

# LINEAR-SIZED INDEPENDENT SETS IN RANDOM COGRAPHS AND INCREASING SUBSEQUENCES IN SEPARABLE PERMUTATIONS

ABSTRACT. This paper is interested in independent sets (or equivalently, cliques) in uniform random cographs. We also study their permutation analogs, namely, increasing subsequences in uniform random separable permutations.

First, we prove that, with high probability as  $n$  gets large, the largest independent set in a uniform random cograph with  $n$  vertices has size  $o(n)$ . This answers a question of Kang, McDiarmid, Reed and Scott. Using the connection between graphs and permutations via inversion graphs, we also give a similar result for the longest increasing subsequence in separable permutations. These results are proved using the self-similarity of the Brownian limits of random cographs and random separable permutations, and actually apply more generally to all families of graphs and permutations with the same limit.

Second, and unexpectedly given the above results, we show that for  $\beta > 0$  sufficiently small, the expected number of independent sets of size  $\beta n$  in a uniform random cograph with  $n$  vertices grows exponentially fast with  $n$ . We also prove a permutation analog of this result. This time the proofs rely on singularity analysis of the associated bivariate generating functions.

## 1. INTRODUCTION

This paper contains both results for graph and permutation models (connected through the mapping associating with a permutation its inversion graph); for simplicity we present these results and the related backgrounds separately.

**1.1. Independent sets in random cographs.** Cographs were introduced in the seventies by several authors independently (under various names), see *e.g.* [Sei74]. 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 whose modular decomposition does not involve any prime graph;
- the inversion graphs of separable permutations;
- the graphs which can be constructed from graphs with one vertex by taking disjoint unions and joins.

The latter characterization is the most useful for our purpose, let us introduce the terminology.

All graphs considered in this paper are *simple* (*i.e.* without multiple edges, nor loops) and not directed. 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, and we denote by  $V_G$  the set of vertices of any graph  $G$ .

---

2010 *Mathematics Subject Classification.* 60C05,05C80,05C69,05A05.

*Key words and phrases.* combinatorial graph theory, combinatorial probability, cographs, random graphs, graphons, self-similarity.

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 1.1.** *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.*

Recall that, for a given graph  $G$ , an *independent set* is a subset of vertices in  $G$  no two of which are adjacent, while a *clique* is a subset of vertices in  $G$  such that every two vertices are adjacent.

The main motivation for studying independent sets in random cographs comes from the series of papers [LRSTT10, KMRS14] on a probabilistic version of the Erdős-Hajnal conjecture.

For a graph  $G$ , a subset of its vertices is called *homogeneous* if it is either a clique or an independent set. It is well-known that every graph of size  $n$  has a homogeneous set of size at least logarithmic in  $n$ , and that this is optimal up to a constant (much work is devoted to get the precise asymptotics; this is equivalent to the computation of diagonal Ramsey numbers, see [Spe75, Sah20] for the better bounds up to date). The conjecture of Erdős and Hajnal states that, assuming that the graphs avoid any given subgraph (as an induced subgraph), homogeneous sets of size polynomial in  $n$  necessarily exist. More precisely, for any  $H$ , there exists a constant  $\varepsilon = \varepsilon(H) > 0$  such that every  $H$ -free graph has a homogeneous set of size  $n^\varepsilon$ .

Despite much effort, the Erdős-Hajnal conjecture is still widely open; see for example the survey [Chu14]. A natural relaxation of the conjecture consists in replacing "every  $H$ -free graph" in the statement above by "almost all  $H$ -free graphs". This weaker version has been established in [LRSTT10]. For a large family of constraints  $H$ , this result was further improved by Kang, McDiarmid, Reed and Scott in [KMRS14]: for those  $H$ , a uniform random  $H$ -free graph has with high probability a homogeneous set of *size linear in  $n$* . When this holds, the graph  $H$  is said to have the *asymptotic linear Erdős-Hajnal property* (see [KMRS14] for a formal definition). Kang, McDiarmid, Reed and Scott ask whether  $H = P_4$  has the asymptotic linear Erdős-Hajnal property, *i.e.* whether a uniform random cograph with  $n$  vertices has a homogeneous set of size linear in  $n$  [KMRS14, Section 5], and this question has remained open until now<sup>1</sup>.

Our first result answers this question in the negative. In the following, for a graph  $G$ , we denote  $\alpha(G)$  the maximum size of an independent set in  $G$ , also called the *independence number* of  $G$ .

**Theorem 1.2.** *Let  $\mathbf{G}_n$  be a uniform random cograph (either labeled or unlabeled) of size  $n$ . The maximum size of an independent set in  $\mathbf{G}_n$  is sublinear in  $n$ , namely  $\frac{\alpha(\mathbf{G}_n)}{n}$  converges to 0 in probability.*

<sup>1</sup> A sketch of proof that  $P_4$  does not have the linear Erdős-Hajnal property was given in 2012 in the PhD thesis of Andreas Würfl [Wür12, Chapter 9], as the result of a joint work in progress with C. Hoppen and M. Noy. However, this proof sketch contains several approximations or inaccuracies, which would need to be fixed for their argument to be a complete proof. It is not clear whether such fixes are possible.

For a discussion on the difference between the labeled and unlabeled settings, we refer to Remark 1.4 below. We recall the standard notation for comparison of random variables:  $\mathbf{X}_n = o_P(\mathbf{Y}_n)$  if  $\mathbf{X}_n/\mathbf{Y}_n$  tends to 0 in probability. The above theorem says that  $\alpha(\mathbf{G}_n)$  is  $o_P(n)$ . By taking complements (see the identity (19)) it also holds that the size of the largest clique in  $\mathbf{G}_n$  is  $o_P(n)$ . Consequently, the size of the largest homogeneous set is also  $o_P(n)$ , answering negatively the question of Kang, McDiarmid, Reed and Scott [KMRS14, Section 5]:  $P_4$  does not have the asymptotic linear Erdős-Hajnal property.

A different approach to study independent sets of size linear in  $n$  in random cographs is the following. Let  $X_k(G)$  be the number of independent sets of size  $k$  in a graph  $G$ . From Theorem 1.2, if  $\mathbf{G}_n$  is a uniform random (labeled) cograph, then the random variable  $\mathbf{X}_{n,k} := X_k(\mathbf{G}_n)$  tends to 0 in probability if  $k \sim \beta n$  for some  $\beta > 0$  as  $n$  tends to infinity. We show that nonetheless, its expectation grows exponentially fast for  $\beta$  small enough. In particular, this indicates that Theorem 1.2 cannot be proved by a naive use of the first moment method. More precisely, we have the following result.

**Theorem 1.3.** *For each  $n \geq 1$ , let  $\mathbf{G}_n$  be a uniform random labeled cograph of size  $n$ , and let  $\mathbf{X}_{n,k}$  be the number of independent sets of size  $k$  in  $\mathbf{G}_n$ . Then there exist some computable functions  $B_\beta > 0$ ,  $C_\beta > 0$  ( $0 < \beta < 1$ ) with the following property. For every fixed closed interval  $[a, b] \subseteq (0, 1)$ , we have*

$$(1) \quad \mathbb{E}[\mathbf{X}_{n,k}] \sim B_{k/n} n^{-1/2} (C_{k/n})^n$$

uniformly for  $an \leq k \leq bn$ . Furthermore,

- i) When  $\beta \rightarrow 0$ , we have  $C_\beta = 1 + \beta |\log(\beta)| + o(\beta \log(\beta))$ .
- ii) Consequently, there exists  $\beta_0 > 0$  such that  $C_\beta > 1$  for every  $\beta \in (0, \beta_0)$ ; numerically, we can estimate

$$\beta_0 \approx 0.522677 \dots$$

We have found no explicit formula for the growth constant  $C_\beta$  but  $C_\beta$  can be computed numerically with arbitrary precision: see Eq. (45). To get an idea of how fast  $\mathbf{X}_{n,k}$  can grow if  $k \sim \beta n$ , we mention that the function  $\beta \mapsto C_\beta$  seems to have a unique maximum on  $(0, 1)$  (see a plot in Fig. 1); denoting  $\beta^*$  its location, we have the following numerical estimates:

$$\beta^* \approx 0.229285 \dots; \quad C_{\beta^*} \approx 1.366306 \dots$$

As additional motivation for Theorem 1.3, let us mention the work of Drmota, Ramos, Requilé and Rué [DRRR20]: they prove (among other things; see their Corollary 2) the exponential growth of the expected number of maximal independent sets in some subcritical graph classes such as trees, cacti, series-parallel graphs, ... (here “maximal independent sets” refers to independent sets that are maximal for inclusion among all independent sets; such sets are not necessarily of maximum size among all independent sets). It could be interesting to adapt our arguments to consider maximal independent sets in cographs instead of independent sets of fixed size.

*Remark 1.4.* There are two different ways to pick a uniform random cograph with  $n$  vertices: taking it uniformly at random among labeled or among unlabeled cographs. Even if the sizes of independent sets are independent from the labelings, this gives two different probability distributions, since some unlabeled cographs have more symmetries and hence fewer distinct labelings than others.

The reader may have noticed that in Theorem 1.2, we consider either labeled or unlabeled uniform random cographs, while Theorem 1.3 only considers the labeled setting. The reason of this choice is given in Section 1.3 when discussing proof methods.

*Remark 1.5.* A natural question is to determine the order of magnitude of  $\alpha(\mathbf{G}_n)$ . A basic lower bound of order  $\sqrt{n}$  is derived as follows. Since cographs are perfect graphs, for any cograph  $G$ , we have (see [Chu14, Theorem 1.4])

$$(2) \quad \max(\alpha(G), \omega(G)) \geq \sqrt{n}$$

where  $\omega(G)$  is the size of the largest *clique* of  $G$ . By symmetry we have that  $\alpha(\mathbf{G}_n) \stackrel{(d)}{=} \omega(\mathbf{G}_n)$  if  $\mathbf{G}_n$  is a uniform (labeled or unlabeled) cograph. Hence:

$$\begin{aligned} 1 &= \mathbb{P}(\max(\alpha(\mathbf{G}_n), \omega(\mathbf{G}_n)) \geq \sqrt{n}) \\ &\leq \mathbb{P}(\alpha(\mathbf{G}_n) \geq \sqrt{n}) + \mathbb{P}(\omega(\mathbf{G}_n) \geq \sqrt{n}) \\ &= 2\mathbb{P}(\alpha(\mathbf{G}_n) \geq \sqrt{n}), \end{aligned}$$

which means that  $\alpha(\mathbf{G}_n)$  is not  $o_P(\sqrt{n})$ .

We have not been able to improve this bound, but we believe it to be far from optimal. In fact, (limited) numerical simulations, as well as the material in [Wür12, Chapter 9] make us believe that  $\alpha(\mathbf{G}_n)$  is of order  $n/\log(n)$ .

**1.2. Increasing subsequences in random separable permutations.** The asymptotic behavior of the length of the longest increasing subsequence  $\text{LIS}(s_n)$  in a uniform random permutation  $s_n$  of size  $n$  is an old and famous problem that led to surprising and deep connections with various areas of pure mathematics (representation theory, combinatorics, linear algebra and operator theory, random matrices, ...). In particular, it is well-known that  $\text{LIS}(s_n)$  is typically close to  $2\sqrt{n}$  and has Tracy-Widom fluctuations of order  $n^{1/6}$ . We refer to [Rom15] for a nice and modern introduction to this topic.

Longest increasing subsequences in random permutations in permutation classes are a much newer topic: see [MRRY20] and references therein. The methods of the present paper allow the proof of the sublinear behavior of the length of the longest increasing subsequence in a uniform random separable permutation. Let us introduce terminology.

Given a permutation  $\sigma$  of size  $n$  (*i.e.* a sequence  $\sigma(1) \dots \sigma(n)$  containing exactly once each integer from 1 to  $n$ ), and given a subset  $I = \{i_1 < \dots < i_k\}$  of  $\{1, \dots, n\}$ , the *pattern* of  $\sigma$  induced by  $I$  is the permutation  $\pi$  of size  $k$  such that  $\pi(\ell) < \pi(m)$  if and only if  $\sigma(i_\ell) < \sigma(i_m)$ . The study of patterns in permutations is an active research topic, particularly in enumerative combinatorics, see *e.g.* [Vat16, Kit11] and references therein. The relation “is a pattern of” is a partial order on the set of all permutations (of all finite sizes), and *permutation classes* are

downsets for this order. Equivalently, permutation classes can be defined as sets of permutations characterized by the avoidance of a (finite or infinite) set of patterns.

**Definition 1.6.** *A separable permutation is a permutation which avoids the patterns 2413 and 3142.*

Separable permutations enjoy many other characterizations, including the following (the related terminology is defined later in this paper if needed, or *e.g.* in [Vat16]):

- they are the permutations whose inversion graph is a cograph;
- they can be obtained from permutations of size 1 by performing direct sums and skew sums;
- no simple permutation appears in their substitution decomposition.

The class of separable permutations is natural, well-studied, and displays many nice properties; we refer the reader to [BBF+18, end of Section 1.1] for a presentation of these properties and a review of literature. We shall also review some of them in Section 5.

We can now state our analog of Theorem 1.2 for separable permutations.

**Theorem 1.7.** *For each  $n \geq 1$ , let  $\sigma_n$  be a uniform random separable permutation of size  $n$ . Then, the maximal length of an increasing subsequence in  $\sigma_n$  is sublinear in  $n$ , namely  $\frac{\text{LIS}(\sigma_n)}{n}$  converges to 0 in probability.*

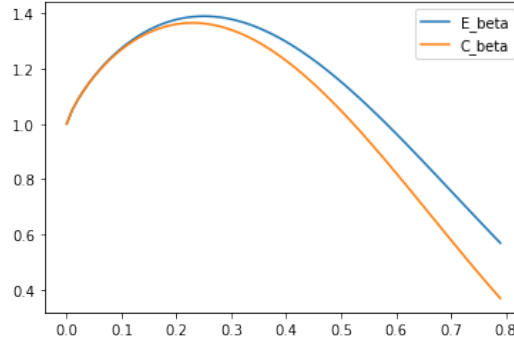
Two remarks about this statement. First, the above sublinearity result does not only apply to separable permutations, but also to any permutation class having a Brownian separable permutation as *permuton* limit – see Section 1.3. Second, as for cographs, we unfortunately did not find a better lower bound for  $\text{LIS}(\sigma_n)$  than the trivial  $\sqrt{n}$  one. The same argument as above applies, where (2) is replaced by Erdős-Szekeres’s Lemma (see *e.g.* [Rom15, Th.1.2]).

We make a further remark about the relation between Theorems 1.2 and 1.7. Recall that for any permutation  $\sigma$  of size  $n$ , its *inversion graph* (denoted  $\text{inv}(\sigma)$ ) is the unlabeled version of the graph with vertex set  $\{1, \dots, n\}$  where there is an edge between  $i$  and  $j$  if and only if  $i$  and  $j$  form an inversion in  $\sigma$ , that is  $(i - j)(\sigma(i) - \sigma(j)) < 0$ . Clearly, through this correspondence, an increasing sequence in  $\sigma$  is mapped to an independent set in  $\text{inv}(\sigma)$ . Nevertheless, Theorem 1.7 is not simply the translation of Theorem 1.2 from the graph setting to the permutation setting. Indeed, since the inversion graph correspondence is not one-to-one, for  $\sigma_n$  a uniform random separable permutation,  $\text{inv}(\sigma_n)$  is not a uniform random unlabeled cograph. (We further note that defining  $\text{inv}(\sigma)$  as a labeled cograph in the obvious manner,  $\text{inv}(\sigma_n)$  would also not be a uniform random labeled cograph.)

We also establish a counterpart of Theorem 1.3 for increasing subsequences in separable permutations.

**Theorem 1.8.** *For each  $n \geq 1$ , let  $\sigma_n$  be a uniform random separable permutation of size  $n$ , and let  $Z_{n,k}$  be the number of increasing subsequences of length  $k$  in  $\sigma_n$ . Then there exist some computable functions  $D_\beta > 0$ ,  $E_\beta > 0$  ( $0 < \beta < 1$ ) with the following property. For every fixed closed interval  $[a, b] \subseteq (0, 1)$ , we have*

$$(3) \quad \mathbb{E}[Z_{n,k}] \sim D_{k/n} n^{-1/2} (E_{k/n})^n$$

FIGURE 1. Plots of  $\beta \mapsto C_\beta$  and  $\beta \mapsto E_\beta$ .

uniformly for  $an \leq k \leq bn$ . Furthermore,

- i) When  $\beta \rightarrow 0$ , we have  $E_\beta = 1 + \beta |\log(\beta)| + o(\beta \log(\beta))$ .
- ii) Consequently, there exists  $\beta_1 > 0$  such that  $E_\beta > 1$  for every  $\beta \in (0, \beta_1)$ ; numerically, we can estimate

$$\beta_1 \approx 0.5827 \dots$$

We observe the same qualitative behavior than for (1):  $E_\beta$  seems to have a unique maximum (numerically estimated at  $\beta \approx 0.2503 \dots$ ). Moreover, it seems from numerical computations that  $E_\beta > C_\beta$  for every  $\beta \in (0, 1)$  (see Fig. 1).

**1.3. Proof methods and universality.** Our sublinearity results are based on limit theorems for uniform random cographs and uniform random separable permutations. We first discuss the graph setting.

It is proved in [BBF+22b, Stu21] that a uniform random (labeled or unlabeled) cograph of size  $n$  converges in the sense of graphons to a limit  $\mathbf{W}^{1/2}$ , called the *Brownian cographon* of parameter  $1/2$  (see also the independent work of Stuffer [Stu21]). We refer to Section 2.2 for details. Moreover, the notion of independence number of a graph has been extended to graphons by Hladký and Rocha [HR20], who proved a semicontinuity property for it (see Section 2.3). Combining these two elements, Theorem 1.2 will follow from the fact that the independence number  $\tilde{\alpha}(\mathbf{W}^{1/2})$  of the Brownian cographon is 0 a.s (see Section 3.3). To prove the latter, we use the explicit construction of the Brownian cographon from a Brownian excursion and some self-similarity property of the Brownian excursion (namely Aldous' decomposition of a Brownian excursion with two independent points into three independent Brownian excursions of random sizes [Ald94]); we deduce from that an *inequation* in distribution for  $\tilde{\alpha}(\mathbf{W}^{1/2})$  (Section 3.1; the use of *inequation* instead of *inequality* is justified there) and we conclude by a fixed point argument (Section 3.2).

An interesting aspect of the proof sketched above is that it relies solely on the fact that uniform random cographs tend to the Brownian cographon; moreover the value  $p = 1/2$  of the parameter of the limit is irrelevant in the proof. Convergence to the Brownian cographon was proved in [BBF+22b, Stu21] both in the labeled and unlabeled settings, so that Theorem 1.2 is proved

simultaneously in both settings. In fact, Theorem 1.2 is proved as a special case of the following theorem.

**Theorem 1.9.** *Let  $\mathbf{G}_n$  be a sequence of random graphs tending to the Brownian cogrAPHON  $\mathbf{W}^p$  for  $p \in [0, 1)$ . Then the maximum size of an independent set in  $\mathbf{G}_n$  is sublinear in  $n$ , namely  $\frac{\alpha(\mathbf{G}_n)}{n}$  converges to 0 in probability.*

By analogy with the realm of permutations (see below), we expect that uniform random graphs in families of graphs well-behaved for the modular decomposition (e.g., a graph class whose modular decomposition trees are all those obtained from a finite set of prime graphs) tend to  $\mathbf{W}^p$ , and hence have a sublinear independence number.

Let us now discuss Theorem 1.7, i.e. the sublinearity of the length of the longest increasing subsequence in a random separable permutation  $\sigma_n$ . It is known that  $\sigma_n$  tends in the permutation topology to a limit  $\mu^{1/2}$ , called the *Brownian separable permuton* of parameter 1/2, see [BBF+18] for the original reference.

As discussed earlier, an increasing subsequence of a permutation corresponds to an independent set of its inversion graph. We remark in Section 2.4 that if a sequence of permutations converges in distribution to the Brownian separable permuton of parameter  $p$ , then the corresponding inversion graphs converge in distribution to the Brownian cogrAPHON  $\mathbf{W}^p$ .

Hence Theorem 1.9 implies the following general result, of which Theorem 1.7 is a particular case (see Section 3.3 for details).

**Theorem 1.10.** *Let  $\sigma_n$  be a sequence of random permutations tending to the Brownian separable permuton  $\mu^p$  for  $p \in [0, 1)$ . Then the maximal length of an increasing subsequence in  $\sigma_n$  is sublinear in  $n$ , namely  $\frac{\text{LIS}(\sigma_n)}{n}$  converges to 0 in probability.*

We note that the Brownian separable permuton  $\mu^p$  has been proved to be a *universal limit* for uniform random permutations in many permutation classes (well-behaved with respect to the substitution decomposition) [BBF+20, BBF+22a, BBFS19], so Theorem 1.10 applies to all these classes.

The technique to prove Theorems 1.3 and 1.8 is completely different. Indeed, the expectation of  $\mathbf{X}_{n,k}$  (resp.  $\mathbf{Z}_{n,k}$ ) for  $k \sim \beta n$  for some  $\beta > 0$  is driven by a set of cogrAPHONS (resp. separable permutations) of small probability and can therefore not be inferred from their limit in distribution. In this case, we use the representation of cogrAPHONS as cotrees, and its analogue for separable permutations through substitution decomposition trees. These tree representations are useful tools in algorithms both for graphs and permutations (see e.g. [HP05, BCH+08] for graphs and [BBL98] for permutations); in the case of permutations, substitution decomposition trees have also been widely used in recent years for enumeration problems (see [Vat16, Section 3.2] and references therein). The tree encoding allows us to write a system of equations for the bivariate generating function of cogrAPHONS with a marked independent set (resp. separable permutations with a marked increasing subsequence). We then obtain our results through singularity analysis.

Unlike for Theorems 1.2 and 1.7, the results we prove are specific to either labeled cographs or separable permutations and do not rely on their Brownian limits. However, our approach should extend to other families of graphs and permutations well-encoded by their (modular or substitution) decomposition trees, but we did not pursue this direction. One such model would be unlabeled cographs: in this model, the analytic equations involve the so-called Pólya operators, making the analysis more technical but we do not expect qualitative differences in the result.

**1.4. Organization of the paper.** The proofs of our two sets of results can be read independently.

- Section 2 provides the necessary background regarding graphons and permutons. Then we prove Theorems 1.9 and 1.10 in Section 3.
- The proofs of Theorems 1.3 and 1.8 are given in Section 4 and Section 5, respectively.

## 2. PRELIMINARIES: GRAPHONS, PERMUTONS, INDEPENDENCE NUMBER AND INCREASING SUBSEQUENCES

We first recall some general material from the theory of graphons (Section 2.1). We present here the strict minimum needed for this paper; an extensive presentation can be found in [Lov12, Chapters 7-16]. Then in Sections 2.2 to 2.4 we review recent material from the literature, used for our proof of Theorems 1.9 and 1.10:

- the convergence of uniform random cographs to the Brownian cographon;
- the notion of independence number of graphons;
- a connection between graphons and the analogue theory for permutations, that of *permutons*.

**2.1. Basics on graphons.** A graphon (contraction for *graph function*) is a measurable symmetric function from  $[0, 1]^2$  to  $[0, 1]$ . Intuitively, we can think of it as the adjacency matrix of an infinite (weighted) graph with vertex set  $[0, 1]$ . A finite graph  $G$  with vertex set  $\{1, \dots, n\}$  can be seen as a graphon  $W_G$  as follows:  $W_G(x, y) = 1$  if the vertices with labels  $\lceil xn \rceil$  and  $\lceil yn \rceil$  are connected in  $G$  ( $\lceil z \rceil$  being the nearest integer above  $z$ , with the unusual convention  $\lceil 0 \rceil = 1$ ) and  $W_G(x, y) = 0$  otherwise.

*Sampling.* Let  $W$  be a graphon and  $k$  a positive extended integer (i.e.  $k \in \mathbb{Z}_{>0} \cup \{+\infty\}$ ). We consider two independent families  $(U_i)_{1 \leq i \leq k}$  and  $(X_{i,j})_{1 \leq i < j \leq k}$  of i.i.d. uniform random variables in  $[0, 1]$ . Given this, we define a random graph  $\mathbf{Sample}_k(W)$  as follows<sup>2</sup>: its vertex set is  $[k] := \{1, \dots, k\}$  and for every  $i, j$ , vertices  $i$  and  $j$  are connected iff  $X_{i,j} \leq W(U_i, U_j)$ . In other words vertices  $i$  and  $j$  are connected with probability  $W(U_i, U_j)$ , independently of each other conditionally on the sequence  $(U_i)_{1 \leq i \leq k}$ .

We note that, for  $k' > k$ , the restriction  $\mathbf{Sample}_{k'}(W)[k]$  of  $\mathbf{Sample}_{k'}(W)$  to the vertex set  $[k]$  has the same distribution as  $\mathbf{Sample}_k(W)$ . In particular, the random graph  $\mathbf{Sample}_\infty(W)$  induces a realization of all  $\mathbf{Sample}_k(W)$  in the same probability space.

*Convergence.* By definition, a sequence of graphons  $(W_n)$  converges to a graphon  $W$  if, for all  $k$ ,  $\mathbf{Sample}_k(W_n)$  converges in distribution to  $\mathbf{Sample}_k(W)$ . It can be shown that this is equivalent to the convergence for the so-called cut-distance; see [Lov12, Theorem 11.5]. We

<sup>2</sup>In [Lov12],  $\mathbf{Sample}_k(W)$  is denoted  $\mathbb{G}(k, W)$ .



note that the graphon limit is unique only up to some equivalence relation, called *weak equivalence* [Lov12, Sections 7.3, 10.7, 13.2]. Moreover, the quotient of the set of graphons by the weak equivalence relation, equipped with the cut-distance metric, is a compact metric space, that we shall call from now on *the space of graphons*. Finally, we say that a sequence of graphs  $(G_n)_{n \geq 1}$  converges to a graphon  $W$  if the associated graphons  $(W_{G_n})$  converge to  $W$  in the space of graphons, and that a sequence of random graphs  $(G_n)_{n \geq 1}$  converges in distribution to a random graphon  $W$ , if  $W_{G_n}$  converges to  $W$  in distribution, as random elements of the space of graphons.

**2.2. Convergence to the Brownian cogrAPHON.** Let  $\epsilon : [0, 1] \rightarrow \mathbb{R}$  denote a Brownian excursion of length one. We recall that, a.s.,  $\epsilon$  has a countable set of local minima, which are all strict and have distinct values<sup>3</sup>. Let us denote  $\{b_i(\epsilon), i \geq 1\}$  an enumeration of the positions of these local minima. It is possible to choose this enumeration in such a way that the  $b_i$ 's and the subsequent functions defined in this section are measurable; we refer to [Maa20, Lemma 2.3] and [BBF+22b, Section 4] for details.

We now choose i.i.d. Bernoulli variables  $s_i$  with  $\mathbb{P}(s_i = 0) = p$ , independent from the foregoing, and write  $\mathbf{S}^p = (s_i)_{i \geq 1}$ . We call  $(\epsilon, \mathbf{S}^p)$  a *decorated Brownian excursion*, thinking of the variable  $s_i$  as a decoration attached to the local minimum  $b_i(\epsilon)$ .

For  $x, y \in [0, 1]$ , we define  $\text{Dec}(x, y; \epsilon, \mathbf{S}^p)$  to be the decoration of the minimum of  $\epsilon$  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,  $\text{Dec}(x, y; \epsilon, \mathbf{S}^p)$  is ill-defined and we take the convention  $\text{Dec}(x, y; \epsilon, \mathbf{S}^p) = 0$ . Note however that, for uniform random  $x$  and  $y$ , this happens with probability 0, so that the object constructed in Definition 2.1 below is independent from this convention.

**Definition 2.1.** *The Brownian cogrAPHON  $W^p$  of parameter  $p$  is the random function*

$$\begin{aligned} W^p : [0, 1]^2 &\rightarrow \{0, 1\}; \\ (x, y) &\mapsto \text{Dec}(x, y; \epsilon, \mathbf{S}^p). \end{aligned}$$

For example, in Fig.2 if the decoration at  $b$  is 0 (resp. 1), then the graphon is constant equal to 0 (resp. 1) on the rectangle  $(a, b) \times (b, c)$ .

The following was proved independently in [Stu21] and [BBF+22b].

**Theorem 2.2.** *Uniform random cogrAPHONS (either labeled or unlabeled) converge in distribution to the Brownian cogrAPHON of parameter  $1/2$ , in the space of graphons.*

**2.3. Independence number of a graphon and semi-continuity.** Let  $W$  be a (deterministic) graphon. Following Hladký and Rocha [HR20], we define an independent set  $I$  of a graphon  $W$  as a measurable subset of  $[0, 1]$  such that  $W(x, y) = 0$  for almost every  $(x, y)$  in  $I \times I$ . The *independence number*  $\tilde{\alpha}(W)$  of a graphon  $W$  is then

$$(4) \quad \tilde{\alpha}(W) = \sup_{\substack{I \subset [0, 1] \\ I \text{ independent set of } W}} \text{Leb}(I),$$

<sup>3</sup>That  $\epsilon$  has a.s. only strict local minima with distinct values is folklore – the interested reader may find a proof in [BBF+18, Appendix A]. This implies readily that the set of local minima is a.s. countable.

where  $\text{Leb}(I)$  denotes the Lebesgue measure of  $I$ . Note that  $\tilde{\alpha}(W)$  is attained by some independent set  $I$  (that is, the supremum in Eq. (4) is in fact a maximum): this follows from *e.g.* [HHP19, Lemma 2.4].

Clearly, for a graph  $G$  we have

$$(5) \quad \tilde{\alpha}(W_G) = \alpha(G)/|V_G|,$$

where  $\alpha(G)$  is the maximum size of an independent set in  $G$ .

Of crucial interest for this paper is the lower semi-continuity of the function  $\tilde{\alpha}$  on the space of graphons [HR20, Corollary 7]. Concretely, this says the following.

**Proposition 2.3.** *Suppose that  $(W_n)_{n \geq 1}$  is a sequence of graphons that converges to some  $W$  in the space of graphons. Then  $\limsup \tilde{\alpha}(W_n) \leq \tilde{\alpha}(W)$ .*

*Remark 2.4.* In the following, we will consider a random variable of the kind  $\tilde{\alpha}(W)$  where  $W$  is a random graphon. For this to make sense, the map  $\tilde{\alpha}$  should be measurable. Since it is defined as a supremum over an uncountable set, its measurability is not a priori clear. However, it is known that any semi-continuous function is measurable, so that Proposition 2.3 implies that  $\tilde{\alpha}$  is indeed measurable. We shall not discuss this point further in the paper.

In the rest of this subsection, we give an alternative definition for  $\tilde{\alpha}(W)$ . This definition is not needed in the rest of the paper (and therefore can be safely skipped); however, it answers a question raised by Hladký and Rocha [HR20, Section 3.2], who asked for a connection between the statistics  $\tilde{\alpha}(W)$  and subgraph densities (or equivalently, samples) of  $W$ .

For a graphon  $W$ , we set

$$(6) \quad \tilde{\alpha}_2(W) = \liminf_{k \rightarrow \infty} \frac{1}{k} \alpha(\mathbf{Sample}_\infty(W)[k]).$$

Since  $\mathbf{Sample}_\infty(W)$  is a random graph, the right-hand side is a priori a random variable. We recall that  $\mathbf{Sample}_\infty(W)$  is constructed from i.i.d. random variables  $\{U_i, X_{i,j}, 1 \leq i < j\}$ . We denote  $\mathcal{G}_n$  the  $\sigma$ -algebra generated by  $\{U_i, X_{i,j}, n < i < j\}$ . It is a simple exercise to see that  $\tilde{\alpha}_2(W)$  is measurable with respect to the *tail*  $\sigma$ -algebra  $\bigcap_{n \geq 1} \mathcal{G}_n$ . By Kolmogorov's 0 – 1 law (easily adapted to our situation with bi-indexed i.i.d. random variables),  $\tilde{\alpha}_2(W)$  is almost surely equal to a constant.

**Lemma 2.5.** *For any graphon  $W$ , we have  $\tilde{\alpha}_2(W) = \tilde{\alpha}(W)$  almost surely, and the  $\liminf$  defining  $\tilde{\alpha}_2(W)$  is almost surely an actual limit.*

*Proof.* We first prove  $\tilde{\alpha}_2(W) \geq \tilde{\alpha}(W)$  almost surely. Let  $I$  be an independent set of  $W$ . For any  $k \geq 1$ , we observe that the set  $\mathbf{J}_k := \{j \leq k : U_j \in I\}$  is a.s. an independent set of  $\mathbf{Sample}_\infty(W)[k]$ .

Hence, a.s.

$$\frac{1}{k} \alpha(\mathbf{Sample}_\infty(W)[k]) \geq \frac{1}{k} |\mathbf{J}_k|.$$

As  $k$  tends to infinity, the law of large numbers asserts that  $|\mathbf{J}_k|/k$  tends a.s. to  $\text{Leb}(I)$ . Therefore we have a.s.  $\tilde{\alpha}_2(W) \geq \text{Leb}(I)$ . Since this holds for any independent set  $I$  of  $W$ , we can consider the independent set  $I$  that realizes the maximum in Eq. (4), proving  $\tilde{\alpha}_2(W) \geq \tilde{\alpha}(W)$  a.s..

Let us prove the converse inequality. It is known that  $(\mathbf{Sample}_\infty(W)[k])$  converges a.s. to  $W$  in the space of graphons (e.g. as a consequence of [Lov12, Lemma 10.16]). Using (5) and Proposition 2.3 this implies that, a.s.,

$$\widetilde{\alpha}_2(W) \leq \limsup_{k \rightarrow \infty} \frac{1}{k} \alpha(\mathbf{Sample}_\infty(W)[k]) = \limsup_{k \rightarrow \infty} \widetilde{\alpha}(\mathbf{Sample}_\infty(W)[k]) \leq \widetilde{\alpha}(W).$$

This concludes the proof that almost surely  $\widetilde{\alpha}_2(W) = \widetilde{\alpha}(W)$ . Moreover in the identity  $\widetilde{\alpha}_2(W) = \liminf_{k \rightarrow \infty} \frac{1}{k} \alpha(\mathbf{Sample}_\infty(W)[k])$  the lim inf is an actual limit.  $\square$

**2.4. The Brownian separable permuton and its relation to the Brownian cogrAPHON.** The theory of permutons (see [GGKK15, HKMRS13]) plays the same role for limits of permutations as the theory of graphons does for dense graphs. A permuton is a probability measure on the unit square with uniform marginals, and the space of permutons equipped with the weak convergence of measures is a compact metric space. We attach to each permutation  $\sigma$  of size  $n \geq 1$  the measure  $\mu_\sigma$  on the unit square with density  $(x, y) \mapsto n \mathbb{1}_{\sigma(\lceil nx \rceil) = \lceil ny \rceil}$ , which is a permuton. This defines a dense embedding of the set of permutations into the space of permutons.

Recall that  $\text{inv}(\sigma)$  denotes the (unlabeled) inversion graph of a permutation  $\sigma$ .

**Proposition 2.6.** *Let  $p \in [0, 1]$  and  $(\sigma_n)_n$  be a sequence of random permutations such that  $\mu_{\sigma_n} \xrightarrow[n \rightarrow \infty]{d} \mu^p$ , where  $\mu^p$  is the Brownian separable permuton of parameter  $p$  defined in [BBF+22a,*

*Definition 3.5]. Let  $\mathbf{G}_n = \text{inv}(\sigma_n)$ . Then we have the convergence in distribution  $W_{\mathbf{G}_n} \xrightarrow[n \rightarrow \infty]{d} W^p$  in the space of graphons, where  $W^p$  is the Brownian cogrAPHON of parameter  $p$ .*

*Remark 2.7.* It was observed in [GGKK15, End of Section 2] that  $\text{inv}$  possesses an extension which is a continuous map  $\widetilde{\text{inv}}$  from the space of permutons to the space of graphons. The above proposition implies that the image of  $\mu^p$  by  $\widetilde{\text{inv}}$  is  $W^p$ .

*Proof.* For every  $k \geq 1$ , denote  $\mathbf{b}_{k,p}$  a uniform random plane binary tree with  $k$  (unlabeled) leaves, whose internal vertices are decorated with independent signs  $\{\oplus, \ominus\}$  such that  $\mathbb{P}(\oplus) = p$ . Before entering the actual proof, we present a useful link between a separable permutation and an unlabeled cogrAPHON constructed from  $\mathbf{b}_{k,p}$ .

Following [BBF+22a, Definition 2.3], we may associate with  $\mathbf{b}_{k,p}$  a separable permutation, denoted  $\text{perm}(\mathbf{b}_{k,p})$ . We do not recall this construction here (for details, see the above reference or the beginning of Section 5), but indicate an important property it enjoys: for  $1 \leq i < j \leq k$ , we have  $\text{perm}(\mathbf{b}_{k,p})(i) > \text{perm}(\mathbf{b}_{k,p})(j)$  if and only if the youngest common ancestor of the  $i$ -th and  $j$ -th leaves (in the left-to-right order) of  $\mathbf{b}_{k,p}$  carries a  $\ominus$  sign.

Similarly, we may also associate with  $\mathbf{b}_{k,p}$  an unlabeled cogrAPHON. We first replace  $\ominus$  by 1 and  $\oplus$  by 0 in all internal nodes and then we forget the plane embedding. We denote by  $\tilde{\mathbf{b}}_{k,p}$  the resulting non-plane and unlabeled decorated tree. With this tree, we associate an unlabeled cogrAPHON  $\text{CogrAPHON}(\tilde{\mathbf{b}}_{k,p})$  as follows: its vertices correspond to the leaves of  $\tilde{\mathbf{b}}_{k,p}$ , and there is an edge between the vertices corresponding to leaves  $\ell$  and  $\ell'$  if and only if the youngest common ancestor of  $\ell$  and  $\ell'$  carries the decoration 1. An alternative recursive presentation of this construction, making it clear that the constructed graph is indeed a cogrAPHON, is given at the beginning of Section 4.

By construction, the equality  $\text{inv}(\text{perm}(\mathbf{b}_{k,p})) = \text{Cograph}(\tilde{\mathbf{b}}_{k,p})$  of unlabeled graphs holds.

Denote  $\sigma_{n,k}$  a uniform random pattern of size  $k$  in  $\sigma_n$ . Theorem 3.1 and Definition 3.5 in [BBF+22a] imply that  $\sigma_{n,k}$  converges in distribution to the random separable permutation  $\text{perm}(\mathbf{b}_{k,p})$ . As this is a convergence in distribution in the discrete space consisting of all permutations of size  $k$ , the map  $\text{inv}$  is continuous, and we obtain the following convergence of unlabeled graphs:

$$(7) \quad \text{inv}(\sigma_{n,k}) \xrightarrow[n \rightarrow \infty]{d} \text{inv}(\text{perm}(\mathbf{b}_{k,p})).$$

It is easy to check that the actions of taking patterns (resp. induced subgraphs) and of computing inversion graphs commute. Namely, for a permutation  $\sigma$  and a subset  $I$  of its indices, the inversion graph of the pattern of  $\sigma$  induced by  $I$  is the subgraph of the inversion graph of  $\sigma$  induced by the vertices corresponding to  $I$ . Therefore,  $\text{inv}(\sigma_{n,k})$  – which appears on the left-hand-side of Eq. (7) – has the same distribution as the subgraph induced by a uniform random subset of  $k$  distinct vertices of  $\mathbf{G}_n$ .

On the right-hand side of Eq. (7), we have already identified  $\text{inv}(\text{perm}(\mathbf{b}_{k,p}))$  as  $\text{Cograph}(\tilde{\mathbf{b}}_{k,p})$ . We recall that  $\tilde{\mathbf{b}}_{k,p}$  is the non-plane version of a uniform random (unlabeled) plane binary tree with independent decorations on its internal nodes. We claim that this has the same distribution as the unlabeled version of a uniform random labeled non-plane binary tree, with the same rule for decorations of the internal nodes (which we denote  $\mathbf{b}_{k,p}^{P,L}$ ). Admitting this claim for the moment, and comparing with [BBF+22b, Proposition 4.3], we get that the right-hand side of Eq. (7) is distributed as  $\text{Sample}_k(\mathbf{W}^p)$ .

With these considerations in hand, we can use [BBF+22b, Theorem 3.8] (more precisely the implication  $(d) \Rightarrow (a)$  and Eq. (4) following this theorem) and conclude from Eq. (7) that  $W_{\mathbf{G}_n} \xrightarrow[n \rightarrow \infty]{d} \mathbf{W}^p$ . This ends the proof of the proposition, up to the above claim.

It remains to prove that  $\tilde{\mathbf{b}}_{k,p} \stackrel{d}{=} \mathbf{b}_{k,p}^{P,L}$ , as non-plane unlabeled trees. Since the rule for the random decorations are the same on both sides, we disregard decorations, and denote the underlying undecorated random trees  $\tilde{\mathbf{b}}_k$  and  $\mathbf{b}_k^{P,L}$  respectively. To prove that  $\tilde{\mathbf{b}}_k \stackrel{d}{=} \mathbf{b}_k^{P,L}$ , we compare both distributions with that of a uniform labeled plane binary tree with  $k$  leaves  $\mathbf{b}_k^{P,L}$ . Since every non-plane labeled binary tree with  $k$  leaves can be embedded in the plane in  $2^{k-1}$  ways, we have  $\mathbf{b}_k^{P,L} \stackrel{d}{=} \mathbf{b}_k^{P,L}$  as non-plane unlabeled trees (there are no symmetry problems, since trees are labeled). On the other hand, since every plane unlabeled binary tree with  $k$  leaves can be labeled in  $k!$  ways, we have  $\tilde{\mathbf{b}}_k \stackrel{d}{=} \mathbf{b}_k^{P,L}$  as non-plane unlabeled trees (again, there are no symmetry problems, since trees are plane). We conclude that  $\tilde{\mathbf{b}}_{k,p} \stackrel{d}{=} \mathbf{b}_{k,p}^{P,L}$ , as wanted.  $\square$

### 3. PROOF OF THE SUBLINEARITY RESULTS THROUGH SELF-SIMILARITY

The main part of the proof of our sublinearity results (Theorems 1.9 and 1.10) is done in the continuous world, proving that the independence number  $\tilde{\alpha}(\mathbf{W}^p)$  of the Brownian cographon is almost surely equal to 0. To this end, we first show that the distribution of  $\tilde{\alpha}(\mathbf{W}^p)$  is solution of a

specific inequation – this is Proposition 3.1. Next, we prove that the only solution of this inequation is the Dirac distribution  $\delta_0$  – this is Proposition 3.2. All results are gathered in Section 3.3, completing the proofs of Theorems 1.9 and 1.10.

**3.1. An inequation in distribution.** We use the standard *stochastic domination order* between real distributions  $\mu$  and  $\nu$ . Namely, we write  $\mu \leq_d \nu$  if  $\mu([x, +\infty)) \leq \nu([x, +\infty))$  for every real  $x$ . By Strassen’s Theorem, this is equivalent to the fact that we can find  $\mathbf{Z}_1$  and  $\mathbf{Z}_2$  defined on the same probability space with distributions  $\mu$  and  $\nu$  respectively, such that  $\mathbf{Z}_1 \leq \mathbf{Z}_2$  almost surely.

Our goal is now to show that the distribution of the random variable  $\tilde{\alpha}(\mathbf{W}^p)$  is stochastically dominated by another distribution defined using some independent copies of  $\tilde{\alpha}(\mathbf{W}^p)$  (we refer to such a relation as an *inequation in distribution*<sup>4</sup>). To this end, we use Aldous’ decomposition of a Brownian excursion into three independent excursions (see Fig. 2). This decomposition has an immediate counterpart, where we decompose a Brownian cogrAPHON into three independent Brownian cogrAPHONS. We then look closely at the behavior of the functional  $\tilde{\alpha}$  along this decomposition.

We introduce the notation needed to state this inequation in distribution. For a random variable  $\mathbf{Y}$ , let us denote by  $\text{Law}(\mathbf{Y})$  its distribution. Recall that, for positive real numbers  $\alpha_1, \dots, \alpha_k$ , the Dirichlet distribution  $\text{Dirichlet}(\alpha_1, \dots, \alpha_k)$  is a probability measure on the simplex  $\{(x_1, \dots, x_k) : x_1 + \dots + x_k = 1, x_i \geq 0 \text{ for all } i\}$ : by definition it has density proportional to  $\prod_{i \leq k} x_i^{\alpha_i}$  with respect to the Lebesgue measure.

Let  $\mu$  be a probability distribution on  $[0, 1]$  and  $p$  a parameter in  $[0, 1]$ . We define the following random variables:

- $(\Delta_0, \Delta_1, \Delta_2)$  is a random vector in  $[0, 1]^3$  with distribution  $\text{Dirichlet}(1/2, 1/2, 1/2)$ ;
- $\mathbf{X}_0^\mu, \mathbf{X}_1^\mu$  and  $\mathbf{X}_2^\mu$  are three independent random variables with distribution  $\mu$ , and independent from  $(\Delta_0, \Delta_1, \Delta_2)$ ;
- $\mathbf{B}$  is a Bernoulli( $1-p$ ) random variable, independent from  $(\Delta_0, \Delta_1, \Delta_2, \mathbf{X}_0^\mu, \mathbf{X}_1^\mu, \mathbf{X}_2^\mu)$ ;
- finally, we set

$$(8) \quad \mathbf{Y}_0^\mu = \Delta_0 \mathbf{X}_0^\mu + \Delta_1 \mathbf{X}_1^\mu + \Delta_2 \mathbf{X}_2^\mu$$

$$(9) \quad \mathbf{Y}_1^\mu = \Delta_0 \mathbf{X}_0^\mu + \max(\Delta_1 \mathbf{X}_1^\mu, \Delta_2 \mathbf{X}_2^\mu)$$

$$(10) \quad \mathbf{Y}_{(p)}^\mu = \mathbf{B} \mathbf{Y}_1^\mu + (1 - \mathbf{B}) \mathbf{Y}_0^\mu$$

$$\Phi_p(\mu) = \text{Law}(\mathbf{Y}_{(p)}^\mu)$$

Then the inequation we are interested in is

$$(11) \quad \mu \leq_d \Phi_p(\mu).$$

**Proposition 3.1.** *For  $p \in [0, 1]$ , the distribution  $\mu$  of  $\tilde{\alpha}(\mathbf{W}^p)$  satisfies the inequation (11).*

*Proof.* Fix  $p \in [0, 1]$  and let  $(\epsilon, \mathbf{S}^p)$  be a decorated Brownian excursion. Let  $\mathbf{U}_1 < \mathbf{U}_2$  be a reordered pair of independent and uniform random variables on  $[0, 1]$ , chosen independently from  $(\epsilon, \mathbf{S}^p)$ . Almost surely, the function  $\epsilon$  reaches its minimum on  $[\mathbf{U}_1, \mathbf{U}_2]$  exactly once, and at

<sup>4</sup>We use the term *inequation* and not *inequality*, because the upper bound also involves the distribution of  $\tilde{\alpha}(\mathbf{W}^p)$ .

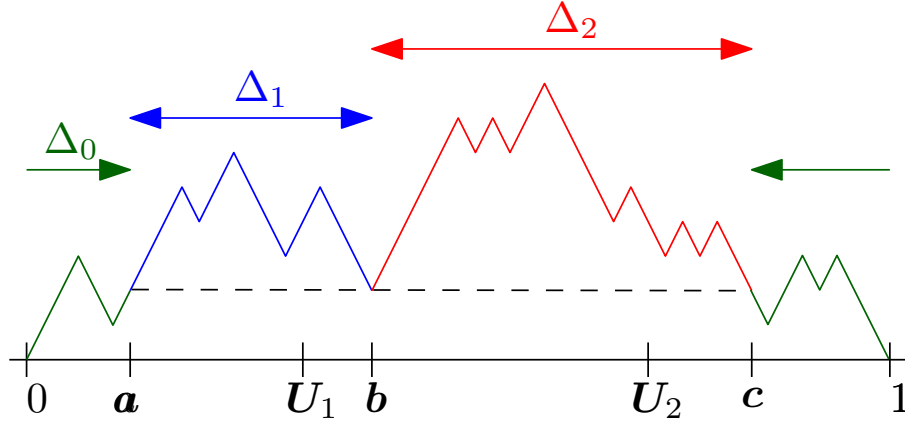


FIGURE 2. A stylized version of a Brownian excursion and the corresponding  $\mathbf{a}, \mathbf{b}, \mathbf{c}, \Delta_0, \Delta_1, \Delta_2$ .

a local minimum. Let us denote by  $s$  the sign of this local minimum in  $S^p$ . Let  $\mathbf{b}$  be the position where this local minimum is reached (see Fig. 2). Let also

$$\mathbf{a} = \max\{t \leq U_1, \epsilon(t) = \epsilon(\mathbf{b})\}, \quad \mathbf{c} = \min\{t \geq U_2, \epsilon(t) = \epsilon(\mathbf{b})\}.$$

Set

$$(12) \quad \Delta_0 = 1 - \mathbf{c} + \mathbf{a}, \quad \Delta_1 = \mathbf{b} - \mathbf{a}, \quad \Delta_2 = \mathbf{c} - \mathbf{b}, \quad X_0 = \frac{\mathbf{a}}{\Delta_0}, \quad B = s.$$

We may now cut the excursion  $\epsilon$  into three excursions, in the manner prescribed by Aldous [Ald94].

$$(13) \quad \begin{aligned} \eta_0(x) &= \Delta_0 x + (1 - \Delta_0)\mathbf{1}_{[x > X_0]}, & x \in [0, 1], & \quad \epsilon_0 = \frac{1}{\sqrt{\Delta_0}} \epsilon \circ \eta_0 \\ \eta_1(x) &= \mathbf{a} + \Delta_1 x, & x \in [0, 1], & \quad \epsilon_1 = \frac{1}{\sqrt{\Delta_1}} (\epsilon \circ \eta_1 - \epsilon(\mathbf{b})) \\ \eta_2(x) &= \mathbf{b} + \Delta_2 x, & x \in [0, 1], & \quad \epsilon_2 = \frac{1}{\sqrt{\Delta_2}} (\epsilon \circ \eta_2 - \epsilon(\mathbf{b})) \end{aligned}$$

Then [Ald94, Corollary 3] states that the random functions  $\epsilon_0, \epsilon_1, \epsilon_2$  are three independent Brownian excursions, independent from the vector  $(\Delta_0, \Delta_1, \Delta_2)$ ; moreover, the latter has distribution Dirichlet(1/2, 1/2, 1/2).

In addition the piecewise affine maps  $(\eta_i)_{0 \leq i \leq 2}$  naturally put the local minima of  $\epsilon$  (except the one at  $t = \mathbf{b}$ ) in bijection with the disjoint union of the local minima of  $\epsilon_0, \epsilon_1$  and  $\epsilon_2$ . (Indeed, almost surely,  $\epsilon$  does not have a local minimum at  $t = \mathbf{a}$  nor at  $t = \mathbf{c}$ .) In particular, this implies (as shown in the proof of Theorem 1.6 in [Maa20], see in particular Observations 5.2 and 5.3 there) that

- $B$  is a Bernoulli( $1 - p$ ) random variable,

- there exist three independent i.i.d. sequences of Bernoulli(1 - p) random variables  $\mathbf{S}_0^p, \mathbf{S}_1^p, \mathbf{S}_2^p$  such that for  $0 \leq x < y \leq 1$  and  $k \in \{0, 1, 2\}$ ,

$$(14) \quad \text{Dec}(x, y; \mathbf{e}_k, \mathbf{S}_k^p) = \text{Dec}(\eta_k(x), \eta_k(y); \mathbf{e}, \mathbf{S}^p),$$

- the random variables  $\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, (\Delta_0, \Delta_1, \Delta_2), \mathbf{B}$  and the sequences  $\mathbf{S}_0^p, \mathbf{S}_1^p, \mathbf{S}_2^p$  are all independent.

Now, let  $\mathbf{W}^p$  be the Brownian cographon (of parameter  $p$ ) associated with  $(\mathbf{e}, \mathbf{S}^p)$  (see Definition 2.1). Similarly, for  $k \in \{0, 1, 2\}$ , consider  $\mathbf{W}_k^p$  the Brownian cographon (also of parameter  $p$ ) constructed from  $(\mathbf{e}_k, \mathbf{S}_k^p)$ , i.e.

$$\begin{aligned} \mathbf{W}_k^p : [0, 1]^2 &\rightarrow \{0, 1\} \\ (x, y) &\mapsto \text{Dec}(x, y; \mathbf{e}_k, \mathbf{S}_k^p). \end{aligned}$$

These three random graphons form a triple of i.i.d. random graphons, independent from the random variables  $\mathbf{B}$  and  $(\Delta_0, \Delta_1, \Delta_2)$ .

Let  $\mathbf{I}$  be an independent set of  $\mathbf{W}^p$ . For each  $k = 0, 1, 2$ , denote by  $\mathbf{A}_k$  the image of  $[0, 1]$  by  $\eta_k$ , namely,  $\mathbf{A}_0 = [0, \mathbf{a}] \cup [\mathbf{c}, 1]$ ,  $\mathbf{A}_1 = [\mathbf{a}, \mathbf{b}]$  and  $\mathbf{A}_2 = [\mathbf{b}, \mathbf{c}]$ . Define  $\mathbf{I}_k$  as follows:

$$\mathbf{I}_k = \eta_k^{-1}(\mathbf{I} \cap \mathbf{A}_k), \quad k \in \{0, 1, 2\}.$$

Since the images of the affine injective maps  $(\eta_k)_{k \in \{0, 1, 2\}}$  partition  $[0, 1]$  up to measure-negligible overlaps,

$$(15) \quad \text{Leb}(\mathbf{I}) = \Delta_0 \text{Leb}(\mathbf{I}_0) + \Delta_1 \text{Leb}(\mathbf{I}_1) + \Delta_2 \text{Leb}(\mathbf{I}_2).$$

Since  $\mathbf{I}$  is an independent set of  $\mathbf{W}^p$ , Eq. (14) implies that  $\mathbf{I}_k$  is an independent set of  $\mathbf{W}_k^p$  for every  $k \in \{0, 1, 2\}$ . In particular,  $\text{Leb}(\mathbf{I}_k) \leq \tilde{\alpha}(\mathbf{W}_k^p)$ . Moreover, we notice that if  $\mathbf{B} = 1$ , then either  $\text{Leb}(\mathbf{I}_1) = 0$  or  $\text{Leb}(\mathbf{I}_2) = 0$  (by definition of independent set in a graphon). Together with Eq. (15), we deduce

$$\text{Leb}(\mathbf{I}) \leq \Delta_0 \tilde{\alpha}(\mathbf{W}_0^p) + \mathbf{B} \max\left(\Delta_1 \tilde{\alpha}(\mathbf{W}_1^p), \Delta_2 \tilde{\alpha}(\mathbf{W}_2^p)\right) + (1 - \mathbf{B})\left(\Delta_1 \tilde{\alpha}(\mathbf{W}_1^p) + \Delta_2 \tilde{\alpha}(\mathbf{W}_2^p)\right).$$

From Eq. (4), taking the supremum over independent sets  $\mathbf{I}$  of  $\tilde{\alpha}(\mathbf{W}^p)$ , one obtains the following a.s. inequality

$$\tilde{\alpha}(\mathbf{W}^p) \leq \Delta_0 \tilde{\alpha}(\mathbf{W}_0^p) + \mathbf{B} \max\left(\Delta_1 \tilde{\alpha}(\mathbf{W}_1^p), \Delta_2 \tilde{\alpha}(\mathbf{W}_2^p)\right) + (1 - \mathbf{B})\left(\Delta_1 \tilde{\alpha}(\mathbf{W}_1^p) + \Delta_2 \tilde{\alpha}(\mathbf{W}_2^p)\right).$$

Since  $\tilde{\alpha}(\mathbf{W}_k^p)$  has the same distribution as  $\tilde{\alpha}(\mathbf{W}^p)$  for  $k \in \{0, 1, 2\}$ , and the three are independent, the right-hand-side is a random variable distributed as  $\mathbf{Y}_{(p)}^{\text{Law}(\tilde{\alpha}(\mathbf{W}^p))}$ , proving that  $\text{Law}(\tilde{\alpha}(\mathbf{W}^p))$  satisfies Eq. (11).  $\square$

### 3.2. Solving the inequation.

**Proposition 3.2.** *For  $p$  in  $[0, 1)$ , the Dirac distribution  $\mu = \delta_0$  is the only probability distribution on  $[0, 1]$  solution of the inequation (11).*

We start by stating and proving a key lemma. Recall the definition of  $\Phi_p(\mu)$  from (11). The map  $\Phi_p$  is a functional from the space  $\mathcal{M}_1([0, 1])$  of probability distributions on  $[0, 1]$ . The space  $\mathcal{M}_1([0, 1])$  can be endowed with the so-called *Wasserstein distance* (also called optimal cost distance, or Kantorovich-Rubinstein distance):

$$d_W(\nu, \nu') = \inf_{\mathbf{X}, \mathbf{X}': \mathbf{X} \sim \nu, \mathbf{X}' \sim \nu'} \mathbb{E}[|\mathbf{X} - \mathbf{X}'|],$$

where the infimum is taken over all pairs  $(\mathbf{X}, \mathbf{X}')$  of random variables defined on the same probability space with distributions  $\nu$  and  $\nu'$ , respectively. We will use below the fact that this infimum is reached (for an explicit expression of the minimizing coupling see e.g. Remark 2.30 in [PC19]).

Furthermore since we are working on a compact space, convergence for  $d_W$  is equivalent to weak convergence of measures (see [Vil08, Sec.6]).

**Lemma 3.3.** *For  $p \in [0, 1]$ , the map  $\Phi_p$  is a weak contraction for  $d_W$ , i.e. for measures  $\mu$  and  $\nu$  in  $\mathcal{M}_1([0, 1])$  with  $\mu \neq \nu$ , we have  $d_W(\Phi_p(\mu), \Phi_p(\nu)) < d_W(\mu, \nu)$ .*

*Proof.* Let  $\mu$  and  $\nu$  be probability distributions on  $[0, 1]$ . We choose a pair  $(\mathbf{X}_0^\mu, \mathbf{X}_0^\nu)$  of random variables of distribution  $\mu$  and  $\nu$  respectively such that  $\mathbb{E}[|\mathbf{X}_0^\mu - \mathbf{X}_0^\nu|] = d_W(\mu, \nu)$  (as mentioned above, such a coupling always exists). We then let  $(\mathbf{X}_1^\mu, \mathbf{X}_1^\nu)$  and  $(\mathbf{X}_2^\mu, \mathbf{X}_2^\nu)$  be independent copies of  $(\mathbf{X}_0^\mu, \mathbf{X}_0^\nu)$ . Finally, we let  $(\Delta_0, \Delta_1, \Delta_2)$  be a random vector with distribution Dirichlet(1/2, 1/2, 1/2) independent from  $(\mathbf{X}_i^\mu, \mathbf{X}_i^\nu)_{i \in \{0,1,2\}}$ , and  $B$  a Bernoulli(1 -  $p$ ) random variable, independent from  $(\Delta_0, \Delta_1, \Delta_2, (\mathbf{X}_i^\mu, \mathbf{X}_i^\nu)_{i \in \{0,1,2\}})$ .

As in Eqs. (8) - (10), we define  $\mathbf{Y}_0^\mu, \mathbf{Y}_0^\nu, \mathbf{Y}_1^\mu, \mathbf{Y}_1^\nu, \mathbf{Y}_{(p)}^\mu$  and  $\mathbf{Y}_{(p)}^\nu$  on the same probability space and coupled in a non-trivial way: we use the same vector  $(\Delta_0, \Delta_1, \Delta_2)$  and Bernoulli variable  $B$  for both  $\mathbf{Y}_{(p)}^\mu$  and  $\mathbf{Y}_{(p)}^\nu$ .

Then we have

$$(16) \quad \mathbb{E}[|\mathbf{Y}_0^\mu - \mathbf{Y}_0^\nu|] \leq \sum_{i=0}^2 \mathbb{E}[\Delta_i] \mathbb{E}[|\mathbf{X}_i^\mu - \mathbf{X}_i^\nu|] = \left( \sum_{i=0}^2 \mathbb{E}[\Delta_i] \right) d_W(\mu, \nu) = d_W(\mu, \nu),$$

where we used successively the fact that  $\Delta_i$  is independent from  $(\mathbf{X}_i^\mu, \mathbf{X}_i^\nu)$ , the fact that the coupling  $\mathbf{X}_i^\mu, \mathbf{X}_i^\nu$  minimizes their  $L^1$  distance and the fact that  $\sum_{i=0}^2 \Delta_i = 1$  almost surely. We also have

$$(17) \quad \mathbb{E}[|\mathbf{Y}_1^\mu - \mathbf{Y}_1^\nu|] \leq \mathbb{E}[\Delta_0] \mathbb{E}[|\mathbf{X}_0^\mu - \mathbf{X}_0^\nu|] + \mathbb{E}\left[ \left| \max(\Delta_1 \mathbf{X}_1^\mu, \Delta_2 \mathbf{X}_2^\mu) - \max(\Delta_1 \mathbf{X}_1^\nu, \Delta_2 \mathbf{X}_2^\nu) \right| \right].$$

We recall the trivial inequality  $|\max(a, b) - \max(c, d)| \leq \max(|a - c|, |b - d|) \leq |a - c| + |b - d|$ . Besides, the second inequality is strict as soon as  $a \neq c$  and  $b \neq d$ . Taking

$$a = \Delta_1 \mathbf{X}_1^\mu, \quad b = \Delta_2 \mathbf{X}_2^\mu, \quad c = \Delta_1 \mathbf{X}_1^\nu, \quad d = \Delta_2 \mathbf{X}_2^\nu,$$

we obtain that, almost surely,

$$\left| \max(\Delta_1 \mathbf{X}_1^\mu, \Delta_2 \mathbf{X}_2^\mu) - \max(\Delta_1 \mathbf{X}_1^\nu, \Delta_2 \mathbf{X}_2^\nu) \right| \leq \Delta_1 |\mathbf{X}_1^\mu - \mathbf{X}_1^\nu| + \Delta_2 |\mathbf{X}_2^\mu - \mathbf{X}_2^\nu|.$$



Moreover, since  $\mu \neq \nu$ , we have that  $\mathbf{X}_1^\mu \neq \mathbf{X}_1^\nu$  with positive probability. The same holds for  $\mathbf{X}_2^\mu \neq \mathbf{X}_2^\nu$ , and, by independence, both inequalities occur simultaneously with positive probability. Since  $\Delta_1$  and  $\Delta_2$  are positive almost surely, we have that  $a \neq c$  and  $b \neq d$  simultaneously with positive probability. We conclude that the above inequality is strict with positive probability. Taking expectation and using Eq. (17), we get

$$(18) \quad \mathbb{E}[|\mathbf{Y}_1^\mu - \mathbf{Y}_1^\nu|] < \sum_{i=0}^2 \mathbb{E}[\Delta_i] \mathbb{E}[|\mathbf{X}_i^\mu - \mathbf{X}_i^\nu|] = d_W(\mu, \nu),$$

where the last equality is taken from (16). Finally,

$$\begin{aligned} d_W(\Phi_p(\mu), \Phi_p(\nu)) &\leq \mathbb{E}[|\mathbf{Y}_{(p)}^\mu - \mathbf{Y}_{(p)}^\nu|] \\ &= \mathbb{P}(\mathbf{B} = 1) \mathbb{E}[|\mathbf{Y}_{(p)}^\mu - \mathbf{Y}_{(p)}^\nu| \mid \mathbf{B} = 1] + \mathbb{P}(\mathbf{B} = 0) \mathbb{E}[|\mathbf{Y}_{(p)}^\mu - \mathbf{Y}_{(p)}^\nu| \mid \mathbf{B} = 0] \\ &= (1-p) \mathbb{E}[|\mathbf{Y}_1^\mu - \mathbf{Y}_1^\nu|] + p \mathbb{E}[|\mathbf{Y}_0^\mu - \mathbf{Y}_0^\nu|]. \end{aligned}$$

The lemma thus follows from Eqs. (16) and (18) and the fact that  $p \neq 1$ .  $\square$

*Proof of Proposition 3.2.* We first note that  $\Phi_p$  is nondecreasing with respect to stochastic domination, namely if  $\mu \leq_d \nu$  then  $\Phi_p(\mu) \leq_d \Phi_p(\nu)$ . Therefore if  $\mu$  is a solution of Inequation (11), i.e.  $\mu \leq_d \Phi_p(\mu)$ , we have  $\Phi_p(\mu) \leq_d \Phi_p^2(\mu)$  and, iterating the application of  $\Phi_p$ , we get  $\mu \leq_d \Phi_p(\mu) \leq_d \dots \leq_d \Phi_p^k(\mu)$  for all  $k \geq 1$ .

Moreover the Dirac distribution  $\delta_0$  is a fixed point of  $\Phi_p$ . Since  $\Phi_p$  is a weak contraction by Lemma 3.3 and since  $\mathcal{M}_1([0, 1])$  is compact, we know from Banach fixed-point theorem that  $\Phi_p^k$  tends to  $\delta_0$  in distribution. Combined with  $\mu \leq_d \Phi_p^k(\mu)$ , this forces  $\mu = \delta_0$  for any probability distribution  $\mu$  on  $[0, 1]$  verifying (11), which is what we wanted to prove.  $\square$

**3.3. Completing the proof of the sublinearity results.** Propositions 3.1 and 3.2 imply the following result, which is the core of the proofs of our sublinearity results (Theorems 1.9 and 1.10).

**Theorem 3.4.** *For  $p$  in  $[0, 1)$ , we have  $\tilde{\alpha}(\mathbf{W}^p) = 0$  almost surely.*

We now proceed with the proofs of our sublinearity results.

*Proof of Theorem 1.9.* Let  $p \in [0, 1)$  and consider a sequence  $(\mathbf{G}_n)$  of random graphs which converges to the Brownian cographon  $\mathbf{W}^p$ . By Skorokhod's representation theorem, we can represent all  $\mathbf{G}_n$  and  $\mathbf{W}^p$  on the same probability space so that  $\mathbf{G}_n$  converges to  $\mathbf{W}^p$  in the cut distance almost surely. Applying Proposition 2.3, we get that, a.s.,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \alpha(\mathbf{G}_n) = \limsup_{n \rightarrow \infty} \tilde{\alpha}(W_{\mathbf{G}_n}) \leq \tilde{\alpha}(\mathbf{W}^p),$$

By Theorem 3.4, the upper bound is 0 a.s. Thus,  $\frac{1}{n} \alpha(\mathbf{G}_n)$  converges to 0 a.s. and hence in probability.  $\square$

*Proof of Theorem 1.10.* Recall that, for any permutation  $\sigma$ , there is a one-to-one correspondence between increasing subsequences of  $\sigma$  and independent sets of  $\text{inv}(\sigma)$ . In particular, one has  $\text{LIS}(\sigma) = \alpha(\text{inv}(\sigma))$ .

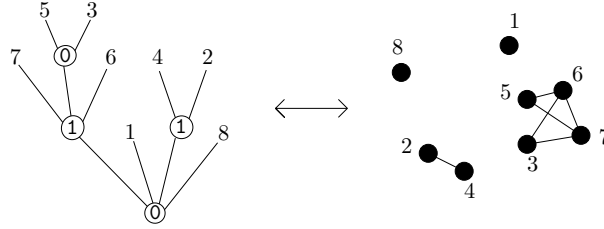


FIGURE 3. Left: A labeled cotree  $t$  with 8 leaves. Right: The associated labeled cograph  $\text{Cograp}(t)$  of size 8.

Consider now a sequence  $\sigma_n$  of random permutations tending to the Brownian separable permutation  $\mu^p$  for  $p \in [0, 1)$ . By Proposition 2.6, the sequence  $\text{inv}(\sigma_n)$  converges to the Brownian cographon  $\mathbf{W}^p$ . Applying Theorem 1.9 gives that  $\frac{\alpha(\text{inv}(\sigma_n))}{n}$  tends to 0 in probability. But  $\frac{\alpha(\text{inv}(\sigma_n))}{n} = \frac{\text{LIS}(\sigma_n)}{n}$  a.s., concluding the proof.  $\square$

#### 4. EXPECTED NUMBER OF INDEPENDENT SETS OF LINEAR SIZE

For  $k \leq n$  let  $\mathbf{X}_{n,k}$  be the random variable given by the number of independent sets of size  $k$  in a uniform labeled cograph of size  $n$ . The goal of this section is to prove Theorem 1.3, i.e. to estimate  $\mathbb{E}[\mathbf{X}_{n,k}]$  in the case where  $k$  grows linearly in  $n$ .

The first step of the proof is to obtain equations for the exponential generating series of cographs with a marked independent set, through symbolic combinatorics. To this aim, it is convenient to encode cographs by their *cotrees*. The asymptotic analysis is then performed via saddle-point analysis.

##### 4.1. Combinatorial preliminaries: cographs and cotrees.

**Definition 4.1.** 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;
- decorations 0 and 1 should alternate along each branch from the root to a leaf.

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

For an unlabeled cotree  $t$ , we denote by  $\text{Cograp}(t)$  the unlabeled graph defined recursively as follows (see an illustration in Fig. 3):

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

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  $\text{Cograp}(t)$ . 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  $\text{Cograph}(t)$ ), with vertex set  $\{1, \dots, n\}$ : each vertex of  $\text{Cograph}(t)$  receives the label of the corresponding leaf of  $t$ .

By construction, for all cotrees  $t$ , the graph  $\text{Cograph}(t)$  is a cograph. Conversely, each cograph can be obtained in this way, and this correspondence is one-to-one. This property is ensured by the alternation of decorations 0 and 1 in cotrees. This was first shown 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].

From a cograph  $G$ , the unique cotree  $t$  such that  $\text{Cograph}(t) = G$  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 under taking induced subgraphs and complement, and a cograph  $G$  of size at least two is not connected exactly when its complement  $\bar{G}$  is connected.

*Remark 4.2.* The transformation which switches every decoration  $1 \leftrightarrow 0$  in a cotree is of course an involution. Moreover, it turns independent sets into cliques in the corresponding cograph (indeed  $\{v, v'\}$  is an edge in  $\text{Cograph}(t)$  if and only if the first common ancestor of the corresponding leaves of  $t$  has decoration 1). This proves that for every  $n$ , if  $\mathbf{G}_n$  denotes a uniform random cograph (either labeled or unlabeled) of size  $n$ , then

$$(19) \quad \alpha(\mathbf{G}_n) \stackrel{(d)}{=} \omega(\mathbf{G}_n),$$

where  $\omega(G)$  is the maximum size of a clique in the graph  $G$ .

**4.2. Proof of Theorem 1.3: Enumeration.** Let  $\mathcal{L}$  be the combinatorial family of labeled cotrees for which we forget decorations, counted by the number of leaves. Let  $L(z)$  denote the corresponding exponential generating function. The series  $L(z) = \sum_{\ell \in \mathcal{L}} z^{|\ell|} / |\ell|!$  is the unique formal power series solution of

$$(20) \quad L(z) = z + e^{L(z)} - 1 - L(z)$$

such that  $L(0) = 0$ . (The enumeration of  $\mathcal{L}$  is provided in [FS09, Example VII.12 p.472] under the name of *labeled hierarchies*, see also Propositions 5.1 and 5.4 of [BBF+22b].)

Next we consider pairs  $(G, I)$ , where  $G$  is a (labeled) cograph and  $I$  an independent set of  $G$ . We see such a pair as a *marked* cograph. Let us consider the associated bivariate generating

function

$$C(z, u) = \sum_{\substack{(G, I): G \text{ cograph,} \\ I \subseteq V_G \text{ independent}}} \frac{1}{|V_G|!} z^{|V_G|} u^{|I|}.$$

Then  $\mathbb{E}[\mathbf{X}_{n,k}] = \frac{[z^n u^k] C(z, u)}{[z^n] C(z, 0)}$ . Our goal is then to find the asymptotics of these coefficients.

Note that if  $G$  is reduced to a single vertex  $\bullet$  we have  $(G, I) = (\bullet, \emptyset)$  or  $(\bullet, \{\bullet\})$ , therefore

$$(21) \quad C(z, u) = z + zu + C_0(z, u) + C_1(z, u),$$

where  $C_0(z, u)$  (resp  $C_1(z, u)$ ) is the bivariate series of the set  $\mathcal{C}_0$  (resp.  $\mathcal{C}_1$ ) of marked cographs (necessarily of size  $\geq 2$ ) for which the root of the associated cotree is decorated with a 0 (resp. a 1). We have

$$L(z) = z + C_0(z, 0) = z + C_1(z, 0).$$

Indeed when the decoration of the root is fixed, the other decorations are then determined by the alternation condition.

**Proposition 4.3** (Functional equations for  $C_0, C_1$ ).

i) A relation between the series  $C_0(z, u)$  and  $C_1(z, u)$  is given by

$$(22) \quad C_0(z, u) = e^{z(1+u)+C_1(z,u)} - 1 - z(1+u) - C_1(z, u),$$

ii) and the series  $C_1(z, u)$  is a solution of

$$(23) \quad C_1(z, u) = e^{L(z)} - 1 - L(z) + (e^{L(z)} - 1) (e^{z(1+u)+C_1(z,u)} - 1 - C_1(z, u) - L(z)).$$

In the proof below, we make use of the notation  $\exp_{\geq k}(x) := \sum_{i \geq k} \frac{x^i}{i!}$ .

*Proof.* When a cotree  $T$  has its root  $r$  decorated by a 0, if we denote by  $(T_i)$  the subtrees rooted at the children of  $r$ , then the cograph  $G$  associated with  $T$  is the disjoint union of the cographs  $G_i$  corresponding to the  $T_i$ . An independent set of  $G$  is then the union of independent sets chosen in each of the  $G_i$ . Recall also that by definition of cotrees,  $r$  has at least two children.

Therefore the marked cographs for which the root of the associated cotree is decorated with a 0 can be described as a multiset of at least two elements chosen between  $(\bullet, \emptyset)$ ,  $(\bullet, \{\bullet\})$  and the elements of  $\mathcal{C}_1$ .

Using the symbolic method for labeled structures [FS09], we get the equation

$$C_0(z, u) = \exp_{\geq 2}(z + zu + C_1(z, u)) = e^{z(1+u)+C_1(z,u)} - 1 - z(1+u) - C_1(z, u),$$

which is Eq. (22).

When on the contrary a cotree  $T$  has its root  $r$  decorated by a 1, if we denote again by  $(T_i)$  the subtrees rooted at the children of  $r$ , then the cograph  $G$  associated with  $T$  is the join of the cographs  $G_i$  corresponding to the  $T_i$ . An independent set of  $G$  must then be an independent set chosen in one of the  $G_i$  only (and the other children of  $r$  do not contribute to this independent set).

Let  $\mathcal{C}_\emptyset$  denote the set of cographs without mark and whose cotree does not have a root decorated by a 1, i.e.  $\mathcal{C}_\emptyset$  is the set consisting in  $(\bullet, \emptyset)$  and the elements of  $\mathcal{C}_0$  marked with an empty independent set. Then, we distinguish two cases to describe the elements of  $\mathcal{C}_1$  (marked cographs

for which the root of the associated cotree is decorated with a 1). Either they are marked with an empty independent set, and in this case they can be described as multisets of at least two elements of  $\mathcal{C}_\emptyset$ . Or they are marked with a nonempty independent set, and they can be described as the pairs consisting of

- a cograph which is either  $(\bullet, \{\bullet\})$  or an element of  $\mathcal{C}_0$  marked with a nonempty independent set (for the graph  $G_i$  containing the independent set); and
- a multiset of at least one element of  $\mathcal{C}_\emptyset$  (for the other graphs  $G_i$ ).

We get the equation

$$C_1(z, u) = \exp_{\geq 2}(z + C_0(z, 0)) + (zu + C_0(z, u) - C_0(z, 0)) \times \exp_{\geq 1}(z + C_0(z, 0)).$$

Thus, by eliminating  $C_0(z, u)$  and using that  $L(z) = z + C_0(z, 0)$ , we obtain Eq. (23).  $\square$

From Eqs. (21) and (22) we have

$$(24) \quad C(z, u) = e^{z(1+u)+C_1(z,u)} - 1.$$

In the following, to get the asymptotics of the coefficients of  $C(z, u)$ , we study  $C_1(z, u)$  using Eq. (23).

**4.3. Proof of Theorem 1.3: main asymptotics - Proof of Eq. (1).** Following Flajolet and Sedgewick [FS09, p. 389], we say that a domain  $\Delta$  is a  $\Delta$ -domain at  $\rho$  if there exist two real numbers  $R > \rho$  and  $0 < \phi < \frac{\pi}{2}$  such that

$$\Delta = \{z \in \mathbb{C} \mid |z| < R, z \neq \rho, |\arg(z - \rho)| > \phi\}$$

and that a power series is  $\Delta$ -analytic if it is analytic in some  $\Delta$ -domain at  $\rho$ , where  $\rho$  is its radius of convergence.

From [FS09, Example VII.12 p.472] the series  $L(z)$  has radius of convergence  $\rho = 2 \log(2) - 1$  and is  $\Delta$ -analytic. Moreover the following expansion holds in a  $\Delta$ -domain at  $z = \rho$ :

$$(25) \quad L(z) \underset{z \rightarrow \rho}{=} \log(2) - \sqrt{\rho} \sqrt{1 - \frac{z}{\rho}} + \mathcal{O}\left(1 - \frac{z}{\rho}\right).$$

Eq. (25) combined with the transfer theorem [FS09, Cor.VI.1] yields

$$(26) \quad [z^n]L(z) \sim \sqrt{\frac{2 \log 2 - 1}{4\pi}} \rho^{-n} n^{-3/2}.$$

This allows to obtain the asymptotics of  $[z^n]C(z, 0)$ . To get the one of  $[z^n u^k]C(z, u)$ , we turn to the study of  $C_1(z, u)$ .

Fix  $u \in \mathbb{C}$ . The overall strategy is to perform saddle-point analysis with  $C_1(z, u)$ . To do so we rewrite Eq. (23) as  $C_1(z, u)$  is solution of  $c = G(z, c, u)$  where

$$G(z, c, u) = e^{L(z)} - 1 - L(z) + (e^{L(z)} - 1) (e^{c+z(1+u)} - 1 - c - L(z)).$$

We will show that this almost fits the settings of the so-called *smooth implicit-function schema* (see [FS09, Sec. VII.4.1]), only the nonnegativity of the coefficients of  $G$  is not satisfied here. Nevertheless, we shall prove that sufficient conditions for the validity of [FS09, Thm. VII.3 p.468] are satisfied.

**Analyticity.**

Observe that for every  $u \in \mathbb{C}$  the bivariate series  $(z, c) \mapsto G(z, c, u)$  is analytic for  $|z| < \rho$  and  $c \in \mathbb{C}$ .

**Solution of the characteristic system.**

We use the notational convention that, for any function  $H$  and variable  $t$ ,  $H_t$  denotes the partial derivative of  $H$  with respect to  $t$ .

We consider the *characteristic system*

$$(27) \quad G(r, s, u) = s, \quad G_c(r, s, u) = 1,$$

namely

$$(28) \quad e^{L(r)} - 1 - L(r) + (e^{L(r)} - 1) (e^{s+r(1+u)} - 1 - s - L(r)) = s,$$

$$(29) \quad (e^{L(r)} - 1) (e^{s+r(1+u)} - 1) = 1.$$

We aim at proving that, for any  $u > 0$ , (27) admits a unique solution  $(r, s) = (r(u), s(u))$  with  $0 < r < \rho$  and  $0 < s$ . Below, we often use that the radius of convergence  $\rho$  of  $L(z)$  satisfies  $\rho = 2 \log(2) - 1$  and  $L(\rho) = \log(2)$ .

We observe that if we substitute Eq. (29) into Eq. (28) we obtain that

$$(30) \quad s = 1 - L(r).$$

Then Eq. (29) can be rewritten as

$$(31) \quad F(r, u) = 0 \quad \text{with} \quad F(x, u) = (e^{L(x)} - 1)(e^{1-L(x)+x(1+u)} - 1) - 1.$$

We have

$$\begin{aligned} F_x(x, u) &= (L'(x)e^{L(x)} + (1 + u - L'(x))(e^{L(x)} - 1)) e^{1-L(x)+x(1+u)} - L'(x)e^{L(x)} \\ &= (1 + u)(e^{L(x)} - 1)e^{1-L(x)+x(1+u)} + L'(x)(e^{1-L(x)+x(1+u)} - e^{L(x)}). \end{aligned}$$

Fix  $u > 0$ . For  $0 < x \leq \rho$ , one has

$$1 - L(x) + x(1 + u) > 1 - L(x) + x \geq L(x)$$

(indeed one has equality for  $x = \rho$  and  $2L(x) - x$  is increasing), so that  $F_x(x, u) > 0$ . Therefore, the function  $F$  is increasing with  $x$  on the interval  $[0, \rho]$ . Since  $F(0, u) = -1$  and  $F(\rho, u) > F(\rho, 0) = 0$ , for any  $u > 0$  Eq. (31) admits a unique solution  $r = r(u)$  in  $[0, \rho]$ , and we have  $0 < r < \rho$ .

From Eq. (30), we have  $s = 1 - L(r)$ . Since  $L$  is increasing and  $r < \rho$ , we have  $s > 1 - L(\rho) = 1 - \log(2) > 0$ .

We conclude that for  $u > 0$ , the characteristic system (27) has a unique solution  $r(u), s(u)$  in  $[0, \rho] \times \mathbb{C}$ , and we have  $0 < r(u) < \rho$  and  $s(u) > 0$ . In particular,  $r(u), s(u)$  belongs to the analyticity domain of  $G$ .

### Locating the singularity of $C_1(z, u)$ .

Fix  $u > 0$ . To obtain the singular behavior of  $C_1(z, u)$  as in [FS09, Thm. VII.3 p.468] despite the negativity of some coefficients of  $G$ , we see from [FS09, Note VII.16 p.471] that it is enough to show the following:  $C_1(z, u)$  has radius of convergence  $r(u)$  and its value at this singularity is given by  $C_1(r(u), u) = s(u)$ , i.e. the dominant singularity of  $C_1(z, u)$  corresponds to the solution of the characteristic system.

The argument to prove this is an adaptation of that in the proof of [FS09, Thm. VII.3 p.468] to our setting where  $G$  has some negative coefficients but a larger analyticity region than what is usually assumed. Namely, our  $G$  is analytic on the whole domain  $\{|z| < \rho, c \in \mathbb{C}\}$ , while the smooth implicit-function schema only assumes analyticity on  $\{|z| < R, |c| < S\}$  for some  $R, S > 0$  (with the notation of [FS09, Sec. VII.4.1]). Let us denote temporarily  $\rho(u)$  the radius of convergence of  $C_1(z, u)$ , which is a singularity of  $C_1(z, u)$  from Pringsheim's theorem.

We first show that  $\rho(u) \geq r(u)$ . We proceed by contradiction, and assume  $\rho(u) < r(u)$ . We set  $\sigma(u) = C_1(\rho(u), u)$  and distinguish two cases.

- Assume  $\sigma(u) < +\infty$ . Then, since  $C_1(z, u)$  is a solution of  $c = G(z, c, u)$ , we have  $\sigma(u) = G(\rho(u), \sigma(u), u)$ . By uniqueness of the solution of the characteristic system (27), we necessarily have  $G_c(\rho(u), \sigma(u), u) \neq 1$ . Therefore, using the analytic implicit function lemma [FS09, Lemma VII.2, p.469],  $C_1(z, u)$  can be extended analytically in a neighborhood of  $\rho(u)$ , contradicting the fact that  $\rho(u)$  is a singularity of  $C_1(z, u)$ .
- If  $\sigma(u) = +\infty$ , one checks easily that the function  $z \mapsto G_c(z, C_1(z, u), u)$  tends to  $+\infty$  when  $z$  tends to  $\rho(u)$ . But for  $z = 0$ , we have  $G_c(0, C_1(0, u), u) = 0$ . The intermediate value theorem ensures the existence of  $z_1$  in  $(0, \rho(u))$  such that  $G_c(z_1, C_1(z_1, u), u) = 1$ . This gives an other solution  $(z_1, C_1(z_1, u))$  of the characteristic system, contradicting the uniqueness of the solution.

We have reached a contradiction in both cases, proving that  $\rho(u) \geq r(u)$ .

This allows us to consider  $C_1(r(u), u)$  (which is possibly infinite), and we assume for the sake of contradiction that  $C_1(r(u), u) \neq s(u)$ . Then for  $a < r(u)$  sufficiently closed to  $r(u)$  the equation  $y = G(a, y, u)$  admits several solutions  $y \in \mathbb{C}$ :

- one is given by  $y = C_1(a, u)$ ,
- and two are obtained evaluating in  $a$  the two functions  $y_1(z)$  and  $y_2(z)$  given by the singular implicit function lemma [FS09, Lemma VII.3, p.469] applied to the point  $(r(u), s(u))$ .

(Note that the applicability of this lemma is guaranteed by the fact that  $(r(u), s(u))$  is a solution of the characteristic system and Eqs. (32) and (40) below.)

From [FS09, Lemma VII.3, p.469], it is clear that the last two solutions above are distinct for  $a$  close enough to  $r(u)$ . The first one is also different from them for  $a$  close enough to  $r(u)$ : indeed, for  $z$  tending to  $r(u)$ ,  $C_1(z, u)$  tends to  $C_1(r(u), u)$  while the two other solutions tend to  $s(u)$ . However, the function  $y \mapsto G(a, y, u)$  is strictly convex (one checks easily that its second derivative is positive) and therefore cannot cross three times the main diagonal. We have reached a contradiction. We conclude that  $C_1(r(u), u) = s(u)$ .

It remains to prove  $\rho(u) = r(u)$ . Since  $(r(u), s(u))$  is a solution of the characteristic system, there is no analytic solution of the equation  $y = G(z, y, u)$  around the point  $(z, y) = (r(u), s(u))$

(see the proof of [FS09, Lemma VII.3, p.469], where it is shown that any solution  $y$  has a series expansion involving a square-root term and hence cannot be analytic). Therefore  $C_1(z, u)$  cannot be extended analytically to a neighborhood of  $r(u)$ . So,  $\rho(u) = r(u)$ , as wanted.

**Derivatives of  $G$ : parametrized expressions and their signs.**

Several derivatives of  $G(z, c, u)$  appear in the computations below, to establish the asymptotic behavior of  $C(z, u)$  as well as estimates (i) and (ii) of Theorem 1.3. We collect useful properties of these derivatives here for convenience. In this paragraph, we also assume  $u > 0$ . Recall that

$$G(z, c, u) = e^{L(z)} - 1 - L(z) + (e^{L(z)} - 1) (e^{c+z(1+u)} - 1 - c - L(z)).$$

First, from the explicit expression of  $G_{cc}$ , it follows that

$$(32) \quad G_{cc}(r(u), s(u), u) > 0.$$

Moving on to  $G_u(r(u), s(u), u)$ , it will be convenient to parametrize the involved quantities by  $y := L(r(u))$ . Eq. (20) becomes

$$(33) \quad y = r(u) + e^y - 1 - y$$

and therefore

$$(34) \quad r(u) = 2y + 1 - e^y.$$

From Eq. (29), we quickly derive

$$(35) \quad e^{s(u)+r(u)(1+u)} = \frac{e^y}{e^y - 1}$$

and we can eliminate  $r(u)$  thanks to Eq. (34): we obtain

$$(36) \quad u = \frac{e^y - 2 - \log(e^y - 1)}{2y + 1 - e^y}.$$

Next we use the definition of  $G$ , and then Eqs. (33) and (35), obtaining

$$(37) \quad \begin{aligned} G_u(r(u), s(u), u) &= (e^{L(r(u))} - 1)e^{s(u)+r(u)(1+u)}r(u) \\ &= (e^y - 1)\frac{e^y}{e^y - 1}(2y + 1 - e^y) = e^y(2y + 1 - e^y). \end{aligned}$$

From the above and Eq. (34), we have in particular

$$(38) \quad G_u(r(u), s(u), u) > 0$$

Finally, we focus on  $G_z(r(u), s(u), u)$ . Using Eqs. (20) and (30), we start by observing that

$$L'(z) = \frac{1}{2 - e^{L(z)}}, \quad 1 + s(u) + y = 2.$$



Therefore, with the shorter notation  $L := L(r(u))$ ,  $L' := L'(r(u))$ ,  $r := r(u)$ ,  $s := s(u)$ , we have

$$\begin{aligned}
 G_z(r(u), s(u), u) &= (e^L - 1)L' + e^L L' (e^{s+r(1+u)} - 1 - s - L) + (e^L - 1) (e^{s+r(1+u)}(1+u) - L') \\
 &= \frac{e^y - 1}{2 - e^y} + \frac{e^y}{2 - e^y} \times \left( \frac{e^y}{e^y - 1} - 2 \right) + (e^y - 1) \times \left( \frac{e^y}{e^y - 1} (1+u) - \frac{1}{2 - e^y} \right) \\
 &= \frac{e^y}{2 - e^y} \times \left( \frac{e^y}{e^y - 1} - 2 \right) + (e^y - 1) \times \left( \frac{e^y}{e^y - 1} (1+u) \right) \\
 (39) \quad &= \frac{e^y}{2 - e^y} \times \frac{2 - e^y}{e^y - 1} + e^y(1+u) = \frac{e^y}{e^y - 1} + e^y(1+u).
 \end{aligned}$$

In particular, this gives

$$(40) \quad G_z(r(u), s(u), u) > 0.$$

### Obtaining the asymptotics.

Recall that we established that  $C_1(z, u)$  has radius of convergence  $r(u)$  and its value at this singularity is given by  $C_1(r(u), u) = s(u)$ . From [FS09, Sec. VII.4.1], we therefore obtain an estimate of  $C_1(z, u)$  as  $z$  approaches  $r(u)$ . Namely, for every  $u > 0$  the series  $C_1(z, u)$  has a square-root singularity at  $r(u)$  and in some  $\Delta$ -domain, we have

$$(41) \quad C_1(z, u) \stackrel{z \rightarrow r(u)}{\sim} s(u) - \gamma_1(u) \sqrt{1 - z/r(u)} + \mathcal{O}(1 - z/r(u))$$

with  $\gamma_1(u) = \sqrt{\frac{2r(u)G_z(r(u), s(u), u)}{G_{cc}(r(u), s(u), u)}}$ . Note that  $G_{cc}(r(u), s(u), u) > 0$  and  $G_z(r(u), s(u), u) > 0$  from Eqs. (32) and (40). The determination of the sign in front of  $\sqrt{1 - z/r(u)}$  uses that  $C_1$  is increasing in  $z$  when  $z$  approaches  $r(u)$  from the left.

To obtain asymptotics for the coefficients of  $C_1(z, u)$ , we have to extend (41) for complex  $u$  around  $u > 0$ . We argue that the solutions  $(r, s) = (r(u), s(u))$  of the characteristic system (27) have analytic continuations in a neighborhood of every  $u > 0$ . Observe that  $G$  is analytic and that the Jacobian matrix of the system is the following determinant (where all derivatives are evaluated at  $(r(u), s(u), u)$ )

$$\begin{vmatrix} G_z & G_c - 1 \\ G_{cz} & G_{cc} \end{vmatrix} = \begin{vmatrix} G_z & 0 \\ G_{cz} & G_{cc} \end{vmatrix} = G_z G_{cc}.$$

It is nonzero for  $u > 0$  from Eqs. (32) and (40). Consequently, there exist analytic functions  $r(u), s(u)$  defined on a neighborhood of the positive real axis, such that, for each  $u$ , the pair  $(r(u), s(u))$  is a solution of the characteristic system for such values of  $u$ .

By continuity we can also ensure that, for  $u$  sufficiently close to the real axis,

- $r(u)$  is the unique singularity of  $C_1(z, u)$  of smallest modulus and  $C_1(r(u), u) = s(u)$ ;
- $G_z$  and  $G_{cc}$  are non-zero at  $(r(u), s(u), u)$ .

We denote by  $U$  the open set of complex numbers  $u$  where these properties hold. Therefore, as stated in [Drm09, Remark 2.20], it follows that the singular representation (41) also holds for complex  $u \in U$  (and for  $z$  in a proper  $\Delta$ -domain depending on  $u$ ).

Combining relation (24) with the above development (41) of  $C_1(z, u)$  near  $z = r(u)$ , we obtain for  $u \in U$

$$C(z, u) \stackrel{z \rightarrow r(u)}{\asymp} e^{r(u)(1+u)+s(u)} - 1 - \gamma(u) \sqrt{1 - z/r(u)} + \mathcal{O}(1 - z/r(u)),$$

where  $\gamma(u)$  is defined by

$$\gamma(u) = \gamma_1(u) \exp(s(u) + r(u)(1 + u)).$$

Moreover since  $C(z, u)$  is aperiodic,  $r(u)$  is the unique dominant singularity of  $C$  and

$$(42) \quad [z^n] C(z, u) \stackrel{n \rightarrow +\infty}{\asymp} \frac{\gamma(u)}{2\sqrt{\pi}} n^{-3/2} (r(u))^{-n} (1 + \mathcal{O}(1/n))$$

uniformly for  $u$  in a compact subset contained in  $U$  (by Transfer Theorem [FS09, Thm.VI.3] and compactness).

Now we can proceed as in [Drm94, Thm.3], with the nonnegativity of the coefficients of  $G$  replaced by the above variant of the smooth-implicit function schema, and obtain by an application of a saddle point integration

$$(43) \quad [z^n u^k] C(z, u) \sim \frac{R_{k/n}}{n^2} (r(u(k/n)) u(k/n)^{k/n})^{-n},$$

uniformly for  $an \leq k \leq bn$  with  $0 < a < b < 1$ , where  $R_\beta$  ( $0 < \beta < 1$ ) is some positive (computable) quantity and  $u = u(\beta)$  is determined by the following equation (which is the rewriting of [Drm94, (2.14)] with our notation):

$$(44) \quad \beta = -\frac{u r'(u)}{r(u)} = \frac{u G_u(r(u), s(u), u)}{r(u) G_z(r(u), s(u), u)}.$$

We explain in Remark 4.4 below why Eq. (44) is indeed invertible.

Finally with Eq. (26) we obtain

$$\mathbb{E}[\mathbf{X}_{n,k}] = \frac{[z^n u^k] C(z, u)}{[z^n] C(z, 0)} \stackrel{\text{for } n \geq 2}{=} \frac{[z^n u^k] C(z, u)}{[z^n] 2L(z)} \sim \frac{B_{k/n}}{\sqrt{n}} (C_{k/n})^n,$$

uniformly for  $an \leq k \leq bn$  for some  $B_\beta > 0$  and with

$$(45) \quad C_\beta := \frac{2 \log(2) - 1}{r(u(\beta)) u(\beta)^\beta}.$$

concluding the proof of Eq. (1).

*Remark 4.4.* Let us justify that Eq. (44) can be inverted to express  $u$  as a function of  $\beta$ .

First, observe that Eq. (36) defines  $u$  as a function of  $y$ . This function is decreasing for  $y \in (0, \log(2))$  and maps bijectively  $(0, \log(2))$  to  $(0, \infty)$ . Therefore Eq. (36) can be inverted to express  $y$  as a function of  $u$ , which is decreasing and maps bijectively  $(0, \infty)$  to  $(0, \log(2))$ .

Second, from the second expression of  $\beta$  in Eq. (44), we obtain an expression of  $\beta$  as a function of  $y$ , substituting Eqs. (34), (36), (37) and (39) into Eq. (44). This gives

$$(46) \quad \beta = \frac{(e^y - 2 - \log(e^y - 1)) (e^y - 1)}{2y + 1 - e^y + (e^y - 1)(2y - 1 - \log(e^y - 1))}$$

This expression defines  $\beta$  as a function of  $y$ . This function is decreasing for  $y \in (0, \log(2))$  and maps bijectively  $(0, \log(2))$  to  $(0, 1)$ .

The function  $\beta = \beta(u)$ , obtained by composition of the above two, is therefore a bijection from  $(0, \infty)$  to  $(0, 1)$ , allowing to define  $u = u(\beta)$ . We observe, in addition, that  $0 < u(\beta) < 1$  for  $\beta > 0$ .

**4.4. Proof of Theorem 1.3: Estimates (i) and (ii).** We now analyze the expression of  $C_\beta$ . Combining Eq. (45) with Eqs. (34), (36) and (46), we can express  $C_\beta$  as an explicit function of  $y$ ; further inverting numerically Eq. (46) gives  $C_\beta$  as a function of  $\beta$ . The graph of the function  $\beta \mapsto C_\beta$  on Fig. 1 was obtained in this way.

From these expressions, we can also perform Taylor expansions (see the jupyter notebook mentioned below). The expansion of Eq. (46) around  $y = \log(2)$  yields

$$(47) \quad \beta = \frac{(y - \log(2))^2}{2 \log(2) - 1} + \mathcal{O}((y - \log(2))^3).$$

Plugging this estimate in the Taylor expansion of Eqs. (34) and (36) around  $y = \log(2)$ , we obtain

$$\begin{aligned} r(u(\beta)) &= 2 \log(2) - 1 - (y - \log(2))^2 + \mathcal{O}((y - \log(2))^3) \\ &= 2 \log(2) - 1 + (1 - 2 \log(2))\beta + \mathcal{O}(\beta^{3/2}); \\ u(\beta) &= \frac{2}{2 \log(2) - 1} (y - \log(2))^2 + \mathcal{O}((y - \log(2))^3) \\ &= 2\beta + \mathcal{O}(\beta^{3/2}). \end{aligned}$$

From Eq. (45) we deduce Theorem 1.3 item (ii):

$$C_\beta = \frac{2 \log(2) - 1}{r(u(\beta))u(\beta)^\beta} = 1 + \beta |\log(\beta)| + o(\beta \log(\beta))$$

when  $\beta \rightarrow 0$ .

In particular, this proves  $C_\beta > 1$  for  $\beta \in (0, \beta_0)$  for some  $\beta_0 > 0$ . Numerical computations give the estimate  $\beta_0 \approx 0.522677\dots$ ; we furthermore observe numerically that  $C_\beta$  reaches its maximum at  $\beta^* \approx 0.229285\dots$  where  $C_{\beta^*} \approx 1.3663055\dots$

Details on the computations above are provided in a jupyter notebook embedded into this pdf (alternatively you can download the source of the arXiv version to get the files). We provide both an html read-only version and an editable ipynb version for the reader's convenience.

## 5. EXPECTED NUMBER OF INCREASING SUBSEQUENCES OF LINEAR SIZE

We now discuss the proof of Theorem 1.8, the analog of Theorem 1.3 for separable permutations. We start with some definitions.

Given two permutations,  $\pi$  of size  $k$  and  $\tau$  of size  $\ell$ , the *direct sum* (resp. *skew sum*) of  $\pi$  and  $\tau$ , denoted  $\oplus[\pi, \tau]$  (resp.  $\ominus[\pi, \tau]$ ) is the permutation  $\sigma$  of size  $k + \ell$  such that

- for  $1 \leq i \leq k$ ,  $\sigma(i) = \pi(i)$  (resp.  $\sigma(i) = \ell + \pi(i)$ ), and
- for  $1 \leq i \leq \ell$ ,  $\sigma(k + i) = k + \tau(i)$  (resp.  $\sigma(k + i) = \tau(i)$ ).

Direct sums and skew sums readily extend to more than two permutations, writing  $\oplus[\pi, \dots, \tau, \rho] = \oplus[\pi, \dots \oplus [\tau, \rho]]$  (and similarly for  $\ominus$ ).

As mentioned in Section 1.2, separable permutations are those which can be obtained from permutations of size 1 performing direct sums and skew sums. This is similar to the characterization of cographs as the graphs obtained using the join and disjoint union constructions, from graphs with one vertex. And similarly to the description of cographs through their cotrees, this allows to associate a tree with each separable permutation. (This is actually a special case of the construction which associate with each permutation, not necessarily separable, its substitution decomposition tree – see *e.g.* [BBF+22a, Section 1.1]).

There are actually several presentations of this correspondence between separable permutations and trees. The one which is suitable here is presented in [BBF+18, Section 2.2], and we borrow our terminology from there.

**Definition 5.1.** *A signed Schröder tree where the signs alternate of size  $n$  is a rooted tree  $t$  with  $n$  leaves such that:*

- $t$  is plane (i.e. the children of every internal node are ordered);
- every internal node has at least two children;
- every internal node in  $t$  is decorated with  $\oplus$  or  $\ominus$ ;
- decorations  $\oplus$  and  $\ominus$  should alternate along each branch from the root to a leaf.

An important difference with cotrees is that the above trees are plane, while cotrees are not plane.

We can associate to a signed Schröder tree where the signs alternate a permutation  $\text{perm}(t)$  of the same size, as follows.

- If  $t$  consists of a single leaf, then  $\text{perm}(t)$  is the permutation of size 1.
- Otherwise, the root of  $t$  has decoration  $\oplus$  or  $\ominus$  and has subtrees  $t_1, \dots, t_d$  attached to it ( $d \geq 2$ ), in this order from left to right. Then, if the root has decoration  $\oplus$ , we let  $\text{perm}(t)$  be  $\oplus[\text{perm}(t_1) \dots, \text{perm}(t_d)]$ . Otherwise, the root has decoration  $\ominus$ , and we let  $\text{perm}(t)$  be  $\ominus[\text{perm}(t_1), \dots, \text{perm}(t_d)]$ .

**Proposition 5.2.** *The correspondence presented above between separable permutations and signed Schröder trees where the signs alternate is one-to-one.*

For a proof of this statement, we refer to [BBF+18, Proposition 2.13] – see also the references given in [BBF+18].

We can now move to the proof of Theorem 1.8. The strategy is the same as in the proof of Theorem 1.3, using the encoding of separable permutations by their signed Schröder trees where the signs alternate, instead of the encoding of cographs by their cotrees. We therefore only sketch the computations here. Details are provided in the attached jupyter notebook.

We denote by  $S := S(z)$  the solution of

$$(48) \quad S = z + S^2/(1 - S).$$

Equivalently,  $S$  is the series of Schröder trees (*i.e.*, plane trees where internal nodes have at least two children) counted by leaves. Unlike the series  $L$  in the case of cographs, the series  $S$  is

explicit here, namely it holds that  $S(z) = (1 + z - \sqrt{1 - 6z + z^2})/4$ . Its radius of convergence is  $\rho = 3 - 2\sqrt{2}$  and we have  $S(\rho) = (2 - \sqrt{2})/2$ .

The proof also involves the generating function  $S_\ominus := S_\ominus(z, u)$  (resp.  $S_\oplus := S_\oplus(z, u)$ ) counting separable permutations which can be decomposed as a skew sum (resp. direct sum) marked with an increasing subsequence. Without marking, from Proposition 5.2, we get

$$S_\ominus(z, 0) = S_\oplus(z, 0) = \frac{S^2}{1 - S} = S - z.$$

The analogs of Eqs. (22) and (23) are then

$$(49) \quad S_\ominus = \frac{S^2}{1 - S} + \left( \frac{(S_\ominus + z + zu)^2}{1 - (S_\ominus + z + zu)} + zu + z - S \right) \times \left( \frac{1}{(1 - S)^2} - 1 \right),$$

$$(50) \quad S_\oplus = \frac{(S_\ominus + z + zu)^2}{1 - (S_\ominus + z + zu)}.$$

Indeed, an increasing subsequence in a direct sum of permutations  $\pi_1 \oplus \dots \oplus \pi_r$  is a union of increasing subsequences in  $\pi_1, \dots$ , and  $\pi_r$ . Hence elements counted by  $S_\oplus$  can be described as sequences of at least two elements chosen between  $(\bullet, \emptyset)$ ,  $(\bullet, \{\bullet\})$  and the elements counted by  $S_\ominus$ . This leads to Eq. (50). On the other hand, a nonempty increasing subsequence in a skew sum of permutations  $\pi_1 \ominus \dots \ominus \pi_r$  is an increasing subsequence in either  $\pi_1, \dots$ , or  $\pi_r$ . Therefore, elements of  $S_\ominus$  marked with a nonempty increasing subsequence correspond to sequences of at least two elements, with exactly one element counted by  $zu + S_\oplus - S_\oplus(z, 0)$  (either  $(\bullet, \{\bullet\})$  or a  $\oplus$ -decomposable permutation with a nonempty marked increasing subsequence) and other elements counted by  $S$ . We need to add a term  $S_\ominus(z, 0) = \frac{S^2}{1 - S}$  for the case of an empty marked increasing subsequence. Substituting Eq. (50) and using  $S_\oplus(z, 0) = S(z) - z$  gives Eq. (49).

Fix  $u \in \mathbb{C}$ . In order to perform saddle-point analysis with  $S_\ominus$ , we rewrite the first equation of the previous system as  $S_\ominus = G(z, S_\ominus, u)$  where

$$(51) \quad G(z, c, u) = \frac{S^2(z)}{1 - S(z)} + \left( \frac{(c + z + zu)^2}{1 - (c + z + zu)} + zu + z - S(z) \right) \times \left( \frac{1}{(1 - S(z))^2} - 1 \right)$$

Again this almost fits the settings of the smooth implicit-function schema, only the nonnegativity of the coefficients of  $G$  is not verified here. And, as we shall see, sufficient conditions for the validity of [FS09, Thm. VII.3 p.468] similar to the cograph case are satisfied.

The bivariate series  $(z, c) \mapsto G(z, c, u)$  is analytic on  $\{(z, c) : |z| < \rho, |c + z + zu| < 1\}$ .

Moreover the characteristic system,

$$(52) \quad G(r, s, u) = s, \quad G_c(r, s, u) = 1,$$

can be worked out and its solutions satisfy either

$$(r_1, s_1) = \left( \frac{1}{1 + u} (2S(r_1) - 2\sqrt{2S(r_1) - S(r_1)^2} + 1), -2S(r_1) + \sqrt{2S(r_1) - S(r_1)^2} \right)$$

or

$$(r_2, s_2) = \left( \frac{1}{1 + u} (2S(r_2) + 2\sqrt{2S(r_2) - S(r_2)^2} + 1), -2S(r_2) - \sqrt{2S(r_2) - S(r_2)^2} \right).$$

Since  $s_2 < 0$  when  $r_2 > 0$ , we focus on solutions of the first kind. We claim that, for any  $u > 0$ , there is a unique  $r_1$  in  $(0, \rho)$  satisfying

$$r_1 = \frac{1}{1+u} (2S(r_1) - 2\sqrt{2S(r_1) - S(r_1)^2} + 1).$$

Indeed, when  $r_1$  goes from 0 to  $\rho = 3 - 2\sqrt{2}$ , the quantity  $S(r_1)$  increases from 0 to  $S(\rho) = (2 - \sqrt{2})/2$  and the right-hand side decreases from  $1/(1+u)$  to  $(3 - 2\sqrt{2})/(1+u) < 3 - 2\sqrt{2}$ .

This proves that for  $u > 0$ , the characteristic system (52) has a unique positive solution  $(r(u), s(u))$ . Moreover, with  $y := S(r(u))$  we have

$$(53) \quad r(u) = \frac{1}{u+1} \left( 2y - 2\sqrt{2y - y^2} + 1 \right),$$

$$(54) \quad s(u) = -2y + \sqrt{2y - y^2},$$

$$(55) \quad u = \frac{2(1-y)\sqrt{2y - y^2} - 1}{2y^2 - y},$$

the equation for  $u$  being a consequence of the one for  $r(u)$  and (48) which gives  $r(u) = y - \frac{y^2}{1-y}$ .

One can show as in the cograph case, but comparing  $S_\ominus(\rho(u), u)$  with  $1 - \rho(u)(1+u)$  instead of  $+\infty$ , that  $S_\ominus(z, u)$  has radius of convergence  $r(u)$  and that its value at this singularity is given by  $S_\ominus(r(u), u) = s(u)$ . Again in an analogous way to the cograph case one can verify that the solutions  $(r, s) = (r(u), s(u))$  of the characteristic system (52) have analytic continuations in a neighborhood of every  $u > 0$ , noting that

$$G_z(r(u), s(u), u) = \frac{6y^2\sqrt{y(2-y)} - 4y^2 - 10y\sqrt{y(2-y)} + 9y + 2\sqrt{y(2-y)} - 2}{y(y-1)(y-2)(2y-1)(2y^2-4y+1)}$$

is positive on the interval  $(0, S(\rho))$ .

Therefore since  $S_\ominus(z, u)$  is aperiodic we can apply [Drm94, Thm.3] to prove Eq. (3) of Theorem 1.8 and we obtain

$$E_\beta = \frac{1}{(3 + 2\sqrt{2})r(u(\beta))u(\beta)^\beta},$$

where  $u(\beta)$  is the inverse function of Eq. (44) (or rather, its permutation counterpart, with the  $G$ ,  $r$  and  $s$  defined in the current section). A complicated expression of  $\beta$  in terms of  $y$  is given in the attached notebook. This can be numerically inverted to get  $y$  in terms of  $\beta$  and thus to compute  $E_\beta$  through Eqs. (53) to (55). One can also perform Taylor expansions around  $\beta = 0$ . It is shown in the notebook that

$$y = 1 - \frac{\sqrt{2}}{2} + \sqrt{\beta}\sqrt{\frac{3}{4}\sqrt{2} - 1} + o(\sqrt{\beta}).$$

From there, a routine computation gives

$$r(u(\beta)) = 3 - 2\sqrt{2} + O(\beta), \quad u(\beta) = \lambda\beta + O(\beta^2),$$

for some explicit constant  $\lambda$ . We finally find

$$E_\beta = 1 + \beta|\log(\beta)| + o(\beta\log(\beta)),$$

as claimed in Theorem 1.8. This implies that there exists  $\beta_1 > 0$  (numerically estimated at  $\beta_1 \approx 0.5827$ ) such that for every  $\beta < \beta_1$  we have  $E_\beta > 1$ . Theorem 1.8 is proved.

#### ACKNOWLEDGEMENTS

The authors are grateful to Marc Noy for stimulating discussions, in particular for bringing to their attention the problem of the maximum size of an independent set in a random cograph and the related literature around the probabilistic version of the Erdős-Hajnal conjecture.

Thanks are also due to an anonymous referee for pointing out the existence of [Wür12], and to Marc Noy and Carlos Hoppen for clarifying the status of the result announced by A. Würfl in [Wür12, Chapter 9].

#### REFERENCES

- [Ald94] D. Aldous. Recursive Self-Similarity for Random Trees, Random Triangulations and Brownian Excursion. *Ann. Probab.* 22(2): 527–545, 1994.
- [BBF+20] F. Bassino, M. Bouvel, V. Féray, L. Gerin, M. Maazoun, A. Pierrot. Universal limits of substitution-closed permutation classes, *J. Eur. Math. Soc.*, vol. 22 (11), pp. 3565–3639, 2020.
- [BBF+22a] F. Bassino, M. Bouvel, V. Féray, L. Gerin, M. Maazoun, A. Pierrot. Scaling limits of permutation classes with a finite specification: a dichotomy. *Advances in Mathematics*, 405: Article 108513, 2022.
- [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+22b] F. Bassino, M. Bouvel, V. Féray, L. Gerin, M. Maazoun, and A. Pierrot. Random cographs: Brownian graphon limit and asymptotic degree distribution. *Random Struct. Algorithms*, 60 (2): 166–200, 2022.
- [BM17] B. Bhattacharya and S. Mukherjee. Degree sequence of random permutation graphs. *Ann. Appl. Probab.*, 27(1): 439–484, 2017.
- [BBFS19] J. Borga, M. Bouvel, V. Féray and B. Stuffer. A decorated tree approach to random permutations in substitution-closed classes. *Electron. J. Probab.*, 25 (67): 1–52, 2020.
- [BBL98] P. Bose, J. Buss and A. Lubiw. Pattern matching for permutations. *Inf. Process. Lett.*, 65: 277–283, 1998.
- [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.
- [Chu14] M. Chudnovsky. The Erdős-Hajnal conjecture – a survey. *J. Graph Theory*, 75(2):178–190, 2014.
- [CLS81] D. G. Corneil, H. Lerchs, L. S. Burlingham. Complement reducible graphs. *Discrete Appl. Math.*, 3(3): 163–174, 1981.
- [Drm94] M. Drmota. Asymptotic distributions and a multivariate Darboux method in enumeration problems. *J. Comb. Theory Ser. A* 67: 169–184, 1994.
- [Drm09] M. Drmota. *Random Trees*. Springer, 2009.
- [DRRR20] M. Drmota, L. Ramos, C. Requilé, and J. Rué. Maximal independent sets and maximal matchings in series-parallel and related graph classes. *Electron. J. Comb.*, 27: P1.5, 2020.
- [FS09] Ph. Flajolet, R. Sedgewick. *Analytic combinatorics*. Cambridge University Press, 2009.
- [GGKK15] R. Glebov, A. Grzesik, T. Klímsová, D. Král’, Finitely forcible graphons and permutations, *J. Combin. Theory Ser. B* vol. 110, p.112–135, 2015.
- [HHP19] J. Hladký, P. Hu, and D. Piguet. Komlós’s tiling theorem via graphon covers, *J. Graph Theory* 90: 24–45, 2019.
- [HP05] M. Habib, C. Paul. A simple linear time algorithm for cograph recognition. *Discrete Appl. Math.*, 145(2): 183–197, 2005.
- [HR20] J. Hladký and I. Rocha. Independent sets, cliques, and colorings in graphons. *Eur. J. Comb.*, 88: 103108, 2020.

- [HKMRS13] C. Hoppen, Y. Kohayakawa, C. G. Moreira, B. Rath, R. M. Sampaio. Limits of permutation sequences. *J. Combin. Theory Ser. B*, vol. 103 (2013) n.1, p.93–113.
- [Kit11] S. Kitaev. *Patterns in permutations and words*. Springer (2011).
- [LRSTT10] M. Loeb, B. Reed, A. Scott, A. Thomason, S. Thomassé. Almost all  $H$ -free graphs have the Erdős–Hajnal property. In *An Irregular Mind, Szemerédi Is 70, Bolyai Soc. Math. Stud.*, 21: 405–414, 2010.
- [KMRS14] R. Kang, C. McDiarmid, B. Reed, and A. Scott. For most graphs  $H$ , most  $H$ -free graphs have a linear homogeneous set. *Random Struct. Algorithms*, 45(3):343–361, 2014.
- [Lov12] L. Lovász. *Large networks and graph limits*. Volume 60 of American Mathematical Society Colloquium Publications, 2012.
- [Maa20] M. Maazoun, On the Brownian separable permuton. *Comb. Probab. Comput.*, 29(2):241–266, 2020.
- [MRRY20] T. Mansour, R. Rastegar, A. Roitershtein, G. Yildirim. The longest increasing subsequence in involutions avoiding 3412 and another pattern. Preprint arXiv:2001:10030, 2020.
- [PC19] G. Peyré, M. Cuturi. Computational Optimal Transport. *Foundations and Trends in Machine Learning*: 11 (5-6), 355–607, 2019.
- [Rom15] D. Romik. *The Surprising Mathematics of Longest Increasing Subsequences*. Cambridge University Press (2015).
- [Sah20] S. Sah. Diagonal Ramsey via effective quasirandomness. Preprint arXiv:2005.0925, 2020.
- [Sei74] S. Seinsche. On a property of the class of  $n$ -colorable graphs. *J. Comb. Theory Ser. B*, 16(2): 191–193, 1974.
- [Spe75] J. Spencer Ramsey’s theorem – a new lower bound. *J. Comb. Theory Ser. A*, 18: 108–115, 1975.
- [Stu21] B. Stuffer. Graphon convergence of random cographs. *Random Struct. Algorithms*, 59: 464–491, 2021.
- [Vat16] V. Vatter. Permutation classes. In M. Bóna, editor, *Handbook of Enumerative Combinatorics*, Discrete Mathematics and its Applications. CRC Press, 2016.
- [Vil08] C. Villani. *Optimal transport: old and new*. Springer (2008).
- [Wür12] A. Würfl. *Spanning subgraphs of growing degree; A generalised version of the blow-up lemma and its applications*. PhD thesis, Technical University of Munich. Available at <https://mediatum.ub.tum.de/doc/1126106/1126106.pdf>.