assembled according to Lemma 2.5 as follows. We sample an $\mathcal{F}$-symmetry (F, σ) following a $\mathbb{P}_{Z_{\mathcal{F}^\omega}, (\tilde{\mathcal{G}}^\nu(\rho), \tilde{\mathcal{G}}^{\nu^2}(\rho^2), \ldots)}$-distribution and let f denote the number of fixpoints of σ. We let $(\mathsf{G}_i)_{i \geqslant 1}$ denote an independent family of $\mathbb{P}_{\tilde{\mathcal{G}}^\nu, \rho}$-distributed $\mathcal{G}$-objects, of which we match the first f to the fixpoints of σ in any canonical order. For example, we may order the fixpoints according to their labels in $\{1, \ldots, |\mathsf{F}|\}$, but any canonical order will do by exchangeability of the G_i. Likewise, for each cycle τ of σ with length $|\tau| \geqslant 2$ we draw a $\mathcal{G}$-object G_τ according to a $\mathbb{P}_{\tilde{\mathcal{G}}^{\nu^{|\tau|}}, \rho^{|\tau|}}$-distribution, and assign to each atom of τ an identical

https://doi.org/10.1017/S0963548319000336 Published online by Cambridge University Press

copy of G_τ. We let H denote the structure obtained from (F, σ) by only attaching the $\mathcal{G}$-objects to atoms of cycles with length at least 2. Then S is fully described by the vector

$$(\mathsf{H}, \mathsf{G}_1, \ldots, \mathsf{G}_f).$$

It holds that

$$|\mathsf{S}| = \sum_{i=1}^{f} g_i + h,$$

where h denotes the number of atoms of H and $g_i = |\mathsf{G}_i|$ for all i. As discussed in Section 2.4, the result of conditioning a Boltzmann-distributed object on having a fixed size gets sampled with probability proportional to its weight among all objects of this size. Hence

$$\mathsf{S}_n \stackrel{d}{=} (\mathsf{S} \mid |\mathsf{S}| = n).$$

Similarly, the $\mathbb{P}_{\widetilde{(\mathcal{F}')^{\omega} \circ \mathcal{G}^{\nu}}, \rho}$-distributed $\mathcal{F}' \circ \mathcal{G}$-object R may, by virtue of Lemma 2.5, be sampled as follows. We draw an $\mathcal{F}$-symmetry (F', σ) following a $\mathbb{P}_{Z_{(\mathcal{F}')^{\omega}}, (\tilde{\mathcal{G}}^{\nu}(\rho), \tilde{\mathcal{G}}^{\nu^2}(\rho^2), \ldots)}$-distribution and let f' denote the number of fixpoints of σ'. Note that σ is a permutation of the non-$*$-atoms of F, hence we do not count the place-holder atom. We let $(\mathsf{G}^i)_{i \geqslant 1}$ denote a list of independent copies of a $\mathbb{P}_{\tilde{\mathcal{G}}^{\nu}, \rho}$-distributed $\mathcal{G}$-object, and match the first f' to the fixpoints of σ' in a canonical way. For each cycle τ of σ' with length $|\tau| \geqslant 2$ we draw a $\mathcal{G}$-object G'_τ according to a $\mathbb{P}_{\tilde{\mathcal{G}}^{\nu^{|\tau|}}, \rho^{|\tau|}}$-distribution, and assign to each atom of τ an identical copy of G_τ. We let H' denote the pruned structure where only atoms of non-fixpoints of σ' receive a $\mathcal{G}$-object. Thus R is fully determined by the vector

$$(\mathsf{H}', \mathsf{G}^1, \ldots, \mathsf{G}^{f'}),$$

and we set $g^i = |\mathsf{G}^i|$ for all i and let h' denote the number of atoms in H'. If R has size less than n, we let $\hat{\mathsf{S}}_n$ denote the result of assigning to the $*$-placeholder atom a random unlabelled $\mathcal{G}$-structure G^* sampled from $\mathscr{U}_{n-|\mathsf{R}|}(\mathcal{G})$ with probability proportional to its weight. If $\mathsf{R} \geqslant n$, we let $\hat{\mathsf{S}}_n$ assume some placeholder value $\hat{\mathsf{S}}_n = \diamond$. We are going to show that

$$d_{\mathrm{TV}}(\mathsf{S}_n, \hat{\mathsf{S}}_n) \to 0, \quad n \to \infty, \quad n \equiv 0 \mod d. \tag{4.13}$$

If $\mathsf{R} < n/2$, then G^* is the largest $\mathcal{G}$-object of $\hat{\mathsf{S}}_n$. Since R is almost surely finite, this event takes place with probability tending to 1 as n becomes large. Hence (4.13) implies that

$$d_{\mathrm{TV}}(\mathsf{R}_n, \mathsf{R}) \to 0.$$

Thus verifying (4.13) is sufficient to conclude the proof.

If we interpret F' as an $\mathcal{F}$-object F'_* (rather than an $\mathcal{F}'$-object), then the permutation σ' extends to an $\mathcal{F}$-automorphism σ'_* of F'_* such that the $*$-vertex is a fixpoint. The distributions of (F, σ) and $(\mathsf{F}'_*, \sigma'_*)$ differ in the fact that σ' always has at least one fixpoint, and that the probability to assume a fixed size is different. However, given integers $m, k \geqslant 1$, it holds that up to relabelling

$$((\mathsf{F}, \sigma) \mid f = k, |\mathsf{F}| = m) \stackrel{d}{=} ((\mathsf{F}'_*, \sigma'_*) \mid f' = k - 1, |\mathsf{F}'| = m - 1). \tag{4.14}$$

This may be verified as follows. The left-hand side gets drawn with probability proportional to its weight from the subset $A_k \subset \mathrm{Sym}(\mathcal{F})[m]$ of all symmetries with $k \geqslant 1$ fixpoints. Likewise, the right-hand side gets drawn with probability proportional to its weight from the subset $B_k \subset \mathrm{Sym}(\mathcal{F}')[m-1]$ of symmetries with k fixpoints in total (counting the $*$-atom). There is a weight-preserving bijection between $\mathrm{Sym}(\mathcal{F}')[m-1]$ and the symmetries in $\mathrm{Sym}(\mathcal{F})[m]$ where the atom m is a fixpoint. It follows that there is a weight-preserving 1 to m correspondence between $\mathrm{Sym}(\mathcal{F}')[m-1]$ and the set of symmetries in $\mathrm{Sym}(\mathcal{F})[m]$ with a distinguished fixpoint.

Now, for each symmetry in A_k, there are precisely k ways to distinguish a fixpoint, hence there is a weight-preserving 1 to km relation between A_k and B_k. Thus (4.14) holds.

Let $x_1, \dots, x_k \geqslant 1$ and $r \geqslant 0$ be given with

$$x_1 + \cdots + x_k + r = n.$$

It follows from (4.14) and the construction of H and H' that

$$(\mathsf{H} \mid f = k, h = r) \overset{d}{=} (\mathsf{H}' \mid f' = k-1, h' = r).$$

If we condition the left-hand side additionally on $g_i = x_i$ for all $1 \leqslant i \leqslant k$, then the distribution of H does not change, and G_i gets drawn from $\tilde{\mathcal{G}}[x_i]$ with probability proportional to its ν-weight. Likewise, if we condition the right-hand side additionally on $g^i = x_i$ for all $1 \leqslant i \leqslant k-1$, then for each i it holds that G^i gets drawn from $\tilde{\mathcal{G}}[x_i]$ with probability proportional to its weight, and G^* gets drawn with probability proportional to its weight among all unlabelled $\mathcal{G}$-objects with $n - r - x_1 - \cdots - x_{k-1} = x_k$ atoms. Thus

$$(\mathsf{S} \mid f = k, h = r, g_1 = x_1, \dots, g_k = x_k) \overset{d}{=} (\hat{\mathsf{S}}_n \mid f' = k-1, h' = r, g^1 = x_1, \dots, g^{k-1} = x_{k-1}). \tag{4.15}$$

We let g denote a random variable that is distributed like the size of a random $\mathcal{G}$-object with a $\mathbb{P}_{\tilde{\mathcal{G}}^\nu,\rho}$ distribution. Since $\tilde{\mathcal{G}}^\nu(z)$ belongs to $\mathscr{S}_d$, it holds that

$$\mathbb{P}(g = n + d) \sim \mathbb{P}(g = n), \quad n \to \infty.$$

This implies that there is a sequence t_n of non-negative integers such that $t_n \to \infty$ and

$$\lim_{n\to\infty} \sup_{\substack{0 \leqslant y \leqslant t_n \\ y \equiv 0 \bmod d}} |\mathbb{P}(g = n + y)/\mathbb{P}(g = n) - 1| = 0. \tag{4.16}$$

Without loss of generality we may assume $t_n < n/2$ for all n. For any sequence $\mathbf{y} = (y_1, \dots, y_{k-1})$ of positive integers, we set

$$D(\mathbf{y}) := y_1 + \cdots , y_{k-1}.$$

For each integer m with $m > D(\mathbf{y})$, we also set

$$\sigma_m(\mathbf{y}) := \{(y_1, \dots, y_{j-1}, m - D(\mathbf{y}), y_j, \dots, y_k) \mid 1 \leqslant j \leqslant k\}.$$

Finally, we set

$$M_n := \{(k, r, \mathbf{y}) \mid k \geqslant 1, r \geqslant 0, \mathbf{y} \in \mathbb{N}^{k-1}, r + D(\mathbf{y}) \leqslant t_n\}.$$

We will show that as n becomes large, it holds uniformly for all $(k, r, \mathbf{y}) \in M_n$ that

$$\begin{aligned} &\mathbb{P}(f = k, h = r, (g_1, \dots, g_k) \in \sigma_{n-r}(\mathbf{y}) \mid g_1 + \cdots + g_f + h = n) \\ &\quad \sim \mathbb{P}(f' = k-1, h' = r, (g^1, \dots, g^{k-1}) = \mathbf{y}). \end{aligned} \tag{4.17}$$

For $D(\mathbf{y}) + r \leqslant t_n < n/2$, the $(g_1, \dots, g_k) \in \sigma_n(\mathbf{y})$ corresponds to k distinct outcomes, depending on the unique location for the maximum of the g_i. Thus the left-hand side in (4.17) divided by the right-hand side equals

$$\frac{k\mathbb{P}(f = k, h = r)\mathbb{P}(g = n - D(\mathbf{y}) - r)}{\mathbb{P}(f' = k-1, h' = r)\mathbb{P}(g_1 + \cdots + g_f + h = n)}.$$

Note that

$$\frac{k\mathbb{P}(f = k, h = r)}{\mathbb{P}(f' = k-1, h' = r)} = \frac{\tilde{\mathcal{G}}^\nu(\rho)\widetilde{(\mathcal{F}')^\omega \circ \mathcal{G}^\nu}(\rho)}{\widetilde{\mathcal{F}^\omega \circ \mathcal{G}^\nu}(\rho)} = \mathbb{E}[f].$$

By Lemma 3.2, it holds that

$$\mathbb{P}(g_1 + \cdots + g_f + h = n) \sim \mathbb{E}[f]\mathbb{P}(g = n).$$

Equation (4.16) and $D(\mathbf{y}) + r \leqslant t_n$ yield that

$$\mathbb{P}(g = n - D(\mathbf{y}) - r) \sim \mathbb{P}(g = n)$$

uniformly for $(k, r, \mathbf{y}) \in M_n$. This verifies the asymptotic equality in (4.17).

As $t_n \to \infty$, it clearly holds that

$$(f' + 1, r', (g^1, \ldots, g^{f'})) \in M_n$$

with probability tending to 1 as n becomes large. Hence it follows from (4.17) that

$$\mathbb{P}((f, h, (g_1, \ldots, g_k)) \in \{\{(k, r)\} \times \sigma_{n-r}(\mathbf{y}) \mid (k, r, \mathbf{y}) \in M_n\}) \to 1$$

as n becomes large. Thus, we have that uniformly for all sets $\mathcal{E}$ of n-sized unlabelled $\mathcal{F} \circ \mathcal{G}$-objects

$$\begin{aligned}\mathbb{P}(\mathsf{S}_n \in \mathcal{E}) &= \mathbb{P}(\mathsf{S} \in \mathcal{E} \mid g_1 + \cdots + g_f + h = n) \\ &= o(1) + \sum_{(k,r,\mathbf{y}) \in M_n} \frac{\mathbb{P}(\mathsf{S} \in \mathcal{E}, (f, h) = (k, r), (g_1, \ldots, g_k) \in \sigma_n(\mathbf{y}))}{\mathbb{P}(g_1 + \cdots + g_f + h = n)}.\end{aligned}$$

The summand for $(k, r, \mathbf{y})$ may be expressed by the product

$$\begin{aligned}&\mathbb{P}(\mathsf{S} \in \mathcal{E} \mid (f, h) = (k, r), (g_1, \ldots, g_k) \in \sigma_n(\mathbf{y})) \\ &\quad \mathbb{P}((f, h) = (k, r), (g_1, \ldots, g_k) \in \sigma_n(\mathbf{y}) \mid g_1 + \cdots + g_f + h = n).\end{aligned}$$

Equation (4.15) yields that the first factor is equal to

$$\mathbb{P}(\hat{\mathsf{S}}_n \in \mathcal{E} \mid f' = k - 1, h' = r, (g^1, \ldots, g^{k-1}) = \mathbf{y}).$$

By equation (4.17), the second factor is asymptotically equivalent to

$$\mathbb{P}(f' = k - 1, h' = r, (g^1, \ldots, g^{k-1}) = \mathbf{y})$$

uniformly for all $(k, r, \mathbf{y}) \in M_n$ as n becomes large. Thus

$$\mathbb{P}(\mathsf{S}_n \in \mathcal{E}) = o(1) + \sum_{(k,r,\mathbf{y}) \in M_n} \mathbb{P}(\hat{\mathsf{S}}_n \in \mathcal{E}, (f', h', (g^i)_i) = (k - 1, r, \mathbf{y})) = o(1) + \mathbb{P}(\hat{\mathsf{S}}_n \in \mathcal{E}).$$

This completes the proof. □

Acknowledgement

I warmly thank the referee for helpful comments and suggestions.

References

[1] Arratia, R., Barbour, A. D. and Tavaré, S. (2003) *Logarithmic Combinatorial Structures: A Probabilistic Approach*, EMS Monographs in Mathematics, European Mathematical Society.

[2] Barbour, A. D. and Granovsky, B. L. (2005) Random combinatorial structures: The convergent case. *J. Combin. Theory Ser. A* **109** 203–220.

[3] Bell, J. P., Bender, E. A., Cameron, P. J. and Richmond, L. B. (2000) Asymptotics for the probability of connectedness and the distribution of number of components. *Electron. J. Combin.* **7** R3.

[4] Bergeron, F., Labelle, G. and Leroux, P. (1998) *Combinatorial Species and Tree-like Structures*, Vol. 67 of Encyclopedia of Mathematics and its Applications, Cambridge University Press.

[5] Bodirsky, M., Fusy, É., Kang, M. and Vigerske, S. (2011) Boltzmann samplers, Pólya theory, and cycle pointing. *SIAM J. Comput.* **40** 721–769.

[6] Chover, J., Ney, P. and Wainger, S. (1973) Functions of probability measures. *J. Analyse Math.* **26** 255–302.

[7] Duchon, P., Flajolet, P., Louchard, G. and Schaeffer, G. (2004) Boltzmann samplers for the random generation of combinatorial structures *Combin. Probab. Comput.* **13** 577–625.

[8] Embrechts, P. (1983) The asymptotic behaviour of series and power series with positive coefficients. *Med. Konink. Acad. Wetensch. België* **45** 41–61.

[9] Embrechts, P. and Omey, E. (1984) Functions of power series. *Yokohama Math. J.* **32** 77–88.

[10] Erlihson, M. M. and Granovsky, B. L. (2008) Limit shapes of Gibbs distributions on the set of integer partitions: The expansive case. *Ann. Inst. Henri Poincaré Probab. Statist.* **44** 915–945.

[11] Flajolet, P., Fusy, É. and Pivoteau, C. (2007) Boltzmann sampling of unlabelled structures. In *Proceedings of the Ninth Workshop on Algorithm Engineering and Experiments and the Fourth Workshop on Analytic Algorithmics and Combinatorics*, SIAM, pp. 201–211.

[12] Flajolet, P. and Sedgewick, R. (2009) *Analytic Combinatorics*, Cambridge University Press.

[13] Foss, S., Korshunov, D. and Zachary, S. (2013) *An Introduction to Heavy-Tailed and Subexponential Distributions*, second edition, Springer Series in Operations Research and Financial Engineering, Springer.

[14] Joyal, A. (1981) Une théorie combinatoire des séries formelles. *Adv. Math.* **42** 1–82.

[15] McDiarmid, C. (2008) Random graphs on surfaces. *J. Combin. Theory Ser. B* **98** 778–797.

[16] McDiarmid, C. (2009) Random graphs from a minor-closed class. *Combin. Probab. Comput.* **18** 583–599.

[17] Mutafchiev, L. (1998) The largest tree in certain models of random forests. *Random Struct. Alg.* **13** 211–228.

[18] Pitman, J. (2006) *Combinatorial Stochastic Processes*, Vol. 1875 of Lecture Notes in Mathematics, Springer.

[19] Stufler, B. (2017) Asymptotic properties of random unlabelled block-weighted graphs. arXiv:1712.01301

[20] Stufler, B. (2018) Gibbs partitions: The convergent case. *Random Struct. Alg.* **53** 537–558.

Cite this article: Stufler B (2020). Unlabelled Gibbs partitions. *Combinatorics, Probability and Computing* **29**, 293–309. https://doi.org/10.1017/S0963548319000336