RANDOM COGRAPHS: BROWNIAN GRAPHON LIMIT
AND ASYMPTOTIC DEGREE DISTRIBUTION

FRÉDÉRIQUE BASSINO, MATHILDE BOUVEL, VALENTIN FÉRAY, LUCAS GERIN,
MICKAËL MAAZOUN, AND ADELINE PIERROT

ABSTRACT. We consider uniform random cographs (either labeled or unlabeled) of large size. Our first main result is the convergence towards a Brownian limiting object in the space of graphons. We then show that the degree of a uniform random vertex in a uniform cograph is of order n, and converges after normalization to the Lebesgue measure on $[0, 1]$. We finally analyze the vertex connectivity (*i.e.* the minimal number of vertices whose removal disconnects the graph) of random connected cographs, and show that this statistics converges in distribution without renormalization. Unlike for the graphon limit and for the degree of a random vertex, the limiting distribution of the vertex connectivity is different in the labeled and unlabeled settings.

Our proofs rely on the classical encoding of cographs via cotrees. We then use mainly combinatorial arguments, including the symbolic method and singularity analysis.

1. INTRODUCTION

1.1. **Motivation.** Random graphs are arguably the most studied objects at the interface of combinatorics and probability theory. One aspect of their study consists in analyzing a uniform random graph of large size n in a prescribed family, *e.g.* perfect graphs [MY19], planar graphs [Noy14], graphs embeddable in a surface of given genus [DKS17], graphs in subcritical classes [PSW16], hereditary classes [HJS18] or addable classes [MSW06, CP19]. The present paper focuses on uniform random *cographs* (both in the labeled and unlabeled settings).

Cographs were introduced in the seventies by several authors independently, see *e.g.* [Sei74] and further references on the Wikipedia page [Wiki]. They enjoy several equivalent characterizations. Among others, cographs are

- the graphs avoiding P_4 (the path with four vertices) as an induced subgraph;
- the graphs which can be constructed from graphs with one vertex by taking disjoint unions and joins;
- the graphs whose modular decomposition does not involve any prime graph;
- the inversion graphs of separable permutations.

Cographs have been extensively studied in the algorithmic literature. They are recognizable in linear time [CPS84, HP05, BCH+08] and many computationally hard problems on general graphs are solvable in polynomial time when restricted to cographs; see [CLS81] and several

subsequent works citing this article. In these works, as well as in the present paper, a key ingredient is the encoding of cographs by some trees, called *cotrees*. These cotrees witness the construction of cographs using disjoint unions and joins (mentioned in the second item above).

To our knowledge, cographs have however not been studied from a probabilistic perspective so far. Our motivation to the study of random cographs comes from our previous work [BBF+18, BBF+20a, BBF+20b, BBF+19] which exhibits a Brownian limiting object for separable permutations (and various other permutation classes). The first main result of this paper (Theorem 1.1) is the description of a Brownian limit for cographs. Although cographs are the inversion graphs of separable permutations, this result is not a consequence of the previous one on permutations: indeed the inversion graph is not an injective mapping, hence a uniform cograph is not the cograph of a uniform separable permutation.

Our convergence result holds in the space of *graphons*. Graphon convergence has been introduced in [BCL+08] and has since then been a major topic of interest in graph combinatorics – see [Lov12] for a broad perspective on the field. The question of studying graphon limits of uniform random graphs (either labeled or unlabeled) in a given class is raised by Janson in [Jan16] (see Remark 1.6 there). Some general results have been recently obtained for hereditary[1] classes in [HJS18]. However, these results (in particular Theorem 3 in [HJS18]) do not apply to cographs, since the class of cographs contain $e^{o(n^2)}$ graphs of size n.

The graphon limit of cographs found here, which we call the *Brownian cographon*, is constructed from a Brownian excursion. By analogy with the realm of permutations [BBF+20a, BBF+20b], we expect that the Brownian cographon (or a one-parameter deformation of it) is a universal limiting object for uniform random graphs in classes of graphs which are small[2] and closed under the substitution operation at the core of the modular decomposition.

1.2. **Main results.**

From now on, for every $n \geq 1$, we let $\boldsymbol{G}_n$ and $\boldsymbol{G}_n^u$ be uniform random labeled and unlabeled cographs of size n, respectively. It is classical (see also Definition 3.2 below) to associate with any graph a graphon, and we denote by $W_{\boldsymbol{G}_n}$ and $W_{\boldsymbol{G}_n^u}$ the graphons associated with $\boldsymbol{G}_n$ and $\boldsymbol{G}_n^u$.

We note that the graphons associated with a labeled graph and its unlabeled version are the same. However, $W_{\boldsymbol{G}_n}$ and $W_{\boldsymbol{G}_n^u}$ have different distributions, since the number of possible labelings of an unlabeled cograph of a given size varies (see Fig. 3 p.8 for an illustration).

Theorem 1.1. *We have the following convergences in distribution as n tends to $+\infty$:*

$$W_{\boldsymbol{G}_n} \to W^{1/2}, \qquad W_{\boldsymbol{G}_n^u} \to W^{1/2},$$

where $W^{1/2}$ is the Brownian cographon introduced below in Definition 4.2.

Roughly speaking, the graphon convergence is the convergence of the rescaled adjacency matrix with an unusual metric, the *cut metric*, see Section 3.1. To illustrate Theorem 1.1, we show on Fig. 1 the adjacency matrix of a large random uniform labeled cograph. Entries 1 in the matrix are represented as black dots, entries 0 as white dots. It was obtained by using the encoding of

---

[1]A class of graphs is hereditary if any induced subgraph of a graph in the class is in the class as well.

[2]A class of labeled (resp. unlabeled) graphs is *small* when its exponential (resp. ordinary) generating series has positive radius of convergence.

cographs by cotrees and sampling a large uniform cotree using Boltzmann sampling [DFL+04] of the equation (9), p. 20. Note that the order of vertices in the axis of Fig. 1 is not the order of labels but is given by the depth-first search of the associated cotree. The fractal aspect of the figure – appearance of imbricated squares at various scale – is consistent with the convergence to a Brownian limiting object, since the Brownian excursion enjoys some self-similarity properties.
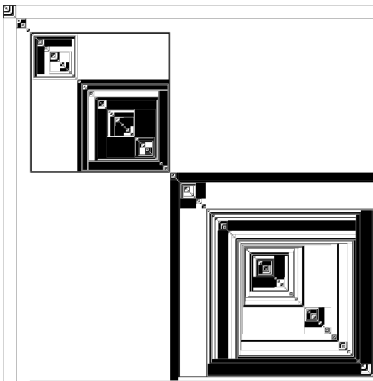

FIGURE 1. The adjacency matrix of a uniform labeled random cograph of size 4482.

We now present further results. It is well-known that the graphon convergence entails the convergence of many graph statistics, like subgraph densities, the spectrum of the adjacency matrix, the normalized degree distribution (see [Lov12] and Section 3 below). Hence, Theorem 1.1 implies that these statistics have the same limit in the labeled and unlabeled cases, and that this limit may (at least in principle) be described in terms of the Brownian cographon. Among these, the degree distribution of the Brownian cographon (or to be precise, its *intensity*[3]) is surprisingly nice: it is simply the Lebesgue measure on $[0, 1]$. We therefore have the following result, where we denote by $\deg_G(v)$ the degree of a vertex v in a graph G.

Theorem 1.2. *For every $n \geq 1$, let $\boldsymbol{v}$ and $\boldsymbol{v}^u$ be uniform random vertices in $\boldsymbol{G}_n$ and $\boldsymbol{G}_n^u$, respectively. We have the following convergences in distribution as n tends to $+\infty$:*

$$\tfrac{1}{n} \deg_{\boldsymbol{G}_n}(\boldsymbol{v}) \to U, \qquad \tfrac{1}{n} \deg_{\boldsymbol{G}_n^u}(\boldsymbol{v}^u) \to U,$$

where U is a uniform random variable in $[0, 1]$.

On the other hand, other graph statistics are not continuous for the graphon topology, and therefore can have different limits in the labeled and unlabeled cases. We illustrate this phenomenon with the *vertex connectivity* κ (defined as the minimum number of vertices whose removal disconnects the graph). Our third result is the following.

---

[3]The degree distribution of a graphon is a measure, and therefore that of the Brownian cographon is a *random* measure. Following Kallenberg [Kal17, Chapter 2], we call the *intensity* of a random measure $\boldsymbol{\mu}$ the (deterministic) measure $I[\boldsymbol{\mu}]$ defined by $I[\boldsymbol{\mu}](A) = \mathbb{E}[\boldsymbol{\mu}(A)]$ for all measurable sets A. In other words, we consider here the "averaged" degree distribution of the Brownian cographon, where we average on all realizations of the Brownian cographon.

Theorem 1.3. *There exist different probability distributions $(\pi_j)_{j\geq 1}$ and $(\pi_j^u)_{j\geq 1}$ such that, for every fixed $j \geq 1$, as n tends to $+\infty$, we have*

(1) $$\mathbb{P}(\kappa(\boldsymbol{G}_n) = j) \to \pi_j, \qquad \mathbb{P}(\kappa(\boldsymbol{G}_n^u) = j) \to \pi_j^u.$$

Formulas for π_j and π_j^u are given in Theorem 7.2.

Remark 1.4. A part of these results (Theorem 1.1) has been independently derived in [Stu19] during the preparation of this paper. The proof method is however different.

1.3. **Proof strategy.** We first discuss the proof of Theorem 1.1. For any graphs g and G of size k and n respectively, we denote by $\mathsf{Dens}(g, G)$ the number of copies of g in G as induced subgraphs normalized by n^k. Equivalently, let $\vec{V}^k$ be a k-tuple of i.i.d. uniform random vertices in G, then $\mathsf{Dens}(g, G) = \mathbb{P}(\mathrm{SubGraph}(\vec{V}^k, G) = g)$, where $\mathrm{SubGraph}(I, G)$ is the subgraph of G induced by the vertices of I. (All subgraphs in this article are induced subgraphs, and we sometimes omit the word "induced".)

From a theorem of Diaconis and Janson [DJ08, Theorem 3.1], the graphon convergence of any sequence of random graphs $(\boldsymbol{H}_n)$ is characterized by the convergence of $\mathbb{E}[\mathsf{Dens}(g, \boldsymbol{H}_n)]$ for all graphs g. In the case of $\boldsymbol{G}_n$ (the uniform random labeled cographs of size n), for any graph g of size k, we have

$$\mathbb{E}[\mathsf{Dens}(g, \boldsymbol{G}_n)] = \frac{\left|\left\{(G, I) : \begin{smallmatrix} G=(V,E) \text{ labeled cograph of size } n, \\ I \in V^k \text{ and } \mathrm{SubGraph}(I,G)=g \end{smallmatrix} \right\}\right|}{|\{G \text{ labeled cograph of size } n\}| \cdot n^k},$$

and a similar formula holds in the unlabeled case.

Both in the labeled and unlabeled cases, the asymptotic behavior of the denominator follows from the encoding of cographs as cotrees, standard application of the symbolic method of Flajolet and Sedgewick [FS09] and singularity analysis (see Propositions 5.4 and 6.5). The same methods can be used to determine the asymptotic behavior of the numerator, counting cotrees with marked leaves inducing a given subtree. This requires more involved combinatorial decompositions, which are performed in Sections 5 and 6.

We note that we already used a similar proof strategy in the framework of permutations in [BBF+20a]. The adaptation to the case of *labeled* cographs does not present major difficulties. The *unlabeled* case is however more subtle, since we have to take care of symmetries when marking leaves in cotrees (see the discussion in Section 6.1 for details). We overcome this difficulty using the $n!$-to-1 mapping that maps a pair (G, a) (where G is a labeled cograph and a an automorphism of G) to the unlabeled version of G. We then make combinatorial decompositions of such pairs (G, a) with marked vertices inducing a given subgraph (or more precisely, of the associated cotrees, with marked leaves inducing a given subtree). Our analysis shows that symmetries have a negligeable influence on the asymptotic behavior of the counting series. This is similar – though we have a different and more combinatorial presentation – to the techniques developed in the papers [PS18, GJW18], devoted to the convergence of unordered unlabeled trees to the *Brownian Continuum Random Tree*.

With Theorem 1.1 in our hands, proving Theorem 1.2 amounts to proving that the intensity of the degree distribution of the Brownian cographon is the Lebesgue measure on $[0, 1]$. Rather than

working in the continuous, we exhibit a discrete approximation $\boldsymbol{G}_n^b$ of the Brownian cographon, which has the remarkable property that the degree of a uniform random vertex in $\boldsymbol{G}_n^b$ is *exactly* distributed as a uniform random variable in $\{0, 1, \cdots, n-1\}$. The latter is proved by purely combinatorial arguments (see Proposition 4.5).

To prove Theorem 1.3, we start with a simple combinatorial lemma, which relates the vertex connectivity of a connected cograph to the sizes of the subtrees attached to the root in its cotree. Based on that, we can use again the symbolic method and singularity analysis as in the proof of Theorem 1.1.

1.4. **Outline of the paper.** Section 2 explains the standard encoding of cographs by cotrees and the relation between taking induced subgraphs and subtrees. Section 3 presents the necessary material on graphons; results stated there are quoted from the literature, except the continuity of the degree distribution, for which we could not find a reference. Section 4 introduces the limit object of Theorem 1.1, namely the Brownian cographon. It is also proved that the intensity of its degree distribution is uniform (which is the key ingredient for Theorem 1.2). Theorems 1.1 and 1.2 are proved in Section 5 for the labeled case and in Section 6 for the unlabeled case. Finally, Theorem 1.3 is proved in Section 7.

2. COGRAPHS, COTREES AND INDUCED SUBGRAPHS

2.1. **Terminology and notation for graphs.** All graphs considered in this paper are *simple* (*i.e.* without multiple edges, nor loops) and not directed. A *labeled graph* G is a pair (V, E), where V is its vertex set (consisting of distinguishable vertices, each identified by its label) and E is its edge set. Two labeled graphs (V, E) and (V', E') are isomorphic if there exists a bijection from V to V' which maps E to E'. Equivalence classes of labeled graphs for the above relation are *unlabeled graphs*.

Throughout this paper, the *size* of a graph is its number of vertices. Note that there are finitely many unlabeled graphs with n vertices, so that the uniform random unlabeled graph of size n is well defined. For labeled graphs, there are finitely many graphs with any given vertex set V. Hence, to consider a uniform random labeled graph of size n, we need to fix a vertex set V of size n. The properties we are interested in do not depend on the choice of this vertex set, so that we can choose V arbitrarily, usually $V = \{1, \ldots, n\}$.

As a consequence, considering a subset (say $\mathcal{C}$) of the set of all graphs, we can similarly define the *uniform random unlabeled graph of size n in $\mathcal{C}$* (resp. the uniform random labeled graph with vertex set $\{1, \ldots, n\}$ in $\mathcal{C}$ – which we simply denote by *uniform random labeled graph of size n in $\mathcal{C}$*). The restricted family of graphs considered in this paper is that of *cographs*.

2.2. **Cographs and cotrees.** Let $G = (V, E)$ and $G' = (V', E')$ be labeled graphs with disjoint vertex sets. We define their *disjoint union* as the graph $(V \uplus V', E \uplus E')$ (the symbol $\uplus$ denoting as usual the disjoint union of two sets). We also define their *join* as the graph $(V \uplus V', E \uplus E' \uplus (V \times V'))$: namely, we take copies of G and G', and add all edges from a vertex of G to a vertex of G'. Both definitions readily extend to more than two graphs (adding edges between any two vertices originating from different graphs in the case of the join operation).

Definition 2.1. *A* labeled cograph *is a labeled graph that can be generated from single-vertex graphs applying join and disjoint union operations. An* unlabeled cograph *is the underlying unlabeled graph of a labeled cograph.*

It is classical to encode cographs by their *cotrees*.

Definition 2.2. *A labeled* cotree *of size n is a rooted tree t with n leaves labeled from 1 to n such that*

- *t is not plane (i.e. the children of every internal node are not ordered);*
- *every internal node has at least two children;*
- *every internal node in t is decorated with a 0 or a 1.*

An unlabeled cotree *of size n is a labeled cotree of size n where we forget the labels on the leaves.*

Remark 2.3. In the literature, cotrees are usually required to satisfy the property that decorations 0 and 1 should alternate along each branch from the root to a leaf. In several proofs, our work needs also to consider trees in which this alternation assumption is relaxed, hence the choice of diverging from the usual terminology. Cotrees which do satisfy this alternation property are denoted *canonical cotrees* in this paper (see Definition 2.4).

For an unlabeled cotree t, we denote by $\mathsf{Cograph}(t)$ the unlabeled graph defined recursively as follows (see an illustration in Fig. 2):

- If t consists of a single leaf, then $\mathsf{Cograph}(t)$ is the graph with a single vertex.
- Otherwise, the root of t has decoration 0 or 1 and has subtrees t_1, ..., t_d attached to it ($d \geq 2$). Then, if the root has decoration 0, we let $\mathsf{Cograph}(t)$ be the *disjoint union* of $\mathsf{Cograph}(t_1)$, ..., $\mathsf{Cograph}(t_d)$. Otherwise, when the root has decoration 1, we let $\mathsf{Cograph}(t)$ be the *join* of $\mathsf{Cograph}(t_1)$, ..., $\mathsf{Cograph}(t_d)$.

Note that the above construction naturally entails a one-to-one correspondence between the leaves of the cotree t and the vertices of its associated graph $\mathsf{Cograph}(t)$. Therefore, it maps the size of a cotree to the size of the associated graph. Another consequence is that we can extend the above construction to a *labeled* cotree t, and obtain a *labeled* graph (also denoted $\mathsf{Cograph}(t)$), with vertex set $\{1, \ldots, n\}$: each vertex of $\mathsf{Cograph}(t)$ receives the label of the corresponding leaf of t.

By construction, for all cotrees t, the graph $\mathsf{Cograph}(t)$ is a cograph. Conversely, each cograph can be obtained in this way, albeit not from a unique tree t. It is however possible to find a canonical cotree representing a cograph G. This tree was first described in [CLS81]. The presentation of [CLS81], although equivalent, is however a little bit different, since cographs are generated using exclusively "complemented unions" instead of disjoint unions and joins. The presentation we adopt has since been used in many algorithmic papers, see *e.g.* [HP05, BCH+08].

Definition 2.4. *A cotree is* canonical *if every child of a node decorated by 0 (resp. 1) is either decorated by 1 (resp. 0) or a leaf.*

Proposition 2.5. *Let G be a labeled (resp. unlabeled) cograph. Then there exists a unique labeled (resp. unlabeled) canonical cotree t such that $\mathsf{Cograph}(t) = G$.*

Example of cographs and their canonical cotree are given in Figs. 2 and 3.

From a graph G, the canonical cotree t such that $\mathsf{Cograph}(t) = G$ is recursively built as follows. If G consists of a single vertex, t is the unique cotree with a single leaf. If G has at least two vertices, we distinguish cases depending on whether G is connected or not.

- If G is not connected, the root of t is decorated with 0 and the subtrees attached to it are the cographs associated with the connected components of G.
- If G is connected, the root of t is decorated with 1 and the subtrees attached to it are the cographs associated with the induced subgraphs of G whose vertex sets are those of the connected components of $\bar{G}$, where $\bar{G}$ is the complement of G (graph on the same vertices with complement edge set).

Important properties of cographs which justify the correctness of the above construction are the following: cographs are stable by induced subgraph and by complement, and a cograph G of size at least two is not connected exactly when its complement $\bar{G}$ is connected.
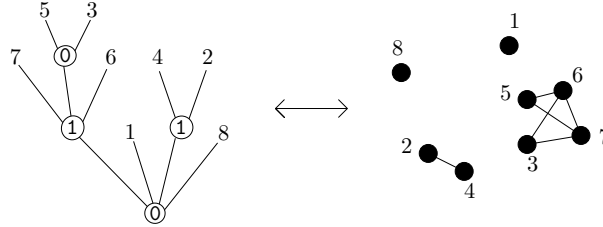

FIGURE 2. Left: A labeled canonical cotree t with 8 leaves. Right: The associated labeled cograph $\mathsf{Cograph}(t)$ of size 8.

2.3. **Subgraphs and induced trees.** Let G be a graph of size n (which may or not be labeled), and let $I = (v_1, \ldots, v_k)$ be a k-tuple of vertices of G. Recall that the *subgraph of G induced by I*, which we denote by $\mathrm{SubGraph}(I, G)$, is the graph with vertex set $\{v_1, \ldots, v_k\}$ and which contains the edge $\{v_i, v_j\}$ if and only if $\{v_i, v_j\}$ is an edge of G. In case of repetitions of vertices in I, we take as many copies of each vertex as times it appears in I and do not connect copies of the same vertex. We always regard $\mathrm{SubGraph}(I, G)$ as an unlabeled graph.

In the case of cographs, the (induced) subgraph operation can also be realized on the cotrees, through *induced trees*, which we now present. We start with a preliminary definition.

Definition 2.6 (First common ancestor). *Let t be a rooted tree, and u and v be two nodes (internal nodes or leaves) of t. The* first common ancestor *of u and v is the node furthest away from the root $\varnothing$ that appears on both paths from $\varnothing$ to u and from $\varnothing$ to v in t.*

For any cograph G, and any vertices i and j of G, the following simple observation allows to read in any cotree encoding G if $\{i, j\}$ is an edge of G.

Observation 2.7. Let $i \neq j$ be two leaves of a cotree t and $G = \mathsf{Cograph}(t)$. We also denote by i and j the corresponding vertices in G. Let v be the first common ancestor of i and j in t. Then $\{i, j\}$ is an edge of G if and only if v has label 1 in t.

FIGURE 3. All unlabeled cographs of size 4 with their corresponding (unlabeled) canonical cotrees and their number of distinct labelings.

Definition 2.8 (Induced cotree). *Let t be a cotree (which may or not be labeled), and $I = (\ell_1, \ldots, \ell_k)$ a k-tuple of distinct leaves of t, which we call the* marked leaves *of t. The tree induced by (t, I), denoted t_I, is the **always labeled** cotree of size k defined as follows. The tree structure of t_I is given by*

- *the leaves of t_I are the marked leaves of t;*
- *the internal nodes of t_I are the nodes of t that are first common ancestors of two (or more) marked leaves;*
- *the ancestor-descendant relation in t_I is inherited from the one in t;*
- *the decoration of an internal node v of t_I is inherited from the one in t;*
- *for each $i \le k$, the leaf of t_I corresponding to leaf ℓ_i in t is labeled i in t_I.*

We insist on the fact that we **always** define the induced cotree t_I as a **labeled** cotree, regardless of whether the original cotree t is labeled or not. The labeling of the induced cotree is related to the order of the marked leaves I (and not to their labels in the case t was labeled); this echoes

the choice of I as k-tuple (*i.e.* an *ordered* collection) of distinct leaves. A detailed example of the induced cotree construction is given in Fig. 4.
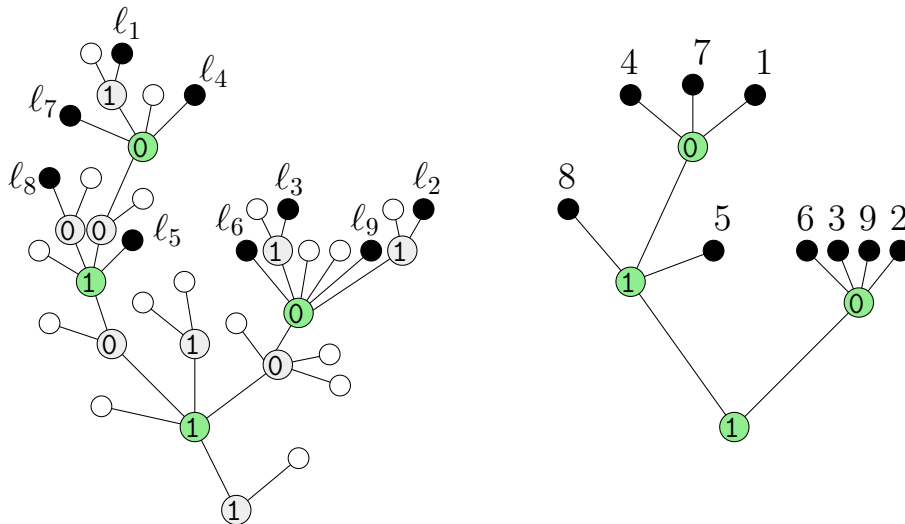

FIGURE 4. On the left: A cotree t of size $n = 26$, where leaves are indicated both by ○ and ●. We also fix a 9-tuple $I = (\ell_1, \ldots, \ell_9)$ of marked leaves (indicated by ●). In green, we indicate the internal nodes of t which are first common ancestors of these 9 marked leaves. On the right: The labeled cotree t_I induced by the 9 marked leaves.

Proposition 2.9. *Let t be a cotree and $G = \mathsf{Cograph}(t)$ the associated cograph. Let I be a k-tuple of distinct leaves in t, which identifies a k-tuple of distinct vertices in G. Then, as unlabeled graphs, we have $\mathrm{SubGraph}(I, G) = \mathsf{Cograph}(t_I)$.*

Proof. This follows immediately from Observation 2.7 and the fact that the induced cotree construction (Definition 2.8) preserves first common ancestors and their decorations. □

Remark 2.10. The reader might be surprised by the fact that we choose induced subtrees to be labeled, while induced subgraphs are not. The reasons are the following. The labeling of induced subtrees avoids symmetry problems when decomposing cotrees with marked leaves inducing a given subtree, namely in Theorems 5.3 and 6.4. On the other hand, the theory of graphons is suited to consider unlabeled subgraphs.

3. GRAPHONS

Graphons are continuum limit objects for sequences of graphs. We present here the theory relevant to our work. We recall basic notions from the literature, following mainly Lovász' book [Lov12], then we recall results of Diaconis and Janson [DJ08] regarding convergence of random graphs to random graphons. Finally, we prove a continuity result for the degree distribution with respect to graphon convergence.

The theory of graphons classically deals with unlabeled graphs. Consequently, unless specified otherwise, all graphs in this section are considered unlabeled. When considering labeled graphs, graphon convergence is to be understood as the convergence of their unlabeled versions.

3.1. **The space of graphons.**

Definition 3.1. *A graphon is an equivalence class of symmetric functions $[0,1]^2 \to [0,1]$, under the equivalence relation $\sim$, where $w \sim u$ if there exists an invertible measurable and Lebesgue measure preserving function $\phi : [0,1] \to [0,1]$ such that $w(\phi(x), \phi(y)) = u(x,y)$ for almost every $x, y \in [0,1]$.*

Intuitively, a graphon is a continuous analogue of the adjacency matrix of a graph, viewed up to relabelings of its continuous vertex set.

Definition 3.2. *The graphon W_G associated to a labeled graph G with n vertices (labeled from 1 to n) is the equivalence class of the function $w_G : [0,1]^2 \to [0,1]$ where*

$$w_G(x,y) = A_{\lceil nx \rceil, \lceil ny \rceil} \in \{0,1\}$$

and A is the adjacency matrix of the graph G.

Since any relabeling of the vertex set of G gives the same graphon W_G, the above definition immediately extends to unlabeled graphs.

We now define the so-called *cut metric*, first on functions, and then on graphons. We note that it is different than usual metrics on spaces of functions (L^1, supremum norms, . . .), see [Lov12, Chapter 8] for details. For a real-valued symmetric function w on $[0,1]^2$, its *cut norm* is defined as

$$\|w\|_\square = \sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} w(x,y) dx dy \right|$$

Identifying as usual functions equal almost-everywhere, this is indeed a norm. It induces the following *cut distance* on the space of graphons

$$\delta_\square(W, W') = \inf_{w \in W, w' \in W'} \|w - w'\|_\square.$$

While the symmetry and triangular inequalities are immediate, this "distance" $\delta_\square$ *does not* separate points, *i.e.* there exist different graphons at distance zero. Call $\widetilde{\mathcal{W}_0}$ the space of graphons, quotiented by the equivalence relation $W \equiv W'$ if $\delta_\square(W, W') = 0$. This is a metric space with distance $\delta_\square$. This definition is justified by the following deep result, see, *e.g.*, [Lov12, Theorem 9.23].

Theorem 3.3. *The metric space $(\widetilde{\mathcal{W}_0}, \delta_\square)$ is compact.*

In the sequel, we think of graphons as elements in $\widetilde{\mathcal{W}_0}$ and convergences of graphons are to be understood with respect to the distance $\delta_\square$.

3.2. **Subgraph densities and samples.**

An important feature of graphons is that one can extend the notion of density of a given subgraph g in a graph G to density in a graphon W. Moreover, convergence of graphons turns to be equivalent to convergence of all subgraph densities.

To present this, we start by recalling from the introduction the definition of subgraph densities in graphs. We recall that, if I is a tuple of vertices of G, then we write $\mathrm{SubGraph}(I, G)$ for the induced subgraph of G on vertex set I.

Definition 3.4 (Density of subgraphs). *The density of an (unlabeled) graph g of size k in a graph G of size n (which may or not be labeled) is defined as follows: let $\vec{V}^k$ be a k-tuple of i.i.d. uniform random vertices in G, then*

$$\mathsf{Dens}(g, G) = \mathbb{P}(\mathrm{SubGraph}(\vec{V}^k, G) = g).$$

We now extend this to graphons. Consider a graphon W and one of its representatives w. We denote by $\mathrm{Sample}_k(W)$ the unlabeled random graph built as follows: $\mathrm{Sample}_k(W)$ has vertex set $\{v_1, v_2, \ldots, v_k\}$ and, letting $\vec{X}^k = (X_1, \ldots, X_k)$ be i.i.d. uniform random variables in $[0, 1]$, we connect vertices v_i and v_j with probability $w(X_i, X_j)$ (these events being independent, conditionally on $(X_1, \cdots, X_k)$). Since the X_i's are independent and uniform in $[0, 1]$, the distribution of this random graph is the same if we replace w by a function $w' \sim w$ in the sense of Definition 3.1. It turns out that this distribution also stays the same if we replace w by a function w' such that $\|w - w'\|_\square = 0$ (it can be seen as a consequence of Theorem 3.6 below), so that the construction is well-defined on graphons.

Definition 3.5. *The density of a graph $g = (V, E)$ of size k in a graphon W is*

$$\mathsf{Dens}(g, W) = \mathbb{P}(\mathrm{Sample}_k(W) = g)$$

$$= \int_{[0,1]^k} \prod_{\{v,v'\} \in E} w(x_v, x_{v'}) \prod_{\{v,v'\} \notin E} (1 - w(x_v, x_{v'})) \prod_{v \in V} dx_v,$$

where, in the second expression, we choose an arbitrary representative w in W.

This definition extends that of the density of subgraphs in (finite) graphs in the following sense. For every finite graphs g and G, denoting by $\vec{V}^k$ a k-tuple of i.i.d. uniform random vertices in G,

$$\mathsf{Dens}(g, W_G) = \mathbb{P}(\mathrm{Sample}_k(W_G) = g) = \mathbb{P}(\mathrm{SubGraph}(\vec{V}^k, G) = g) = \mathsf{Dens}(g, G).$$

The following theorem is a prominent result in the theory of graphons, see *e.g.* [Lov12, Theorem 11.5].

Theorem 3.6. *Let W_n (for all $n \geq 0$) and W be graphons. Then the following are equivalent:*

(a) *(W_n) converges to W (for the distance $\delta_\square$);*
(b) *for any fixed graph g, we have $\mathsf{Dens}(g, W_n) \to \mathsf{Dens}(g, W)$.*

Classically, when (H_n) is a sequence of graphs, we say that (H_n) converges to a graphon W when (W_{H_n}) converges to W.

3.3. **Random graphons.**

We now discuss convergence of a sequence of random graphs $\boldsymbol{H}_n$ (equivalently, of the associated random graphons $W_{\boldsymbol{H}_n}$) towards a possibly random limiting graphon $\boldsymbol{W}$. In this context, the densities $\mathsf{Dens}(g, \boldsymbol{H}_n)$ are random variables. This was studied in [DJ08] and it turns out that it is enough to consider the expectations $\mathbb{E}\big[\mathsf{Dens}(g, \boldsymbol{H}_n)\big]$ and $\mathbb{E}\big[\mathsf{Dens}(g, \boldsymbol{W})\big]$ to extend Theorem 3.6 to this random setting. Note first that $\mathbb{E}\big[\mathsf{Dens}(g, \boldsymbol{H}_n)\big] = \mathbb{P}(\mathrm{SubGraph}(\vec{V}^k, \boldsymbol{H}_n) = g)$, where both $\boldsymbol{H}_n$ and $\vec{V}^k$ are random, and that similarly $\mathbb{E}\big[\mathsf{Dens}(g, \boldsymbol{W})\big] = \mathbb{P}(\mathrm{Sample}_k(\boldsymbol{W}) = g)$, where the randomness comes both from $\boldsymbol{W}$ and the operation Sample_k.

A first result states that the distributions of random graphons are characterized by expected subgraph densities.

Proposition 3.7 (Corollary 3.3 of [DJ08])**.** *Let* $\mathbf{W}, \mathbf{W}'$ *be two random graphons, seen as random variables in* $\widetilde{\mathcal{W}_0}$. *The following are equivalent:*

- $\mathbf{W} \overset{d}{=} \mathbf{W}'$;
- *for every finite graph* g, $\mathbb{E}[\mathsf{Dens}(g, \mathbf{W})] = \mathbb{E}[\mathsf{Dens}(g, \mathbf{W}')]$;
- *for every* $k \geq 1$, $\mathrm{Sample}_k(\boldsymbol{W}) \overset{d}{=} \mathrm{Sample}_k(\boldsymbol{W}')$.

The next result, which is essentially [DJ08, Theorem 3.1], characterizes the convergence in distribution of random graphs to random graphons.

Theorem 3.8. *For any* n, *let* $\boldsymbol{H}_n$ *be a random graph of size* n. *Denote by* $W_{\boldsymbol{H}_n}$ *the graphon associated to* $\boldsymbol{H}_n$ *by Definition 3.2. The following assertions are equivalent:*

(a) *The sequence of random graphons* $(W_{\boldsymbol{H}_n})_n$ *converges in distribution to some random graphon* $\boldsymbol{W}$.

(b) *The random infinite vector* $\big(\mathsf{Dens}(g, \boldsymbol{H}_n)\big)_{g\,\text{finite graph}}$ *converges in distribution in the product topology to some random infinite vector* $(\boldsymbol{\Lambda}_g)_{g\,\text{finite graph}}$.

(c) *For every finite graph* g, *there is a constant* $\Delta_g \in [0, 1]$ *such that*

$$\mathbb{E}[\mathsf{Dens}(g, \boldsymbol{H}_n)] \xrightarrow{n\to\infty} \Delta_g.$$

(d) *For every* $k \geq 1$, *denote by* $\vec{V'}^k = (V_1', \ldots, V_k')$ *a uniform* k-*tuple of* distinct *vertices of* $\boldsymbol{H}_n$. *Then the induced graph* $\mathrm{SubGraph}(\vec{V'}^k, \boldsymbol{H}_n)$ *converges in distribution to some random graph* $\boldsymbol{g}_k$.

Whenever these assertions are verified, we have

$$(2) \qquad\qquad (\boldsymbol{\Lambda}_g)_{g\,\text{finite graphs}} \overset{d}{=} (\mathsf{Dens}(g, \boldsymbol{W}))_{g\,\text{finite graphs}}.$$

and, for every graph g *of size* k,

$$(3) \qquad\qquad \Delta_g = \mathbb{E}[\boldsymbol{\Lambda}_g] = \mathbb{E}[\mathsf{Dens}(g, \boldsymbol{W})] = \mathbb{P}(\boldsymbol{g}_k = g).$$

Using the identity $\mathbb{E}\big[\mathsf{Dens}(g, \boldsymbol{W})\big] = \mathbb{P}(\mathrm{Sample}_k(\boldsymbol{W}) = g)$, we note that Eq. (3) implies that, for all $k \geq 1$, we have

$$(4) \qquad\qquad \mathrm{Sample}_k(\boldsymbol{W}) \overset{d}{=} \boldsymbol{g}_k$$

Proof. The equivalence of the first three items, Eq. (2) and the first two equalities in Eq. (3) are all proved in [DJ08]; see Theorem 3.1 there. Thus, we only prove (c) $\Leftrightarrow$ (d) and the related equality $\mathbb{P}(\boldsymbol{g}_k = g) = \Delta_g$.

For any graphs g, G of respective sizes $k \leq n$, we define their *injective density* $\mathsf{Dens}^{\mathrm{inj}}(g, G) = \mathbb{P}(\mathrm{SubGraph}(\vec{V'}^k, G) = g)$ where $\vec{V'}^k$ is a uniform k-tuple of *distinct* vertices of G. As explained in [DJ08] (and standard in the graphon literature), Assertion (c) is equivalent, for the same limits (Δ_g), to its analogue with injective densities, which is: for every graph g,

$$(5) \qquad \mathbb{E}[\mathsf{Dens}^{\mathrm{inj}}(g, \boldsymbol{H}_n)] \xrightarrow{n \to \infty} \Delta_g.$$

Moreover, we note that, if $(\boldsymbol{H}_n)$ is a sequence of random graphs, then, for any graph g of size k,

$$(6) \qquad \mathbb{E}\big[\mathsf{Dens}^{\mathrm{inj}}(g, \boldsymbol{H}_n)\big] = \mathbb{P}(\mathrm{SubGraph}(\vec{V'}^k, \boldsymbol{H}_n) = g),$$

where both $\vec{V'}^k$ and $\boldsymbol{H}_n$ are random. Since $\mathrm{SubGraph}(\vec{V'}^k, \boldsymbol{H}_n)$ takes value in a finite set, its convergence in distribution (Assertion (d)) is equivalent to the convergence of its point probabilities, *i.e.* of the right-hand side of Eq. (6). Recalling Eq. (5), this proves the equivalence of Assertions (c) and (d). Futhermore, when these assertions hold, we have

$$\mathbb{P}(\boldsymbol{g}_k = g) = \lim_{n \to \infty} \mathbb{P}\big[\mathrm{SubGraph}(\vec{V'}^k, \boldsymbol{H}_n) = g\big] = \lim_{n \to \infty} \mathbb{E}\big[\mathsf{Dens}^{\mathrm{inj}}(g, \boldsymbol{H}_n)\big] = \Delta_g,$$

as wanted. $\square$

We finally collect an immediate corollary.

Lemma 3.9. *If $\boldsymbol{W}$ is a random graphon, then $W_{\mathrm{Sample}_n(\boldsymbol{W})}$ converges in distribution to $\boldsymbol{W}$ as $n \to \infty$.*

Proof. Recall that $\mathrm{Sample}_n(\boldsymbol{W})$ is the random graph on vertex set $\{v_1, \cdots, v_n\}$ obtained by taking $X_1, \ldots, X_n$ i.i.d. uniform in $[0, 1]$ and joining v_i and v_j with probability $w(X_i, X_j)$ where w is a representative of W. Fix k in $\{1, \cdots, n\}$. As in the previous theorem, let $\vec{V'}^k = (v_{h_1}, \cdots, v_{h_k})$ be a uniform random k-tuple of distinct vertices of $\mathrm{Sample}_n(\boldsymbol{W})$. Then $\mathrm{SubGraph}(\vec{V'}^k, \mathrm{Sample}_n(\boldsymbol{W}))$ is the random graph on vertex set $\{v_{h_1}, \cdots, v_{h_k}\}$ obtained by taking $X_{h_1}, \ldots, X_{h_k}$ i.i.d. uniform in $[0, 1]$ and joining v_{h_i} and v_{h_j} with probability $w(X_{h_i}, X_{h_j})$. Up to renaming v_{h_i} as v_i and X_{h_i} as X_i, this matches the construction of $\mathrm{Sample}_k(\boldsymbol{W})$. Therefore we have the following equality in distribution of random unlabeled graphs:

$$\mathrm{SubGraph}(\vec{V'}^k, \mathrm{Sample}_n(\boldsymbol{W})) \overset{d}{=} \mathrm{Sample}_k(\boldsymbol{W}).$$

Thus, Assertion (d) of Theorem 3.8 is fulfilled for the graph sequence $(\mathrm{Sample}_n(\boldsymbol{W}))_n$ and for $\boldsymbol{g}_k \overset{d}{=} \mathrm{Sample}_k(\boldsymbol{W})$. Therefore Assertion (a) holds and the graphon sequence $(W_{\mathrm{Sample}_n(\boldsymbol{W})})_n$ has a limit in distribution $\boldsymbol{W}'$, which satisfies, for all k (see Eq. (4)):

$$\mathrm{Sample}_k(\boldsymbol{W}') \overset{d}{=} \boldsymbol{g}_k \overset{d}{=} \mathrm{Sample}_k(\boldsymbol{W}).$$

From Proposition 3.7, we have $\boldsymbol{W} \overset{d}{=} \boldsymbol{W}'$, concluding the proof of the lemma. $\square$

3.4. **Graphons and degree distribution.** In this section, we consider the degree distribution of a graphon, as introduced by Diaconis, Holmes and Janson in [DHJ08, Section 4]. This defines a continuous functional from the space of graphons to that of probability measures on $[0, 1]$ (equipped with the weak topology) [DHJ08, Theorem 4.2]. For self-completeness we provide a proof here of this simple fact. Our approach is different from that of Diaconis, Holmes and Janson: we prove that the functional is 2-Lipschitz with respect to natural metrics, while they express moments of the degree distribution in terms of subgraph densities to prove the continuity.

At the end of the section, we also discuss the degree distribution of random graphs and random graphons. This is a preparation for Section 4.4 where we shall study the degree distribution of random cographs and of the Brownian cographon. We also note that other works [BCL11, DDS18] study of the degree distributions of specific random graph models and their convergence to that of their graphon limit.

The *degree distribution* of a graphon W is the measure D_W on $[0, 1]$ defined as follows (see [DHJ08, Theorem 4.4]): for every continuous bounded function $f : [0, 1] \to \mathbb{R}$, we have

$$\int_{[0,1]} f(x) D_W(dx) = \int_{[0,1]} f\left(\int_{[0,1]} w(u, v) dv\right) du,$$

where w is, as usual, an arbitrary representative of W (the resulting measure does not depend on the chosen representative).

For the graphon W_G associated to a graph G of size n, the measure D_{W_G} is simply the empirical distribution of the rescaled degrees:

$$D_{W_G} = \frac{1}{n} \sum_{v \in G} \delta_{\deg_G(v)/n}$$

where δ_u is the Dirac measure concentrated at u.

The next lemma implies that graphon convergence entails weak convergence of degree distributions. To state a more precise result, it is useful to endow the space $\mathcal{M}_1([0, 1])$ of Borel probability measures on $[0, 1]$ with the so-called Wasserstein metric (see *e.g.* [Ros11, Section 1.2]), defined as

$$d_{\text{Wass}}(\mu, \nu) = \inf_f \left| \int_{[0,1]} f(x) \mu(dx) - \int_{[0,1]} f(x) \nu(dx) \right|,$$

where the infimum runs over all 1-Lipschitz functions f from $[0, 1]$ to $\mathbb{R}$. We recall that this distance metrizes weak convergence (see *e.g.* [Bog07, Sec. 8.3]).

Lemma 3.10. *The map $W \mapsto D_W$ from $(\widetilde{\mathcal{W}_0}, \delta_\square)$ to $(\mathcal{M}_1([0, 1]), d_{\text{Wass}})$ is 2-Lipschitz. Consequently, if (W_n) converges to W in $\widetilde{\mathcal{W}_0}$, then the sequence of associated measures (D_{W_n}) converges weakly to D_W.*

Proof. Let W and W' be graphons with representatives w and w'. Let $f : [0,1] \to \mathbb{R}$ be 1-Lipschitz. We have

$$d_{\text{Wass}}(D_W, D_{W'}) \le \left| \int_{[0,1]} f(x) D_W(dx) - \int_{[0,1]} f(x) D_{W'}(dx) \right|$$

$$= \left| \int_{[0,1]} f\left(\int_{[0,1]} w(u,v) dv \right) - f\left(\int_{[0,1]} w'(u,v) dv \right) du \right|$$

$$\le \int_{[0,1]} \left| \int_{[0,1]} (w(u,v) - w'(u,v)) dv \right| du$$

$$= \int_S \int_{[0,1]} (w(u,v) - w'(u,v)) dv du - \int_{[0,1] \setminus S} \int_{[0,1]} (w(u,v) - w'(u,v)) dv du$$

where $S = \left\{ u \in [0,1] : \int_{[0,1]} (w(u,v) - w'(u,v)) dv \ge 0 \right\}$. But, from the definition of $\|\cdot\|_\square$, each of the two summands has modulus bounded by $\|w - w'\|_\square$. We finally get

$$d_{\text{Wass}}(D_W, D_{W'}) \le 2 \|w - w'\|_\square.$$

which ends the proof by definition of $\delta_\square$ since the choice of representatives w, w' was arbitrary. $\square$

Remark that when W is a random graphon, D_W is a random measure. We recall, see *e.g.* [Kal17, Lemma 2.4], that given a random measure μ on some space B, its *intensity measure* $I[\mu]$ is the deterministic measure on B defined by $I[\mu](A) = \mathbb{E}[\mu(A)]$ for any measurable subset A of B.

To get an intuition of what $I[D_W]$ is for a random graphon W, it is useful to consider the case where $W = W_G$ is the graphon associated with a random graph G of size n. In this case, for any measurable subset A of $[0,1]$,

$$D_{W_G}(A) = \mathbb{P}(\tfrac{1}{n} \deg_G(v) \in A \mid G),$$

where v is a uniform random vertex in G. Therefore

$$I[D_{W_G}](A) = \mathbb{E}\big[D_{W_G}(A) \big] = \mathbb{P}(\tfrac{1}{n} \deg_G(v) \in A),$$

so that $I[D_{W_G}]$ is the law of the normalized degree of a uniform random vertex v in the random graph G.

We sum up the results of this section into the following proposition.

Proposition 3.11. *Let H_n be a random graph of size n for every n, and W be a random graphon, such that $W_{H_n} \xrightarrow[n \to \infty]{d} W$. Then we have the following convergence in distribution of random measures:*

$$\frac{1}{n} \sum_{v \in H_n} \delta_{\deg_{H_n}(v)/n} \xrightarrow[n \to \infty]{d} D_W.$$

Furthermore, denoting v_n a uniform random vertex in H_n and Z a random variable with law $I[D_W]$,

$$\tfrac{1}{n} \deg_{H_n}(v_n) \xrightarrow[n \to \infty]{d} Z.$$

Proof. From Lemma 3.10, we immediately obtain $D_{W_{H_n}} \xrightarrow{d} D_W$, which is by definition of D_{W_G} exactly the first of the stated convergences. The second one follows from the first, combining Lemma 4.8 and Theorem 4.11 of [Kal17][4]. □

4. THE BROWNIAN COGRAPHON

4.1. **Construction.** Let $\mathfrak{e}$ denote a Brownian excursion of length one. We start by recalling a technical result on the local minima of $\mathfrak{e}$: the first two assertions below are well-known, we refer to [Maa17, Lemma 2.3] for the last one.

Lemma 4.1. *With probability one, the following assertions hold. First, all local minima of $\mathfrak{e}$ are strict, and hence form a countable set. Moreover, the values of $\mathfrak{e}$ at two distinct local minima are different. Finally, there exists an enumeration $(b_i)_i$ of the local minima of $\mathfrak{e}$, such that for every $i \in \mathbb{N}, x, y \in [0,1]$, the event $\{b_i \in (x,y), \mathfrak{e}(b_i) = \min_{[x,y]} \mathfrak{e}\}$ is measurable.*

Let $\boldsymbol{S}^p = (\boldsymbol{s}_1, \ldots)$ be a sequence of i.i.d. random variables in $\{0, 1\}$, independent of $\mathfrak{e}$, with $\mathbb{P}(\boldsymbol{s}_1 = 0) = p$ (in the sequel, we simply speak of i.i.d. decorations of bias p). We call $(\mathfrak{e}, \boldsymbol{S}^p)$ a *decorated Brownian excursion*, thinking of the decoration $\boldsymbol{s}_i$ as attached to the local minimum b_i. For $x, y \in [0, 1]$, we define $\mathrm{Dec}(x, y; \mathfrak{e}, \boldsymbol{S}^p)$ to be the decoration of the minimum of $\mathfrak{e}$ on the interval $[x, y]$ (or $[y, x]$ if $y \leq x$; we shall not repeat this precision below). If this minimum is not unique or attained in x or y and therefore not a local minimum, $\mathrm{Dec}(x, y; \mathfrak{e}, \boldsymbol{S}^p)$ is ill-defined and we take the convention $\mathrm{Dec}(x, y; \mathfrak{e}, \boldsymbol{S}^p) = 0$. Note however that, for uniform random x and y, this happens with probability 0, so that the object constructed in Definition 4.2 below is independent from this convention.

Definition 4.2. *The Brownian cographon $\boldsymbol{W}^p$ of parameter p is the equivalence class of the random function[5]*

$$\boldsymbol{w}^p : \begin{array}{ccc} [0,1]^2 & \to & \{0,1\}; \\ (x,y) & \mapsto & \mathrm{Dec}(x, y; \mathfrak{e}, \boldsymbol{S}^p). \end{array}$$

In most of this article, we are interested in the case $p = 1/2$; in particular, as claimed in Theorem 1.1, $W^{1/2}$ is the limit of uniform random (labeled or unlabeled) cographs, justifying its name.

4.2. **Sampling from the Brownian cographon.** We now compute the distribution of the random graph $\mathrm{Sample}_k(\boldsymbol{W}^p)$.

Proposition 4.3. *If $\boldsymbol{W}^p$ is the Brownian cographon of parameter p, then for every $k \geq 2$, $\mathrm{Sample}_k(\boldsymbol{W}^p)$ is distributed like the unlabeled version of $\mathsf{Cograph}(\boldsymbol{b}_k^p)$, where the cotree $\boldsymbol{b}_k^p$ is a uniform labeled binary tree with k leaves equipped with i.i.d. decorations of bias p.*

---

[4]Theorem 4.11 tells us that if random measures $(\boldsymbol{\xi}_n)$ converge in distribution to $\boldsymbol{\xi}$ then, for any compactly supported continuous function f, we have $\boldsymbol{\xi}_n f \xrightarrow[n \to \infty]{d} \boldsymbol{\xi} f$. But since those variables are bounded (by $\|f\|_\infty$), this convergence also holds in L^1, *i.e.* $\boldsymbol{\xi}_n \xrightarrow{L^1} \boldsymbol{\xi}$ in the notation of [Kal17]. By Lemma 4.8, this implies the convergence of the corresponding intensity measures.

[5]Of course, in the image set of $\boldsymbol{w}^p$, the *real values* 0 and 1 correspond to the *decorations* 0 and 1 respectively.

Let us note that $\boldsymbol{b}_k^p$ is not necessarily a canonical cotree.

Proof. We use a classical construction (see [LG05, Section 2.5]) which associates to an excursion e and real numbers $x_1, \cdots, x_k$ a plane tree, denoted $\mathrm{Tree}(e; x_1, \ldots, x_k)$, which has the following properties:

- its leaves are labeled with $1, \cdots, k$ and correspond to $x_1, \ldots, x_k$ respectively;
- its internal nodes correspond to the local minima of e on intervals $[x_i, x_j]$;
- the first common ancestor of the leaves i and j corresponds to the local minimum of e on $[x_i, x_j]$.

The tree $\mathrm{Tree}(e; x_1, \ldots, x_k)$ is well-defined with probability 1 when e is a Brownian excursion and $x_1, \cdots, x_k$ i.i.d. uniform random variables in $[0, 1]$. Moreover, in this setting, it has the distribution of a uniform random plane and labeled binary tree with k leaves [LG05, Theorem 2.11]. Forgetting the plane structure, it is still uniform among binary trees with k labeled leaves, because the number of plane embeddings of a *labeled binary tree* depends only on its size.

We now let $(\mathfrak{e}, \boldsymbol{S})$ be a decorated Brownian excursion, and $X_1, \ldots, X_k$ denote a sequence of i.i.d. uniform random variables in $[0, 1]$, independent from $(\mathfrak{e}, \boldsymbol{S})$. We make use of the decorations $\boldsymbol{S}$ of the local minima of $\mathfrak{e}$ to turn $\mathrm{Tree}(\mathfrak{e}; X_1, \ldots, X_k)$ into a cotree. Namely, since its internal nodes correspond to local minima of $\mathfrak{e}$, we can simply report these decorations on the tree, and we get a decorated tree $\mathrm{Tree}_{0/1}(\mathfrak{e}, \boldsymbol{S}; X_1, \ldots, X_k)$. When the decorations in $\boldsymbol{S}$ are i.i.d. of bias p, then $\mathrm{Tree}_{0/1}(\mathfrak{e}, \boldsymbol{S}, X_1, \ldots, X_k))$ is a uniform labeled binary tree with k leaves, equipped with i.i.d. decorations of bias p.

Finally, recall that $\mathrm{Sample}_k(\boldsymbol{W}^p)$ is built by considering $X_1, \ldots, X_k$ i.i.d. uniform in $[0, 1]$ and connecting vertices v_i and v_j if and only if $\boldsymbol{w}^p(X_i, X_j) = 1$ (since a representative $\boldsymbol{w}^p$ of $\boldsymbol{W}^p$ takes value in $\{0, 1\}$, there is no extra randomness here). By definition of $\boldsymbol{w}^p$, $\boldsymbol{w}^p(X_i, X_j) = 1$ means that the decoration of the minimum of $\mathfrak{e}$ on $[X_i, X_j]$ is 1. But, by construction of $\mathrm{Tree}_{0/1}(\mathfrak{e}, \boldsymbol{S}; X_1, \ldots, X_k)$, this decoration is that of the first common ancestor of the leaves i and j in $\mathrm{Tree}_{0/1}(\mathfrak{e}, \boldsymbol{S}; X_1, \ldots, X_k)$. So it is equal to 1 if and only if i and j are connected in the associated cograph (see Observation 2.7). Summing up, we get the equality of unlabeled random graphs

$$\mathrm{Sample}_k(\boldsymbol{W}^p) = \mathsf{Cograph}\big(\mathrm{Tree}_{0/1}(\mathfrak{e}, \boldsymbol{S}, X_1, \ldots, X_k))\big),$$

ending the proof of the proposition. $\qquad\square$

4.3. **Criterion of convergence to the Brownian cographon.**
The results obtained so far yield a very simple criterion for convergence to the Brownian cographon. For simplicity and since this is the only case we need in the present paper, we state it only in the case $p = 1/2$.

Lemma 4.4. *Let $\boldsymbol{t}^{(n)}$ be a random cotree of size n for every n (which may be labeled or not). For $n \geq k \geq 1$, denote by $\boldsymbol{t}_k^{(n)}$ the subtree of $\boldsymbol{t}^{(n)}$ induced by a uniform k-tuple of distinct leaves. Suppose that for every k and for every labeled binary cotree t_0 with k leaves,*

$$(7) \qquad\qquad\qquad \mathbb{P}(\boldsymbol{t}_k^{(n)} = t_0) \xrightarrow[n \to \infty]{} \frac{(k-1)!}{(2k-2)!}.$$

Then $W_{\mathsf{Cograph}(\boldsymbol{t}^{(n)})}$ converges as a graphon to $\boldsymbol{W}^{1/2}$.

Proof. We first remark that $\frac{(k-1)!}{(2k-2)!} = \frac{1}{\text{card}(\mathcal{C}_k)}$, where $\mathcal{C}_k$ is the set of labeled binary cotrees with k leaves. Indeed the number of plane labeled binary trees with k leaves is given by $k!\,\text{Cat}_{k-1}$ where $\text{Cat}_{k-1} = \frac{1}{k}\binom{2k-2}{k-1}$ is the $(k-1)$-th Catalan number. Decorations on internal nodes induce the multiplication by a factor 2^{k-1} while considering non-plane trees yields a division by the same factor in order to avoid symmetries. Therefore $\text{card}(\mathcal{C}_k) = k!\,\text{Cat}_{k-1} = \frac{(2k-2)!}{(k-1)!}$.

Consequently, Eq. (7) states that $\boldsymbol{t}_k^{(n)}$ converges in distribution to a uniform element of $\mathcal{C}_k$. Morever, a uniform element of $\mathcal{C}_k$ is distributed as $\boldsymbol{b}_k^{1/2}$ where $\boldsymbol{b}_k^{1/2}$ is a uniform labeled binary tree with k leaves equipped with i.i.d. decorations of bias $1/2$. Hence, as n tends to $+\infty$, we have the following convergence of random labeled graphs of size k,

$$\mathsf{Cograph}(\boldsymbol{t}_k^{(n)}) \xrightarrow{d} \mathsf{Cograph}(\boldsymbol{b}_k^{1/2}).$$

Forgetting the labeling, the left-hand side is $\text{SubGraph}(\vec{V'}^k, \mathsf{Cograph}(\boldsymbol{t}^{(n)}))$, where $\vec{V'}^k$ is a uniform tuple of k distinct vertices of $\mathsf{Cograph}(\boldsymbol{t}^{(n)})$; see the definition of $\boldsymbol{t}_k^{(n)}$ in the statement of the lemma and Proposition 2.9. Moreover, thanks to Proposition 4.3, forgetting again the labeling, the right-hand side has the same distribution as $\text{Sample}_k(\boldsymbol{W}^{1/2})$. This proves the lemma, using Theorem 3.8 (namely, the implication $(d) \Rightarrow (a)$, and Eq. (4) together with Proposition 3.7 to identify the limit in item (a) with $\boldsymbol{W}^{1/2}$). $\qquad\square$

4.4. **The degree distribution of the Brownian cographon.** In this section we are interested in the degree distribution $D_{\boldsymbol{W}^p}$ of the Brownian cographon. It turns out that, in the special case $p = 1/2$, the intensity $I[D_{\boldsymbol{W}^{1/2}}]$ is particularly simple.

Proposition 4.5. $I[D_{\boldsymbol{W}^{1/2}}] \overset{d}{=} U$, *where U is the Lebesgue measure on $[0,1]$.*

Proof. Rather than working in the continuous, we exhibit a discrete approximation $\boldsymbol{G}_n^b$ of the Brownian cographon, which has the remarkable property that the degree of a uniform random vertex $\boldsymbol{v}_n$ in $\boldsymbol{G}_n^b$ is *exactly* distributed as a uniform random variable in $\{0, 1, \cdots, n-1\}$.

To construct $\boldsymbol{G}_n^b$, we let $\boldsymbol{b}_n$ be a uniform $0/1$-decorated plane labeled binary tree with n leaves. Forgetting the plane structure, it is still uniform among labeled binary cotrees with n leaves. Set $\boldsymbol{G}_n^b = \mathsf{Cograph}(\boldsymbol{b}_n)$. From Proposition 4.3, $\boldsymbol{G}_n^b$ has the same distribution as $\text{Sample}_n(\boldsymbol{W}^{1/2})$, so that $W_{\boldsymbol{G}_n^b}$ converges in distribution to $\boldsymbol{W}^{1/2}$ (Lemma 3.9).

Consider a uniform random vertex $\boldsymbol{v}_n$ in $\boldsymbol{G}_n^b$. Thanks to Proposition 3.11, $\text{Law}\left(\frac{1}{n}\deg_{\boldsymbol{G}_n^b}(\boldsymbol{v}_n)\right)$ converges to $I[D_{\boldsymbol{W}^{1/2}}]$. Proving the following claim will therefore conclude the proof of the proposition.

Claim. The law of $\deg(\boldsymbol{v}_n)$ in $\boldsymbol{G}_n^b$ is the uniform law in $\{0, 1, \cdots, n-1\}$.

Proof of the claim. We start by defining two operations for deterministic $0/1$-decorated plane labeled binary trees b.

- First, we consider a (seemingly unnatural[6]) order on the leaves of b. To compare two leaves ℓ and r, we look at their first common ancestor u and assume w.l.o.g. that ℓ and

---

[6]This order is actually very natural if we interpret b as the separation tree of a separable permutation (see [BBF+18] for the definition). It is simply the value order on the elements of the permutation.

r are descendants of its left and right children, respectively. If u has decoration 0, we declare ℓ to be smaller than r; if it has decoration 1, then r is smaller than ℓ. It is easy to check that this defines a total order on the leaves of b (if we flip the left and right subtrees of internal nodes with decoration 1, this is simply the left-to-right depth-first order of the leaves). We write $\mathrm{rk}_b(\ell)$ for the rank of a leaf ℓ in this order.

- Second, we define an involution Φ on the set of $0/1$-decorated plane labeled binary trees b with a distinguished leaf ℓ. We keep the undecorated structure of the tree, and simply flip the decorations of all the ancestors of ℓ which have ℓ as a descendant *of their right child*. This gives a new decorated plane labeled binary tree b' and we set $\Phi(b, \ell) = (b', \ell)$.

Consider b as above, with two distinguished leaves ℓ and $\tilde{\ell}$. The corresponding vertices v and $\tilde{v}$ in $G = \mathsf{Cograph}(b)$ are connected if and only if the first common ancestor u of ℓ and $\tilde{\ell}$ in b has decoration 1. Setting $\Phi(b, \ell) = (b', \ell)$, this happens in two cases:

- either ℓ is a descendant of the left child of u, and u has decoration 1 in b';
- or ℓ is a descendant of the right child of u, and u has decoration 0 in b';

This corresponds exactly to ℓ being bigger than $\tilde{\ell}$ in the order associated to b'. Consequently, $\deg_G(v)$ is the number of leaves smaller than ℓ in that order, *i.e.*

$$\deg_G(v) = \mathrm{rk}_{b'}(\ell) - 1. \tag{8}$$

Recall now that $\boldsymbol{G}_n^b = \mathsf{Cograph}(\boldsymbol{b}_n)$, where $\boldsymbol{b}_n$ is a uniform $0/1$-decorated plane labeled binary tree with n leaves. The uniform random vertex $\boldsymbol{v}_n$ in $\boldsymbol{G}_n^b$ corresponds to a uniform random leaf $\boldsymbol{\ell}_n$ in $\boldsymbol{b}_n$. Set $(\boldsymbol{b}_n', \boldsymbol{\ell}_n) = \Phi(\boldsymbol{b}_n, \boldsymbol{\ell}_n)$. Since Φ is an involution, $(\boldsymbol{b}_n', \boldsymbol{\ell}_n)$ is a uniform $0/1$-decorated plane labeled binary tree of size n with a uniform random leaf $\boldsymbol{\ell}_n$. Conditioning on $\boldsymbol{b}_n'$, the rank $\mathrm{rk}_{\boldsymbol{b}_n'}(\boldsymbol{\ell}_n)$ is a uniform random variable in $\{1, \cdots, n\}$. The same holds taking $\boldsymbol{b}_n'$ at random, and, using Eq. (8), we conclude that $\deg_{\boldsymbol{G}_n^b}(\boldsymbol{v}_n)$ is a uniform random variable in $\{0, \cdots, n-1\}$.

This proves the claim, and hence the proposition. $\qquad\square$

Remark 4.6. It seems likely that this result can also be proved by working only in the continuous. In particular, using a result of Bertoin and Pitman [BP94, Theorem 3.2], the degree $D(x) = \int_y \boldsymbol{W}^{1/2}(x, y) dy$ of a uniform random x in $[0, 1]$ in the Brownian cographon corresponds to the cumulated length of a half of the excursions in a Brownian bridge.

5. CONVERGENCE OF LABELED COGRAPHS TO THE BROWNIAN COGRAPHON

In this section, we are interested in labeled cographs with n vertices, which are in one-to-one correspondence with labeled canonical cotrees with n leaves (Proposition 2.5).

To study these objects, we use the framework of labeled combinatorial classes, as presented in the seminal book of Flajolet and Sedgewick [FS09, Chapter II]. In this framework, an object of size n has n atoms, which are labeled bijectively with integers from 1 to n. For us, the atoms are simply the leaves of the trees, which is consistent with Definition 2.2.

We will also consider (co)trees with marked leaves and, here, more care is needed. Indeed, in some instances, those marked leaves have a label (and thus should be seen as atoms and counted in the size of the objects), while, in other instances, they do not have a label (and are therefore not counted in the size of the object). To make the distinction, we will refer to marked leaves

of the latter type (*i.e.* without labels) as *blossoms* and reserve *marked leaves* for those carrying labels.

5.1. **Enumeration of labeled canonical cotrees.** Let $\mathcal{L}$ be the family of non-plane labeled rooted trees in which internal nodes have at least two children. For $n \geq 1$, let ℓ_n be the number of trees with n leaves in $\mathcal{L}$. Let $L(z)$ denote the corresponding exponential generating function:

$$L(z) = \sum_{n \geq 1} \frac{\ell_n}{n!} z^n.$$

Proposition 5.1. *The series $L(z)$ is the unique formal power series without constant term solution of*

$$(9) \qquad\qquad L(z) = z + \exp(L(z)) - 1 - L(z).$$

Proof. (This series is treated in [FS09, Example VII.12 p.472].) A tree in $\mathcal{L}$ consists of

- either a single leaf (counted by z) ;
- or a root to which is attached an unordered sequence of at least two trees of $\mathcal{L}$ (counted by $\sum_{k \geq 2} L^k/k! = e^L - 1 - L$).

This justifies that $L(z)$ satisfies Eq. (9). The uniqueness is straightforward, since Eq. (9) determines for every n the coefficient of z^n in $L(z)$ from those of z^k for $k < n$. $\qquad\square$

Computing the first coefficients, we find

$$L(z) = z + \frac{z^2}{2!} + 4\frac{z^3}{3!} + 26\frac{z^4}{4!} + 236\frac{z^5}{5!} + 2752\frac{z^6}{6!}\mathcal{O}(z^7).$$

These numbers correspond to the fourth Schröder's problem (see Sequence A000311 in [OEIS]).

Let m_n be the number of labeled canonical cotrees with n leaves. We have $m_1 = 1$ and $m_n = 2\,\ell_n$ for $n \geq 2$. Indeed to each tree of $\mathcal{L}$ containing internal nodes (*i.e.*, with at least two leaves) there correspond two canonical cotrees: one with the root decorated by 0 and one with the root decorated by 1 (the other decorations are then determined by the alternation condition). The exponential generating series $M(z) = \sum_{n \geq 1} \frac{m_n}{n!} z^n$ of labeled canonical cotrees (or equivalently of labeled cographs) thus satisfies $M(z) = 2L(z) - z$. Combining this with Proposition 5.1, we have

$$(10) \qquad\qquad M(z) = \exp(L(z)) - 1.$$

It is standard (and easy to see) that the series

$$L'(z) = \sum_{n \geq 1} \frac{\ell_n}{(n-1)!} z^{n-1} \quad \text{and} \quad L^\bullet(z) = zL'(z) = \sum_{n \geq 1} \frac{\ell_n}{(n-1)!} z^n$$

counts trees of $\mathcal{L}$ with a blossom or a marked leaf, repectively. In the subsequent analysis we need to consider the generating function L^{even} (resp. L^{odd}) counting trees of $\mathcal{L}$ having a blossom at even (resp. odd) distance from the root. Obviously, $L^{\text{even}} + L^{\text{odd}} = L'$.

Proposition 5.2. *We have the following identities*

(11)
$$L^{even} = \frac{1}{e^L(2 - e^L)},$$

(12)
$$L^{odd} = \frac{e^L - 1}{e^L(2 - e^L)}.$$

Proof. We first claim that

(13)
$$\begin{cases} L^{\text{even}} & = 1 + (e^L - 1)L^{\text{odd}}, \\ L^{\text{odd}} & = (e^L - 1)L^{\text{even}}. \end{cases}$$

We prove the first identity, the second one is proved similarly. A tree counted by L^{even} is

- either reduced to a blossom (therefore the tree has size 0, *i.e.* is counted by 1);
- or has a root to which are attached
  - a tree with a blossom at odd height (counted by L^{odd}), and
  - an unordered sequence of at least one unmarked trees (counted by $\sum_{k \geq 1} L^k/k! = e^L - 1$).

We obtain the proposition by solving Eq. (13). □

5.2. **Enumeration of canonical cotrees with marked leaves inducing a given cotree.**
For a labeled (not necessarily canonical) cotree t_0 of size k, we consider the family $\mathcal{M}_{t_0}$ of tuples $(t; \ell_1, \ldots, \ell_k)$ where

- t is a labeled canonical cotree;
- $(\ell_1, \ldots, \ell_k)$ is a k-tuple of distinct leaves of t;
- the subtree of t induced by $(\ell_1, \ldots, \ell_k)$ is t_0.

We denote by M_{t_0} the associated exponential generating function.

Theorem 5.3. *Let t_0 be a labeled cotree with k leaves. Denote by n_v its number of internal nodes, by $n_=$ its number of edges of the form $0 - 0$ or $1 - 1$, and by $n_{\neq}$ its number of edges of the form $0 - 1$ or $1 - 0$. We have the identity*

(14)
$$M_{t_0} = (L')(\exp(L))^{n_v}(L^\bullet)^k(L^{odd})^{n_=}(L^{even})^{n_{\neq}}.$$

Proof. (Main notations of the proof are summarized in Fig. 5.)

Let $(t; \ell_1, \ldots, \ell_k) \in \mathcal{M}_{t_0}$. There is a correspondence between the nodes of t_0 and some nodes of t, mapping leaves to marked leaves and internal nodes to first common ancestors of marked leaves. These first common ancestors of marked leaves in t will be refered to as *branching nodes* below. In order to prove Eq. (14) we will decompose each such t into subtrees, called *pieces*, of five different types: *pink, blue, yellow, green* and *gray* (see the color coding[7] in Fig. 5). Our decomposition has the following property: to reconstruct an element of $\mathcal{M}_{t_0}$, we can choose each piece independently in a set depending on its color only, so that the generating series of $\mathcal{M}_{t_0}$ writes as a product of the generating series of the pieces.

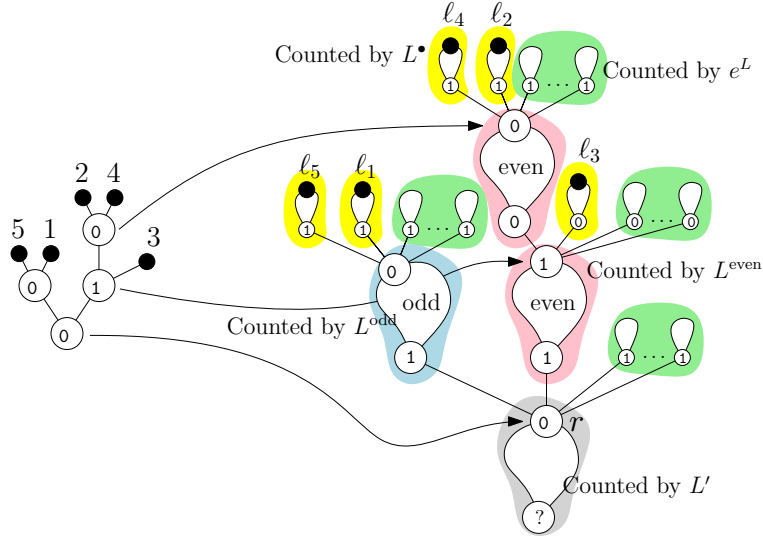[7]We apologize to the readers to whom only a black-and-white version were available.

FIGURE 5. On the left: a (non-canonical) labeled cotree t_0 of size 5. On the right: a schematic view of a canonical cotree in $\mathcal{M}_{t_0}$.

In this decomposition, there is exactly one gray piece obtained by pruning t at the node r of t corresponding to the root of t_0. In this piece, r is replaced by a blossom. We note that, by definition of induced cotree, the decoration of r has to match that of the root of t_0. Since decorations in canonical cotrees must alternate, this determines all decorations in the gray piece. Possible choices for the gray piece are therefore counted by the same series as undecorated trees with a blossom, that is L'.

For the rest of the decomposition, we consider branching nodes v of t (including r), and look at all children w of such nodes v.

- If such a node w has exactly one descendant (possibly, w itself) which is a marked leaf, we build a piece, colored yellow, by taking the fringe subtree rooted at w. Yellow pieces are labeled canonical cotrees with one marked leaf. However, the decoration within the yellow piece is forced by the alternation of decorations in t and by the decoration of the parent v of w, which has to match the decoration of the corresponding node in t_0 (see Fig. 5). So the generating function for yellow pieces is $L^\bullet$.

  Of course, we have a yellow piece for each marked leaf of t, *i.e.* for each leaf of t_0.

- If a node w child of a branching node in t has at least two marked leaves among its descendants, it must also have a descendant (possibly equal to w) that is a branching node. We define v' as the branching node descending from w (possibly equal to it) which is the closest to w. This implies that the node of t_0 corresponding to v' (denoted v'_0) is a child of the one corresponding to v (denoted v_0). We build a piece rooted at w, which corresponds to the edge (v_0, v'_0) of t_0. This piece is the fringe subtree rooted at w pruned at v', *i.e.* where v' is replaced by a blossom. We color it blue if the blossom is at odd distance from w, pink otherwise. The generating functions for blue and pink

pieces are therefore L^{odd} and L^{even}, respectively (since again all decorations in the piece are dertermined by the one of v_0).

Because of the alternation of decoration, the piece is blue if and only if w and v' have different decorations in t, which happens if and only if v and v' (or equivalently, v_0 and v_0') have the same decoration. We therefore have a blue piece for each internal edge of t_0 with extremities with the same decoration, and a pink piece for each internal edge of t_0 with extremities with different decorations.

- All other nodes w have no marked leaf among their descendants. We group all such nodes w that are siblings to build a single green piece, attached to their common parent v. Namely, for each branching node v, we consider all its children w having no marked leaf as a descendant (possibly, there are none), and we define the green piece attached to v as the (possibly empty) forest of fringe subtrees of t rooted at these nodes w. Green pieces are forest, *i.e.* unordered set of labeled canonical cotrees. The decoration within the green piece is forced by the alternation of decorations in t and by the decoration of v, which as before has to match the decoration of the corresponding node in t_0. Therefore, choosing a green piece amounts to choosing an unordered set of undecorated trees in $\mathcal{L}$. We conclude that possible choices for each green piece are counted by e^L.

Finally, we recall that there is one (possibly empty) green piece for each branching node of t, *i.e.* for each internal node of t_0.

Since t_0 is a labeled cotree, leaves / internal nodes / edges of t_0 can be ordered in a canonical way. Since yellow (resp. green, resp. blue and pink) pieces in the above decomposition are indexed by leaves (resp. internal nodes, resp. edges) of t_0, they can be ordered in a canonical way as well. Moreover, the correspondence between marked trees $(t; \ell_1, \cdots, \ell_k)$ in $\mathcal{M}_{t_0}$ and tuples of colored pieces is one-to-one. This completes the proof of Eq. (14). $\qquad\square$

5.3. Asymptotic analysis.
Following Flajolet and Sedgewick [FS09, p. 389], we say that a power series is Δ-analytic if it is analytic in some Δ-domain $\Delta(\phi, \rho)$, where ρ is its radius of convergence. This is a technical hypothesis, which enables to apply the transfer theorem [FS09, Corollary VI.1 p. 392]; all series in this paper are Δ-analytic.

Proposition 5.4. *The series $L(z)$ has radius of convergence $\rho = 2\log(2) - 1$ and is Δ-analytic. Moreover, the series L is convergent at $z = \rho$ and we have*

$$(15) \qquad L(z) \underset{z \to \rho}{=} \log(2) - \sqrt{\rho}\sqrt{1 - \tfrac{z}{\rho}} + \mathcal{O}(1 - \tfrac{z}{\rho}).$$

Proof. Using Proposition 5.1, Proposition 5.4 is a direct application of [BMN16, Theorem 1]. $\qquad\square$

It follows from Proposition 5.4 that L', $\exp(L)$, L^{even} and L^{odd} also have radius of convergence $\rho = 2\log(2) - 1$, are all Δ-analytic and that their behaviors near ρ are

$$(16) \qquad L'(z) \underset{z \to \rho}{\sim} \frac{1}{2\sqrt{\rho}}\left(1 - \tfrac{z}{\rho}\right)^{-1/2}; \qquad \exp(L(z)) \underset{z \to \rho}{\sim} 2;$$

$$(17) \qquad L^{\mathrm{even}}(z) \underset{z \to \rho}{\sim} \frac{1}{4\sqrt{\rho}}\left(1 - \tfrac{z}{\rho}\right)^{-1/2}; \qquad L^{\mathrm{odd}}(z) \underset{z \to \rho}{\sim} \frac{1}{4\sqrt{\rho}}\left(1 - \tfrac{z}{\rho}\right)^{-1/2}.$$

Indeed, the first estimate follows from Eq. (15) by singular differentiation [FS09, Thm. VI.8 p. 419], while the third and fourth ones are simple computations using Eq. (11) and Eq. (12).

5.4. **Distribution of induced subtrees of uniform cotrees.** We take a uniform labeled canonical cotree $\boldsymbol{t}^{(n)}$ with n leaves. We also choose uniformly at random a k-tuple $(\boldsymbol{\ell}_1, \cdots, \boldsymbol{\ell}_k)$ of distinct leaves of $\boldsymbol{t}^{(n)}$. Equivalently, $(\boldsymbol{t}^{(n)}; \boldsymbol{\ell}_1, \cdots, \boldsymbol{\ell}_k)$ is chosen uniformly at random among labeled canonical cotrees of size n with k marked leaves. We denote by $\mathbf{t}_k^{(n)}$ the labeled cotree induced by the k marked leaves.

Proposition 5.5. *Let $k \geq 2$, and let t_0 be a labeled binary cotree with k leaves. Then*

$$(18) \qquad \mathbb{P}(\mathbf{t}_k^{(n)} = t_0) \underset{n \to +\infty}{\to} \frac{(k-1)!}{(2k-2)!}.$$

Proof. We fix a labeled binary cotree t_0 with k leaves. From the definitions of $\mathbf{t}_k^{(n)}$, M and M_{t_0} we have for $n \geq k$ (we use the standard notation $[z^n]A(z)$ for the n-th coefficient of a power series A)

$$(19) \qquad \mathbb{P}(\mathbf{t}_k^{(n)} = t_0) = \frac{n! [z^n] M_{t_0}(z)}{n \cdots (n-k+1) \, n! [z^n] M(z)}.$$

Indeed, the denominator counts the total number of labeled canonical cotrees $(t; \ell_1, \cdots, \ell_k)$ of size n with k marked leaves. The numerator counts those tuples, for which $(\ell_1, \cdots, \ell_k)$ induce the subtree t_0. The quotient is therefore the desired probability.

By Theorem 5.3, and using the notation introduced therein, we have

$$M_{t_0} = (L')(\exp(L))^{n_v} (L^{\bullet})^k (L^{\text{odd}})^{n_=} (L^{\text{even}})^{n_{\neq}}.$$

Since t_0 is binary, we have $n_v = k-1$ and $n_= + n_{\neq} = k-2$. We now consider the asymptotics around $z = \rho$. Using Eq. (16) and (17) and recalling that $L^{\bullet}(z) = zL'(z)$, we get

$$M_{t_0}(z) \underset{z \to \rho}{\sim} \rho^k \left(\frac{1}{2\sqrt{\rho}} \left(1 - \frac{z}{\rho} \right)^{-1/2} \right)^{k+1} 2^{k-1} \left(\frac{1}{4\sqrt{\rho}} \left(1 - \frac{z}{\rho} \right)^{-1/2} \right)^{k-2}$$

$$\underset{z \to \rho}{\sim} \frac{\rho^{1/2}}{2^{2k-2}} \left(1 - \frac{z}{\rho} \right)^{-(k-1/2)}.$$

By the transfer theorem ([FS09, Corollary VI.1 p.392]) we obtain

$$[z^n] M_{t_0}(z) \underset{n \to +\infty}{\sim} \frac{\rho^{1/2}}{2^{2k-2} \rho^n} \frac{n^{k-3/2}}{\Gamma(k-1/2)} = \frac{(k-1)!}{\sqrt{\pi}(2k-2)!} \frac{n^{k-3/2}}{\rho^{n-1/2}}.$$

Applying again the transfer theorem to $M(z) = 2L(z) - z$ whose asymptotics is given in Eq. (15), we also have

$$n(n-1) \ldots (n-k+1)[z^n] M(z) \underset{n \to +\infty}{\sim} n^k (-2\sqrt{\rho}) \frac{n^{-3/2}}{\rho^n \Gamma(-1/2)} \sim \frac{n^{k-3/2}}{\rho^{n-1/2}\sqrt{\pi}}.$$

Finally, $\mathbb{P}(\mathbf{t}_k^{(n)} = t_0) \to \frac{(k-1)!}{(2k-2)!}$. □

5.5. **Proof of Theorems 1.1 and 1.2 in the labeled case.** Since labeled canonical cotrees and labeled cographs are in bijection, $\mathsf{Cograph}(\boldsymbol{t}^{(n)})$ is a uniform labeled cograph of size n, *i.e.* is equal to $\boldsymbol{G}_n$ in distribution. Thus Theorem 1.1 follows from Lemma 4.4 and Proposition 5.5. Theorem 1.2 is now a consequence of Theorem 1.1, combined with Propositions 3.11 and 4.5.

6. CONVERGENCE OF UNLABELED COGRAPHS TO THE BROWNIAN COGRAPHON

6.1. **Reducing unlabeled canonical cotrees to labeled objects.** In this section, we are interested in unlabeled cographs. They are in one-to-one correspondence with unlabeled canonical cotrees. We denote by $\overline{\mathcal{V}}$ the class of unlabeled canonical cotrees and by $\overline{\mathcal{U}}$ the class of rooted non-plane unlabeled trees with no unary nodes, counted by the number of leaves. If V and U are their respective ordinary generating functions, then clearly, $V(z) = 2U(z) - z$.

The class $\overline{\mathcal{U}}$ may be counted using the multiset construction and the Pólya exponential [FS09, Thm. I.1]: a tree of $\overline{\mathcal{U}}$ is either a single leaf or a multiset of cardinality at least 2 of trees of $\overline{\mathcal{U}}$, yielding the following equation:

$$(20) \qquad U(z) = z + \exp\left(\sum_{r \geq 1} \frac{1}{r} U(z^r) \right) - 1 - U(z).$$

As in the labeled case, we want to count the number of pairs (t, I) where t is a cotree of $\overline{\mathcal{V}}$ with n leaves, and I is a k-tuple of leaves of t (considered labeled by the order in which they appear in the tuple), such that the subtree induced by I in t is a given labeled cotree t_0.

To that end, we would need to refine Eq. (20) to count trees with marked leaves, inducing a given subtree, in a similar spirit as in Theorem 5.3. There is however a major difficulty here, which we now explain. There are two ways of looking at tuples of marked leaves in unlabeled trees.

- We consider pairs (t, I), where t is a labeled tree and I a k-tuple of leaves of t. Then we look at orbits $\overline{(t, I)}$ of such pairs under the natural relabeling action.
- Or we first consider orbits $\overline{t}$ of labeled trees under the relabeling action, *i.e.* unlabeled trees. For each such orbit we fix a representative and consider pairs $(\overline{t}, I)$, where I is a k-tuple of leaves of the representative of $\overline{t}$.

In the second model, every unlabeled tree has exactly $\binom{n}{k}$ marked versions, which is not the case in the first model[8]. Consequently, if we take an element uniformly at random in the second model, the underlying unlabeled tree is a uniform unlabeled tree, while this property does not hold in the first model.

Our goal is to study uniform random unlabeled cographs of size n, where we next choose a uniform random k-tuple of leaves. This corresponds exactly to the second model.

The problem is that combinatorial decompositions of unlabeled combinatorial classes are suited to study the first model (unlabeled objects are orbits of labeled objects under relabeling). In particular, Theorem 5.3 has an easy analogue for counting unlabeled trees with marked leaves inducing a given labeled cotree in the first sense, but not in the second sense.

---

[8]*E.g.*, the tree with three leaves all attached to the root, two of which are marked, has only one marked version in the first model.

To overcome this difficulty, we consider the following labeled combinatorial class:

$$\mathcal{U} = \{(t, a) : t \in \mathcal{L}, a \text{ a root-preserving automorphism of } t\}$$

where $\mathcal{L}$ is the family of non-plane labeled rooted trees in which internal nodes have at least two children, studied in Section 5. We define the size of an element (t, a) of $\mathcal{U}$ as the number of leaves of t. This set is relevant because of the following easy but key observation.

Proposition 6.1. *Let Φ denote the operation of forgetting both the labels and the automorphism. Then, $\Phi(\mathcal{U}) = \overline{\mathcal{U}}$ and every $t \in \overline{\mathcal{U}}$ of size n has exactly $n!$ preimages by Φ. As a result, the ordinary generating series U of $\overline{\mathcal{U}}$ equals the exponential generating function of $\mathcal{U}$ and the image by Φ of a uniform random element of size n in $\mathcal{U}$ is a uniform random element of size n in $\overline{\mathcal{U}}$.*

Proof. The number of preimages of $t \in \overline{\mathcal{U}}$ is the number of automorphisms of t times the number of distinct labelings of t, which equals $n!$ by the orbit-stabilizer theorem. The other claims follow immediately. $\qquad\square$

Working with $\mathcal{U}$ instead of $\overline{\mathcal{U}}$ solves the issue raised above concerning marking, since we have labeled objects. However the additional structure (the automorphism) has to be taken into account in combinatorial decompositions, but this turns out to be tractable (at least asymptotically).

6.2. **Combinatorial decomposition of $\mathcal{U}$.** We first describe a method for decomposing pairs (t, a) in $\mathcal{U}$ at the root of t, which explains *combinatorially* why the exponential generating function U of $\mathcal{U}$ satisfies Eq. (20). This combinatorial interpretation of Eq. (20) is necessary for the refinement with marked leaves done in the next section.

Let $(t, a) \in \mathcal{U}$. Then t is a non-plane rooted labeled tree with no unary nodes and a is one of its root-preserving automorphisms. Assuming t is of size at least two, we denote by $v_1, \ldots v_d$ the children of the root, and $t_1, \ldots, t_d$ the fringe subtrees rooted at these nodes, respectively.

Because a is a root-preserving automorphism, it preserves the set of children of the root, hence there exists a permutation $\pi \in \mathfrak{S}_d$ such that $a(v_i) = v_{\pi(i)}$ for all $1 \leq i \leq d$. Moreover, we have necessarily $a(t_i) = t_{\pi(i)}$ for all $1 \leq i \leq d$.

Let $\pi = \prod_{s=1}^{p} c_s$ be the decomposition of π into disjoint cycles, including cycles of length one. Let $c_s = (i_1, \ldots, i_r)$ be one of them. We consider the forest $t(c_s)$ formed by the trees $t_{i_1}, \ldots, t_{i_r}$. Then the pair $(t(c_s), a_{|t(c_s)})$ lies in the class $\mathcal{C}_r$ of pairs (f, a), where f is a forest of r trees and a an automorphism of f acting cyclically on the components of f.

The tree t can be recovered by adding a root to $\biguplus_{s=1}^{p} t(c_s)$. Moreover, a is clearly determined by $(a_{|t(c_s)})_{1 \leq s \leq p}$. So we can recover (t, a) knowing $(t(c_s), a_{|t(c_s)})_{1 \leq s \leq p}$. Recall that the cycles c_s indexing the latter vector are the cycles of the permutation π, which has size at least 2 (the root of t has degree at least 2). Since permutations π are sets of cycles, we get the following decomposition of $\mathcal{U}$ (using as usual $\mathcal{Z}$ for the atomic class, representing here the single tree with one leaf):

$$(21) \qquad\qquad \mathcal{U} = \mathcal{Z} \uplus \text{SET}_{\geq 1}\left(\biguplus_{r \geq 1} \mathcal{C}_r \right) \setminus \mathcal{C}_1,$$

We then relate $\mathcal{C}_r$ to $\mathcal{U}$ to turn Eq. (21) into a recursive equation. Let (f, a) be an element of $\mathcal{C}_r$, and t be one of the component of f. We write $f = \{t_1, \cdots, t_r\}$ such that $t_1 = t$ and a acts on these components by $t_1 \xrightarrow{a} t_2 \xrightarrow{a} \cdots \xrightarrow{a} t_r \xrightarrow{a} t_1$ (this numbering of the components of f is uniquely determined by t). We then encode (f, a) by a unique tree $\widehat{t}$ isomorphic to t_1, with multiple labelings, *i.e.* each leaf $\widehat{v} \in \widehat{t}$, corresponding to $v \in t_1$, is labeled by $(v, a(v), a^2(v), \ldots, a^{r-1}(v))$. Finally, a^r induces an automorphism of $\widehat{t}$. Consequently, $(\widehat{t}, a^r)$ is an element of the combinatorial class $\mathcal{U} \circ \mathcal{Z}^r$, *i.e.* an element of $\mathcal{U}$ where each atom (here, each leaf of the tree) carries a vector of r labels; the size of an element of $\mathcal{U} \circ \mathcal{Z}^r$ is the total number of labels, *i.e.* r times the number of leaves of $\widehat{t}$. The forest f and its marked component t are trivially recovered from $(\widehat{t}, a^r)$. Since a forest automorphism is determined by its values on leaves, we can recover a as well.

This construction defines a size-preserving bijection between triples (f, a, t), where (f, a) is in $\mathcal{C}_r$ and t one of the component of f, and elements of $\mathcal{U} \circ \mathcal{Z}^r$. Forgetting the marked component t, this defines an r-to-1 size-preserving correspondence from $\mathcal{C}_r$ to $\mathcal{U} \circ \mathcal{Z}^r$. Together with Eq. (21), this gives the desired combinatorial explanation to the fact that the exponential generating function of $\mathcal{U}$ satisfies Eq. (20).

We now introduce the combinatorial class $\mathcal{D}$ of trees in $\mathcal{U}$ with size ≥ 2 such that no child of the root is fixed by the automorphism. This means that there is no cycle of size 1 in the above decomposition of π into cycles. Therefore, the exponential generating function of $\mathcal{D}$ satisfies

$$(22) \qquad D(z) = \exp\left(\sum_{r \geq 2} \frac{1}{r} U(z^r)\right) - 1.$$

Note that introducing the series D is classical when applying the method of singularity analysis on unlabeled unrooted structures (aka Pólya structures), see, *e.g.*, [FS09, p 476]. However, interpreting it combinatorially with objects of $\mathcal{D}$ is not standard, but necessary for our purpose.

In the sequel, for $k \geq 0$, we write $\exp_{\geq k}(z) = \sum_{z \geq k} \frac{z^k}{k!}$. Algebraic manipulations from Eq. (20) allow to rewrite the equation for U as

$$(23) \qquad U = z + \exp_{\geq 2}(U) + D \exp(U).$$

Moreover, Eq. (23) has a combinatorial interpretation. Indeed, pairs (t, a) in $\mathcal{U}$ of size at least 2 can be split into two families as follows.

- The first family consists in pairs (t, a), for which all children of the root are fixed by the automorphism a; adapting the above combinatorial argument, we see that the generating series of this family is $\exp_{\geq 2}(U)$ (recall that the root has at least 2 children).
- The second family consists in pairs (t, a), where some children of the root are moved by the automorphism a. Taking the root, its children moved by a and their descendants give a tree t_1 such that $(t_1, a_{|t_1})$ is in $\mathcal{D}$. Each child c of the root fixed by a with its descendants form a tree t_c such that $(t_c, a_{|t_c})$ is in $\mathcal{U}$. We have a (possibly empty) unordered set of such children. Therefore, elements in this second family are described as pairs consisting of an element of $\mathcal{D}$ and a (possibly empty) unordered set of elements of $\mathcal{U}$, so that the generating series of this second family is $D \exp(U)$.

Bringing the two cases together, we obtain a combinatorial interpretation of Eq. (23). Again, this combinatorial interpretation will be important later, when refining with marked leaves.

We can now turn to defining the combinatorial classes that will appear in our decomposition. Similarly to the case of labeled cographs, we will need to consider objects of $\mathcal{U}$ (recall that these are *labeled* objects) where some leaves are marked. Here again, we need to distinguish marked leaves carrying a label (contributing to the size of the objects), and leave not carrying any label (not counted in the size). We keep the terminology of our section on labeled cographs, namely we call *blossoms* marked leaves of the latter type (*i.e.* without labels) and we reserve *marked leaves* for those carrying labels.

We let $\mathcal{U}^\bullet$ (resp. $\mathcal{U}'$) be the combinatorial class of pairs (t, a) in $\mathcal{U}$ with a marked leaf (resp. blossom) in t. Their exponential generating functions are respectively $zU'(z)$ and $U'(z)$ (the derivative of $U(z)$). We also define $\mathcal{U}^\star \subset \mathcal{U}'$ as the class of pairs (t, a) in $\mathcal{U}$ with a blossom in t *which is fixed by* a. Finally, we decompose $\mathcal{U}^\star$ as $\mathcal{U}^\star = \mathcal{U}^{\text{even}} \uplus \mathcal{U}^{\text{odd}}$, according to the parity of the distance from the root to the blossom. We denote by $U^\star$, U^{even} and U^{odd} the exponential generating functions of these classes, respectively.

Proposition 6.2. *We have the following equations:*

$$(24) \qquad U^\star = 1 + U^\star \exp_{\geq 1}(U) + U^\star D \exp(U),$$

$$(25) \qquad \begin{cases} U^{\text{even}} &= 1 + U^{\text{odd}} \exp_{\geq 1}(U) + U^{\text{odd}} D \exp(U), \\ U^{\text{odd}} &= U^{\text{even}} \exp_{\geq 1}(U) + U^{\text{even}} D \exp(U). \end{cases}$$

Proof. Note that if a blossom is required to be fixed by the automorphism, then all of its ancestors are also fixed by the automorphism. Then, the equation of $U^\star$ is obtained by the same decomposition as for Eq. (23), imposing that the blossom belongs to a subtree attached to a child of the root which is fixed by the automorphism. The other two equations follow immediately. $\qquad \square$

6.3. **Enumeration of canonical cotrees with marked leaves inducing a given cotree.** We first define $\mathcal{V}$ as the class of pairs (t, a), where t is a labeled canonical cotree and a a root-preserving automorphism of t. As for $\mathcal{U}$ and $\overline{\mathcal{U}}$, we have a $n!$-to-1 map from $\mathcal{V}$ to $\overline{\mathcal{V}}$ and V can be seen either as the ordinary generating function of $\overline{\mathcal{V}}$ or the exponential generating function of $\mathcal{V}$.

We would like to find a combinatorial decomposition of pairs in $\mathcal{V}$ with marked leaves inducing a given cotree. It turns out that it is simpler and sufficient for us to work with a smaller class, which we now define.

Definition 6.3. *Let t_0 be a labeled cotree of size k. Let $\mathcal{V}_{t_0}$ be the class of tuples $(t, a; \ell_1, \ldots, \ell_k)$, where (t, a) is in $\mathcal{V}$ and $\ell_1, \ldots, \ell_k$ are distinct leaves of t (referred to as* marked leaves*) such that*

- *the marked leaves induce the subtree t_0;*
- *the following nodes are fixed by a: all first common ancestors of the marked leaves, and their children leading to a marked leaf.*

We note that, because of the second item in the above definition, not all tuples $(t, a; \ell_1, \ldots, \ell_k)$ (where (t, a) is in $\mathcal{V}$ and $\ell_1, \ldots, \ell_k$ are leaves of t) belong to some $\mathcal{V}_{t_0}$. However, we will see below (as a consequence of Proposition 6.7) that asymptotically almost all tuples $(t, a; \ell_1, \ldots, \ell_k)$

do belong to some $\mathcal{V}_{t_0}$ (even if we restrict to binary cotrees t_0, which is similar to the previous section).

Let V_{t_0} be the exponential generating series of $\mathcal{V}_{t_0}$; it is given by the following result.

Theorem 6.4. *Let t_0 be a labeled cotree with k leaves, n_v internal nodes, $n_=$ edges of the form $0 - 0$ or $1 - 1$, $n_{\neq}$ edges of the form $0 - 1$ or $1 - 0$. We have the identity*

$$V_{t_0} = (U^\star)(2U + 1 - z)^{n_v}(U^\bullet)^k(U^{\text{odd}})^{n_=}(U^{\text{even}})^{n_{\neq}}.$$

Proof. Let $(t, a; \ell_1, ..., \ell_k)$ be a tree in $\mathcal{V}_{t_0}$. The tree t with its marked leaves $\ell_1, ..., \ell_k$ can be decomposed in a unique way as in the proof of Theorem 5.3 into pieces: pink trees, blue trees, yellow trees, gray trees and green forests.

As soon as a node of t is fixed by the automorphism a, then the set of its descendants is stable by a. Therefore, the second item of Definition 6.3 ensures that each colored piece in the decomposition of t is stable by a, so that a can be decomposed uniquely into a collection of automorphisms, one for each colored piece. Consequently, from now on, we think at pieces as trees/forests *with an automorphism*.

As in Theorem 5.3, each piece can be chosen independently in a set depending on its color. Moreover, since t_0 is labeled, the pieces can be ordered in a canonical way, so that the generating series of V_{t_0} is the product of the generating series of the pieces.

- The gray subtree is a tree with an automorphism and a blossom which is fixed by the automorphism (because of the second item of Definition 6.3). As in Theorem 5.3, the decoration is forced by the context, so that we can consider the gray subtree as not decorated. The possible choices for the gray subtrees are therefore counted by $U^\star$.
- The possible choices for each green forest (and its automorphism) are counted by $1 + U + (U - z)$: the first term corresponds to the empty green piece, the second one to exactly one tree in the green forest, and the third one to a set of at least two green trees (which can be seen as a non-trivial tree in $\mathcal{U}$ by adding a root).
- The possible choices for each yellow piece are counted by $U^\bullet$, since these trees have a marked leaf which is not necessarily fixed by the automorphism.
- The possible choices for each pink piece are counted by U^{even}: the blossom must be at even distance from the root of the piece (for the same reason as in Theorem 5.3) and must be fixed by the automorphism (because of the second item of Definition 6.3).
- Similarly, the possible choices for each blue piece are counted by U^{odd}.

Bringing everything together gives the formula in the theorem. $\square$

6.4. **Asymptotic analysis.** Let ρ be the radius of convergence of U. It is easily seen that we have $0 < \rho < 1$, see, *e.g.*, [Gen16], where the numerical approximation $\rho \approx 0.2808$ is given.

Proposition 6.5. *The series $U, U', U^\star, U^{\text{even}}, U^{\text{odd}}$ all have the same radius of convergence ρ, are Δ-analytic and admit the following expansions around ρ:*

$$U(z) \underset{z \to \rho}{=} \frac{1 + \rho}{2} - \beta\sqrt{\rho - z} + o(\sqrt{\rho - z}), \qquad U'(z) \underset{z \to \rho}{\sim} \frac{\beta}{2\sqrt{\rho - z}},$$

$$2U^{\text{even}}(z) \sim 2U^{\text{odd}}(z) \sim U^\star(z) \underset{z \to \rho}{\sim} \frac{1}{2\beta\sqrt{\rho - z}},$$

for some constant $\beta > 0$.

To prove the proposition, we need the following lemma, which is standard in the analysis of Pólya structures.

Lemma 6.6. *The radius of convergence of D is $\sqrt{\rho} > \rho$.*

Proof. Since U has no constant term, for every $x \geq 1$ and $0 < z < 1$ we have $U(z^x) \leq U(z)z^{x-1}$. Hence for $0 < t < \rho$,

$$D(\sqrt{t}) = \exp_{\geq 1}\left(\sum_{r \geq 2} \frac{1}{r} U(t^{r/2})\right) \leq \exp\left(\sum_{r \geq 2} U(t)t^{r/2-1}\right) \leq \exp\left(U(t)\frac{1}{1-\sqrt{t}}\right) < \infty.$$

This implies that the radius of convergence of D is at least $\sqrt{\rho}$. Looking at Eq. (22), we see that D termwise dominates $\frac{1}{2}U(z^2)$, whose radius of convergence is $\sqrt{\rho}$. Therefore, the radius of convergence of D is exactly $\sqrt{\rho}$. $\qquad\square$

Proof of Proposition 6.5. Set $F(z, u) = z + \exp_{\geq 2}(u) + D(z)\exp(u)$. Then U verifies the equation $U = F(z, U)$, which is the setting of [BBF+19, Theorem A.6][9] (in the one dimensional case, which is then just a convenient rewriting of [FS09, Theorem VII.3]). The only non-trivial hypothesis to check is the analyticity of F at $(\rho, U(\rho))$. This holds because $\exp$ has infinite radius of convergence, while D has radius of convergence $\sqrt{\rho} > \rho$ from Lemma 6.6.

From items vi) and vii) of [BBF+19, Theorem A.6], we have that U and $(1 - \partial_u F(z, U(z)))^{-1}$ have radius of convergence ρ, are Δ-analytic and that $\partial_u F(\rho, U(\rho)) = 1$. Moreover,

$$U(z) \underset{z \to \rho}{=} U(\rho) - \frac{\beta}{\zeta}\sqrt{\rho - z} + o(\sqrt{\rho - z}), \qquad U'(z) \underset{z \to \rho}{\sim} \frac{\beta}{2\zeta\sqrt{\rho - z}}$$

$$(1 - \partial_u F(z, U(z)))^{-1} \underset{z \to \rho}{\sim} \frac{1}{2\beta\zeta\sqrt{\rho - z}},$$

where $\beta = \sqrt{\partial_z F(\rho, U(\rho))}$ and $\zeta = \sqrt{\frac{1}{2}\partial_u^2 F(\rho, U(\rho))}$.

We have $\partial_u F(z, u) = \exp_{\geq 1}(u) + D(z)\exp(u) = F(z, u) + u - z$. Hence $\partial_u F(z, U(z)) = 2U(z) - z$. Recalling that $\partial_u F(\rho, U(\rho)) = 1$, we get $U(\rho) = \frac{1+\rho}{2}$. In addition, $\partial_u^2 F(z, u) = \exp(u) + D(z)\exp(u) = \partial_u F(z, u) + 1$. Therefore, $\partial_u^2 F(\rho, U(\rho)) = 2$ and $\zeta = 1$. The asymptotics of U and U' follow.

Regarding $U^\star$, Eq. (24) implies that $U^\star = (1 - \partial_u F(z, U(z)))^{-1}$. Similarly solving the system of equations (25) we get $U^{\text{even}} = (1 - (\partial_u F(z, U(z)))^2)^{-1}$ and $U^{\text{odd}} = \partial_u F(z, U(z))U^{\text{even}}$. By the daffodil lemma [FS09, Lemma IV.1, p.266], we have $|\partial_u F(z, U(z))| < 1$ for $|z| \leq \rho$ and $z \neq \rho$. In particular, $\partial_u F(z, U(z))$ avoids the value 1 and -1 for such z. Therefore $U^\star$, U^{even} and U^{odd} are Δ-analytic. The asymptotics of $U^\star$ follows from the above results. Finally, since $\partial_u F(\rho, U(\rho)) = 1$, we have $U^{\text{even}} \sim U^{\text{odd}}$ when z tends to ρ. And, since $U^\star = U^{\text{even}} + U^{\text{odd}}$, their asymptotics follow. $\qquad\square$

---

[9]We warn the reader that the function U appearing in [BBF+19, Theorem A.6] is unrelated to the quantity $U(z)$ in the present article (which corresponds instead to $Y(z)$ in [BBF+19, Theorem A.6]).

6.5. **Distribution of induced subtrees of uniform cotrees.** We take a uniform unlabeled canonical cotree $\boldsymbol{t}^{(n)}$ with n leaves, *i.e.* a uniform element of size n in $\overline{\mathcal{V}}$. We also choose uniformly at random a k-tuple of distinct leaves of $\boldsymbol{t}^{(n)}$. We denote by $\mathbf{t}_k^{(n)}$ the labeled cotree induced by the k marked leaves.

Proposition 6.7. *Let $k \geq 2$, and let t_0 be a labeled binary cotree with k leaves. Then*

$$(26) \qquad \mathbb{P}(\mathbf{t}_k^{(n)} = t_0) \underset{n \to +\infty}{\to} \frac{(k-1)!}{(2k-2)!}.$$

Proof. We take a uniform random pair $(\boldsymbol{T}^{(n)}, \boldsymbol{a})$ of $\mathcal{V}$ of size n with a k-tuple of distinct leaves of $\boldsymbol{T}^{(n)}$, also chosen uniformly. We denote by $\mathbf{T}_k^{(n)}$ the cotree induced by the k marked leaves. Since the forgetting map from $\mathcal{V}$ to $\overline{\mathcal{V}}$ is $n!$-to-1, $\boldsymbol{T}_k^{(n)}$ is distributed as $\boldsymbol{t}_k^{(n)}$. Hence, similarly to Eq. (19), we have

$$\mathbb{P}(\mathbf{t}_k^{(n)} = t_0) = \mathbb{P}(\mathbf{T}_k^{(n)} = t_0) \geq \frac{n![z^n]V_{t_0}(z)}{n \ldots (n-k+1)\, n![z^n]V(z)}.$$

The inequality comes from the fact that $\mathcal{V}_{t_0}$ does not consist of all pairs in $\mathcal{V}$ with a k-tuple of marked leaves inducing t_0, but only of some of them (see the additional constraint in the second item of Definition 6.3).

From Theorem 6.4, we have

$$V_{t_0} = (U^\star)(2U + 1 - z)^{n_v}(U^\bullet)^k(U^{\mathrm{odd}})^{n_=}(U^{\mathrm{even}})^{n_{\neq}}.$$

Recalling that $U^\bullet(z) = zU'(z)$, we use the asymptotics for $U, U', U^\star, U^{\mathrm{even}}, U^{\mathrm{odd}}$ (given in Proposition 6.5) and furthermore the equalities $n_v = k - 1$ and $n_= + n_{\neq} = k - 2$ (which hold since t_0 is binary) to obtain

$$V_{t_0}(z) \underset{z \to \rho}{\sim} \frac{1}{2\beta}2^{k-1}\left(\frac{\beta}{2} \cdot \rho\right)^k \left(\frac{1}{4\beta}\right)^{k-2}(\rho - z)^{-(k-1/2)}$$

$$\underset{z \to \rho}{\sim} \frac{\beta\rho^k}{2^{2k-2}}(\rho - z)^{-(k-1/2)} = \frac{\beta\sqrt{\rho}}{2^{2k-2}}(1 - \tfrac{z}{\rho})^{-(k-1/2)}.$$

By the transfer theorem ([FS09, Corollary VI.1 p.392]) we have

$$[z^n]V_{t_0}(z) \underset{n \to +\infty}{\sim} \frac{\beta\sqrt{\rho}}{2^{2k-2}\rho^n}\frac{n^{k-3/2}}{\Gamma(k-1/2)} = \beta\frac{(k-1)!}{\sqrt{\pi}(2k-2)!}\frac{n^{k-3/2}}{\rho^{n-1/2}}$$

Besides, using $V(z) = 2U(z) - z$, Proposition 6.5, and the transfer theorem as above, we have

$$n(n-1)\ldots(n-k+1)[z^n]V(z) \underset{n \to +\infty}{\sim} n^k(-2\beta\sqrt{\rho})\frac{n^{-3/2}}{\rho^n\Gamma(-1/2)} \sim \beta\frac{n^{k-3/2}}{\rho^{n-1/2}\sqrt{\pi}}.$$

Finally, $\liminf_{n\to\infty}\mathbb{P}(\mathbf{t}_k^{(n)} = t_0) \geq \frac{(k-1)!}{(2k-2)!}$. To conclude, recall (as seen in the proof of Lemma 4.4) that summing the right-hand-side over all labeled binary cotrees t_0 of size k gives 1, from which the proposition follows. $\qquad\square$

6.6. **Proof of Theorems 1.1 and 1.2 in the unlabeled case.** The argument is identical to the labeled case. Recall that $\boldsymbol{t}^{(n)}$ is a uniform unlabeled canonical cotree of size n, so that $\mathsf{Cograph}(\boldsymbol{t}^{(n)})$ is a uniform unlabeled cograph of size n, *i.e.* has the same ditribution as $\boldsymbol{G}_n^u$. Thus Theorem 1.1 follows from Lemma 4.4 and Proposition 6.7, and Theorem 1.2 is then a consequence of Theorem 1.1 and Propositions 3.11 and 4.5.

7. VERTEX CONNECTIVITY

A connected graph G is said to be k-connected if it does not contain a set of $k - 1$ vertices whose removal disconnects the graph. The *vertex connectivity* $\kappa(G)$ is defined as the largest k such that G is k-connected.

Throughout this section, $\boldsymbol{G}_n$ (resp. $\boldsymbol{G}_n^u$) is a uniform random labeled (resp. unlabeled) cograph of size n, *conditioned to be connected*. The aim of this section is to prove that the random variable $\kappa(\boldsymbol{G}_n)$ (resp. $\kappa(\boldsymbol{G}_n^u)$) converges in distribution to a non-trivial random variable (without renormalizing). The limiting distributions in the labeled and unlabeled cases are different.

A cograph G (of size at least 2) is connected if and only if the root of its canonical cotree is decorated by 1. (This implies that in both cases a uniform cograph of size n is connected with probability $1/2$ for every n.) Therefore, any connected cograph G (of size at least 2) can be uniquely decomposed as the join of $F_1, \ldots, F_k$ where each F_i is either a disconnected cograph or a one-vertex graph. Moreover, the cographs F_i are those whose canonical cotrees are the fringe subtrees attached to the root of the canonical cotree of G. Throughout this section, we refer to the F_i's as the *components* of G. The following lemma, illustrated by Fig. 6, gives a simple characterization of $\kappa(G)$ when G is a cograph.

Lemma 7.1. *Let G be a connected cograph which is not a complete graph. Let $F_1, \ldots, F_k$ be the components of G. It holds that*

$$\kappa(G) = |G| - \max_{1 \leq i \leq k}\{|F_i|\}.$$

Proof. We reorder the components such that $|F_1| = \max_i |F_i|$. Because G is not a complete graph, F_1 is not a one-vertex graph, and therefore is disconnected. Let us denote by $v_1, \ldots, v_r$ the vertices of $F_2 \cup F_3 \cup \cdots \cup F_k$. We have to prove that $\kappa(G) = r$.
Proof of $\kappa(G) \leq r$. If we remove all vertices $v_1, \ldots, v_r$ then we are left with F_1 which is disconnected.
Proof of $\kappa(G) \geq r$. If we remove only $r - 1$ vertices then there remains at least one v_j among $v_1, \ldots, v_r$. Let us denote by F_i the component of v_j. There also remains at least a vertex $v \notin F_i$ (or $|F_i|$ would be larger than $|F_1|$). Consequently, v and v_j are connected by an edge, and every remaining vertex is connected to v_j (when not in F_i) or to v (when not in the component containing v), so that G remains connected. Therefore we must remove at least r points to disconnect G. $\qquad\square$

Theorem 7.2. *Let $M(z)$ (resp. $V(z)$) be the exponential (resp. ordinary) generating series of labeled (resp. unlabeled) cographs. Their respective radii of convergence are $\rho = 2\log(2) - 1$*
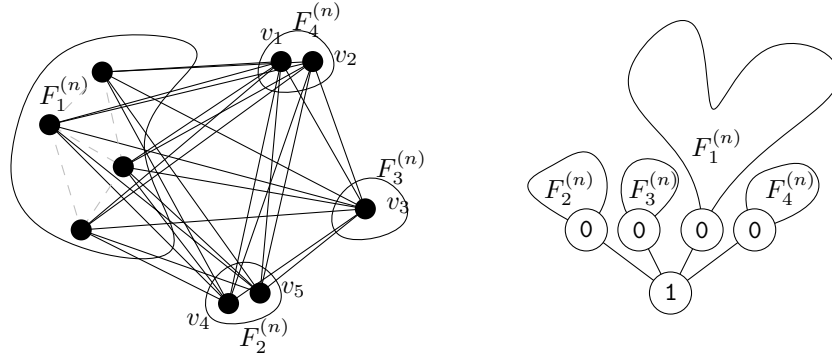
FIGURE 6. A connected cograph and the corresponding cotree. The connectivity degree of this graph is $|F_2| + |F_3| + |F_4| = 2 + 2 + 1 = 5$.

and $\rho_u \approx 0.2808$. For $j \geq 1$, set

$$\pi_j = \rho^j [z^j] M(z), \qquad \pi_j^u = \rho_u^j [z^j] V(z).$$

Then $(\pi_j)_{j \geq 1}$ and $(\pi_j^u)_{j \geq 1}$ are probability distributions and, for every fixed $j \geq 1$,

(27) $$\mathbb{P}(\kappa(\boldsymbol{G}_n) = j) \underset{n \to +\infty}{\to} \pi_j, \qquad \mathbb{P}(\kappa(\boldsymbol{G}_n^u) = j) \underset{n \to +\infty}{\to} \pi_j^u.$$

Remark 7.3. Readers acquainted with Boltzmann samplers may note that $(\pi_j)_{j \geq 1}$ and $(\pi_j^u)_{j \geq 1}$ are distributions of sizes of Boltzmann-distributed random labeled and unlabeled cographs, respectively. The Boltzmann parameters are chosen to be the radii of convergence. We do not have a direct explanation of this fact.

Proof. Recall from Sections 5 and 6 that $M(z) = 2L(z) - z$ and $V(z) = 2U(z) - z$. It follows from Propositions 5.4 and 6.5 that $\rho = 2\log(2) - 1$ and $\rho_u \approx 0.2808$ are their respective radii of convergence. We first prove that (π_j) (resp. (π_j^u)) sum to one:

$$\sum_{j \geq 1} \pi_j = \sum_{j \geq 1} \rho^j [z^j] M(z) = M(\rho) = 2L(\rho) - \rho = 1,$$

$$\sum_{j \geq 1} \pi_j^u = \sum_{j \geq 1} \rho_u^j [z^j] V(z) = V(\rho_u) = 2U(\rho_u) - \rho_u = 1,$$

using Propositions 5.4 and 6.5 for the last equalities.

For the remaining of the proof, we fix $j \geq 1$. In the labeled case, let $\boldsymbol{T}_n$ be the canonical cotree of $\boldsymbol{G}_n$. Since $\boldsymbol{G}_n$ is conditioned to be connected, $\boldsymbol{T}_n$ is a uniform labeled canonical cotree of size n conditioned to have root decoration 1. Forgetting the decoration, we can see it as a uniform random element of size n in $\mathcal{L}$.

Let $n > 2j$. As the components of $\boldsymbol{G}_n$ correspond to the subtrees attached to the root of $\boldsymbol{T}_n$, using Lemma 7.1 we have $\kappa(\boldsymbol{G}_n) = j$ if and only if $\boldsymbol{T}_n$ is composed of a tree of $\mathcal{L}$ of size $n - j$ and $k \geq 1$ trees of $\mathcal{L}$ of total size j, all attached to the root. Since $n > 2j$, the fringe subtree of size $n - j$ is uniquely defined, and there is only one such decomposition. Therefore, for every

fixed $j \geq 1$ and $n > 2j$, we have

$$\mathbb{P}(\kappa(\boldsymbol{G}_n) = j) = \frac{[z^{n-j}]L(z)\,[z^j]\big(e^{L(z)} - 1\big)}{[z^n]L(z)}.$$

From Proposition 5.4, the series $L(z)$ has radius of convergence ρ, is Δ-analytic and has a singular expansion amenable to singularity analysis. Thus, the transfer theorem ensures that $\frac{[z^{n-j}]L(z)}{[z^n]L(z)}$ tends to ρ^j, so that

$$\mathbb{P}(\kappa(\boldsymbol{G}_n) = j) \underset{n \to +\infty}{\to} \rho^j\,[z^j]\big(e^{L(z)} - 1\big) = \pi_j,$$

where we used $M(z) = e^{L(z)} - 1$ (see Eq. (10)).

In the unlabeled case, let $\boldsymbol{T}_n^u$ be the canonical cotree of $\boldsymbol{G}_n^u$. Like in the labeled case, forgetting the decoration, it is a uniform element of $\overline{\mathcal{U}}$ of size n. Let $n > 2j$. We have $\kappa(\boldsymbol{G}_n^u) = j$ if and only if $\boldsymbol{T}_n^u$ has a fringe subtree of size $n - j$ at the root.

Let us count the number of trees of $\overline{\mathcal{U}}$ of size n that have a fringe subtree of size $n - j$ at the root. Since $n - j > n/2$, there must be exactly one such fringe subtree, and there are $[z^{n-j}]U(z)$ choices for it. Removing it, the rest of the tree contains j leaves, and is either a tree of $\overline{\mathcal{U}}$ of size ≥ 2 (if the root still has degree at least 2), or a tree formed by a root and a single tree of $\overline{\mathcal{U}}$ attached to it. So the number of choices for the rest is $[z^j](2U(z) - z)$. We deduce that for $j \geq 1$ and $n > 2j$,

$$\mathbb{P}(\kappa(\boldsymbol{G}_n^u) = j) = \frac{[z^{n-j}]U(z)\,[z^j](2U(z) - z)}{[z^n]U(z)}.$$

From Proposition 6.5, the series $U(z)$ has radius of convergence ρ_u, is Δ-analytic and has a singular expansion amenable to singularity analysis. The transfer theorem ensures that $\frac{[z^{n-j}]U(z)}{[z^n]U(z)}$ tends to ρ_u^j, so that

$$\mathbb{P}(\kappa(\boldsymbol{G}_n^u) = j) \underset{n \to +\infty}{\to} \rho_u^j\,[z^j](2U(z) - z) = \pi_j^u$$

where we used $V(z) = 2U(z) - z$. $\hfill\square$

Remark 7.4. In the labeled case, we could have used Lemma 7.1 and local limit results for trees instead of the generating series approach above. Indeed, the canonical cotree of $\boldsymbol{G}_n$ (without its decorations) is distributed as a Galton-Watson tree with an appropriate offspring distribution conditioned on having n leaves. Such conditioned Galton-Watson trees converge in the local sense near the root towards a Kesten's tree [AD15, Section 2.3.13]. Since Kesten's trees have a unique infinite path from the root, this convergence implies the convergence (without renormalization) of the sizes of all components of $\boldsymbol{G}_n$ but the largest one. Therefore the sum $\kappa(\boldsymbol{G}_n)$ of these sizes also converges (without renormalization); the limit can be computed (at least in principle) using the description of Kesten's trees.

In the unlabeled case, the canonical cotree of $\boldsymbol{G}_n^u$ (without its decorations) belongs to the family of random *Pólya* trees. Such trees are *not* conditioned Galton-Watson trees. For scaling limits, it has been proven they can be approximated by conditioned Galton-Watson trees and hence converge under suitable conditions to the Brownian Continuum Random Tree [PS18], but we are not aware of any local limit result for such trees.

References

[AD15] R. Abraham, J.-F. Delmas. An introduction to Galton-Watson trees and their local limits. Preprint arXiv:1506.05571 (2015).

[BBF+18] F. Bassino, M. Bouvel, V. Féray, L. Gerin, A. Pierrot. The Brownian limit of separable permutations. *Ann. Probab.*, 46(4): 2134–2189 (2018).

[BBF+19] F. Bassino, M. Bouvel, V. Féray, L. Gerin, M. Maazoun, A. Pierrot. Scaling limits of permutation classes with a finite specification: a dichotomy. Preprint arXiv:1903.07522 (2019).

[BBF+20a] F. Bassino, M. Bouvel, V. Féray, L. Gerin, M. Maazoun, A. Pierrot. Universal limits of substitution-closed permutation classes, *J. Eur. Math. Soc.* , 22 (11), pp. 3565–3639 (2020).

[BBF+20b] J. Borga, M. Bouvel, V. Féray, B. Stufler. A decorated tree approach to random permutations in substitution-closed classes. *Electron. J. Probab.*, 25, paper no. 67: 1–52 (2020).

[BCH+08] A. Bretscher, D. Corneil, M. Habib, C. Paul. A simple linear time LexBFS cograph recognition algorithm. *SIAM J. Discrete Math.*, 22(4): 1277–1296 (2008).

[BCL+08] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, K. Vesztergombi. Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6): 1801–1851 (2008).

[BCL11] P. J. Bickel, A. Chen, E. Levina. The method of moments and degree distributions for network models *Ann. Statist.* 39(5): 2280–2301 (2011).

[BMN16] M. Bouvel, M. Mishna, C. Nicaud. Some families of trees arising in permutation analysis. *Electron. J. Combin.*, 27(2), paper no. 2.20: 1–296 (2020).

[Bog07] V.I.Bogachev. *Measure theory (Vol. II)*. Springer (2007).

[BP94] J. Bertoin, J. Pitman. Path transformations connecting Brownian bridge, excursion and meander. *Bull. Sci. Math.*, 118(2): 147–166 (1994).

[CLS81] D. G. Corneil, H. Lerchs, L. Stewart Burlingham. Complement reducible graphs. *Discrete Appl. Math.*, 3(3): 163–174 (1981).

[CP19] G. Chapuy, G. Perarnau, Connectivity in bridge-addable graph classes: The McDiarmid–Steger–Welsh conjecture. *J. Combin. Theory, Series B*, vol. 136, p. 44–71 (2019).

[CPS84] D. G. Corneil, Y. Perl, L. K. Stewart. A linear recognition algorithm for cographs. *SIAM J. Comput.*, 14(4): 926–934 (1984).

[DDS18] J.-F. Delmas, J.-S. Dhersin, M. Sciauveau. Asymptotic for the cumulative distribution function of the degrees and homomorphism densities for random graphs sampled from a graphon. *Random Struct. Algor.*, vol. 58, p. 94–149 (2021).

[DFL+04] Ph. Duchon, Ph. Flajolet, G. Louchard, G. Schaeffer. Boltzmann Samplers for the Random Generation of Combinatorial Structures *Combinatorics, Probability and Computing*, 13(4–5): 577–625, 2004.

[DHJ08] P. Diaconis, S. Holmes, S. Janson. Threshold Graph Limits and Random Threshold Graphs. *Internet Math.*, 5 (3): 267 – 320 (2008).

[DJ08] P. Diaconis, S. Janson. Graph limits and exchangeable random graphs. *Rendiconti di Matematica*, vol. 28, p. 33–61 (2008).

[DKS17] C. Dowden, M. Kang, P. Sprüssel. The evolution of random graphs on surfaces. *Electron. Notes in Discrete Math.*, vol. 61, p. 367–373 (2017).

[FS09] Ph. Flajolet, R. Sedgewick. *Analytic combinatorics*. Cambridge University Press (2009).

[Gen16] A.Genitrini. Full asymptotic expansion for Pólya structures. Proceedings of *27th International Meeting on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms* (AofA'16), p. 151–162 (2016).

[GJW18] B. Gittenberger, E. Y. Jin, M. Wallner. On the shape of random Pólya structures. *Discrete Math.*, 341(4): 896–911 (2018).

[HJS18] H. Hatami, S. Janson, B. Szegedy. Graph properties, graph limits and entropy. *J Graph Theory*. 87: 208–229 (2018).

[HKM+13] C. Hoppen, Y. Kohayakawa, C. G. Moreira, B. Rath, R. M. Sampaio. Limits of permutation sequences. *J. Combin. Theory, Series B*, 103(1): 93–113 (2013).

[HP05] M. Habib, C. Paul. A simple linear time algorithm for cograph recognition. *Discrete Appl. Math.*, 145(2): 183–197 (2005).

[Jan12] S. Janson. Simply generated trees, conditioned Galton–Watson trees, random allocations and condensation. *Probab. Surv.*, vol. 9, p. 103–252 (2012).

[Jan16] S. Janson. Graph limits and hereditary properties. *European J. Combin.*, vol. 52 , p. 321–337 (2016).

[Kal17] O. Kallenberg. *Random Measures, Theory and Applications*. Vol. 77 of *Probability Theory and Stochastic Modelling*, 680 p. +xxviii, Springer (2017).

[LG05] J.-F. Le Gall. Random trees and applications. *Probab. Surv.*, vol. 2, p. 245–311 (2005).

[Lov12] L. Lovász. *Large networks and graph limits*. American Mathematical Society (2012).

[LS06] L. Lovász, B. Szegedy. Limits of dense graph sequences, *J. Combin. Theory, Series B*, 96(6): 933-957 (2006).

[Maa17] M. Maazoun, On the Brownian separable permuton. *Combin. Probab. Comput.*, 29(2): 241–266 (2020).

[MSW06] C. McDiarmid, A. Steger, D. Welsh. Random graphs from planar and other addable classes. In *Topics in Discrete Mathematics*, p. 231-–246. Springer (2006).

[MY19] C. McDiarmid, N. Yolov. Random perfect graphs. *Random Struct. Algor.*, vol. 54, p. 148–186 (2019).

[Noy14] M. Noy. Random planar graphs and beyond. In *Proceedings of the International Congress Mathematicians (ICM)*, Seoul 2014, vol. IV, p. 407–430 (2014).

[OEIS] N. J. A. Sloane. The Encyclopedia of Integer Sequences. `https://oeis.org/`.

[PS18] K. Panagiotou, B. Stufler. Scaling limits of random Pólya trees. *Probab. Theory Relat. Fields*, 170(3-4): 801–820 (2018).

[PSW16] K. Panagiotou, B. Stufler, and K. Weller. Scaling limits of random graphs from subcritical classes. *Ann. Probab.*, 44(5): 3291–3334 (2016).

[Ros11] N. Ross. Fundamentals of Stein's method. *Probab. Surv.*, vol. 8, p. 210–293 (2011).

[Sei74] S. Seinsche. On a property of the class of n-colorable graphs. *J. Combin. Theory, series B*, 16(2): 191–193 (1974).

[Stu19] B. Stufler. Graphon convergence of random cographs. *Random Struct. Algor.* To appear.

[Wiki] Wikipedia. Page entitled *Cograph*, `https://en.wikipedia.org/wiki/Cograph`, Accessed on July 1st, 2019.

(FB) Université Paris 13, Sorbonne Paris Cité, LIPN, CNRS UMR 7030, F-93430 Villetaneuse, France

E-mail address: bassino@lipn.univ-paris13.fr

(MB) Institut für Mathematik, Universität Zürich, Winterthurerstr. 190, CH-8057 Zürich, Switzerland, and, Université de Lorraine, CNRS, Inria, LORIA, F 54000 Nancy, France

E-mail address: mathilde.bouvel@loria.fr

(VF) Université de Lorraine, CNRS, IECL, F 54000 Nancy, France

E-mail address: valentin.feray@univ-lorraine.fr

(LG) CMAP, École Polytechnique, CNRS, Route de Saclay, 91128 Palaiseau Cedex, France

E-mail address: gerin@cmap.polytechnique.fr

(MM) École Normale Supérieure de Lyon, UMPA UMR 5669 CNRS, 46 allée d'Italie, 69364 Lyon Cedex 07, France

E-mail address: mickael.maazoun@ens-lyon.fr

(AP) LISN, Université Paris-Saclay, Bat. 650 Ada Lovelace, 91405 Orsay Cedex, France

E-mail address: adeline.pierrot@lri.fr