

## Abstract

We consider a finite relational vocabulary  $\sigma$  and a first order theory  $T$  for  $\sigma$  composed of symmetry and anti-reflexivity axioms. We define a binomial random model of finite  $\sigma$ -structures that satisfy  $T$  and show that first order properties have well defined asymptotic probabilities in the sparse case. It is also shown that those limit probabilities are well-behaved with respect to some parameters that represent edge densities. An application of these results to the problem of random Boolean satisfiability is presented afterwards. We show that there is no first order property of  $k$ -CNF formulas that implies unsatisfiability and holds for almost all typical unsatisfiable formulas when the number of clauses is linear.

# Introduction

Since the work of Erdős and R enyi on the evolution of random graphs [1] the study of the asymptotic properties of random structures has played a relevant role in combinatorics and computer science. A central theme in this topic is, given a succession  $(G_n)_n$  of random structures of some sort and a property  $P$ , to determine the limit probability that  $G_n$  satisfies  $P$  or to determine whether that limit exists.

One approach that has proven to be useful is to classify the properties  $P$  according to the logical languages they can be defined in. We say that the succession  $(G_n)_n$  obeys a convergence law with respect to some logical language  $\mathcal{L}$  if for any given property  $P$  expressible in  $\mathcal{L}$  the probability that  $G_n$  satisfies  $P$  tends to some limit as  $n$  grows to infinity. We say that  $(G_n)_n$  obeys a zero-one law with respect to  $\mathcal{L}$  if that limit is always either zero or one. The seminal theorem on this topic, due to Fagin [2] and Glebskii et al. [3] independently, states that if  $G_n$  denotes a labeled graph with  $n$  vertices picked uniformly at random among all  $2^{\binom{n}{2}}$  possible then  $(G_n)_n$  satisfies a zero-one law with respect to the first order (FO) language of graphs.

Originally this result was proven in the broader context of relational structures but it was in the theory of random graphs where the study of other zero-one and convergence laws became more prominent. In particular, the asymptotic behavior of FO logic in the binomial model of random graphs  $G(n, p)$  has been extensively studied. In this model, introduced by Gilbert [4], a random graph is obtained from  $n$  labeled vertices by adding each possible edge with probability  $p$  independently. When  $p = 1/2$  this distribution of random graphs coincides with the uniform one, mentioned above. In general, for the case where  $p$  is a constant probability a slight generalization of the proofs in [2] and [3] works and  $G(n, p)$  satisfies a zero-one law for FO logic. If we consider  $p(n)$  a decreasing function of the form  $n^{-\alpha}$  we can ask the question of what are the values of  $\alpha$  for which  $G(n, p(n))$  obeys a zero-one or a convergence law for FO logic. In [5] Shelah and Spencer gave a complete answer for the range  $\alpha \in (0, 1)$ . Among other results, they proved that if  $\alpha$  is an irrational number in this interval then  $G(n, p(n))$  obeys a zero-one law for FO logic, while if  $\alpha$  is a rational number in the same range then  $G(n, p(n))$  does not even satisfy a convergence law for FO logic. The case  $\alpha = 1$  was later solved by Lynch in [6]. A weaker form of the main theorem in that article states the following:

**Theorem 0.1.** *For any FO sentence  $\phi$ , the function  $F_\phi : (0, \infty) \rightarrow [0, 1]$  given by*

$$F_\phi(\beta) = \lim_{n \rightarrow \infty} \Pr(G(n, \beta/n) \text{ satisfies } \phi)$$

*is well defined and analytic. In particular, for any  $\beta \geq 0$  the model  $G(n, \beta/n)$  obeys a convergence law for FO logic.*

The analyticity of these asymptotic probabilities with respect to the parameter  $\beta$  imply that FO properties cannot "capture" sudden changes that occur in the random graph  $G(n, \beta/n)$  as  $\beta$  changes. Given  $p(n)$  a probability,  $P$  a property of graphs, and  $Q$  a sufficient condition for  $P$  - i.e., a property that implies  $P$  -, we say that  $Q$  explains  $P$  if  $G(n, p(n))$  satisfies the converse implication  $P \implies Q$  asymptotically almost surely (a.a.s.). A notable example of this phenomenon happens in the range  $p(n) = \log(n)/n + \beta/n$  with  $\beta$  constant. Erdős and R enyi [1] showed that for probabilities of this form  $G(n, p(n))$  a.a.s. is disconnected only if it contains an isolated vertex. An observation by Albert Atserias is the following:

**Theorem 0.2.** *Let  $c$  be a real constant such that  $\lim_{n \rightarrow \infty} \Pr(G(n, c/n) \text{ is not 3-colorable}) > 0$ . Then there is no FO graph property that explains non-3-colorability for  $G(n, c/n)$ .*

The short proof of this theorem is as follows: It is a known fact that there are positive constants  $c_0 \leq c_1$  such that  $G(n, c/n)$  is a.a.s 3-colorable if  $c < c_0$  and it is a.a.s non 3-colorable if  $c > c_1$  REFERENCES NEEDED. Suppose that  $P$  is a FO graph property that implies non-3-colorability. Then, because of this implication, for all values of  $c$

$$\lim_{n \rightarrow \infty} \Pr(G(n, c/n) \text{ satisfies } P) \leq \lim_{n \rightarrow \infty} \Pr(G(n, c/n) \text{ is not 3-colorable}).$$

In consequence the asymptotic probability that  $G(n, c/n)$  satisfies  $P$  is zero when  $c < c_0$ . By Lynch's theorem, if  $P$  is definable in FO logic then this asymptotic probability varies analytically with  $c$ . Using the fact that any analytic function that takes value zero in a non-empty interval must equal zero everywhere, we obtain that  $G(n, c/n)$  a.a.s does not satisfy  $P$  for any value of  $c$ . As a consequence the theorem follows.

The aim of this work is to extend Lynch's result to arbitrary relational structures where the relations are subject to some predetermined symmetry and anti-reflexivity axioms. This was originally motivated by an application to the study of random  $k$ -CNF formulas. Since [7] it is known that for each  $k$  there are constants  $c_0, c_1$  such that a random  $k$ -CNF formula with  $cn$  clauses over  $n$  variables

## 1 Preliminaries

### 1.1 General notation

Given a positive natural number  $n$ , we will write  $[n]$  to denote the set  $1, 2, \dots, n$ .

Given a set  $S$  and a natural number  $k \in \mathbb{N}$  we will use  $\binom{S}{k}$  to denote the set of subsets of  $S$  whose size is  $k$ .

Let  $S$  be a set,  $a$  a positive natural number, and  $\Phi$  a group of permutations over  $[a]$ . Then  $\Phi$  acts naturally over  $S^a$  in the following way: Given  $g \in \Phi$  and  $(x_1, \dots, x_a)$  we define  $g(x_1, \dots, x_a) = (x_{g(1)}, \dots, x_{g(a)})$ . We will denote by  $S^a/\Phi$  the quotient of the set  $S^a$  by this action. Given an element  $(x_1, \dots, x_a) \in S^a$  we will denote its equivalence class in  $S^a/\Phi$  by  $[x_1, \dots, x_a]$ . Thus, for any  $g \in \Phi$ , by definition  $[x_1, \dots, x_a] = [x_{g(1)}, \dots, x_{g(a)}]$ . The notation  $(x_1, \dots, x_a)$  will be reserved to ordered tuples while  $[x_1, \dots, x_a]$  will denote an ordered tuple modulo the action of some arbitrary group of permutations. Which group is this will depend on the ambient set where  $[x_1, \dots, x_a]$  belongs and it should either be clear from context or not be relevant.

We will denote ordered lists of elements by  $\bar{x} := x_1, \dots, x_a$ . This way, expressions like  $(\bar{x})$  or  $[\bar{x}]$  would mean  $(x_1, \dots, x_a)$  and  $[x_1, \dots, x_a]$  respectively. Sometimes we will directly write  $\bar{x}$  without specifying the list it names nor its length when it is understood or not relevant.

Given two real functions over the natural numbers  $f, g : \mathbb{N} \rightarrow \mathbb{R}$  we will write  $f = O(g)$  to mean that there exists some constant  $C \in \mathbb{R}$  such that  $f(n) \leq Cg(n)$  for  $n$  sufficiently large, as usual. If  $g(n) \neq 0$  for sufficiently large values of  $n$  then we will write  $f \simeq g$  if  $\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 1$ .

### 1.2 Logical preliminaries

We will assume a certain degree of familiarity with the concepts. For a more complete exposition of the topics presented here one can consult [8].

A relational vocabulary  $\sigma$  is a collection of relation symbols  $(R_1, \dots, R_m, \dots)$  where each relation symbol  $R_i$  has associated a natural  $a_i$  number called its arity. A  $\sigma$ -structure  $\mathfrak{A}$  is composed of a set  $A$ , called the universe of  $\mathfrak{A}$ , equipped with relations  $R_1^{\mathfrak{A}} \subseteq A^{a_1}, \dots, R_m^{\mathfrak{A}} \subseteq A^{a_m}$ . When  $\sigma$  is understood we may refer to  $\sigma$ -structures as relational structures or simply as structures. A structure is called finite if its universe is a finite set.

In the first order language  $FO[\sigma]$  with signature  $\sigma$  formulas are formed by variables  $x_1, \dots, x_i, \dots$ , the relation symbols in  $\sigma$ , the equal symbol  $=$ , the usual Boolean connectives  $\neg, \wedge, \vee, \dots$ , the existential and universal quantifiers  $\exists, \forall$ , and the parentheses  $), ($ . Then formulas in  $FO[\sigma]$  are defined as follows.

- The expression  $R(y_1, \dots, y_a)$ , where the  $y_i$ 's are variables and  $R$  is a relation symbol in  $\sigma$  with arity  $a$ , belongs to  $FO[\sigma]$ .
- The expression  $y_1 = y_2$ , where  $y_1, y_2$  are variables, belongs to  $FO[\sigma]$ .
- Given formulas  $\phi, \psi \in FO[\sigma]$ , any Boolean combination of them  $\neg(\phi), (\phi \wedge \psi), (\phi \vee \psi), \dots$  belongs to  $FO[\sigma]$  as well.
- Given a formula  $\phi \in FO[\sigma]$  and  $x$  a variable that does not appear bounded by a quantifier in  $\phi$ , the expressions  $\forall x(\phi)$  and  $\exists x(\phi)$  belong both to  $FO[\sigma]$ .

We will write  $\forall y_1, y_2, \dots, y_m$  or simply  $\forall \bar{y}$  instead of  $\forall y_1, \forall y_2, \dots, \forall y_m$  and likewise for the quantifier  $\exists$ .

We define the set of free variables of a formula as usual. We will use the notation  $\phi(\bar{y})$  to refer to a formula  $\phi \in FO[\sigma]$  to denote that its free variables are the ones in  $\bar{y}$ . Formulas with no free variables are called sentences.

The quantifier rank of a formula  $\phi$ , denoted by  $qr(\phi)$ , is defined as the maximum number of nested quantifiers in  $\phi$ .

Sentences in  $FO[\sigma]$  are interpreted over  $\sigma$ -structures in the natural way. Given an structure  $\mathcal{A}$ , and a sentence  $\phi \in FO[\sigma]$  we write  $\mathcal{A} \models \phi$  to denote that  $\mathcal{A}$  satisfies  $\phi$ . If  $\psi(\bar{y})$  is a formula,  $\bar{a}$  are elements in the universe of  $\mathcal{A}$ , and  $\bar{y}$  and  $\bar{a}$  are lists of the same size, then we write  $\mathcal{A} \models \psi(\bar{a})$  to mean that  $\mathcal{A}$  satisfies  $\psi$  when the free variables in  $\bar{y}$  are interpreted as the elements in  $\bar{a}$ .

Deberá al menos mencionar un par de trabajos que estudien random k-SAT y propiedades a.a.s suficientes para no-satisfacibilidad

### 1.3 Structures as multi-hypergraphs

For the rest of the article consider fixed:

- Positive natural numbers  $t, \bar{a} = a_1, \dots, a_t$ , with all the  $a_i$ 's greater than 1.
- A relational vocabulary  $\sigma = \{\bar{R}\}$ , with  $\bar{R} = R_1, \dots, R_t$  such that  $a_i$  is the arity of  $R_i$ .
- Groups  $\bar{\Phi} = \Phi_1, \dots, \Phi_t$  such that each  $\Phi_i$  consists of permutations on  $[a_i]$  with the usual composition as its operation.
- Sets  $\bar{P} = P_1, \dots, P_t$  satisfying  $P_i \subseteq \binom{[a_i]}{2}$

We will only consider relational structures where the relations are of arity at least two. This restriction is not necessary, but it makes notation easier.

Quizás debería añadir un anexo dando alguna indicación sobre cómo tratar las relaciones unarias?

We can think of structures in  $\mathcal{C}_{\bar{\Phi}, \bar{P}}^\sigma$  as "multi-hypergraphs" with  $t$  edge sets whose edges are of sizes  $\bar{a}$  respectively, are invariant under permutations in  $\bar{\Phi}$  resp., and do not contain repetitions of vertices in the positions given by  $\bar{P}$  resp. We make this observation formal in the following definitions:

We define the class  $\mathcal{C}$  as the class of  $\sigma$ -structures that satisfy the following axioms:

- *Symmetry axioms*: For each  $1 \leq s \leq t$  and each  $g \in \Phi_s$ :

$$\forall x_1, \dots, x_{a_s} (R_s(x_1, \dots, x_{a_s}) \iff R_s(x_{g(1)}, \dots, x_{g(a_s)}))$$

- *Anti-reflexivity axioms*: For each  $1 \leq s \leq t$  and  $\{i, j\} \in P_s$

$$\forall x_1, \dots, x_{a_s} ((x_i = x_j) \implies \neg R_s(x_1, \dots, x_{a_s}))$$

We can think of structures in  $\mathcal{C}_{\bar{\Phi}, \bar{P}}^\sigma$  as "multi-hypergraphs" with  $t$  edge sets whose edges are of sizes  $\bar{a}$  respectively, are invariant under permutations in  $\bar{\Phi}$  resp., and do not contain repetitions of vertices in the positions given by  $\bar{P}$  resp. We make this observation formal in the following definitions:

**Definition 1.1.** Let  $V$  be a set,  $a$  be a positive natural number,  $\Phi$  be a group of permutations over  $[a]$  and  $P \subseteq \binom{[a]}{2}$ . We define the **total edge set over  $V$  with edge size  $a$ , symmetry group  $\Phi$  and anti-reflexivity restrictions  $P$**  as the set

$$E_{a, \Phi, P}^V = (V^a / \Phi) \setminus \{[x_1, \dots, x_a] \mid x_1, \dots, x_a \in V \wedge x_i = x_j \text{ for some } \{i, j\} \in P\}.$$

That is,  $E_{a, \Phi, P}^V$  contains all the "tuples modulo the permutations in  $\Phi$ " excluding those that contain some repetition of vertices in the positions given by  $P$ .

**Definition 1.2.** A **multi-hypergraph with  $t$  edge sets, edge sizes given by  $\bar{a}$ , symmetry groups  $\bar{\Phi}$ , and anti-reflexivity restrictions  $\bar{P}$**  is a pair  $G = (V(G), \bar{E}(G))$ , where  $\bar{E}(G) = E_1(G), \dots, E_t(G)$  and for each  $i$ ,  $E_i(G) \subseteq E_{a_i, \Phi_i, P_i}^V$ .

For the sake of word economy the expression "multi-hypergraph with  $t$  edge sets, with edge sizes given by  $\bar{a}$ , symmetry groups  $\bar{\Phi}$ , and anti-reflexivity restrictions  $\bar{P}$ " will be replaced simply by "hypergraph". The word "hypergraph" will not hold any other meaning than this for the rest of this writing except for the places where it is explicitly stated.

Hypergraphs, as we have defined them, can be naturally interpreted as structures from  $\mathcal{C}_{\bar{\Phi}, \bar{P}}^\sigma$  in the following way: given  $G = (V, \bar{E})$ , we consider  $V$  to be the universe of  $G$ , and for any  $i$  we define  $R_i^G \subseteq V^{a_i}$  as the set such that  $(\bar{x}) \in V^{a_i}$ ,  $(\bar{x}) \in R_i^G$  if and only if  $[\bar{x}] \in E_i$ . Under this interpretation hypergraphs, by definition, satisfy the symmetry and anti-reflexivity axioms given above. It is also easy to see that this interpretation induces a one-to-one identification between structures in  $\mathcal{C}_{\bar{\Phi}, \bar{P}}^\sigma$  and hypergraphs.

We will use standard nomenclature and notation from graph theory. This way, we will call vertex set to  $V(G)$  and vertices to its elements. Likewise, each of the  $E_i(G)$ 's will be called an edge set and its elements, edges. Given an edge set  $E_i(G)$ , the index  $i$  will be called its color, and the number  $a_i$  its size. Thus, we will say that an edge  $e \in E_i(G)$  has color  $i$  and size  $a_i$ .

Given a set of vertices  $U \subseteq V(G)$ , we will denote by  $G[U]$  the hypergraph induced by  $G$  on  $U$ . That is,  $G[U]$  is an hypergraph  $H = (V(H), E(H))$  such that  $V(H) = U$  and for any list  $\bar{v}$  of vertices in  $U$ ,  $[\bar{v}] \in E_i(H)$  if and only if  $[\bar{v}] \in E_i(G)$ .

An homomorphism between two hypergraphs  $G$  and  $H$  a map  $f : V(G) \rightarrow V(H)$  that sends edges from  $G$  to edges in  $H$  of the same color. That is, if vertices  $v_1, \dots, v_{a_i}$  form an edge  $[v_1, \dots, v_{a_i}] \in E_i(G)$ , then  $[f(v_1), \dots, f(v_{a_i})] \in E_i(H)$ . If  $f$  is injective then it is called a monomorphism. If  $f$  is bijective and its inverse is also an homomorphism between  $H$  and  $G$  then  $f$  is called an isomorphism.

The group of automorphisms  $Aut(G)$  of an hypergraph  $G$  is the group of isomorphisms between  $G$  and itself.

Given two hypergraphs  $G$  and  $H$ , a copy of  $H$  in  $G$  is a sub-hypergraph  $H_2 \subseteq G$  isomorphic to  $H$ . The copy is called induced if  $H_2$  is an induced sub-hypergraph. We will call a labeled copy of  $H$  in  $G$  to a monomorphism  $f : H \rightarrow G$ . It is satisfied that the number of labeled copies of  $H$  in  $G$  is  $|Aut(H)|$  times the number of copies of  $H$  in  $G$ .

The excess  $ex(G)$  of an hypergraph  $G$  is the number

$$ex(G) = \left( \sum_{i=1}^t (a_i - 1) |E_i(G)| \right) - |V(G)|.$$

That is, the excess of  $G$  is its "weighted number of edges" minus its number of vertices. Here the "weighted number of edges" in  $G$  the sum of the weight of all its edges, where the weight of an edge is its size minus one.

Before moving on we need to introduce some additional notation. NOTACION

Given an hypergraph  $G$  we define the following metric,  $d$ , over  $V(G)$ :

$$d^G(v, u) = \min_{\substack{H \text{ subgraph of } G \\ H \text{ connected} \\ v, u \in V(H)}} |V(H)| - 1.$$

That is, the distance between  $v$  and  $u$  is the minimum size of a connected graph  $H$  containing both vertices, minus one. If such graph does not exist we define  $d^G(u, v) = \infty$ . This definition extends naturally to subsets  $X, Y \subseteq V(G)$ :

$$d^G(X, Y) = \min_{\substack{x \in X \\ y \in Y}} d^G(x, y).$$

As usual, when  $X = \{x\}$  we will omit the brackets and write  $d^G(x, Y)$  instead of  $d^G(\{x\}, Y)$ , for example. When  $G$  is understood or not relevant we will usually simply denote the distance by  $d$  instead of  $d^G$ .

Given set of vertices vertex,  $X \subseteq V(G)$ , we denote by  $N^G(X; r)$  the  $r$ -neighborhood of  $X$  in  $G$ . That is,  $N^G(X; r) = G[Y]$ , where  $Y \subseteq V(G)$  is the set:

$$Y := \{u \in V(G) \mid d(X, u) \leq r\}.$$

In particular, when  $X$  is a singleton  $\{v\}$ , we will write  $N^G(v; r)$  instead of  $N^G(\{v\}; r)$ . As before, we will usually drop the " $G$ " from our notation when  $G$  is understood or not relevant.

## 1.4 The random model

Let  $p_1, \dots, p_t$  real numbers between zero and one, and let  $\bar{p} = p_1, \dots, p_t$ . The random model  $G^\mathcal{C}(n, \bar{p})$  is the discrete probability space that assigns to each hypergraph  $G$  whose vertex set  $V(G)$  is  $[n]$  the following probability:

$$\Pr(G) = \prod_{i=1}^t p_i^{|E_i(G)|} (1 - p_i)^{|E_{a_i, \Phi_i, P_i}^{[n]}| - |E_i(G)|}.$$

Equivalently, this is the probability space obtained by assigning to each edge with color  $i$ ,  $e \in E_{a_i, \Phi_i, P_i}^{[n]}$  probability  $p_i$  independently.

As in the case of Lynch's theorem, we are interested in the "sparse regime" of  $G^\mathcal{C}(n, \bar{p})$ , where the expected number of edges each color is linear. This is achieved when each of the  $p_i$ 's are of the form  $\beta_i/n^{a_i-1}$  for some non-negative real numbers  $\beta_1, \dots, \beta_t$ . Let  $\bar{\beta} := \beta_1, \dots, \beta_t$ . We will write  $\bar{p}(n, \bar{\beta})$  to denote the list  $\beta_1/n^{a_1-1}, \dots, \beta_t/n^{a_t-1}$ . We will treat the parameters  $\beta_1, \dots, \beta_t$  as fixed real constants for the most part and will abbreviate  $\bar{p}(n, \bar{\beta})$  as  $\bar{p}(n)$ .

Our goal is to prove the following theorem:

**Theorem 1.1.** *Let  $\phi$  be a sentence in  $FO[\sigma]$ . Then the function  $F_\phi : [0, \infty)^t \rightarrow \mathbb{R}$  given by*

$$\bar{\beta} \mapsto \lim_{n \rightarrow \infty} \Pr(G^\mathcal{C}(n, \bar{p}(n, \bar{\beta})) \models \phi)$$

*is well defined and analytic.*

## 1.5 Ehrenfeucht-Fraisse Games

Let  $G_1$  and  $G_2$  be hypergraphs. We define the  $k$  round Ehrenfeucht-Fraisse game on  $G_1$  and  $G_2$ , denoted by  $\text{EHR}_k(G_1, G_2)$ , as follows: The game is played between two players, Spoiler and Duplicator, and the number of rounds,  $k$ , is known for both from the start. At the beginning of each round Spoiler chooses a vertex from either  $V(G_1)$  or  $V(G_2)$  and Duplicator responds by choosing a vertex from the other set. Let us denote by  $v_i$ , resp.  $u_i$  the vertex from  $G_1$ , resp. from  $G_2$ , chosen in the  $i$ -th round, for  $i \in [k]$ . At the end of the  $k$ -th round Duplicator wins if the following holds:

- For any  $i, j \in [k]$ ,  $v_i = v_j \iff u_i = u_j$ .
- Given indices  $i_1, \dots, i_a \in [k]$ , and a color  $c \in [t]$ ,  $[v_{i_1}, \dots, v_{i_a}] \in E_c(G_1) \iff [u_{i_1}, \dots, u_{i_a}] \in E_c(G_2)$ .

We define the equivalence relation  $=_k$  between hypergraphs as follows: We say that  $G_1 =_k G_2$  if for any sentence  $\phi \in FO[\sigma]$  with  $qr(\phi) \leq k$  then  $G_1 \models \phi$  if and only if  $G_2 \models \phi$ .

The following is satisfied:

**Theorem 1.2** (Ehrenfeut, 9). *Let  $G_1$  and  $G_2$  be hypergraphs. Then Duplicator wins  $\text{EHR}_k(G_1, G_2)$  if and only if  $G_1 =_k G_2$ .*

Now consider  $\bar{v}$ , and  $\bar{u}$  lists of vertices of the same length,  $l$ , from  $G_1$  and  $G_2$  respectively. We define the  $k$  round Ehrenfeucht-Fraisse game on  $G_1$  and  $G_2$  with initial position given by  $\bar{v}$

and  $\bar{u}$ , denoted by  $\text{EHR}_k(G_1, \bar{v}, G_2, \bar{u})$ , the same way as  $\text{EHR}_k(G_1, G_2)$ , but in this case the game has  $l$  extra rounds at the beginning where the vertices in  $\bar{v}$  and  $\bar{u}$  are played successively. After this,  $k$  more rounds are played normally.

We also define the  $k$ -round distance Ehrenfeucht-Fraïssé game on  $G_1$  and  $G_2$ , denoted by  $d\text{EHR}_k(G_1, G_2)$ , the same way as  $\text{EHR}_k(G_1, G_2)$ , but now, in order for Duplicator to win the game, the following additional condition has to be satisfied at the end of the  $k$ -th round:

- For any  $i, j \in [k]$ ,  $d^{G_1}(v_i, v_j) = d^{G_2}(u_i, u_j)$ .

Given  $\bar{v}$ , and  $\bar{u}$  lists of vertices of the same length, from  $G_1$  and  $G_2$  respectively we define the game  $d\text{EHR}_k(G_1, \bar{v}, G_2, \bar{u})$  analogously to  $\text{EHR}_k(G_1, \bar{v}, G_2, \bar{u})$ .

## 1.6 Outline of the proof

We show now an outline of the proof.

We show that for any quantifier rank  $k$  there are some classes of hypergraphs  $C_1^k, \dots, C_{n_k}^k$  such that

- (1) a.a.s the rank  $k$  type of any two graphs in the same class coincide,
- (2) a.a.s. any random graph belongs to some of them, and
- (3) the limit probability of random graph belonging to any of them is an analytic expression on the parameters  $\underline{\beta}$ .

After this is archived the theorem follows easily.

The objective of next sections will be to define the classes  $C_1, \dots, C_{n_k}$  and to show that they satisfy properties (1), (2) and (3).

Explicar esto mejor

## 1.7 Some winning strategies for Duplicator

The aim of this section is to show the winning strategy for Duplicator that is going to be used in our proofs.

Let  $G_1$  and  $G_2$  be hypergraphs, and let  $\bar{v} \subseteq V(G_1), \bar{u} \subseteq V(G_2)$  be lists of vertices of the same size. We say that  $N(\bar{v}; r)$  and  $N(\bar{u}; r)$  are  $k$ -similar, or that  $\bar{v}$  and  $\bar{u}$  have  $k$ -similar  $r$ -neighborhoods, if Duplicator wins  $d\text{EHR}_k(N(\bar{v}; r), \bar{v}, N(\bar{u}; r), \bar{u})$ .

If  $X \subseteq V(G_1)$  and  $Y \subseteq V(G_2)$  are sets of vertices we say that  $X$  and  $Y$  have  $k$ -similar  $r$ -neighborhoods if we can order their vertices to form lists  $\bar{v}$ , resp.  $\bar{u}$  such that  $N(\bar{v}; r)$  and  $N(\bar{u}; r)$  are  $k$ -similar.

Now suppose that  $X \subseteq V(G_1)$  and  $Y \subseteq V(G_2)$  can be partitioned into lists  $X = \bar{v}_1 \cup \dots \cup \bar{v}_a$  and  $Y = \bar{u}_1 \cup \dots \cup \bar{u}_b$  such that  $N(\bar{v}_i; r)$ 's, and the  $N(\bar{u}_i; r)$ 's, are connected and disjoint. We say that  $N(X; r)$  and  $N(Y; r)$  are  $k$ -agreeable, or that they have  $k$ -agreeable neighborhoods, if any  $\bar{w}$  among the  $\bar{v}_i$ 's or among the  $\bar{u}_i$ 's satisfies:

- The number of  $\bar{v}_i$ 's and the number of  $\bar{u}_i$ 's satisfying that “ $\bar{v}_i$  (resp.  $\bar{u}_i$ ) and  $\bar{w}$  have  $k$ -similar  $r$ -neighborhoods” are the same or are both greater or equal than  $k$ .



The main theorem of this section, which is a slight strengthening of Theorem 2.6.7 from [10], is the following:

**Theorem 1.3.** *Set  $r = (3^k - 1)/2$ . Let  $G_1, G_2$  be hypergraphs, and suppose there exist sets  $X \subseteq V(G_1)$ ,  $Y \subseteq V(G_2)$  with the following properties:*

- (1)  $N(X; r)$  and  $N(Y; r)$  are  $k$ -agreeable.
- (2) Let  $r' \leq r$ . Let  $v \in V(G_1)$  such that  $d(v, X) > 2r' + 1$ , and let  $u_1, \dots, u_{k-1} \in V(G_2)$ . Then there exists a vertex  $u \in V(G_2)$  with  $u, v$  having  $k$ -similar  $r'$ -neighborhoods and satisfying  $d(u, u_i) > 2r' + 1$  for all  $u_i$ 's as well as  $d(u, Y) > 2r' + 1$ .
- (3) Let  $r' \leq r$ . Let  $u \in V(G_2)$  such that  $d(u, Y) > 2r' + 1$ , and let  $v_1, \dots, v_{k-1} \in V(G_1)$ . Then there exists a vertex  $v \in V(G_1)$  with  $v, u$  having  $k$ -similar  $r'$ -neighborhoods and satisfying  $d(v, v_i) > 2r' + 1$  for all  $v_i$ 's as well as  $d(v, X) > 2r' + 1$ .

Then Duplicator wins  $\text{EHR}_k(G_1, G_2)$ .

In order to prove this theorem we need to make two observations and prove a previous lemma.

**Observation 1.1.** *Let  $H_1, H_2$  be hypergraphs and  $\bar{v}, \bar{u}$ , be lists of vertices from  $V(H_1)$  and  $V(H_2)$  respectively. Suppose that Duplicator wins  $d\text{EHR}_k(H_1, \bar{v}, H_2, \bar{u})$ . Then, for any  $r$  Duplicator also wins  $d\text{EHR}_k(N(\bar{v}; r), \bar{v}, N(\bar{u}; r), \bar{u})$ . In particular, given hypergraphs  $G_1, G_2$  and sets  $X \subseteq V(G_1)$ ,  $Y \subseteq V(G_2)$  such that  $N(X; r)$  and  $N(Y; r)$  are  $k$ -similar, then for any  $r' \leq r$  the graphs  $N(X; r')$  and  $N(Y; r')$  are  $k$ -similar as well.*

**Observation 1.2.** *Let  $H_1, H_2$  be hypergraphs and  $\bar{v}, \bar{u}$ , be lists of vertices from  $V(H_1)$  and  $V(H_2)$  respectively. Suppose Duplicator wins  $d\text{EHR}_k(H_1, \bar{v}, H_2, \bar{u})$ . Let  $v' \in V(H_1), u' \in V(H_2)$  be vertices played in the first round of an instance of the game where Duplicator is following a winning strategy. Then Duplicator also wins  $d\text{EHR}_{k-1}(H_1, \bar{v}_2, H_2, \bar{u}_2)$ , where  $\bar{v}_2 := \bar{v}, v'$  and  $\bar{u}_2 := \bar{u}, u'$ .*

**Lemma 1.1.** *Let  $G_1, G_2$  be hypergraphs and  $\bar{v}, \bar{u}$ , be lists of vertices from  $V(G_1)$  and  $V(G_2)$  respectively. Let  $r$  be greater than zero. Suppose that  $N(\bar{v}; 3r + 1)$  and  $N(\bar{u}; 3r + 1)$  are  $k$ -similar. Let  $v' \in V(G_1), u' \in V(G_2)$  be vertices played in the first round of an instance of  $d\text{EHR}_k(N(\bar{v}; 3r + 1), \bar{v}, N(\bar{u}; 3r + 1), \bar{u})$  where Duplicator is following a winning strategy. Further suppose that  $d(\bar{v}, v_2) \leq 2r + 1$  (and in consequence  $d(\bar{u}, u_2) \leq 2r + 1$  as well). Let  $\bar{v}_2 := \bar{v}, v'$  and  $\bar{u}_2 := \bar{u}, u'$ . Then  $N(\bar{v}_2; r)$  and  $N(\bar{u}_2; r)$  are  $(k - 1)$ -similar*

*Proof.* Using observation 1.2 we get that Duplicator wins

$$d\text{EHR}_k(N^{G_1}(\bar{v}; 3r + 1), \bar{v}_2, N^{G_2}(\bar{u}; 3r + 1), \bar{u}_2)$$

as well. Call  $H_1 = N^{G_1}(\bar{v}; 3r + 1)$ ,  $H_2 = N^{G_2}(\bar{u}; 3r + 1)$ . Then by observation 1.2 Duplicator wins

$$d\text{EHR}_k(N^{H_1}(\bar{v}_2; r), \bar{v}_2, N^{H_2}(\bar{u}_2; r), \bar{u}_2).$$

Because of this if we prove  $N^{G_1}(\bar{v}_2; r) = N^{H_1}(\bar{v}_2; r)$  and  $N^{G_2}(\bar{u}_2; r) = N^{H_2}(\bar{u}_2; r)$ , then we are finished. Let  $z \in N^{G_1}(\bar{v}_2; r)$ . Then  $d(z, \bar{v}) \leq d(z, v') + d(v', \bar{v}) = 3r + 1$ . In consequence,  $N^{G_1}(\bar{v}_2; r) \subseteq H_1$ . Thus,  $N^{G_1}(\bar{v}_2; r) \subseteq H_1$ , and  $N^{G_1}(\bar{v}_2; r) = N^{H_1}(\bar{v}_2; r)$ . Analogously we obtain  $N^{G_2}(\bar{u}_2; r) = N^{H_2}(\bar{u}_2; r)$ , as we wanted.  $\square$

Now we are in conditions to prove theorem 1.3.

*Proof of theorem 1.3.* Define  $r_0 = 0$  and  $r_i = 3r_{i-1} + 1$  for  $i > 0$ . Let us denote by  $w_i$  and  $z_i$  the vertices played in  $G_1$  and  $G_2$  respectively during the  $i$ -th round of  $\text{EHR}_k(G_1, G_2)$ .

Let  $\bar{v}_1, \dots, \bar{v}_a$  and  $\bar{u}_1, \dots, \bar{u}_b$  be lists forming partitions of  $X$  and  $Y$  respectively, and assume they are as in the definition of  $k$ -agreeability. Set

$$\mathcal{X}[0] = \{\bar{v}_1, \dots, \bar{v}_a\}, \quad \mathcal{Y}[0] = \{\bar{u}_1, \dots, \bar{u}_b\}.$$

That is,  $\mathcal{X}[0]$  and  $\mathcal{Y}[0]$  are the whose elements are the  $\bar{v}_i$ 's and  $\bar{u}_i$ 's respectively. At the end of the  $s$ -th round  $\mathcal{X}[s-1]$ , resp.  $\mathcal{Y}[s-1]$ , will be updated into  $\mathcal{X}[s]$ , resp.  $\mathcal{Y}[s]$ , by performing on it some of the following operations: adding a new list to it, appending one vertex to an existing list, and marking a list with the index  $s$ . Duplicator will keep track of the sets  $\mathcal{X}[s]$  and  $\mathcal{Y}[s]$ .

We show first an strategy for Duplicator and will prove its correctness afterwards. The strategy is as follows: At the beginning of the  $s$ -th round suppose Spoiler plays  $w_s$  in  $G_1$ . The case where they play  $z_s$  in  $G_2$  is symmetric. Call  $r = r_{k-s}$ . There are three possibilities.

- Case 1: The vertex  $w_s$  satisfies  $d(w_s, \bar{v}) > 2r + 1$  for all  $\bar{v} \in \mathcal{X}[s-1]$ . Then Duplicator can find a vertex  $z_s$  in  $G_2$  such that  $d(z_s, \bar{u}) > 2r + 1$  for all  $\bar{u} \in \mathcal{Y}[s-1]$  satisfying that  $w_s$  and  $z_s$  have  $(k-s)$ -similar  $r$ -neighborhoods. To form  $\mathcal{X}[s]$  and  $\mathcal{Y}[s]$ , add to  $\mathcal{X}[s-1]$  and  $\mathcal{Y}[s-1]$  the lists consisting of only  $w_s$  and only  $z_s$  respectively, and mark them with the number  $s$ .
- Case 2: The vertex  $w_s$  satisfies  $d(w_s, \bar{v}) \leq 2r + 1$  for a unique  $\bar{v} \in \mathcal{X}[s-1]$ , and  $\bar{v}$  is marked. In this case, find the list  $\bar{u} \in \mathcal{Y}[s-1]$  with the same mark. Duplicator then can chose  $z_s \in N(\bar{u}, 2r + 1)$  in response to  $w_s$  according to a winning strategy for

$$d\text{EHR}_{k-s}(N(\bar{v}, 3r + 1), \bar{v}, N(\bar{u}, 3r + 1), \bar{u}).$$

To form  $\mathcal{X}[s]$  and  $\mathcal{Y}[s]$ , append  $w_s$  and  $z_s$  to  $\bar{v}$  and  $\bar{u}$  respectively.

- Case 3: The vertex  $w_s$  satisfies  $d(w_s, \bar{v}) \leq 2r + 1$  for a unique  $\bar{v} \in \mathcal{X}[s-1]$ , and  $\bar{v}$  is not marked. In this case we can find a non-marked list  $\bar{u} \in \mathcal{Y}[s-1]$  such that  $\bar{v}$  and  $\bar{u}$  have  $(k-s)$ -similar  $(3r + 1)$ -neighborhoods. Duplicator then can chose  $z_s \in N(\bar{u}, 2r + 1)$  in response to  $w_s$  according to a winning strategy for

$$d\text{EHR}_{k-s}(N(\bar{v}, 3r + 1), \bar{v}, N(\bar{u}, 3r + 1), \bar{u}).$$

To form  $\mathcal{X}[s]$  and  $\mathcal{Y}[s]$ , append  $w_s$  and  $z_s$  to  $\bar{v}$  and  $\bar{u}$  respectively, and mark those lists with the number  $s$ .

All that is left now is to prove the correctness of the strategy. We show that at the end the  $s$ -th round, if two lists  $\bar{v} \in \mathcal{X}[s]$  and  $\bar{u} \in \mathcal{Y}[s]$  have the same mark then  $\bar{v}$  and  $\bar{u}$  have  $(k-s)$ -similar  $r_{k-s}$ -neighborhoods. This happens trivially at the end of the zeroth round -i.e., the beginning of the game- as there are no marked lists. Assume the statement holds up to the end of the  $(s-1)$ -th round, where  $s > 0$ .

- Case 1: Notice that the lists in  $\mathcal{Y}[s-1]$  only contain the vertices previously played in  $G_2$  and the ones from  $Y$ . Thus, assumption (3) of the theorem, (or assumption (2) in the symmetric case where Spoiler plays in  $G_2$ ) assures us that Duplicator can always find such  $z_s$  sufficiently far away from all the other lists. In this case, the only new marked lists in  $\mathcal{X}[s]$  and  $\mathcal{Y}[s]$  are the ones consisting of  $w_s$  and  $z_s$  respectively. By assumption  $w_s$  and  $z_s$  have  $(k-s)$ -similar  $r_{k-s}$ -neighborhoods.
- Case 2: Notice that by the induction hypothesis  $\bar{v}$  and  $\bar{u}$  have  $(k-s+1)$ -similar  $r_{s-k+1}$ -neighborhoods, and in consequence a winning strategy for Duplicator exists. Using lemma 1.1 we obtain that the extended lists  $\bar{v}, w_s$  and  $\bar{u}, z_s$  have  $(k-s)$ -similar  $r_{s-k}$ -neighborhoods.
- Case 3: This case is analogous to the previous one. The definition of  $k$ -agreeability implies that there is such an unmarked list  $\bar{u}$  available. Using lemma 1.1 we obtain that the extended lists  $\bar{v}, w_s$  and  $\bar{u}, z_s$  have  $(k-s)$ -similar  $r_{s-k}$ -neighborhoods.

In the three cases, if  $\bar{v}$  and  $\bar{v}$  are lists in  $\mathcal{X}[s-1]$  and  $\mathcal{Y}[s-1]$  respectively that share the same mark and remain unmodified in  $\mathcal{X}[s]$  and  $\mathcal{Y}[s]$ , then by the induction hypothesis  $\bar{v}$  and  $\bar{v}$  have  $(k-s+1)$ -similar  $r_{k-s+1}$ -neighborhoods. This easily implies that they also have  $(k-s)$ -similar  $r_{k-s}$ -neighborhoods.

At the end of the game, if  $\bar{v} \in \mathcal{X}[k]$  and  $\bar{u} \in \mathcal{Y}[k]$  are lists with the same mark then the natural mapping between  $\bar{v}$  and  $\bar{u}$  defines an isomorphism between  $G_1[\bar{v}]$  and  $G_2[\bar{u}]$ .  $\square$

Quizás reordenando esta demostración se puede acortar o se entiende mejor.

## 1.8 Types of trees

We define tree  $T$  as a connected hypergraph such that  $ex(T) = -1$ . We define a vertex-rooted tree  $(T, v)$  as a tree  $T$  with a distinguished vertex  $v \in V(T)$  called its root. We will usually omit the root when it is not relevant and write just  $T$  instead of  $(T, v)$ . We define the set of initial edges of a vertex-rooted tree  $(T, v)$  as the set of edges in  $T$  that contain  $v$ .

Given a rooted tree  $(T, v)$ , and a vertex  $u \in V(T)$ , we define  $Tree_{(T,v)}(u)$  as the tree  $T[X]$  induced on the set  $X := \{w \in V(T) \mid d(v, w) = d(v, u) + d(u, w)\}$ , to which we assign  $u$  as the root. That is,  $Tree_{(T,v)}(u)$  is the tree consisting of those vertices whose only path to  $v$  contains  $u$ .

We define the radius of a vertex-rooted, or edge-rooted, tree as the maximum distance between its marked vertex and any other one.

Fix a natural number  $k$ . We will define two equivalence relations, one between rooted trees and another between pairs  $(T, e)$  of rooted trees  $T$  and initial edges  $e \in E(T)$ . We will name both relations  $k$ -equivalence relations and denote them by  $\simeq_k$ . They are defined recursively as follows:

- Any two trees with radius zero are  $k$ -equivalent. Notice that those trees consist only of one vertex, their respective roots.
- Suppose that the  $k$ -equivalence relation has been defined for rooted trees with radius at most  $r$ .

- Let  $T_1$  and  $T_2$  be rooted trees with radius at most  $r + 1$ , and let  $e_1, e_2$  be initial edges of  $T_1$  and  $T_2$  respectively. Then  $(T_1, e_1) \simeq_k (T_2, e_2)$  if  $e_1$  and  $e_2$  have the same color and there is a bijection  $f : e_1 \rightarrow e_2$  between the vertices in  $e_1$  and  $e_2$  such that:
  - \* If  $e_1 = [u_1, \dots, u_a]$  then  $e_2 = [f(u_1), \dots, f(u_a)]$ .
  - \* If  $v_1$  and  $v_2$  are the roots of  $T_1$  and  $T_2$  respectively, then  $v_2 = f(v_1)$ .
  - \* For any vertex different from the root  $u \in e_1$ , it is satisfied that

$$Tree_{(T_1, v_1)}(u) \simeq_k Tree_{(T_2, v_2)}(f(u)).$$

- Let  $T_1$  and  $T_2$  be rooted trees with radius at most  $r + 1$ . Then  $T_1 \simeq_k T_2$  if for any chosen  $T = T_1$  or  $T = T_2$  and any initial edge  $e \in E(T)$ , the "quantity of initial edges  $e_1$  from  $T_1$  that satisfy  $(T_1, e_1) \simeq_k (T, e)$ " and the "quantity of initial edges  $e_2$  from  $T_2$  that satisfy  $(T_2, e_2) \simeq_k (T, e)$ " are the same or are both greater than  $k - 1$ .

We want prove the following

**Theorem 1.4.** *Let  $(T_1, v_1)$  and  $(T_2, v_2)$  be rooted trees. Then, if they are  $k$ -equivalent Duplicator wins  $dEHR_k(T_1, v_1, T_2, v_2)$ .*

Before proceeding with the proof that we need an auxiliary result. Let  $(T, v)$  be a rooted tree and  $e$  an initial edge of  $T$ . We define  $Tree_{(T, v)}(e)$  as the induced tree  $T[X]$  on the set  $X := \{v\} \cup \{u \in V(T) \mid d(v, u) = |e| + d(e, v)\}$ , to which we assign  $v$  as the root. In other words,  $Tree_{(T, v)}(e)$  is the tree formed of  $v$  and all the vertices in  $T$  whose only path to  $v$  contain  $e$ . Now we can check the following:

**Lemma 1.2.** *Fix  $r > 0$ . Suppose that theorem 1.4 holds for rooted trees with radii at most  $r$ . Let  $(T_1, v_1)$  and  $(T_2, v_2)$  be rooted trees with radii at most  $r + 1$ . Let  $e_1$  and  $e_2$  be initial edges of  $T_1$  and  $T_2$  respectively satisfying  $(T_1, e_1) \simeq_k (T_2, e_2)$ . Name  $T'_1 = Tree_{(T_1, v_1)}(e_1)$  and  $T'_2 = Tree_{(T_2, v_2)}(e_2)$ . Then Duplicator wins  $dEHR_k(T'_1, v_1, T'_2, v_2)$ .*

*Proof.* We show a winning strategy for Duplicator. Suppose that in the  $i$ -th round of the game Spoiler plays on  $T'_1$ . The other case is symmetric. Let  $f : e_1 \rightarrow e_2$  be a bijection as in the definition of  $(T_1, e_1) \simeq_k (T_2, e_2)$ . There are two possibilities:

- If Spoiler plays a vertex  $v$  on  $e_1$  then Duplicator can play  $f(v)$  on  $e_2$ .
- Otherwise, Spoiler plays a vertex  $v$  that belongs to some  $Tree_{(T'_1, v_1)}(u)$  for a unique  $u \in e_1$  different from the root  $v_1$ . By the definition of  $(T_1, e_1) \simeq_k (T_2, e_2)$ ,  $Tree_{(T'_1, v_1)}(u) \simeq_k Tree_{(T'_2, v_2)}(f(u))$ . As both these trees have radii at most  $r$ , by assumption Duplicator has a winning strategy between them and they can follow it.

□

Now we can prove the main theorem of this section:

*Proof of theorem 1.4.*

Notice that, as  $T_1 \simeq_k T_2$ , both  $T_1$  and  $T_2$  have the same radius  $r$ . We prove the result by induction on  $r$ . If  $r = 0$  then both  $T_1$  and  $T_2$  consist of only one vertex and we are done.

Now let  $r > 0$  and assume that the statement is true for all lesser values of  $r$ . We will show that there is a winning strategy for Duplicator in  $d\text{EHR}_k(T_1, v_1, T_2, v_2)$ . At the start of the game, set all the initial edges in  $T_1$  and  $T_2$  as non-marked. Suppose that in the  $i$ -th round Spoiler plays in  $T_1$ . The other case is symmetric.

- If Spoiler plays  $v_1$  then Duplicator plays  $v_2$ .
- Otherwise, the vertex played by Spoiler belongs to  $\text{Tree}_{(T_1, v_1)}(e_1)$  for a unique initial edge  $e_1$  of  $T_1$ . There are two possibilities:
  - If  $e_1$  is not marked yet, mark it with the index  $i$ . In this case, there is a non-marked initial edge  $e_2$  in  $T_2$  satisfying  $(T_1, e_1) \simeq_k (T_2, e_2)$ . Mark  $e_2$  with the index  $i$  as well. Because of lemma 1.2, Duplicator has a winning strategy in

$$d\text{EHR}_k(\text{Tree}_{T_1}(e_1), v_1, \text{Tree}_{T_1}(e_2), v_2)$$

and can play according to it.

- If  $e_1$  is already marked then there is a unique initial edge  $e_2$  in  $T_2$  marked with the same mark as  $e_1$  and  $(T_1, e_1) \simeq_k (T_2, e_2)$ . Again, Because of lemma 1.2, Duplicator has a winning strategy in

$$d\text{EHR}_k(\text{Tree}_{T_1}(e_1), v_1, \text{Tree}_{T_1}(e_2), v_2)$$

and can continue playing according to it.

Then Duplicator can find an initial edge  $e_2$  of  $T_2$  such that  $(T_1, e_1) \simeq_k (T_2, e_2)$ . Because of lemma 1.2, Duplicator has a winning strategy in  $d\text{EHR}_k(\text{Tree}_{T_1}(e_1), v_1, \text{Tree}_{T_1}(e_2), v_2)$  and can play according to it.

□

probablemente con algún dibujo sencillo esta demostración se entienda mejor

## 1.9 Probabilistic results

Given a natural numbers  $n$  and  $l$  we will use  $(n)_l$  to denote  $n(n-1)\cdots(n-l+1)$  or 1 if  $l = 0$ .

Our main tool for computing probabilities will be the following multivariate version of Brun's Sieve CITA REQUERIDA

**Theorem 1.5.** Fix  $k \in \mathbb{N}$ . For each  $n \in \mathbb{N}$ , let  $X_{1,n}, \dots, X_{k,n}$  be non-negative random integer variables over the same probability space. Let  $\lambda_1, \dots, \lambda_k$  be real numbers. Suppose that for any  $r_1, \dots, r_l \in \mathbb{N}$

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[ \prod_{i=1}^k (X_{i,n})_{r_i} \right] = \prod_{i=1}^k \frac{\lambda_i}{r_i!}.$$

Then the  $X_{1,n}, \dots, X_{k,n}$  converge in distribution to independent Poisson variables with means  $\lambda_1, \dots, \lambda_k$  respectively.

## 1.10 Almost all hypergraphs are simple

We say that a connected hypergraph  $G$  is **dense** if  $ex(G) > 0$ . Given  $r \in \mathbb{N}$ , we say that  $G$  is  **$r$ -simple** if  $G$  does not contain any dense subgraph  $H$  such that  $diam(H) \leq r$ . The goal of this section is to show that, for any fixed  $r$ , a.a.s  $G_n$  is  $r$ -simple.

**Lemma 1.3.** *Let  $H$  be an hypergraph. Then  $E[\# \text{ copies of } H \text{ in } G_n] = \Theta(n^{-ex(H)})$  as  $n$  tends to infinity.*

*Proof.* Indeed, let  $n \in \mathbb{N}$  and let  $G_n$  be a random hypergraph sampled from  $G^{\mathcal{C}}(n, \bar{p})$ . Let  $\bar{h}$  be an ordering of the vertices in  $V(H)$ . For any list  $\bar{v} \in [n]_{|V(H)|}$  (recall that  $[n] = V(G)$ ), let  $X_{\bar{v}}$  be the random indicator variable that takes value 1 if the natural map  $f : \bar{h} \rightarrow \bar{v}$  defines an homomorphism between  $H$  and  $G$ . That is,  $X_{\bar{v}}$  takes value 1 with probability  $\prod_{i=1}^t (\beta_i n^{1-a_i})^{|E_i(H)|}$ , and zero otherwise. The sum of all  $X_{\bar{v}}$ 's for all  $\bar{v} \in [n]_{|V(H)|}$  counts how many labelled copies of  $H$  are in  $G_n$ . Thus,

$$E[\# \text{ labelled copies of } H \text{ in } G_n] = \sum_{\bar{v} \in [n]_{|V(H)|}} \prod_{i=1}^t (\beta_i n^{1-a_i})^{|E_i(H)|} = (n)_{|V(H)|} \prod_{i=1}^t (\beta_i n^{1-a_i})^{|E_i(H)|}.$$

The expected number of copies of  $H$  in  $G_n$  is then this last expectation divided by the number of automorphisms of  $H$ ,  $|Aut(H)|$ . In consequence,

$$\begin{aligned} \lim_{n \rightarrow \infty} E[\# \text{ copies of } H \text{ in } G_n] &= \lim_{n \rightarrow \infty} \frac{(n)_{|V(H)|}}{|Aut(H)|} \prod_{i=1}^t (\beta_i n^{1-a_i})^{|E_i(H)|} = \\ &= \lim_{n \rightarrow \infty} \left[ (n)_{|V(H)|} \prod_{i=1}^t n^{(1-a_i)|E_i(H)|} \right] \frac{\prod_{i=1}^t \beta_i^{|E_i(H)|}}{|Aut(H)|} \simeq \lim_{n \rightarrow \infty} n^{-ex(H)} \frac{\prod_{i=1}^t \beta_i^{|E_i(H)|}}{|Aut(H)|}, \end{aligned}$$

and last expression belongs to  $\theta(n^{-ex(H)})$ , as desired. □

As a corollary of last result we get the following:

**Lemma 1.4.** *Let  $H$  be an hypergraph such that  $ex(H) > 0$ . Then a.a.s there are no copies of  $H$  in  $G_n$ .*

*Proof.* Because of the previous fact,  $E[\# \text{ copies of } H \text{ in } G_n] \xrightarrow{n \rightarrow \infty} 0$ . Then an application of the first moment method yields the desired result. □

The main theorem of this section is the following

**Lemma 1.5.** *Let  $H$  be a dense hypergraph of radius  $r$ . Then  $H$  contains a dense sub-hypergraph  $H'$  with size no greater than  $(a+2)(r+1) + 2a$ , where  $a$  is the largest edge size in  $H$ .*

*Proof.* Choose  $x \in V(H)$ . Successively remove from  $G$  edges  $e$  such that  $d(x, e)$  is maximum until the resulting graph  $H'$  has excess no greater than 0. We have two cases:

- $ex(H') = -1$ . Let  $e = [x_1, \dots, x_b]$  be the last removed edge and  $e \cap H' = \{x_{i_1}, \dots, x_{i_d}\}$ . For any  $j = 1, \dots, d$  choose  $P_j$  a path of size no greater than  $r+1$  joining  $x$  and  $x_{i_j}$  in  $H'$ . Then  $P_1 \cup \dots \cup P_d \cup e$  is a dense sub-hypergraph of  $H$  of size less than  $a(r+1) + a < (a+2)(r+1) + 2a$ .

- $ex(H') = 0$ . Let  $e_1 = [x_1, \dots, x_{b_1}]$  be the last removed edge. Continue removing the edges of  $G'$  that are at maximum distance from  $x$  until you obtain  $H''$  with  $ex(H'') = -1$ . Let  $e_2 = [y_1, \dots, y_{b_2}]$  be the last removed edge this time. As before, let  $e_1 \cap H' = \{x_{i_1}, \dots, x_{i_d}\}$  and for  $j = 1, \dots, d$  let  $P_j$  a path of size no greater than  $r + 1$  joining  $x$  and  $x_{i_j}$  in  $H'$ . Then  $e_2 \cup H'' = \{y_{i_1}, y_{i_2}\}$ . Let  $Q_1, Q_2$  be paths size no greater than  $r + 1$  from  $x$  to  $y_{i_1}$  and  $y_{i_2}$  in  $H''$ . Then  $Q_1 \cup Q_2 \cup e_2$  is a graph of likelihood 0 and size less than  $2r + 2 + a$ , and  $Q_1 \cup Q_2 \cup P_1 \cup \dots \cup P_d \cup e_1 \cup e_2$  is a critical graph with size less than  $(2 + a)(r + 1) + 2a$

□

**Theorem 1.6.** *Let  $r \in \mathbb{N}$ . Then a.a.s  $G_n$  is  $r$ -simple.*

*Proof.* The first moment method alone is not sufficient to prove our claim because the amount of dense hypergraphs  $H$  such that  $diam(H) \leq r$  is not finite in general. Thus, we need to prove that it suffices to prohibit a finite amount of dense sub-hypergraphs in order to guarantee that  $G_n$  is  $r$ -simple. □

## 1.11 Probabilities of trees

During this section we want to study the asymptotic probability that the  $r$ -neighborhood of a given vertex  $v \in \mathbb{N}$  in  $G_n$  is a tree that belongs to a given  $k$ -equivalence class of trees  $\mathcal{T}$  with radius at most  $r$ . That is, we want to know

$$\lim_{n \rightarrow \infty} \Pr(T := N^{G_n}(v; r) \text{ is a tree, and } (T, v) \in \mathcal{T}).$$

Denote this limit by  $\Pr[r, \mathcal{T}]$ . Notice that the definition of  $\Pr[r, \mathcal{T}]$  does not depend by the choice of  $v$ .

We define  $\Lambda$  and  $M$  as the minimal families of expressions with arguments  $\bar{\beta}$  that satisfy the conditions: **(1)**  $1 \in \Lambda$ , **(2)** for any  $b, i \in \mathbb{N}$  with  $1 \leq i \leq c$ ,  $b > 0$ , and  $\lambda_1, \dots, \lambda_{a_i-1} \in \Lambda$ , the expression  $(\beta_i/b) \prod_{j=1}^{a_i-1} \lambda_j$  belongs to  $M$ , **(3)** for any  $\mu \in M$  and any  $n \in \mathbb{N}$  both  $\text{Pois}_\mu(n)$  and  $\text{Pois}_\mu(\geq n)$  are in  $\Lambda$ , and **(4)** for any  $\lambda_1, \lambda_2 \in \Lambda$ , the product  $\lambda_1 \lambda_2$  belongs to  $\Lambda$  as well.

The goal of this section is to show want to show that  $\Pr[r, \mathcal{T}]$ , as an expression with parameters  $\bar{\beta}$ , belongs to  $\Lambda$  for any choice of  $r$  and  $\mathcal{T}$ . The argument is as follows: let  $T := N^{G_n}(v; r)$ .

**Lemma 1.6.** *Let  $X \subset \mathbb{N}$  be a finite set of fixed vertices, and let  $r \in \mathbb{N}$ . Define  $G'_n = G_n \setminus E(X)$ . Then (1) a.a.s  $G'_n$  does not contain a path of size  $r + 1$  between any two vertices  $u, v \in X$ , and (2) a.a.s  $G'_n$  does not contain a cycle containing any vertex  $u \in X$ .*

*Proof.* To prove the first part of the statement it suffices to show that the expected number of such paths tends to zero as  $n$  tends to infinity. Let  $P$  be a path of size  $r + 1$  between two vertices  $w, z$ , and let  $u, v$  be two vertices from  $X$ . Given  $f$  □

Let  $\bar{v} \in \mathbb{N}^*$  a list of vertices. For each  $n, r \in \mathbb{N}$  and  $v_i \in \bar{v}$  we define the random hypergraph  $T_n(v_i, \bar{v}; r)$  in the following way: Let  $G'_n := G_n \setminus E(G_n[\bar{v}])$ . Then  $T_n(v_i, \bar{v}; r) := N^{G'_n}(v_i; r)$ . That is,  $T_n(v_i, \bar{v}; r)$  is the  $r$ -neighborhood of  $v_i$  in “ $G_n$  minus the edges induced on  $\bar{v}$ ”.

**Theorem 1.7.** *The following are satisfied:*

- (1) Let  $r \in \mathbb{N}$  and let  $\mathcal{T}$  be a  $k$ -equivalence class for trees with radii at most  $r$ . Then  $\Pr[r, \mathcal{T}]$  exists and is an expression in  $\Lambda$ .
- (2) Let  $\bar{w} \in \mathbb{N}^*$  be a list of different fixed vertices, and let  $\phi[\bar{x}] \in FO[\sigma]$  be an open sentence (i.e. with no bounded variables) such that  $\bar{x}$  and  $\bar{w}$  have the same size. Let  $v_1, \dots, v_k \in \mathbb{N}$  be another list of different vertices that are contained in  $\bar{w}$ . Let  $r_1, \dots, r_k \in \mathbb{N}$  and let  $\mathcal{T}_1, \dots, \mathcal{T}_k$  be  $k$ -equivalence classes for trees such that each  $\mathcal{T}_i$  has radii bounded by  $r_i$ . Then

$$\lim_{n \rightarrow \infty} \Pr\left(\bigwedge_{i=1}^k T_{n,i} = T_n(v_i, \bar{w}; r_i) \text{ is a tree and } (T_{n,i}, v_i) \in \mathcal{T}_i \mid \sigma(\bar{w})\right) = \prod_{i=1}^k \Pr[r_i, \mathcal{T}_i].$$

*Proof.* Let  $R$  be an upper bound of the radius of  $r$  and the  $r_i$ 's in the statement. We will prove (1) and (2) together by induction on  $R$ .

Assume  $R = 0$ . We start by showing that (1) holds. Recall that all trees with radius zero are  $k$ -equivalent. Thus, if  $\mathcal{T}$  is the unique  $k$ -equivalence class of trees with radius zero and  $v \in \mathbb{N}$  is a fixed vertex, then

$$\Pr[0; \mathcal{T}] = \lim_{n \rightarrow \infty} \Pr(T := N^{G_n}(v; 0) \text{ is a tree, and } (T, v) \in \mathcal{T}) = 1,$$

Indeed,  $N^{G_n}(v; 0)$  consists of a single vertex for all  $n \geq v$ , and the above equation follows. The expression 1 belongs to  $\Lambda$ , so (1) holds.

The case of (2) is analogous. As  $R = 0$ , then  $\mathcal{T}_1 = \dots = \mathcal{T}_k$  are the unique  $k$ -equivalence class of trees with radius zero. Then, given  $\sigma, \bar{w}, v_1, \dots, v_k$  as in the statement,

$$\lim_{n \rightarrow \infty} \Pr\left(\bigwedge_{i=1}^k T_{n,i} = T_n(v_i, \bar{w}; 0) \text{ is a tree and } (T_{n,i}, v_i) \in \mathcal{T}_i \mid \sigma(\bar{w})\right) = 1.$$

Because of (1),  $\Pr[0, \mathcal{T}_i] = 1$  for all  $i$ 's, and (2) holds. □

Let  $\bar{\mathcal{T}}$  be a  $\bar{u} \in N$



## References

- [1] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 5(1):17–60, 1960.
- [2] Ronald Fagin. Probabilities on finite models 1. *The Journal of Symbolic Logic*, 41(1):50–58, 1976.
- [3] Yu V Glebskii, Do I Kogan, MI Liogon’kiI, and VA Talanov. Range and degree of realizability of formulas in the restricted predicate calculus. *Cybernetics and Systems Analysis*, 5(2):142–154, 1969.
- [4] Edgar N Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.
- [5] Saharon Shelah and Joel Spencer. Zero-one laws for sparse random graphs. *Journal of the American Mathematical Society*, 1(1):97–115, 1988.
- [6] James F Lynch. Probabilities of sentences about very sparse random graphs. *Random Structures & Algorithms*, 3(1):33–53, 1992.
- [7] Vašek Chvátal and Bruce Reed. Mick gets some (the odds are on his side)(satisfiability). In *Proceedings., 33rd Annual Symposium on Foundations of Computer Science*, pages 620–627. IEEE, 1992.
- [8] H-D Ebbinghaus, Jörg Flum, and Wolfgang Thomas. *Mathematical logic*. Springer Science & Business Media, 2013.
- [9] Andrzej Ehrenfeucht. An application of games to the completeness problem for formalized theories. *Fund. Math*, 49(129-141):13, 1961.
- [10] Joel Spencer. *The strange logic of random graphs*, volume 22. Springer Science & Business Media, 2013.