**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

INCOMPLETA, PENDIENTE DE SER REVISADA

Since the work of Erdös and Rènyi 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 $\mathscr{L}$ if for any given property $P$ expressible in $\mathscr{L}$ 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 $\mathscr{L}$ 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 were 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 $F_\phi : (0, \infty) \to [0, 1]$ given by*

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

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

The analyticity of these asymptotic probabilities with respect to the parameter $\beta$ implies that FO properties cannot "capture" sudden changes that occur in the random graph $\mathrm{G}(n, \beta/n)$ 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 $\mathrm{G}(n, p(n))$ 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. Erdös and Rènyi [1] showed that for probabilities of this form $\mathrm{G}(n, p(n))$ 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 $\lim_{n\to\infty} \Pr\big(G(n,c/n)$ is not 3-colorable $\big) > 0$. 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 $c_0 \le c_1$ 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 $c$

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

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 were 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 $k$.

Given numbers, $n,m \in \mathbb{N}$ with $m \le n$ we define $(n)_m := n\cdot(n-1)\cdots(n-m+1)$ when $m \neq 0$ and $(n)_0 := 1$. Given a set $S$ and a number $n \in \mathbb{N}$ with $n \le |S|$ we define $(S)_n$ as the subset of $S^n$ consisting of the $n$-tuples whose coordinates are all different. We also define $S^* := \bigcup_{n=0}^{\infty} S^n$ and $(S)_* := \bigcup_{n\le|S|}(S)_n$. Given two tuples $\bar{x},\bar{y} \in S^*$ we write $\bar{x}^\frown\bar{y}$ to denote their concatenation. Given a tuple $\bar{x} \in S^*$ and an element $x \in S$ the expression $x \in \bar{x}$ will mean that $x$ appears as some coordinate in $\bar{x}$. We will at times make an abuse of notation and treat the tuple $\bar{x}$ as the set of elements $x \in \bar{x}$. We will use the convention that over-lined variables , like $\bar{x}$, denote ordered tuples of arbitrary length. Given an ordered tuple $\bar{x}$ we define the number $len(\bar{x})$ as its length. Given a map $f : X \to Y$ between two sets $X,Y$ and an ordered tuple $\bar{x} := (x_1,\dots,x_a) \in X^*$ we define $f(\bar{x}) \in Y^*$ as the tuple $(f(x_1),\dots,f(x_a))$.

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 $\bar{x} := (x_1,\dots,x_a) \in S^a$ we define $g\cdot\bar{x}$ as the tuple $(x_{g(1)},\dots,x_{g(a)})$. We will denote by $S^a/\Phi$ to the quotient of the set $S^a$ by this action. Given an element $\bar{x} := (x_1,\dots,x_a) \in S^a$ we will denote its equivalence class in $S^a/\Phi$ by $[x_1,\dots,x_a]$ or $[\bar{x}]$. Thus, for any $g \in \Phi$, by definition $[x_1,\dots,x_a] = [x_{g(1)},\dots,x_{g(a)}]$. The notations $\bar{x}$ and $(x_1,\dots,x_a)$ will be reserved to ordered tuples while $[\bar{x}]$ and $[x_1,\dots,x_a]$ will denote ordered tuples 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.

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

## 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 $\{R_1,\dots,R_m\}$ where each relation symbol $R_i$ has associated a natural number $ar(R_i)$ 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 $\sigma$ such that $ar(R)=a$, 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 $\forall\overline{y}$, where $\overline{y}:=(y_1,\dots,y_m)$ instead of $\forall y_1,\forall y_2,\dots,\forall y_m$ and likewise for the quantifier $\exists$. Also, given $\overline{y}:=(y_1,\dots,y_a)$, we may write simply $R(\overline{y})$ instead of $R(y_1\dots,y_a)$.

For the remaining of this article we will reserve the names $x,y,z$ for the variables in our first order formulas.

We define the set of free variables of a formula as usual. Given a formula $\phi\in FO[\sigma]$ we will use the notation $\phi(\overline{y})$ to denote that $\overline{y}$ is a tuple of (different) variables that contains all free variables in $\phi$ and none of its bounded variables, although it may contain variables which not appear in $\phi$.

Formulas with no free variables are called sentences and formulas with no quantifiers are called open formulas.

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 $\mathscr{A}$, and a sentence $\phi\in FO[\sigma]$ we write $\mathscr{A}\models\phi$ to denote that $\mathscr{A}$ satisfies $\phi$. If $\psi(\overline{y})$ is a formula, $\overline{a}$ are elements in the universe of $\mathscr{A}$, and $\overline{y}$ and $\overline{a}$ are lists of the same size, then we write $\mathscr{A} \models \psi(\overline{a})$ to mean that $\mathscr{A}$ satisfies $\psi$ when the free variables in $\overline{y}$ are interpreted as the elements in $\overline{a}$ (i.e., the $i$-th variable of $\overline{y}$ is interpreted as the $i$-th element of $\overline{a}$).

## 1.3 Structures as multi-hypergraphs

For the rest of the article consider fixed:

- A relational vocabulary $\sigma$ such that all the relations $R \in \sigma$ satisfy $ar(R) \geq 2$.
- Groups $\{\Phi_R\}_{R\in\sigma}$ such that each $\Phi_R$ is consists of permutations on $[ar(R)]$ with the usual composition as its operation.
- Sets $\{P_R\}_{R\in\sigma}$ satisfying that for all $R \in \sigma$, $P_R \subseteq \binom{[ar(R)]}{2}$

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

- *Symmetry axioms*: For each $R \in \sigma$ and each $g \in \Phi_R$:

$$\forall \overline{x} := x_1, \dots, x_{ar(R)} \big(R(\overline{x}) \iff R(g \cdot \overline{x})\big)$$

- *Anti-reflexivity axioms*: For each $R \in \sigma$ and $\{i, j\} \in P_R$

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

We can think any structure $G$ in $\mathscr{C}$ as a "multi-hypergraph" whose vertices are the elements of the universe of $G$. Each relation $R \in \sigma$, can be represented over $G$ as an "edge set" formed by tuples vertices of size $ar(R)$ modulo the permutations in $\Phi_R$. Furthermore, repetitions of vertices in the positions given by $P_R$ are not allowed in these tuples.

The following definitions make this ideas formal. They depend on our choices of $\sigma$, $\{\phi_R\}_{R\in\sigma}$ and $\{P_R\}_{R\in\sigma}$ but as those are fixed we can allow ourselves to omit those dependencies for the sake of readability.

**Definition 1.1.** Let $V$ be a set, and let $R \in \sigma$. We define the **total edge set over $V$ given by $R$** as

$$E_R[V] = (V^{ar(R)}/\Phi_R) \quad \setminus \quad X,$$

where

$$X = \Big\{[x_1, \dots, x_{ar(R)}] \;\Big|\; x_1, \dots, x_{ar(R)} \in V, \text{ and } x_i = x_j \text{ for some } \{i, j\} \in P_R\Big\}.$$

Also, we will say that the **sort** of the elements $e \in E_R[V]$ is $R$.

That is, $E_R[V]$ contains all the "$ar(R)$-tuples of elements in $V$ modulo the permutations in $\phi_R$" excluding those that contain some repetition of elements in the positions given by $P_R$.

The fact that the elements $e \in E_R(V)$ are of sort $R$ is a technical detail introduced so that for any different relation symbols $R_1$ and $R_2$ it holds $E_{R_1}(V) \cap E_{R_2}(V) = \emptyset$ even in the case that $ar(R_1) = ar(R_2)$, $\Phi_{R_1} = \Phi_{R_2}$ and $P_{R_1} = P_{R_2}$.

In the case where $V = [n]$ we will write simply $E_R[n]$ instead of $E_R[[n]]$.

**Definition 1.2.** We call $\mathscr{C}$-**hypergraph**, or simply hypergraph, to a pair $G = \Big(V(G), \{E_R(G)\}_{R\in\sigma}\Big)$, where for each $R$, $E_R(G) \subseteq E_R[V]$.

Hypergraphs, as we have defined them, can be naturally interpreted as structures from $\mathscr{C}$ in the following way: given $G = (V, \{E_R\}_{R\in\sigma})$, we consider $V$ to be the universe of $G$, and for any $R \in \sigma$ we define $R^G \subseteq V^{ar(R)}$ as the set of tuples $\bar{x} \in V^{ar(R)}$ such that $[\bar{x}] \in E_R$. Under this interpretation hypergraphs satisfy by definition the symmetry and anti-reflexivity axioms given at the beginning of this section. It is also easy to see that this interpretation induces a one-to-one identification between structures in $\mathscr{C}$ and hypergraphs.

## 1.4 Hypergraph notation and nomenclature

We will use standard nomenclature and notation from graph theory with some additions. Given an hypergraph $G$ we will call its **vertex set** to $V(G)$ and **vertices** to the elements $v \in V(G)$. Likewise, each of the $E_R(G)$'s will be called an **edge set** and its elements, **edges**. Given an edge, $e \in V(G)$ we will denote by $V(e)$ the set of all vertices that participate in $e$.

Given an hypergraph $G$ we define the set $E(G)$ as the union $\cup_{R\in\sigma} E_R(G)$. Notice that this union is disjoint because elements from different $E_R(G)$'s are of different sorts. Thus, $|E(G)| = \sum_{R\in\sigma} |E_R(G)|$. Analogously, given a set $V$, we define $E[V] := \cup_{R\in\sigma} E_R[V]$. Given an edge $e \in E[V]$ we define $R(e)$ as the sort of $e$, i.e., the unique relation symbol $R(e) \in \sigma$ such that $e \in E_{R(e)}[V]$.

Given two hypergraphs $H$ and $G$ we say that $H$ is a **sub-hypergraph** of $G$, which we write as $H \subset G$, if $V(H) \subset V(G)$ and $E(H) \subset E(G)$ (notice that this is equivalent to $E_R(H) \subset E_R(G)$ for all $R \in \sigma$, since the edges are sorted).

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 $H = (V(H), \{E(H)_R\}_{R\in\sigma})$ such that $V(H) = U$ and for any $R \in \sigma$ each edge $e \in E_R(G)$ belongs to $E_R(H)$ as well if and only if the vertices involved in $e$ are in $U$ (i.e. $e \in E_R[U]$).

We define the **excess** $ex(G)$ of an hypergraph $G$ as the number

$$ex(G) := \Big(\sum_{R\in\sigma} (ar(R)-1)|E_R(G)|\Big) - |V(G)|.$$

That is, the excess of $G$ is its "weighted number of edges" minus its number of vertices.

An hypergraph $G$ is **connected** if for any two vertices $v, u \in V(G)$ there is a sequence of edges $e_1, \dots, e_m \in E(G)$ such that $v \in V(e_1), u \in V(e_m)$ and for each $i \in [m-1]$, $V(e_i) \cap V(e_{i+1}) \neq \emptyset$. It holds that $ex(G) \geq -1$ for any connected hypergraph.

Given a connected hypergraph $G$ and vertices $v, u \in V(G)$, we say that $G$ is a **path** between $v$ and $u$ if $G$ does not contain any connected proper sub-hypergraph containing both $v, u$.

A connected hypergraph $G$ is a **tree** if $ex(G) = -1$.

A connected hypergraph $G$ is called **dense** if $ex(G) > 0$.

A connected hypergraph $G$ with $ex(G) \geq 0$ is called **saturated** if for any non-empty proper sub-hypergraph $H \subset G$ it holds $ex(H) < ex(G)$.

A connected hypergraph $G$ with $ex(G) = 0$ is called **unicycle**. A saturated unicycle is called a **cycle**.

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

$$d^G(u,v) = \min_{\substack{H \subset G \\ H \text{ connected} \\ u,v \in V(H)}} |E(H)|.$$

That is, the distance between $v$ and $u$ is the minimum number of edges necessary to connect $v$ and $u$. If such number does not exist we define $d^G(u,v) = \infty$.

As usual, define $d^G(X,Y)$ for sets $X,Y \subset V(G)$ as the minimum distance $d(u,v)$ where $u \in X$ and $v \in Y$. 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.

We will often write tuples of vertices instead of sets inside of $d(\cdot,\cdot)$ and $N(\cdot;r)$. In those cases we are treating those tuples as sets as specified in section 1.1.

## 1.5 Copies of Hypergraphs

Given two hypergraphs $H_1, H_2$ an **isomorphism** $f$ between them is a bijection $f : V(H_1) \to V(H_2)$ such that for any $R \in \sigma$, a tuple $\overline{v} \in V(H_1)^*$ forms an edge $[\overline{v}] \in E_R(H_1)$ if and only if $[f(\overline{v})] \in E_R(H_2)$. We say that $H_1$ and $H_2$ are isomorphic, written as $H_1 \simeq H_2$, if there exists some isomorphism between them.

Let $H$ be an hypergraph. An **automorphism** of $H$ is an isomorphism from $H$ to itself. We will denote by $\text{aut}(H)$ the number of such automorphisms.

Let, $\mathbb{H}$ be an isomorphism class of hypergraphs. Let $H \in \mathbb{H}$ be a representative of the class. Then we define $ex(\mathbb{H})$ as $ex(H)$. Clearly this definition does not depend on the choice of $H$. The expressions $\text{aut}(\mathbb{H})$, $|V(\mathbb{H})|$, $|E(\mathbb{H})|$ or $|E_R(\mathbb{H})|$ for $R \in \sigma$ are defined in an analogous way.

Let $H$ be an hypergraph and let $V$ be a set. We define the set $Copies(H,V)$ as the set of hypergraphs $H'$ such that $V(H') \subset V$ and $H \simeq H'$. Analogously, if $\mathbb{H}$ is an isomorphism class of hypergraphs we define $Copies(\mathbb{H},V)$ as the set of hypergraphs $H'$ that satisfy $V(H') \subset V$ as well as $H' \in \mathbb{H}$.

In our proofs it will be useful to consider colorings over hypergraphs as a way to assign extra information to their vertices.

**Definition 1.3.** Let $\Sigma$ be a set. A **$\Sigma$-hypergraph** is a pair $(H,\chi)$ where $H$ is an hypergraph and $\chi : V(H) \to \Sigma$ is a map called **$\Sigma$-coloring** of $H$.

Given two $\Sigma$-hypergraphs $(H_1,\chi_1)$ and $(H_2,\chi_2)$, an **isomorphism** between them is a bijection $f : V(H_1) \to V(H_2)$ satisfying that $f$ is an hypergraph isomorphism between $H_1$ and $H_2$ as well as $\chi_2(f(v)) = \chi_1(v)$ for all $v \in V(H_1)$. We say that $(H_1,\chi_1)$ is isomorphic to $(H_2,\chi_2)$, written as $(H_1,\chi_1) \simeq (H_2,\chi_2)$, if there exists an isomorphism between them.

Let $(H,\chi)$ be a $\Sigma$-hypergraph, an **automorphism** of $(H,\chi)$ is an isomorphism from it into itself. We will denote by $\text{aut}(H,\chi)$ the number of such automorphisms.

As with the case of hypergraphs, if $\mathbb{H}$ is an isomorphism class of $\Sigma$-hypergraphs then the expressions $\text{aut}(\mathbb{H})$ and $|V(\mathbb{H})|$ are defined via representatives.

Let $(H,\chi)$ be a $\Sigma$-hypergraph and let $V$ be a set. We define the set $Copies\big((H,\chi),\, V\big)$ as the set of $\Sigma$-hypergraphs $(H',\chi')$ satisfying $V(H') \subset V$ and $(H,\chi) \simeq (H',\chi')$. Let $\mathbb{H}$ be an isomorphism class of $\Sigma$-hypergraphs. Then the set $Copies(\mathbb{H},V)$ is defined as the set of $\Sigma$-hypergraphs $(H',\chi')$ such that $V(H') \subset V$ and $(H',\chi') \in \mathbb{H}$. Let $v \in V$ and $s \in \Sigma$. We define the set $Copies\big(\mathbb{H},V;\, (v,s)\big)$ as the set of $\Sigma$-hypergraphs $(H',\chi') \in Copies(\mathbb{H},V)$ that satisfy $v \in V(H')$ as well as $\chi'(v) = s$.

## 1.6 The random model

For each $R \in \sigma$ let $p_R$ be a real number between zero and one. Let $\overline{p} := \{p_R\}_{R\in\sigma}$. The random model $G^{\mathscr{C}}(n,\overline{p})$ is the discrete probability space that assigns to each hypergraph $G$ whose vertex set $V(G)$ is $[n]$ the following probability:

$$\Pr(G) = \prod_{R\in\sigma} p_R^{|E_R(G)|}(1-p_R)^{\left|E_R[n]\right| - \left|E_R(G)\right|}.$$

Equivalently, this is the probability space obtained by assigning to each edge $e \in E_R[n]$ probability $p_R$ independently.

As in the case of Lynch's theorem, we are interested in the "sparse regime" of $G^{\mathscr{C}}(n,\overline{p})$, were the expected number of edges each sort is linear. This is achieved when each of the $p_R$'s are of the form $\beta_R/n^{ar(R)-1}$ for some positive real numbers $\{\beta_R\}_{R\in\sigma}$. From now on we will write $G_n\left(\{\beta_R\}_{R\in\sigma}\right)$ to denote a random sample of $G^{\mathscr{C}}\left(n, \left\{\frac{\beta_R}{n^{ar(R)-1}}\right\}_{R\in\sigma}\right)$. When the choice of $\{\beta_R\}_{R\in\sigma}$ is not relevant we will write $G_n$ instead of $G_n\left(\{\beta_R\}_{R\in\sigma}\right)$.

Our goal is to prove the following theorem:

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

$$\{\beta_R\}_{R\in\sigma} \mapsto \lim_{n\to\infty} Pr\big(G_n\left(\{\beta_R\}_{R\in\sigma}\right) \models \phi\big)$$

*is well defined and analytic.*

## 1.7 Ehrenfeucht-Fraisse Games

Let $H_1$ and $H_2$ be hypergraphs. We define the $k$ round Ehrenfeucht-Fraisse (EF) game on $H_1$ and $H_2$, denoted by $\text{EHR}_k(H_1;H_2)$, 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 $V(H_1)$ or $V(H_2)$ and Duplicator responds by choosing a vertex from the other set. Let us denote by $v_i$, resp. $u_i$ the vertex from $H_1$, resp. from $H_2$, 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 a relation symbol $R \in \sigma$ and indices $i_1, \dots, i_{ar(R)} \in [k]$, $[v_{i_1}, \dots, v_{i_{ar(R)}}] \in E_R(H_1) \iff [u_{i_1}, \dots, u_{i_{ar(R)}}] \in E_R(H_2)$.

The following is satisfied:

**Theorem 1.2** (Ehrenfeut, 9). *Let $H_1$ and $H_2$ be hypergraphs. Then Duplicator wins* $\text{EHR}_k(H_1;H_2)$ *if and only if $H_1$ and $H_2$ satisfy the same sentences $\phi \in FO[\sigma]$ with $qr(\phi) \leq k$.*

Let $\overline{v} \in V(H_1)^*$, and $\overline{u} \in V(H_2)^*$ be lists of vertices of the same length, $l = len(\overline{v}) = len(\overline{u})$. We define the $k$ round Ehrenfeucht-Fraisse game on $H_1$ and $H_2$ with initial position given by $\overline{v}$ and $\overline{u}$, denoted by $\text{EHR}_k(H_1,\overline{v};H_2,\overline{u})$, the same way as $\text{EHR}_k(H_1;H_2)$, but in this case the game has $l$ extra rounds at the beginning where the vertices in $\overline{v}$ and $\overline{u}$ are played successively. After this, $k$ more rounds are played normally.

We also define the $k$-round distance Ehrenfeucht-Fraisse game on $H_1$ and $H_2$, denoted by $d\text{EHR}_k(H_1;H_2)$, the same way as $\text{EHR}_k(H_1;H_2)$, but now, in order for Duplicator to win the game, the following additional condition has to be satisfied at the end of the game:

- For any $i, j \in [k]$, $d^{H_1}(v_i, v_j) = d^{H_2}(u_i, u_j)$.

Given $\overline{v} \in V(H_1)^*$, and $\overline{u} \in V(H_2)^*$ lists of vertices of the same length, we define the game $d\text{EHR}_k(H_1,\overline{v};H_2,\overline{u})$ analogously to $\text{EHR}_k(H_1,\overline{v};H_2,\overline{u})$.

## 1.8 Outline of the proof

We show now an outline of the proof of theorem 1.1.

In essence the arguments mirror the ones in the proof of theorem 2.1 in [6], adapted to fit our context. Fix $r \in \mathbb{N}$. The following facts hold:

- A.a.s $G_n$ does not contain any dense hypergraph of diameter at most $2r+1$ (theorem 3.1). In consequence, all $r$-neighborhoods in $G_n$ a.a.s are either trees or unicycles.
- Given any fixed vertices $v_1, \dots, v_m \in \mathbb{N}$, a.a.s $d(v_i, v_j) > 2r+1$ for any two $v_i, v_j$ and a.a.s all the $N(v_i;r)$'s are trees (lemma 3.6).

For any given $k \in \mathbb{N}$, we define an equivalence relation $\sim_k$ for hypergraphs (section 2.2.1, and section 2.2.2) in a way that $H_1 \sim_k H_2$ implies that Duplicator wins $d\text{EHR}_k(H_1;H_2)$.

Given $r \in \mathbb{N}$, we define the ***r*-core** of an hypergraph $H$, written as $Core(H;r)$, the union of the $r$-neighborhoods of all saturated sub-hypergraphs of $H$ whose diameter is at most $2r+1$. We say that $H$ is $r$-simple if all connected components of $Core(H;r)$ are unicycles (definition 2.9)

For any given $k,r \in \mathbb{N}$ we say that $H_1 \approx_{k,r} H_2$ for two hypergraphs if $Core(H_1;r)$ and $Core(H_2;r)$ contain "the same number up to $k$" of connected components of each $\sim_k$ class (definition 2.10).

Given $k,r \in \mathbb{N}$ we say that an hypergraph $H$ is $(k,r)$-rich if, informally, it has enough "$r$-neighborhoods that are trees" of each $\sim_k$ class which are sufficiently far from each other and sufficiently far from any small saturated graph (definition 2.12).

We prove the following facts:

(1) Let $k \in \mathbb{N}$ and let $r := (3^k - 1)/2$. Let $H_1$ and $H_2$ be $(k,r)$-rich hypergraphs satisfying $H_1 \approx_{k,r} H_2$. Then Duplicator wins $\text{EHR}_k(H_1;H_2)$ (theorem 2.4).

(2) Let $r \in \mathbb{N}$. Then a.a.s $G_n$ is $r$-simple (corollary 3.1).

(3) Let $k,r \in \mathbb{N}$. Then a.a.s $G_n$ is $(k,r)$-rich (theorem 3.4).

(4) Let $k,r \in \mathbb{N}$. Let $\mathscr{O}$ be a $\approx_{k,r}$ class of $r$-simple hypergraphs. Then

$$\lim_{n\to\infty} \Pr\left(G_n\left(\{\beta_R\}_{R\in\sigma}\right) \in \mathscr{O}\right)$$

exists and is an analytic expression in $\{\beta_R\}_{R\in\sigma}$ (theorem 3.5).

Then an sketch of the proof of the main theorem theorem 1.1, given in section 4, is the following: Let $\Phi \in FO[\sigma]$ be a sentence and let $k := qr(\Phi)$, $r := (3^k - 1)/2$. Because of (1) and (3) it holds that for any $\approx_{k,r}$ class $\mathscr{O}$

$$\lim_{n\to\infty} \Pr\left(G_n \models \Phi \,\middle|\, G_n \in \mathscr{O}\right) = \text{ 0 or 1 }.$$

This together with (3) and the fact that there are a finite number of $\approx_{k,r}$-classes of $r$-simple hypergraphs imply that $\lim_{n\to\infty} \Pr\big(G_n \models \Phi\big)$ equals a finite sum of limits of the form $\lim_{n\to\infty} \Pr\big(G_n \in \mathscr{O}\big)$ where $\mathscr{O}$ is some $\approx_{k,r}$-class of $r$-simple hypergraphs. Finally, using (4) we get that $\lim_{n\to\infty} \Pr\big(G_n \models \Phi\big)$ exists and is an analytic expression in $\{\beta_R\}_{R\in\sigma}$, as we wanted.

# 2 Model theoretic results

## 2.1 Some winning strategies for Duplicator

**Definition 2.1.** Let $H_1$ and $H_2$ be hypergraphs, and let $V_1 := V(H_1), V_2 := V(H_2)$. Let $\overline{v} \in V_1^*, \overline{u} \in V_2^*$ be tuples of the same length. We say that $\overline{v}$ and $\overline{u}$ have $k$-**similar** $r$-neighborhoods, written as $(H_1,\overline{v}) \simeq_{k,r} (H_2\overline{u})$, if Duplicator wins $d\text{EHR}_k\big(N(\overline{v};r),\overline{v};N(\overline{u};r),\overline{u}\big)$. Given sets of vertices $X \subseteq V_1$ and $Y \subseteq V_2$ we say that $X$ and $Y$ have $k$-similar $r$-neighborhoods, written as $X \simeq_{k,r} Y$, if we can order their elements to form lists $\overline{v}$, resp. $\overline{u}$ such that $(H_1,\overline{v}) \simeq_{k,r} (H_2,\overline{u})$. Given sets of vertices $X \in V_1$, $Y \in V_2$ and tuples of the same length $\overline{v} \in V_1^*$ and $\overline{u} \in V_2^*$ we will say that $(X,\overline{v})$ and $(Y,\overline{u})$ have $k$-similar $r$-neighborhoods, written as $\big(H_1,(X,\overline{v})\big) \simeq_{k,r} \big(H_2,(Y,\overline{u})\big)$, if the elements of $X$ and $Y$ can be ordered in lists $\overline{w}$, $\overline{z}$ such that $(H_1,\overline{w}^\frown\overline{v}) \simeq_{k,r} (H_2,\overline{z}^\frown\overline{u})$.

**Definition 2.2.** Fix $r \in \mathbb{N}$. Suppose that $X \subseteq V_1$ and $Y \subseteq V_2$ can be partitioned into sets $X = X_1 \cup \cdots \cup X_a$ and $Y = Y_1 \cup \cdots \cup Y_b$ such that $N(X_i;r)$'s, and the $N(Y_i;r)$'s, are connected and disjoint.

We say that $X$ and $Y$ have $k$-**analogous** $r$-neighborhoods, written as $(H_1, X) \cong_{k,r} (H_2, Y)$, if for any set $Z \subset V_\delta$, with $\delta \in \{1,2\}$, among the $X_i$'s or the $Y_i$'s it is satisfied that "the number of $X_i$'s such that $(H_\delta, Z) \simeq_{k,r} (H_1, X_i)$" and "the number of $Y_i$'s such that $(H_\delta, Z) \simeq_{k,r} (H_2, Y_i)$" are both equal or are both greater than $k-1$.

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

**Theorem 2.1.** *Set $r = (3^k - 1)/2$. Let $H_1, H_2$ be hypergraphs and let $V_1 := V(H_1)$, $V_2 := V(H_2)$. Suppose there exist sets $X \subseteq V_1$, $Y \subseteq V_2$ with the following properties:*

*(1)* $(H_1, X) \cong_{k,r} (H_2, Y)$.

*(2)*
- *Let $r' \leq r$. Let $v \in V_1$ be a vertex such that $d(X, v) > 2r' + 1$. Let $\overline{u} \in (V_2)^{k-1}$ be a tuple of vertices. Then there exists $u \in V_2$ such that $d(u, \overline{u}) > 2r' + 1$, $d(Y, u) > 2r' + 1$ and $(H_1, v) \simeq_{k,r'} (H_2, u)$.*
- *Let $r' \leq r$. Let $u \in V_2$ be a vertex such that $d(Y, u) > 2r' + 1$. Let $\overline{v} \in (V_1)^{k-1}$ be a tuple of vertices. Then there exists $v \in V_1$ such that $d(v, \overline{v}) > 2r' + 1$, $d(X, v) > 2r' + 1$ and $(H_1, v) \simeq_{k,r'} (H_2, u)$*

*Then Duplicator wins* $\mathrm{EHR}_k\big(H_1; H_2\big)$.

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

**Observation 2.1.** *Let $H_1$, $H_2$ be hypergraphs and $\overline{v} \in V(H_1)^*$, $\overline{u} \in V(H_2)^*$, be lists of vertices. Suppose that Duplicator wins $d\mathrm{EHR}_k(H_1, \overline{v}; H_2, \overline{u})$. Then, for any $r \in \mathbb{N}$, $(H_1, \overline{v}) \simeq_{k,r} (H_2, \overline{u})$. A direct consequence of this fact is that given hypergraphs $H_1, H_2$ and sets $X \subseteq V(H_1)$, $Y \subseteq V(H_2)$ such that $(H_1, X) \simeq_{k,r} (H_2, Y)$ for some $r \in \mathbb{N}$, then for any $r' \leq r$ it also holds $(H_1, X) \simeq_{k,r'} (H_2, Y)$*

**Observation 2.2.** *Let $H_1$, $H_2$ be hypergraphs and let $\overline{v} \in V(H_1)^*$, $\overline{u} \in V(H_2)^*$ be lists of vertices. Suppose Duplicator wins $d\mathrm{EHR}_k(H_1, \overline{v}; H_2, \overline{u})$. 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 $d\mathrm{EHR}_{k-1}(H_1, \overline{v_2}; H_2, \overline{u_2})$, where $\overline{v_2} := \overline{v}^\frown v$ and $\overline{u_2} := \overline{u}^\frown u$.*

**Lemma 2.1.** *Let $H_1$, $H_2$ be hypergraphs and let $V_1 := V(H_1), V_2 := V(H_2)$. Let $\overline{v} \in V_1^*$ and $\overline{u} \in V_2^*$ be lists of vertices. Let $r \in \mathbb{N}$ be greater than zero. Suppose that $(H_1, \overline{v}) \simeq_{k,3r+1} (H_2, \overline{u})$. Let $v \in V_1$ and $u \in V_2$ be vertices played in the first round of an instance of*

$$d\mathrm{EHR}_k\big(N(\overline{v}; 3r+1), \overline{v}; \quad N(\overline{u}; 3r+1), \overline{u}\big)$$

*where Duplicator is following a winning strategy. Further suppose that $d(\overline{v}, v) \leq 2r + 1$ (and in consequence $d(\overline{u}, u) \leq 2r + 1$ as well). Let $\overline{v_2} := \overline{v}^\frown v$ and $\overline{u_2} := \overline{u} \frown u$. Then $(H_1, \overline{v_2}) \simeq_{k-1,r}$ $(H_2, \overline{u_2})$.*

*Proof.* Using observation 2.2 we get that Duplicator wins

$$d\mathrm{EHR}_{k-1}\big(N(\overline{v}; 3r+1), \overline{v_2}; \quad N(\overline{u}; 3r+1), \overline{u_2}\big)$$

as well. Call $H_1' = N(\overline{v}; 3r+1)$, $H_2' = N(\overline{u}; 3r+1)$. Then by observation 2.2 Duplicator wins

$$d\textsc{Ehr}_{k-1}\big(N^{H_1'}(\overline{v_2};r),\overline{v_2};\quad N^{H_2'}(\overline{u_2};r),\overline{u_2}\big).$$

Because of this if we prove $N^{H_1}(\overline{v_2};r) = N^{H_1'}(\overline{v_2};r)$ and $N^{H_2}(\overline{u_2};r) = N^{H_2'}(\overline{u_2};r)$, then we are finished. Let $z \in N^{H_1}(v';r)$. Then $d(z,\overline{v}) \le d(z,v') + d(v',\overline{v}) = 3r+1$. In consequence, $N^{H_1}(v;r) \subset H_1'$. Thus, $N^{H_1}(\overline{v_2};r) \subseteq H_1'$, and $N^{H_1}(\overline{v_2};r) = N^{H_1'}(\overline{v_2};r)$. Analogously we obtain $N^{H_2}(\overline{u_2};r) = N^{H_2'}(\overline{u_2};r)$, as we wanted. $\square$

---

Now we are in conditions to prove theorem 2.1.

*Proof of theorem 2.1.* Let $X_1,\ldots,X_a$ and $Y_1,\ldots,Y_b$ be partitions of $X$ and $Y$ respectively as in the definition of $\cong_{k,r}$. Define $r_0 = (3^k - 1)/2$ and $r_i = (r_{i-1} - 1)/3$ for each $1 \le i \le k$. Let us denote by $v_i^1$ and $v_i^2$ the vertices played in $H_1$ and $H_2$ respectively during the $i$-th round of $\textsc{Ehr}_k(H_1,H_2)$. We will show a winning strategy for Duplicator in $d\textsc{Ehr}_k\big(H_1;H_2\big)$. For each $0 \le i \le k$, Duplicator will keep track of some marked sets of vertices $T \subset V_1$, $S \subset V_2$. For $\delta = 1,2$ each marked set $T \subset V_\delta$ will have associated a tuple of vertices $\overline{v}(T) \in V_\delta^*$ consisting of the vertices played in $H_\delta$ so far that were "appropriately close" to $T$ when chosen, ordered according to the rounds they where played in. The game will start with no sets of vertices marked and at the end of the $i$-th round Duplicator will perform one of the two following operations:

- Mark two sets $S \subset V_1$ and $T \subset V_2$ and define $\overline{v}(S) := v_i^1$ and $\overline{v}(T) := v_i^2$.
- Given two sets $S \subset V_1$, $T \subset V_2$ that were previously marked during the same round, append $v_i^1$ and $v_i^2$ to $\overline{v}(S)$ and $\overline{v}(T)$ respectively.

In particular this means that at the end of the $i$-th round the marked sets $S \subset V_1$ , $T \subset V_2$ and their respective lists $\overline{v}(S)$, $\overline{v}(T)$ satisfy

(i) For $\delta = 1,2$, each vertex played so far $v_j^\delta \in V_\delta$ belongs to $\overline{v}(S)$ for a unique marked set $S \subset V_\delta$.

(ii) Let $S \subset V_1$ and $T \subset V_2$ be sets marked during the same round. Then any previously played vertex $v_j^1$ occupies a position in $\overline{v}(S)$ if and only if $v_j^2$ occupies the same position in $\overline{v}(T)$.

The following conditions will also be satisfied at the end of the $i$-th round

(iii)
- Let $S \subset V_1$ be a marked set. Then for any different marked $S' \subset V_1$ of any different $S'$ among $X_1,\ldots,X_a$ it holds $d(S,S') > 2r_i + 1$.
- Let $T \subset V_2$ be a marked set. Then for any different marked $T' \subset V_2$ or any different $T'$ among $Y_1,\ldots,Y_b$ it holds $d(T,T') > 2r_i + 1$.

(iv) Let $S \subset V_1, T \subset V_2$ be sets marked during the same round. Then

$$\big(H_1,(S,\overline{v}(S))\big) \simeq_{k-i,r_i} \big(H_2,(T,\overline{v}(T))\big).$$

In particular, if conditions (i) to (iv) are satisfied this means that if $\overline{v}^1 := (v_1^1, \dots, v_i^1)$ and $\overline{v}^2 := (v_1^2, \dots, v_i^2)$ are the vertices played so far then Duplicator wins

$$d\text{EHR}_{k-i}\big(N(\overline{v}^1; r_i), \overline{v}^1; \quad N(\overline{v}^2; r_i), \overline{v}^2\big),$$

And at the end of the $k$-th round Duplicator will have won $\text{EHR}(H_1; H_2)$.

The game $d\text{EHR}_k(H_1; H_2)$ proceeds as follows. Clearly properties (i) to (iv) hold at the beginning of the game. Suppose that Duplicator can play in such a way that properties (i) to (iv) hold until the beginning of the $i$-th round. Suppose that during the $i$-th round Spoiler chooses $v_i^1 \in V_1$ (the case where they play in $V_2$ is symmetric). There are three possible cases:

- For some unique previously marked set $S \subset V_1$ it holds that $d(S \cup \overline{v}, \quad v_i^1) \leq 2r_i + 1$. In this case let $T \subset V_2$ be the set in $H_2$ marked in the same round as $T$. By hypothesis
$$\big(H_1, (S, \overline{v}(S))\big) \simeq_{k-i+1, 3r_i+1} \big(H_2, (T, \overline{v}(T))\big).$$
Then, by definition, for some orderings $\overline{w}$, $\overline{z}$ of the vertices in $S$ and $T$ respectively it holds that Duplicator wins
$$d\text{EHR}_{k-i+1}\big(N(\overline{w}^\frown \overline{v}(S); 3r_i + 1), \overline{w}^\frown \overline{v}(S); \quad N(\overline{z}^\frown \overline{v}(T); 3r_i + 1), \overline{z}^\frown \overline{v}(T)\big).$$
Thus Duplicator can choose $v_i^2 \in V_2$ according to the winning strategy in that game. After this Duplicator sets $\overline{v}(S) := \overline{v}(S)^\frown v_i^1$, and $\overline{v}(T) := \overline{v}(T)^\frown v_i^2$. Notice that because of lemma 2.1 now
$$\big(H_1, (S, \overline{v}(S))\big) \simeq_{k-i, r_i} \big(H_2, (T, \overline{v}(T))\big).$$
- For all marked sets $S \subset V_1$ it holds $d(S \cup \overline{v}(S), \quad v_i^1) > 2r_i + 1$, but there is a unique $S$ among $X_1, \dots, X_a$ such that $d(S, v_i^1) \leq 2r_i + 1$. In this case from condition (1) of the statement follows that there is some non-marked set $T$ among $Y_1, \dots, Y_b$ such that
$$(H_1, S) \simeq_{k-i+1, 3r_i+1} (H_2, T).$$
Thus, by definition, for some orderings $\overline{w}$, $\overline{z}$ of the vertices in $S$ and $T$ respectively it holds that Duplicator wins
$$d\text{EHR}_{k-i+1}\big(N(\overline{w}; 3r_i + 1), \overline{w}; \quad N(\overline{z}; 3r_i + 1), \overline{z}\big).$$
Then Duplicator can choose $v_i^2 \in V_2$ according to a winning strategy for this game. After this Duplicator marks both $S$ and $T$ and sets $\overline{v}(S) := v_i^1$, and $\overline{v}(T) := v_i^2$. Notice that because of lemma 2.1 now
$$\big(H_1, (S, \overline{v}(S))\big) \simeq_{k-i, r_i} \big(H_2, (T, \overline{v}(T))\big).$$
- For all marked sets $S \subset V_1$ it holds $d(S \cup \overline{v}(S), \quad v_i^1) > 2r_i + 1$, and for all sets $S$ among $X_1, \dots, X_a$ it also holds $d(S, v_i^1) > 2r_i + 1$. In this case from condition (2) of the statement follows that Duplicator can choose $v_i^2 \in V_2$ such that (A) $d(T \cup \overline{v}(T), \quad v_i^2) > 2r_i + 1$ for all marked sets $T \subset V_2$, (B) $d(T, v_i^2) > 2r_i + 1$ for all sets $T$ among $Y_1, \dots, Y_b$, and (C) $(H_1, v_i^1) \simeq_{k-i, r_i} (H_2, v_i^2)$. After this Duplicator marks both $S = \{v_i^1\}$ and $T = \{v_i^2\}$ and sets $\overline{v}(S) := v_i^1$, and $\overline{v}(T) := v_i^2$.

The fact that conditions (i) to (iv) still hold at the end of the round follows from comparing $r_{i-1}$ and $r_i$ as well as applying observation 2.1 and observation 2.2.

□

## 2.2 k-Equivalence relation

### 2.2.1 k-Equivalent trees

**Definition 2.3.** A **rooted tree** $(T,v)$ is 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 $(T,v)$. The **initial edges** of a rooted tree $(T,v)$ are the edges in $T$ that contain $v$.

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

We define the radius of a rooted tree as the maximum distance between its root and any other vertex.

**Definition 2.4.** Fix a natural number $k$. We define the $k$**-equivalence** relation over rooted trees, written as $\sim_k$, by induction over their radii as follows:

- Any two trees with radius zero are $k$-equivalent. Notice that those trees consist only of one vertex: their respective roots.
- Let $r > 0$. Suppose that the $k$-equivalence relation has been defined for rooted trees with radius at most $r-1$. Let $\Sigma_{k,r-1}$ be the set consisting of the $k$-equivalence classes of trees with radius at most $r-1$. Let $\rho$ be an special symbol called the **root symbol**. Given a rooted tree $(T,v)$ with whose radius is $r$ we define its **canonical** $(\Sigma_{k,r-1} \cup \{\rho\})$**-coloring** as the map $\tau_{(T,v)} : V(T) \to \Sigma_{k,r-1} \cup \{\rho\}$ satisfying that $\tau_{(T,v)}(u)$ is the $\sim_k$ class of $\mathrm{Tr}(T,u;\, v)$ for any $u \neq v$, and $\tau_{(T,v)}(v) = \tau$.

  Let $T_1$ and $T_2$ be rooted trees with radius $r$. We say that $(T_1,v_1) \sim_k (T_2,v_2)$ if for any $\delta = 1,2$, given any initial edge $e$ from $(T_\delta, v_\delta)$ the "quantity of initial edges $e_1 \in E(T_1)$ such that
  $$\big(e, \tau_{(T_\delta,v_\delta)}\big) \simeq \big(e_1, \tau_{(T_1,v_1)}\big)\text{"}$$
  and the "quantity of initial edges $e_2 \in E(T_2)$ such that
  $$\big(e, \tau_{(T_\delta,v_\delta)}\big) \simeq \big(e_2, \tau_{(T_2,v_2)}\big)\text{"}$$
  are equal or are both greater than $k-1$.

Here we are making an slight abuse of notation in the expressions of the type $\big(e_1, \tau_{(T_1,v_1)}\big)$. There $e$ denotes the sub-hypergraph of $T_1$ whose vertices are the ones in $V(e)$ and whose only edge is $e$. This way $\big(e_1, \tau_{(T_1,v_1)}\big)$ is a $(\Sigma_{(k,r-1)} \cup \{\rho\})$-hypergraph, as defined in definition 1.3.

**Definition 2.5.** Let $\Sigma_{(k,r-1)}$ be the set of $\sim_k$ classes of rooted trees with radii at most $r-1$. Then we call a $(k,r)$-**pattern** to an isomorphism class of $(\Sigma_{(k,r-1)} \cup \{\rho\})$-hypergraphs $(e,\tau)$ that consist of only one edge and satisfy $\tau(v) = \rho$ for exactly one vertex $v \in V(e)$. We will denote by $P(k,r)$ the set of $(k,r)$-patterns.

The following is a way of characterizing $\sim_k$ classes of rooted trees with radii at most $r$ that will be useful later.

**Observation 2.3.** *Let $\mathscr{T}$ be a $\sim_k$ class of rooted trees with radii at most $r$. Then there is a partition $E^1_{\mathscr{T}}, E^2_{\mathscr{T}}$ of $P(k,r)$ and natural numbers $a_\varepsilon < k$ for each $\varepsilon \in E^2_{\mathscr{T}}$ that only depend on $\mathscr{T}$ such that any rooted tree $(T,v)$ belongs to $\mathscr{T}$ if and only if it holds that*

*(1) For any pattern $\varepsilon \in E^1_{\mathscr{T}}$ there are at least $k$ initial edges $e \in E(T)$ such that $(e, \tau_{(T,v)}) \in \varepsilon$.*

*(2) For any pattern $\varepsilon \in E^2_{\mathscr{T}}$ there are exactly $a_\varepsilon$ initial edges $e \in E(T)$ such that $(e, \tau_{(T,v)}) \in \varepsilon$.*

**Observation 2.4.** *Using last characterization of $\sim_k$ classes it is easy to show that for any $r \in \mathbb{N}$ the quantity of $\sim_k$ classes of trees with radii at most $r$ is finite. We proceed by induction. For $r = 0$ there is only one $\sim_k$ class. Now let $r > 0$ and suppose that the statement holds for $r-1$. Then the number of $(k,r)$-patterns is finite and so is the number of $\sim_k$ classes of trees with radii at most $r$.*

We want prove the following

**Theorem 2.2.** *Let $(T_1,v_1)$ and $(T_2,v_2)$ be rooted trees such that $(T_1,v_1) \sim_k (T_2,v_2)$. Then Duplicator wins $d\text{EHR}_k(T_1,v_1;\ T_2,v_2)$.*

Before proceeding with the proof we need an auxiliary result. Let $(T,v)$ be a rooted tree and $e$ an initial edge of $T$. We define $\mathrm{Tr}(T,v;\,e)$ as the induced tree $T[X]$ on the set $X := \{v\} \cup \{\, u \in V(T) \,|\, d(v,u) = 1 + d(e,u)\,\}$, to which we assign $v$ as the root. In other words, $\mathrm{Tr}\big(T,v;\,e\big)$ 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 2.2.** *Fix $r > 0$. Suppose that theorem 2.2 holds for rooted trees with radii at most $r$. Let $(T_1,v_1)$ and $(T_2,v_2)$ be rooted trees with radius $r+1$. Let $\tau_{(T_1,v_1)}$ and $\tau_{(T_2,v_2)}$ be colorings over $T_1$ and $T_2$ as in the definition of k-equivalence. Let $e_1$ and $e_2$ be initial edges of $T_1$ and $T_2$ respectively satisfying $(e_1, \tau_{(T_1,v_1)}) \simeq (e_2, \tau_{(T_2,v_2)})$. Name $T_1' := \mathrm{Tr}(T_1,v_1;\ e_1)$ and $T_2' := \mathrm{Tr}(T_2,v_2;\ e_2)$. Then Duplicator wins $d\text{EHR}_k(T_1',v_1;\ T_2',v_2)$.*

*Proof.* We show a winning strategy for Duplicator. At the beginning of the game fix $f : V(e_1) \to V(e_2)$ an isomorphism between $(e_1, \tau_{(T_1,v_1)})$ and $(e_2, \tau_{(T_2,v_2)})$. Suppose that in the $i$-th round of the game Spoiler plays on $T_1'$. The other case is symmetric. There are two possibilities:

- If Spoiler plays $v_1$ then Duplicator chooses $v_2$.
- Otherwise, Spoiler plays a vertex $v$ that belongs to some $\mathrm{Tr}(T_1',v_1;\ u)$ for a unique $u \in V(e_1)$ different from the root $v_1$. Set $T_1'' := \mathrm{Tr}\big(T_1',v_1;\ u\big)$ and $T_2'' := \mathrm{Tr}\big(T_2',v_2;\ f(u)\big)$ Then, as $\tau_{(T_1,v_1)}\big(u\big) = \tau_{(T_2,v_2)}\big(f(u)\big)$, we obtain $\big(T_1'',u\big) \sim_k \big(T_2'',f(u)\big)$. As both these trees have radii at most $r$, by assumption Duplicator has a winning strategy in

$$d\text{EHR}_k\big(T_1'',u;\quad T_2'',f(u)\big)$$

and they can follow it considering the previous plays in $T_1''$ and $T_2''$.

□

Now we can prove the main theorem of this section:

*Proof of theorem 2.2.*

Notice that, as $(T_1,v_1) \sim_k (T_2,v_2)$, 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 $r$. Let $\tau_{(T_1,v_1)}$ and $\tau_{(T_2,v_2)}$ be the colorings over $T_1$ and $T_2$ defined as in the definition of $\sim_k$. 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 $\text{Tr}(T_1,v_1;\, e_1)$ for a unique initial edge $e_1$ of $T_1$. There are two possibilities:
  - If $e_1$ is not marked yet, mark it. In this case, there is a non-marked initial edge $e_2$ in $T_2$ satisfying
  $$\big(e_1,\tau_{(T_1,v_1)}\big) \simeq \big(e_2,\tau_{(T_2,v_2)}\big).$$
  Mark $e_2$ as well. Set $T_1' := \text{Tr}(T_1,v_1;\, e_1)$ and $T_2' := \text{Tr}(T_2,v_2;\, e_2)$ Because of lemma 2.2, Duplicator has a winning strategy in $d\text{EHR}k(T_1',v_1;\, T_2',v_2)$ and can play according to it.
  - If $e_1$ is already marked then there is a unique initial edge $e_2$ in $T_2$ that was marked during the same round as $e_1$ and it satisfies
  $$\big(e_1,\tau_{(T_1,v_1)}\big) \simeq \big(e_2,\tau_{(T_2,v_2)}\big).$$
  Again, Because of lemma 2.2, Duplicator has a winning strategy in $d\text{EHR}k(T_1',v_1;\, T_2',v_2)$ and can continue playing according to it taking into account the plays made previously in $T_1'$ and $T_2'$.

$\square$

### 2.2.2 k-Equivalent hypergraphs

**Definition 2.6.** The **center** of an hypergraph $H$, written as $Center(H)$, is the union of all saturated sub-hypergraphs of $H$.

Let $H$ be an hypergraph, let $\overline{v} \in V(H)^*$. If $H$ is connected then we define the graph $Center(H,\overline{v})$ as the minimal connected sub-hypergaph of $H$ that contains both $Center(H)$ and the vertices in $\overline{v}$. Otherwise, if $H$ is a general hypergraph we define $Center(H,\overline{v})$, as the union, for all connected components $H' \subset H$, of the minimal connected sub-hypergraph of $H'$ that contains both $Center(H')$ and the vertices in $\overline{v}$ that belong to $V(H')$.

**Definition 2.7.** Let $H$ be an hypergraph, let $\overline{v} \in V(H)^*$ and let $v \in H$. We define $\text{Tr}\big(H,\overline{v};\, v\big)$ in the following way:

- If $d\big(Center(H,\overline{v}),\, v\big) = \infty$ then $v$ belongs to a connected component $T$ of $H$ which is a tree and does not contain any vertex in $\overline{v}$. In this case $\text{Tr}\big(H,\overline{v};\, v\big)$ is the tree $T$ rooted at $v$.

- Otherwise $\mathrm{Tr}\big(H,\overline{v};\ v\big)$ is the tree $H[X]$ induced on the set

$$X := \big\{u \in V(H)\ \big|\ d\big(Center(H,\overline{v}),\ u\big) = d\big(Center(H,\overline{v}),\ v\big) + d(v,u)\big\},$$

to which we assign $v$ as a root. That is, $\mathrm{Tr}\big(H,\overline{v};\ v\big)$ is the tree formed of all vertices whose only path to $Center(H,\overline{v})$ contains $v$.

In the case that $\overline{v}$ is the empty list we will write simply $\mathrm{Tr}(H;\ v)$ instead of $\mathrm{Tr}(H,\ ;\ v)$. Notice that in the case that $(T,u)$ is a rooted tree then the definition of $\mathrm{Tr}(T,u;\ v)$ given in section 2.2.1 coincides with the one we have given now, so no confusion should arise.

**Definition 2.8.** Let $H_1$ and $H_2$ be connected hypergraphs which are not trees. Set $H_1' := Center(H_1)$ and $H_2' := Center(H_2)$. We say that $H_1$ and $H_2$ are $k$-equivalent, written as $H_1 \sim_k H_2$, if there is an hypergraph isomorphism $f : V(H_1') \to V(H_2')$ such that for all vertices $v \in V(H_1')$ it holds that

$$\mathrm{Tr}\big(H_1;\ v\big) \sim_k \mathrm{Tr}\big(H_2;\ f(v)\big).$$

Let $H$ be a non-tree connected hypergraph. We define the canonical coloring $\tau_H$ over $Center(H)$ as the one that assigns to each vertex $v \in V\big(Center(H)\big)$ the $\sim_k$ class of the tree $\mathrm{Tr}(H,v)$.

An equivalent definition of $\sim_k$ for non tree connected hypergraphs is the following: Let $H_1$ and $H_2$ be connected hypergraphs which are not trees. Set $H_1' := Center(H_1)$ and $H_2' := Center(H_2)$. Then we say that $H_1 \sim_k H_2$ if $\big(H_1', \tau_{H_1}\big) \simeq \big(H_2', \tau_{H_2}\big)$.

The main theorem of this section is the following

**Theorem 2.3.** *Let $H_1$ and $H_2$ be non-tree connected hypergraphs satisfying $H_1 \sim_k H_2$. Set $H_1' := Center(H_1)$ and $H_2' := Center(H_2)$. Let $f$ be an isomorphism between $\big(H_1', \tau_{H_1}\big)$ and $\big(H_2', \tau_{H_2}\big)$. Let $\overline{v_1}$ be an ordering of the vertices of $H_1'$ and let $\overline{v_2} := f(\overline{v_1})$ be the corresponding ordering of the vertices of $H_2'$. Then Duplicator wins*

$$d\mathrm{EHR}_k\big(H_1', \overline{v_1};\quad H_2', \overline{v_2}\big).$$

*Proof.* The winning strategy for Duplicator is as follows. Suppose that at the beginning of the $i$-th round Spoiler plays in $H_1$ (the case where they play in $H_2$ is symmetric). Then Spoiler has chosen a vertex that belongs to $\mathrm{Tr}(H_1;\ u)$ for a unique $u \in H_1'$. Set $T_1 := \mathrm{Tr}\big(H_1;\ u\big)$ and $T_2 := \mathrm{Tr}\big(H_2;\ f(u)\big)$. By hypothesis $(T_1,u) \sim_k (T_2, f(u))$. Then because of theorem 2.2 we have that Duplicator has a winning strategy in

$$d\mathrm{EHR}_k\big(T_1, u;\quad T_2, f(u)\big),$$

and they can follow it taking into account the previous plays made in $T_1$ and $T_2$. In particular, if Spoiler has chosen $u$ then Duplicator will necessarily choose $f(u)$.

One can easily check that distances are preserved following this strategy. □

## 2.3 $r$-Cores

**Definition 2.9.** Let $H$ be an hypergraph. We define its $r$**-core**, written as $Core(H;r)$ as $N(X;r)$, where $X$ is the set of vertices $v \in V(H)$ that belong to any saturated sub-hypergraph of $H$ whose diameter is at most $2r+1$. Equivalently, $Core(H;r) = N\big(Center(H;r);r\big)$. Given $\overline{v} \in V(G)^*$, we define $Core(H,\overline{v};r)$ as $N(Y;r)$, where $Y$ is the set of vertices $v \in V(H)$ that either belong to $\overline{v}$ or belong to any saturated sub-hypergraph of $H$ whose diameter is at most $2r+1$.

We say that $H$ is $r$-**simple** if all connected components of $Core(H;r)$ are unicycles.

**Definition 2.10.** Let $H_1$ and $H_2$ be hypergraphs and let $r \in \mathbb{N}$. Let $H_1' := Core(H_1;r)$ and $H_2' := Core(H_2;r)$. We say that $H_1$ and $H_2$ are $(k,r)$-agreeable, written as $H_1 \approx_{k,r} H_2$ if for any $\sim_k$ class $\mathscr{H}$ "the number of connected components in $H_1'$ that belong to $\mathscr{H}$" and "the number of connected components in $H_2'$ that belong to $\mathscr{H}$" are the same or are both greater than $k-1$.

**Definition 2.11.** Let $\Sigma_{(k,r-1)}$ be the set of $\sim_k$ classes of rooted trees with radii at most $r-1$. Then we call a $(k,r)$-**cycle** to an isomorphism class of $\Sigma_{(k,r-1)}$-hypergraphs $(H,\tau)$ that are cycles of diameter at most $2r+1$. We will denote by $C(k,r)$ the set of $(k,r)$-cycles.

The following is a way of characterizing $\approx_{k,r}$ classes of $r$-simple hypergraphs.

**Observation 2.5.** *Let $\mathscr{O}$ be a $\approx_{k,r}$ class of r-simple hypergraphs. Then there is a partition $U^1_{\mathscr{O}}, U^2_{\mathscr{O}}$ of $C(k,r)$ and natural numbers $a_\omega < k$ for each $\omega \in U^2_{\mathscr{O}}$ that only depend on $\mathscr{O}$ such that any r-simple hypergraph G belongs to $\mathscr{O}$ if and only if it holds that*

*(1) For any $\omega \in U^1_{\mathscr{O}}$ there are at least k connected components $H \subset Core(H;r)$ whose cycle $H' = Center(H)$ satisfies that $(H',\tau_H) \in \omega$.*

*(2) For any $\omega \in U^2_{\mathscr{O}}$ there are exactly $a_\omega$ connected components $H \subset Core(H;r)$ whose cycle $H' = Center(H)$ satisfies that $(H',\tau_H) \in \omega$.*

**Lemma 2.3.** *Let $r \in \mathbb{N}$ and let $H_1,H_2$ be hypergraphs such that $H_1 \approx_{k,r} H_2$. Let X and Y be the sets of vertices in $H_1$, resp. $H_2$, that belong to any saturated sub-hypergraph of diameter at most $2r+1$. Then $(H_1,X) \cong_{k,r} (H_2,Y)$ in the sense of definition 2.2.*

*Proof.* Let $X_1,\dots,X_a$ and $Y_1,\dots,Y_b$ be partitions of $X$ and $Y$ such that each $N(X_i;r)$ and $N(Y_i;r)$ is a connected component of $Core(H_1;r)$, resp. $Core(H_2;r)$. Using the definition of $H_1 \approx_{k,r} H_2$ as well as the fact that because of theorem 2.3 $N(X_i;r) \sim_k N(Y_j;r)$ implies $(H_1,X_i) \simeq_{k,r} (H_2,Y_j)$ the result follows. □

---

**Definition 2.12.** Let $H$ be an hypergraph and let $r \in \mathbb{N}$. Let $X \subset V(H)$ be the set of vertices in $H$ belonging to some saturated sub-hypergraph of diameter at most $2r+1$. We say that $H$ is $(k,r)$-**rich** if for any $r' \leq r$, any vertices $v_1,\dots,v_k$ and any $\sim_k$ class $\mathscr{T}$ of trees with radius at most $r'$ it holds that there exists a vertex $v \in V(H)$ such that $d(v,X) > 2r'+1$, $d(v,v_i) > 2r'+1$ for all $v_i$'s and that $T := N(v;r')$ is a tree such that $(T,v) \in \mathscr{T}$ (notice that $T$ is a tree necessarily. Otherwise $T$ contains a saturated sub-hypergraph with diameter lesser or equal than $2r'+1$).

**Theorem 2.4.** *Let $H_1$, $H_2$ be hypergraphs. Let $r := (3^k-1)/2$. Suppose that both $H_1$ and $H_2$ are $(k,r)$-rich and $H_1 \approx_{k,r} H_2$. Then Duplicator wins* $\textsc{Ehr}_k(H_1,H_2)$.

*Proof.* Because of the previous lemma we can apply theorem 2.1 with $X \subset V(H_1)$ and $Y \subset V(H_2)$ the sets of vertices that belong to some saturated sub-hypergraph of $H_1$ or $H_2$ respectively with diameter at most $2r+1$. □

# 3 Probabilistic results

## 3.1 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 $r$-**sparse** 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 $r$-sparse.

**Lemma 3.1.** *Let $H$ be an hypergraph. Then $\mathrm{E}\big[\#\text{ copies of } H \text{ in } G_n\big] \sim C \cdot (n^{-ex(H)})$ for some constant $C \in \mathbb{R}$ as $n$ tends to infinity.*

*Proof.* It holds

$$\mathrm{E}\big[\#\text{ copies of } H \text{ in } G_n\big] = \sum_{H' \in Copies(H,[n])} \Pr\big(H' \subset G_n\big).$$

We have that $|Copies(H,[n])| = \frac{(n)_{v(H)}}{\mathrm{aut}(H)}$. Also, for any $H' \in Copies(H,[n])$ it is satisfied

$$\Pr\big(H' \subset G_n\big) = \prod_{R \in \sigma} \left(\frac{\beta_R}{n^{ar(R)-1}}\right)^{|E_R(H)|}.$$

Substituting in the first equation we get

$$\mathrm{E}\big[\#\text{ copies of } H \text{ in } G_n\big] = \frac{(n)_{v(H)}}{\mathrm{aut}(H)} \cdot \prod_{R \in \sigma} \left(\frac{\beta_R}{n^{ar(R)-1}}\right)^{|E_R(H)|} \sim n^{-ex(H)} \cdot \frac{\prod_{R \in \sigma} \beta_R^{|E_R H|}}{\mathrm{aut}(H)}.$$

□

---

As a corollary of last result we get the following:

**Lemma 3.2.** *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, $\mathrm{E}\big[\#\text{ copies of } H \text{ in } G_n\big] \xrightarrow{n\to\infty} 0$ . An application of the first moment method yields the desired result. □

---

A similar result that will be useful later is the following:

**Lemma 3.3.** *Let $H$ be an hypergraph. Let $\overline{v} \in (\mathbb{N})_*$ be a list of vertices with $len(\overline{v}) \leq |V(H)|$. For each $n \in \mathbb{N}$ let $X_n$ be the random variable that counts the copies of $H$ in $G_n$ that contain the vertices in $\overline{v}$. Then*

$$\mathrm{E}\big[X_n\big] = \Theta(n^{-ex(H)-len(\overline{v})}).$$

*In particular, given any fixed $r \in \mathbb{N}$, a.a.s the vertices $v \in \overline{v}$ satisfy that the $N(v;r)$'s are disjoint trees.*

*Proof.* It holds that the number of hypergraphs $H' \in Copies(H,[n])$ that contain all vertices in $\overline{v}$ is $\Theta\left(n^{|V(H)|-len(\overline{v})}\right)$. Then for some constant $C$,

$$\mathrm{E}\left[X_n\right] \sim Cn^{|V(H)|-len(\overline{v})} \cdot \prod_{R\in\tau}\left(\frac{\beta_R}{n^{ar(R)-1}}\right)^{e_R(H)} = n^{-ex(H)-len(\overline{v})} \cdot C \cdot \prod_{R\in\tau}(\beta_R)^{e_R(H)}.$$

□

---

The main theorem of this section is the following

**Theorem 3.1.** *Let $r \in \mathbb{N}$. Then a.a.s $G_n$ is $r$-sparse.*

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 $r$-sparse.

Given an hypergraph $H$ and an edge $e \in E(H)$ we define the operation of **cutting** the edge $e$ as removing $e$ from $H$ and then removing any isolated vertices from the resulting hypergraph.

**Lemma 3.4.** *Let $G$ be a dense hypergraph with diameter at most $r$. And let $H \subset G$ be a connected sub-hypergraph with $ex(H) < ex(G)$. Then there is a connected sub-hypergraph $H' \subset G$ satisfying $H \subset H'$, $ex(H) < ex(H')$ and that $|E(H')| \leq |E(H)| + 2\cdot(r)+1$,*

*Proof.* Suppose that there is some edge $e \in E(G) \setminus E(H)$ with and $ex(e) \geq 0$. Let $P$ be a path of length at most $r$ joining $H$ and $e$ in $G$. Then $H' := H \cup P \cup e$ satisfies the conditions of the statement.

Otherwise, all edges $e \in E(G) \setminus E(H)$ satisfy $ex(e) = -1$. In this case we successively cut edges $e$ from $G$ such that $d(e,H)$ is the maximum possible (notice that this always yields a connected hypergraph) until we obtain an hypergraph $G'$ with $ex(G') < ex(G)$. Let $e$ be the edge that was cut last. Then $V(G') \cap V(e) = ex(G) - ex(G') + 1 \geq 2$. Let $v_1, v_2 \in V(G') \cap V(e)$, and let $P_1$, $P_2$ be paths of length at most $r$ that join $H$ with $v_1$ and $v_2$ respectively in $G'$. Then the hypergraph $H' := H \cup e \cup P^1 \cup P^2$ satisfies the conditions in the statement. □

---

**Lemma 3.5.** *Let $G$ be a dense hypergraph of diameter at most $r$. Then $G$ contains a connected dense sub-hypergraph $H$ with $|E(H)| \leq 4r+2$.*

*Proof.* Apply the previous lemma twice in a row starting with $G$ and taking as $H$ a sub-hypergraph of $G$ consisting of a single vertex and no edges. □

---

In particular, if we define $l := \max_{R\in\sigma} ar(R)$ then last lemma implies that if $G$ is a dense hypergraph whose diameter is at most $r$ then $G$ contains a dense sub-hypergraph $H$ with $|V(H)| \leq l\cdot(4r+2)$.

Now we are in conditions to prove theorem 3.1.

*Proof.* Because of last lemma there is a constant $R$ such that "$G$ does not contain dense hypergraphs of size bounded by $R$" implies that "$G$ is $r$-sparse". Thus,

$$\lim_{n\to\infty} \Pr\big(G_n \text{ is } r\text{-sparse}\big) \geq \lim_{n\to\infty} \Pr\big(G_n \text{ does not contain dense hypergraphs of size bounded by } R\big).$$

Because of lemma 3.2, given any individual dense hypergraph, the probability that there are no copies of it in $G_n$ tends to 1 as $n$ goes to infinity. Using that there are a finite number of $\sim$ classes of dense hypergraphs whose size bounded by $R$ we deduce that the RHS of last inequality tends to 1. □

---

**Corollary 3.1.** *Let $r \in \mathbb{N}$. Then a.a.s $G_n$ is r-simple.*

*Proof.* If some connected component of $Core(G_n; r)$ is not a cycle that means that either $G_n$ contains a dense hypergraph of diameter at most $4r+1$ or that $G_n$ contains two cycles of diameter at most $2r+1$ that are at a distance at most $2r+1$. In the second case, considering the two cycles and the path joining them, $G_n$ contains a dense hypergraph of diameter bounded by $6r+3$. In consequence the fact that $G_n$ is $(6r+3)$-sparse implies that $G_n$ is $r$-simple. Because of the previous theorem $G_n$ is a.a.s $(6r+3)$-sparse and the result follows. □

**Lemma 3.6.** *Let $\overline{v} \in (\mathbb{N})_*$ and let $r \in \mathbb{N}$. Then a.a.s, for all vertices $v \in \overline{v}$ the neighborhoods $N(v;r)$'s are all trees and they are all disjoint.*

*Proof.* An application of the first moment method together with lemma 3.3 and the fact that there are a finite number of $\sim$ classes of paths whose length is at most $2r+1$ implies that a.a.s the $N(v;r)$'s are disjoint.

Also, because of theorem 3.1 a.a.s the $N(v;r)$'s are either trees or unicycles. But if any of the $N(v;r)$'s was an unicycle then that would mean that in $G_n$ there exists a path $P$ of length at most $2r+1$ joining some vertex $v \in \overline{v}$ with a cycle $C$ of diameter at most $2r+1$. Using lemma 3.3 again as well as the fact that the number of $\sim$ classes of the possible hypergraphs $P \cup C$ is finite we obtain that a.a.s no such $P$ and $C$ exist. In consequence all the $N(v;r)$'s are disjoint trees as we wanted to prove.

□

## 3.2 Convergence to Poisson variables

Our main tool for computing probabilities in the following sections will be the following multivariate version of Brun's Sieve (Theorem 1.23, [11]).

**Theorem 3.2.** *Fix $k \in \mathbb{N}$. For each $n \in \mathbb{N}$, let $X_{n,1}, \dots, X_{n,l}$ be non-negative random integer variables over the same probability space. Let $\lambda_1, \dots, \lambda_l$ be real numbers. Suppose that for any $r_1, \dots, r_l \in \mathbb{N}$*

$$\lim_{n\to\infty} \mathrm{E}\left[\prod_{i=1}^{l} \binom{X_{n,i}}{r_i}\right] = \prod_{i=1}^{l} \frac{\lambda_i}{r_i!}.$$

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

To make use of last theorem we will need to employ the following observation.

**Observation 3.1.** *Let $X_1,\dots,X_l$ be non negative random integer variables over the same space. Let $r_1,\dots,r_l \in \mathbb{N}$. Suppose that each $X_i$ is the sum of various indicator random variables (i.e. variables that only take the values 0 and 1) $X_i = \sum_{j=1}^{a_i} Y_{i,j}$. Define $\Omega := \prod_{i=1}^{l} \binom{[a_i]}{r_i}$. That is, the elements $\{S_i\}_{i\in[l]} \in \Omega$ represent all the possible unordered choices of $r_i$ indicator variables $Y_{i,j}$ for each $i \in [l]$. Then*

$$\mathrm{E}\left[\prod_{i=1}^{l}\binom{X_i}{r_i}\right] = \sum_{\{S_i\}_{i\in[l]}} \Pr\left(\bigwedge_{\substack{i\in[l]\\ j\in S_i}} Y_{i,j}=1\right).$$

## 3.3 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 $\mathscr{T}$ with radius at most $r$. That is, we want to know

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

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

**Definition 3.1.** We define $\Lambda$ and $M$ as the minimal families of expressions with arguments $\{\beta_R\}_{R\in\sigma}$ that satisfy the conditions: **(1)** $1\in\Lambda$, **(2)** for any $R\in\sigma$, any positive $b\in\mathbb{N}$, and $\overline{\lambda}\in\Lambda^*$, the expression $(\beta_R/b)\prod_{\lambda\in\overline{\lambda}}\lambda$ belongs to $M$, **(3)** for any $\mu\in M$ and any $n\in\mathbb{N}$ both $\mathrm{Poiss}_\mu(n)$ and $\mathrm{Poiss}_\mu(\geq n)$ 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 that $\Pr[r,\mathscr{T}]$, as an expression with parameters $\{\beta_R\}_{R\in\sigma}$, belongs to $\Lambda$ for any choice of $r$ and $\mathscr{T}$.

Before we proceed it will be useful to define the following abbreviation

**Definition 3.2.** Let $H$ be an hypergraph, $\overline{v}\in V(H)^*$, $v\in V(H)$ and $r\in\mathbb{N}$. Then we define $\mathrm{Tr}(H,\overline{v};v;r)$ as

$$\mathrm{Tr}\Big(Core(H,\overline{v};\, r),\overline{v};\, v\Big).$$

**Lemma 3.7.** *Let $\overline{v}\subset\mathbb{N}*$ be a finite set of fixed vertices and let $\pi(\overline{x})$ be an edge sentence such that $len(\overline{x}) = len(\overline{v})$. Define $G'_n = G_n\setminus E[\overline{v}]$ (i.e. $G_n$ minus all the edges induced over $\overline{v}$). Fix $R\in\mathbb{N}$.*

- *Let $A_n$ be the event that $G'_n$ contains a path of length at most $R$ between any two vertices $u,w\in\overline{v}$.*
- *Let $B_n$ be the event that $G'_n$ contains a cycle of diameter at most $R$ at distance at most $R$ to some vertex $u\in\overline{v}$.*

*Then* $\lim_{n\to\infty}\Pr\big(A_n \,|\, \pi(\overline{v})\big) = 0$, *and* $\lim_{n\to\infty}\Pr\big(B_n \,|\, \pi(\overline{v})\big) = 0$.

*Proof.* Notice that the events $A_n$ and $B_n$ do not concern the possible edges induced over $\overline{v}$. In consequence, because edges are independent in our random model, $\Pr\big(A_n \,|\, \pi(\overline{v})\big) = \Pr(A_n)$ and $\Pr\big(B_n \,|\, \pi(\overline{v})\big) = \Pr(B_n)$.

Now both $\lim_{n\to\infty} \Pr(A_n) = 0$ and $\lim_{n\to\infty} Pr(B_n) = 0$ follow from lemma 3.6 using that $G'_n \subset G_n$. □

---

**Theorem 3.3.** *Fix $r \in \mathbb{N}$. The following are satisfied:*

*(1) Let $\mathscr{T}$ be a k-equivalence class for trees with radii at most r. Let $v \in \mathbb{N}$ be a vertex. Then*

$$\Pr[r, \mathscr{T}] := \lim_{n\to\infty} \Pr\big(Tr\big(G_n, v;\ v;\ r\big) \in \mathscr{T}\big)$$

*exists, is positive for all choices of $\{\beta_R\}_{R\in\varepsilon} \in [0,\infty)^{|\sigma|}$, and is an expression in $\Lambda$ that depends only on the choice of r and $\mathscr{T}$.*

*(2) Let $\overline{u} \in (\mathbb{N})_*$ be a list of different fixed vertices, and let $\pi(\overline{x}) \in FO[\sigma]$ be a consistent edge sentence such that $len(\overline{x}) = len(\overline{u})$. Let $\overline{v} \in (\mathbb{N})_*$ be vertices contained in $\overline{u}$. For each $v \in \overline{v}$ let $\mathscr{T}_v$ be a k-equivalence class of trees with radii at most r. Then*

$$\lim_{n\to\infty} \Pr\Big(\bigwedge_{v\in\overline{v}} Tr\big(G_n, \overline{u};\ v;\ r\big) \in \mathscr{T}_v \,|\, \pi(\overline{u})\Big) = \prod_{v\in\overline{v}} \Pr[r, \mathscr{T}_v].$$

We will devote the rest of this section to proving this theorem. The proof will be done by induction on $r$.

**Lemma 3.8.** *Conditions (1) and (2) of theorem 3.3 are satisfied for $r = 0$.*

*Proof.* Recall that all trees with radius zero are $k$-equivalent. Thus, the limits appearing in conditions (1) and (2) are both equal to 1. □

---

Now, for the induction step we have to prove that conditions (1) and (2) of theorem 3.3 hold for any $r > 0$, given that they hold for all $r' < r$. Notice that (1) is, in fact, a particular case of (2). Proving only (2) would be the "mathematically" reasonable way to proceed, but the proof of (2) does not contain any new ideas that do not appear in the proof of (1) and is, in turn, more convoluted notation-wise. In consequence we will offer the complete proof of (1) and later give indications about what changes in the proof are necessary in order to show (2).

**Lemma 3.9.** *Let $r \in \mathbb{N}$, $r > 0$. Let $\mathscr{T}$ be a $\sim_k$ class of trees with radii at most r and let $v \in \mathbb{N}$ be any vertex. Suppose that theorem 3.3 holds for $r - 1$. Then*

$$\Pr\big[r, \mathscr{T}\big] := \lim_{n\to\infty} \Pr\big(Tr\big(G_n, v;\ v;\ r\big) \in \mathscr{T}\big)$$

*exists, is positive for all choices of $\{\beta_R\}_{R\in\varepsilon} \in [0,\infty)^{|\sigma|}$, and is and expression in $\Lambda$ that depends only on the choice of r and $\mathscr{T}$.*

*Proof.* For any $(k,r)$-pattern $\varepsilon$ let $X_{n,\varepsilon}$ be the random variable that counts the initial edges $e \in E(T_n)$ whose pattern is $\varepsilon$ (i.e. $(e, \tau_{(T_n,v)}) \in \varepsilon$). Fix a pattern $\varepsilon \in P(k,r)$. we define the expressions $\lambda_{r,\varepsilon}$ and $\mu_{r,\varepsilon}$ as follows: let $(e,\tau)$ be a representative of $\varepsilon$ whose root is $v$. Then

$$\lambda_{r,\varepsilon} = \prod_{\substack{u\in V(e)\\ u\neq v}} \Pr\left[r-1, \tau(u)\right], \quad \text{and} \quad \mu_{r,\varepsilon} = \frac{\beta_{R(e)}}{\mathrm{aut}(\varepsilon)} \cdot \lambda_{r,\varepsilon}.$$

Clearly the definitions of $\lambda_{r,\varepsilon}$ and $\mu_{r,\varepsilon}$ are independent from the representative $(e,\tau)$ and by hypothesis depend only on the choice of $r$ and $\varepsilon$. By hypothesis it also holds that $\mu_{r,\varepsilon}$ is positive for all values of $\{\beta_R\}_{R\in\sigma} \in [0,\infty)^{|\sigma|}$. Furthermore, $\mu_{r,\varepsilon}$ belongs to $M$.

First we are going to show that for any $(k,r)$-pattern $\varepsilon$ it holds that

$$\lim_{n\to\infty} E[X_{n,\varepsilon}] = \mu_{r,\varepsilon}.$$

This step is not necessary for the proof as a whole but serves as a simple example that showcases the methods used.

Fix a $(k,r)$-pattern $\varepsilon$. By definition $X_{n,\varepsilon}$ counts the colored edges $(e,\tau) \in Copies(\varepsilon,[n],v)$ such that $e$ is an initial edge in $T_n$ satisfying that for any $u \in V(e)$ with $u \neq v$, it holds $Tr(T_n(v),u) \in \tau(u)$. Thus,

$$\mathrm{E}\left[X_{n,\varepsilon}\right] = \sum_{(e,\tau)\in Copies(\varepsilon,[n];(v,\rho))} \Pr\left( e \in E(T_n) \bigwedge_{\substack{u\in V(e)\\ u\neq v}} Tr(T_n,v;u) \in \tau(u) \right).$$

Because of the symmetry of our random model the probability in the RHS of last equation is the same for all $(e,\tau) \in Copies(\varepsilon,[n];(v,\rho))$. Let $(e,\tau) \in Copies(\varepsilon,\mathbb{N};(v,\rho))$ be fixed. Using that $|Copies(\varepsilon,[n];(v,\rho))| = \frac{(n)_{|e|-1}}{\mathrm{aut}(\varepsilon)}$ we obtain

$$\mathrm{E}\left[X_{n,\varepsilon}\right] = \frac{(n)_{|e|-1}}{|Aut(\varepsilon)|} \Pr\left( e \in E(T_n) \bigwedge_{\substack{u\in V(e)\\ u\neq v}} Tr(T_n,v;u) \in \tau(u) \right).$$

Also, it is satisfied

$$\Pr\left( e \in E(T_n) \bigwedge_{\substack{u\in V(e)\\ u\neq v}} Tr(T_n,v;u) \in \tau(u) \right) =$$

$$\Pr\big(e \in E(G_n)\big) \cdot \Pr\left( e \in E(T_n) \bigwedge_{\substack{u\in V(e)\\ u\neq v}} Tr(T_n,v;u) \in \tau(u) \,\middle|\, e \in E(G_n) \right)$$

Because of lemma 3.7, a.a.s $T_n = N(v;r)$ so a.a.s if $e \in E(G_n)$ and $v \in V(e)$, then $e \in E(T_n)$. Also, $\Pr\big(e \in E(G_n)\big) = \frac{\beta_{R(e)}}{n^{|e|-1}}$. In consequence the RHS of last equation is asymptotically equivalent to

$$\frac{\beta_{R(e)}}{n^{|e|-1}} \cdot \Pr\left( \bigwedge_{\substack{u\in V(e)\\ u\neq v}} Tr(T_n,v;u) \in \tau(u) \,\middle|\, e \in E(G_n) \right)$$

Fix $\overline{u} \in (\mathbb{N})_*$ a list that contains exactly the vertices in $e$. Notice that the event $e \in E(G_n)$ clearly can be described by an edge sentence over the vertices in $\overline{u}$. Because of lemma 3.7,

$$\lim_{n\to\infty} \Pr\Big(N^{G_n}(v;r) \text{ is a tree } \Big|\, e \in E(G_n)\Big) = 1.$$

Set $G_n' := G_n \setminus e$. In the case that $T_n = N(v;r)$ and $e \in E(G_n)$ then the following chain of equalities holds for all $u \in \overline{u}$ different from $v$:

$$Tr(T_n, v, u) = N^{G_n'}(u; r-1) = Tr(G_n, \overline{u}; u; r-1).$$

Thus,

$$\Pr\left( \bigwedge_{\substack{u\in V(e)\\ u\neq v}} Tr(T_n, v; u) \in \tau(u) \,\middle|\, e \in E(G_n) \right) \sim$$
$$\Pr\left( \bigwedge_{\substack{u\in V(e)\\ u\neq v}} Tr(G_n, \overline{u}; u; r-1) \in \tau(u) \,\middle|\, e \in E(G_n) \right) \tag{1}$$

By hypothesis, the RHS of last equality is asymptotically equivalent to $\prod_{\substack{u\in V(e)\\ u\neq v}} \Pr\big[r-1, \tau(u)\big] = \lambda_{r,\varepsilon}$. Finally, joining everything we obtain

$$\lim_{n\to\infty} \mathrm{E}\big[X_{n,\varepsilon}\big] = \lim_{n\to\infty} \frac{(n)_{|e|-1}}{\mathrm{aut}(\varepsilon)} \cdot \frac{\beta_{R(e)}}{n^{|e|-1}} \prod_{\substack{u\in V(e)\\ u\neq v}} \Pr\big[r-1, \tau(u)\big] = \frac{\beta_{R(e)}}{\mathrm{aut}(\varepsilon)} \cdot \lambda_{r,\varepsilon} = \mu_{r,\varepsilon},$$

as we wanted.

Now we are going to prove that the variables $X_{n,\varepsilon}$ converge in distribution to independent Poisson variables with mean values $\mu_{r,\varepsilon}$ respectively. For each $\varepsilon \in P(k,r)$ let $b_\varepsilon \in \mathbb{N}$. We want to show that

$$\lim_{n\to\infty} \mathrm{E}\left[ \prod_{\varepsilon\in P(k,r)} \binom{X_{n,\varepsilon}}{b_\varepsilon} \right] = \prod_{\varepsilon\in P(k,r)} \frac{(\mu_{r,\varepsilon})^{b_\varepsilon}}{b_\varepsilon!}. \tag{2}$$

For each $n \in \mathbb{N}$ define

$$\Omega_n := \Big\{ (E_\varepsilon)_{\varepsilon\in P(k,r)} \quad \Big| \quad \forall \varepsilon \in P(k,r) \quad E_\varepsilon \subset Copies(\varepsilon, [n], (v,\rho)), \quad |E_\varepsilon| = b_\varepsilon \Big\}$$

We also define $\Omega_{\mathbb{N}}$ substituting writing $\mathbb{N}$ instead of $[n]$ in the definition of $\Omega_n$. Informally, elements of $\Omega_n$ represent choices of $b_\varepsilon$ possible initial edges of $T_n$ whose $k$- pattern is $\varepsilon$ for all $(k,r)$-patterns $\varepsilon$. Using observation 3.1 we obtain

$$\mathrm{E}\left[ \prod_{\varepsilon\in P(k,r)} \binom{X_{n,\varepsilon}}{b_\varepsilon} \right] = \sum_{(E_\varepsilon)_{\varepsilon\in P(k,r)} \in \Omega_n} \Pr\left( \bigwedge_{\substack{\varepsilon\in P(k,r)\\ (e,\tau)\in E_\varepsilon}} \left( e \in E(T_n) \bigwedge_{\substack{u\in V(e)\\ u\neq v}} Tr(T_n, v; u) \in \tau(u) \right) \right).$$

We say that an element $(E_\varepsilon)_{\varepsilon\in P(k,r)}$ of $\Omega_n$ is **disjoint** each vertex $w \in [n] \setminus \{v\}$ belongs to at most one edge $(e,\tau) \in \bigcup_{\varepsilon\in P(k,r)} E_\varepsilon$. Notice that if we want the probability in the last sum to be greater than 0 for a particular $(E_\varepsilon)_{\varepsilon\in P(k,r)} \in \Omega_n$ then necessarily $(E_\varepsilon)_{\varepsilon\in P(k,r)}$ is disjoint. Indeed, suppose that a vertex $w \in [n]\setminus\{v\}$ belongs to two different edges $(e_1,\tau_1)$, $(e_2,\tau_2) \in \bigcup_{\varepsilon\in P(k,r)} E_\varepsilon$. In consequence $e_1$ and $e_2$ form a cycle of diameter 1, as they both contain $v$ and $w$. This implies that $e_1, e_2 \notin E(T_n)$.

For each $n \in \mathbb{N}$ let $\Omega_n' \subset \Omega_n$ be the set of disjoint elements in $\Omega_n$. Then

$$\mathrm{E}\left[\prod_{\varepsilon\in P(k,r)} \binom{X_{n,\varepsilon}}{b_\varepsilon}\right] = \sum_{(E_\varepsilon)_{\varepsilon\in P(k,r)}\in\Omega_n'} \Pr\left(\bigwedge_{\substack{\varepsilon\in P(k,r)\\(e,\tau)\in E_\varepsilon}} \left(e\in E(T_n) \bigwedge_{\substack{u\in V(e)\\u\neq v}} Tr(T_n,v;u)\in\tau(u)\right)\right).$$

Also, because of the symmetry of the random model, for all disjoint elements $(E_\varepsilon)_{\varepsilon\in P(k,r)}$ the probability in last sum is the same. In consequence, if we fix $(E_\varepsilon)_{\varepsilon\in P(k,r)} \in \Omega_{\mathbb{N}}'$ we obtain

$$\mathrm{E}\left[\prod_{\varepsilon\in P(k,r)} \binom{X_{n,\varepsilon}}{b_\varepsilon}\right] = |\Omega_n'|\cdot\Pr\left(\bigwedge_{\substack{\varepsilon\in P(k,r)\\(e,\tau)\in E_\varepsilon}} \left(e\in E(T_n) \bigwedge_{\substack{u\in V(e)\\u\neq v}} Tr(T_n,v;u)\in\tau(u)\right)\right). \tag{3}$$

Counting vertices and automorphisms we get that

$$|\Omega_n'| = (n)_{\sum_{\varepsilon\in P(k,r)}(|\varepsilon|-1)\cdot b_\varepsilon} \prod_{\varepsilon\in P(k,r)} \frac{1}{b_\varepsilon!}\cdot\left(\frac{1}{\mathrm{aut}(\varepsilon)}\right)^{b_\varepsilon}. \tag{4}$$

Because of lemma 3.7 a.a.s if $e\in E(G_n)$ and $v\in V(e)$, then $e\in E(T_n)$. In consequence:

$$\Pr\left(\bigwedge_{\substack{\varepsilon\in P(k,r)\\(e,\tau)\in E_\varepsilon}} \left(e\in E(T_n) \bigwedge_{\substack{u\in V(e)\\u\neq v}} Tr(T_n;u)\in\tau(u)\right)\right) \sim \tag{5}$$

$$\Pr\left(\bigwedge_{\substack{\varepsilon\in P(k,r)\\(e,\tau)\in E_\varepsilon}} e\in E(G_n)\right)\cdot\Pr\left(\bigwedge_{\substack{\varepsilon\in P(k,r)\\(e,\tau)\in E_\varepsilon\\u\in V(e)\\u\neq v}} Tr(T_n;u)\in\tau(u) \;\middle|\; \bigwedge_{\substack{\varepsilon\in P(k,r)\\(e,\tau)\in E_\varepsilon}} e\in E(G_n)\right).$$

Let $\overline{w}\in(\mathbb{N})_*$ be a list containing exactly the vertices $u\in V(e)$ for all $e\in\bigcup_{\varepsilon\in P(k,r)} E_\varepsilon$. The event $\bigwedge_{\substack{\varepsilon\in P(k,r)\\(e,\tau)\in E_\varepsilon}} e\in E(G_n)$ clearly can be described via an edge sentence whose variables are interpreted as vertices in $\overline{w}$. Thus, analogously to eq. (1) we obtain

$$\Pr\left(\bigwedge_{\substack{\varepsilon\in P(k,r)\\(e,\tau)\in E_\varepsilon}} \left(e\in E(T_n) \bigwedge_{\substack{u\in V(e)\\u\neq v}} Tr(T_n,v;u)\in\tau(u)\right) \;\middle|\; \bigwedge_{\substack{\varepsilon\in P(k,r)\\(e,\tau)\in E_\varepsilon}} e\in E(G_n)\right) \sim \tag{6}$$

$$\Pr\left(\bigwedge_{\substack{\varepsilon\in P(k,r)\\(e,\tau)\in E_\varepsilon}} \left(e\in E(T_n) \bigwedge_{\substack{u\in V(e)\\u\neq v}} Tr(G_n,\overline{w};u;r-1)\in\tau(u)\right) \;\middle|\; \bigwedge_{\substack{\varepsilon\in P(k,r)\\(e,\tau)\in E_\varepsilon}} e\in E(G_n)\right)$$

By hypothesis last probability is asymptotically equivalent to

$$\prod_{\varepsilon \in P(k,r)} (\lambda_{r,\varepsilon})^{b_\varepsilon}.$$

Joining this with eq. (3), eq. (4), eq. (5) and eq. (5) we obtain

$$\begin{aligned}\lim_{n\to\infty} \mathrm{E}\left[\prod_{\varepsilon\in P(k,r)} \binom{X_{n,\varepsilon}}{b_\varepsilon}\right] &= \lim_{n\to\infty} \frac{(n)_{\sum_{\varepsilon\in P(k,r)}(|\varepsilon|-1)\cdot b_\varepsilon}}{n^{\sum_{\varepsilon\in P(k,r)}(|\varepsilon|-1)\cdot b_\varepsilon}} \cdot \prod_{\varepsilon\in P(k,r)} \frac{1}{b_\varepsilon!}\cdot\left(\frac{\beta_{R(\varepsilon)}}{\mathrm{aut}(\varepsilon)}\right)^{b_\varepsilon}\cdot(\lambda_{r,\varepsilon})^{b_\varepsilon} \\ &= \prod_{\varepsilon\in P(k,r)} \frac{(\mu_{r,\varepsilon})^{b_\varepsilon}}{b_\varepsilon!},\end{aligned}$$

as we wanted. In consequence, by theorem 3.2, given $a_\varepsilon \in \mathbb{N}$ for all $\varepsilon \in P(k,r)$ it holds

$$\lim_{n\to\infty} \Pr\left(\bigwedge_{\varepsilon\in P(k,r)} X_{n,\varepsilon} = a_\varepsilon\right) = \prod_{\varepsilon\in P(k,r)} e^{-\mu_{r,\varepsilon}} \frac{(\mu_{r,\varepsilon})^{a_\varepsilon}}{a_\varepsilon!}.$$

Finally, using observation 2.3 we get that for some partition $E^1_{\mathscr{T}}, E^2_{\mathscr{T}}$ of $P(k,r)$ and some natural numbers $a_\varepsilon < k$ for each $\varepsilon \in E^2_{\mathscr{T}}$ it holds that

$$\Pr\big([\,\big)r,\mathscr{T}] = \lim_{n\to\infty}\Pr\big(T_n\in\mathscr{T}\big) = \left(\prod_{\varepsilon\in E^1_{\mathscr{T}}} \mathrm{Poiss}_{\geq k}(\mu_{r,\varepsilon})\right)\cdot\left(\prod_{\varepsilon\in E^2_{\mathscr{T}}} \mathrm{Poiss}_{a_\varepsilon}(\mu_{r,\varepsilon})\right).$$

And last expression belongs to $\Lambda$ as we wanted to prove. Furthermore, as the $\mu_{r,\varepsilon}$'s are positive, this expression is also positive for all values of $\{\beta_R\}_{R\in\sigma} \in [0,\infty)^{|\sigma|}$. □

---

In order for theorem 3.3 to be completely proven the following result is needed

**Lemma 3.10.** *Let $r \in \mathbb{N}$ be a positive number. Let $\overline{u} \in (\mathbb{N})_*$ be a list of different fixed vertices, and let $\pi(\overline{x}) \in FO[\sigma]$ be an edge sentence such that $len(\overline{x}) = len(\overline{u})$. Let $\overline{v} \in (\mathbb{N})_*$ be vertices contained in $\overline{u}$. For each $v \in \overline{v}$ let $\mathscr{T}_v$ be a k-equivalence class of trees with radii at most r. Suppose that theorem 3.3 holds for $r-1$. Then*

$$\lim_{n\to\infty}\Pr\big(\bigwedge_{v\in\overline{v}} Tr\big(G_n,\overline{u};\ v;\ r\big)\in\mathscr{T}_v\,|\,\pi(\overline{u})\big) = \prod_{v\in\overline{v}}\Pr[r,\mathscr{T}_v].$$

*Sketch of the proof.* The proof is completely analogous to the one of the previous lemma but now with more random variables. For each $v \in \overline{v}$ we define $T_{n,v} := Tr\big(G_n,\overline{u};\ v;\ r\big)$. Given a $(k,r)$-pattern $\varepsilon \in P(k,r)$ and a vertex $v \in \overline{v}$ we define the random variable $X_{n,v,\varepsilon}$ as the one that counts the number of initial edges $e \in E(T_{n,v})$ whose pattern is $\varepsilon$. Similarly to last lemma one can show that the $X_{n,v,\varepsilon}$ are asymptotically distributed like independent Poisson variables whose respective means are the $\mu_{r,\varepsilon}$ defined in the previous lemma. As before, this is done using theorem 3.2, by computing the binomial moments of the $X_{n,v,\varepsilon}$'s. Once the asymptotic distribution of those variables is computed the identity

$$\lim_{n\to\infty}\Pr\big(\bigwedge_{v\in\overline{v}} Tr\big(G_n,\overline{u};\ v;\ r\big)\in\mathscr{T}_v\,|\,\pi(\overline{u})\big) = \prod_{v\in\overline{v}}\Pr[r,\mathscr{T}_v]$$

follows easily using the definition of $\Pr[r,\mathscr{T}_v]$ provided at the end of last proof. □

## 3.4 Almost all graphs are (k,r)-rich

**Theorem 3.4.** *Let $r \in \mathbb{N}$. Then a.a.s $G_n$ is $(k,r)$-rich.*

*Proof.* Let $\Sigma$ be the set of all $\sim_k$-classes of rooted trees with radii at most $r$. For each $\mathscr{T} \in \Sigma$ let $\overline{v}(\mathscr{T}) \in (\mathbb{N})_k$ be a $k$-tuple of vertices such that all the $\overline{v}(\mathscr{T})$'s are disjoint. Given any $\overline{v} \in (N)_*$ event $\phi_n\big(\overline{v}(\mathscr{T})\big)$ as the event that for any $v \in \overline{v}$, $N(v;r) \cup Core(G_n;r)\emptyset$ (thus $N(v;r)$ is a tree), and that for any two $v_1, v_2 \in \overline{v}$, $d^{G_n}(v_1,v_2) > 2r+1$. For each $n \in \mathbb{N}$ let

$$A_n := \bigwedge_{\mathscr{T}\in\Sigma} \exists\, \overline{v}(\mathscr{T}) \in ([n])_k \left( \phi_n(\overline{v}(\mathscr{T})) \bigwedge_{v\in\overline{v}(\mathscr{T})} Tr\big(G_n;\, v;\, r\big) \in \mathscr{T} \right).$$

Then the event $A_n$ that $G_n$ is $(k,r)$-rich. Indeed, suppose that $G_n$ satisfies $A_n$. Let $\mathscr{T} \in \Sigma$ be a $\sim_k$ class and let $v_1,\dots,v_{k-1} \in [n]$ be any vertices in $G_n$. Then because of $\phi_n(\overline{v}(\mathscr{T}))$ there is at least one vertex $v \in \overline{v}(\mathscr{T})$ such that $d(v,v_i) \geq 2r+1$ for all $v_i$'s. It also holds that $T := N(v;r)$ is a tree, and because of $A_n$, $(T,v) \in \mathscr{T}$. In consequence $G_n$ is $(k,r)$-rich.

Now we show that $\lim_{n\to\infty} \Pr\big(A_n\big) = 1$. For that we will prove that for each $\mathscr{T} \in \Sigma$, a.a.s. the following event holds:

$$B_{n,\mathscr{T}} := \exists\, \overline{v}(\mathscr{T}) \in ([n])_k \left( \phi_n(\overline{v}(\mathscr{T})) \bigwedge_{v\in\overline{v}(\mathscr{T})} Tr\big(G_n;\, v;\, r\big) \in \mathscr{T} \right).$$

Fix $\mathscr{T} \in \Sigma$ and fix $\varepsilon > 0$ an arbitrarily small real number. Let $m \in \mathbb{N}$, and let $\overline{v} \in (\mathbb{N})_m$ be a $m$-tuple of vertices. Let $X_{n,\overline{v}}$ be the random variable that counts the number of vertices $v \in \overline{v}$ such that $\mathrm{Tr}(G_n;\, v;\, r)$ belongs to $\mathscr{T}$. Because of theorem 3.3, $X_{n,\sqsubseteq}$ converges in distribution to a binomial variable with parameters $m$ and $\Pr[r,\mathscr{T}]$. That is, for each $l \in [m]$,

$$\lim_{n\to\infty} \Pr\big(X_{n,\overline{v}} = l\big) = \binom{m}{l} \Pr[r,\mathscr{T}]^l \cdot (1-\Pr[r,\mathscr{T}])^{m-l}.$$

In particular, as $\Pr[r,\mathscr{T}]$ is greater than zero, for $m$ sufficiently big it holds that

$$\lim_{n\to\infty} \Pr\big(X_{n,\overline{v}} \geq k\big) > 1-\varepsilon.$$

Also, because of lemma 3.7 we have that $\lim_{n\to\infty} \phi_n(\overline{v}) = 1$, and in consequence for $m$ sufficiently big

$$\lim_{n\to\infty} \Pr\left((X_{n,\overline{v}} \geq k) \wedge \phi(\overline{v})\right) > 1-\varepsilon.$$

As $(X_{n,\overline{v}} \geq k) \wedge \phi(\overline{v})$ implies $B_{n,\mathscr{T}}$ we have that $\lim_{n\to\infty} \Pr\big(B_{n,\mathscr{T}}\big) = 1$. As this holds for all $\mathscr{T} \in \Sigma$ and $\Sigma$ is a finite set, then $\lim_{n\to\infty} \Pr\big(A_n\big) = 1$ as well and the result follows. $\square$

## 3.5 Probabilities of cycles

**Definition 3.3.** We define $\Gamma$ and $\Upsilon$ as the minimal families of expressions with arguments $\{\beta_R\}_{R\in\sigma}$ that satisfies the conditions:

(1) For each $R \in \sigma$ let $a_R \in \mathbb{N}$ be any natural number. Then, given any positive number $b \in \mathbb{N}$ and any $\lambda \in \Lambda$ the expression $\frac{\lambda}{b} \cdot \prod_{R\in\sigma} \beta_R^{a_R}$ belongs to $\Gamma$.

(2) Given any $\gamma \in \Gamma$ and any $a \in \mathbb{N}$, the expressions $\mathrm{Poiss}_a(\gamma)$ and $\mathrm{Poiss}_{\geq a}(\gamma)$ both belong to $\Upsilon$.

(3) If $\upsilon_1, \upsilon_2 \in \Upsilon$ then $\upsilon_1 \cdot \upsilon_2 \in \Upsilon$ as well.

**Theorem 3.5.** *Let $\mathscr{O}$ be a simple k-agreeability class of hypergraphs. Then $\lim_{n\to\infty} \Pr\big(G_n \in \mathscr{O}\big)$ exists and is an expression in $\Upsilon$.*

*Proof.* Define $r := 3^k$. For each $O \in C(k,r)$ let $X_{n,O}$ be the random variable that counts the number of cycles in $Core(G_n;r)$ whose $k$-type is $O$. Fix $O \in C(k,r)$. For any $O \in C(k,r)$ we define $\lambda_O$ and $\gamma_O$ in the following way. Let $(H,\tau)$ be a representative of $O$. Then

$$\lambda_O := \prod_{v\in V(H)} \Pr\left[r, \tau(v)\right],$$

and

$$\gamma_O := \frac{\prod_{R\in\sigma} \beta_R^{|E_R(H)|}}{\mathrm{aut}(H,\tau)} \cdot \lambda_O.$$

As in the proof of lemma 3.9 one can show that

$$\lim_{n\to\infty} \mathrm{E}\big[X_{n,O}\big] = \gamma_O$$

. Notice that the expression $\gamma_O$ both belongs to $\Gamma$ and does only depend on the $(k,r)$-cycle $O$.

We are going to prove that the variables $X_{n,O}$ converge in distribution as $n$ tends to infinity to independent Poisson variables whose respective means are the $\gamma_O$. For that we are going to use again the factorial moments method. For each $O \in C(k,r)$ fix a number $b_O \in \mathbb{N}$. We want to prove

$$\lim_{n\to\infty} \mathrm{E}\left[\prod_{O\in C(k,r)} \binom{X_{n,O}}{b_O}\right] = \prod_{O\in C(k,r)} \frac{(\gamma_O)^{b_O}}{b_O!}.$$

For each $n \in \mathbb{N}$ we define

$$\Omega_n := \left\{ (F_O)_{O\in C(k,r)} \;\middle|\; \forall O \in C(k,r) \quad F_O \subset Copies(O,[n]), \quad |F_O| = b_O \right\}.$$

We also define $\Omega_{\mathbb{N}}$ by substituting $[n]$ for $\mathbb{N}$ in the definition of $\Omega_n$. Informally, an element of $\Omega_n$ represents a choice of an unordered $b_O$-tuple of possible cycles over $[n]$ whose $(k,r)$-type is $O$, for each $(k,r)$-cycle $O$. Using observation observation 3.1 we obtain

$$\mathrm{E}\left[\prod_{O\in C(k,r)} \binom{X_{n,O}}{b_O}\right] = \sum_{(F_O)_{O\in C(k,r)}\in\Omega_n} \Pr\left( \bigwedge_{\substack{O\in C(k,r)\\ (H,\tau)\in F_O}} \left( H \subset G_n \bigwedge_{v\in V(H)} Tr(G_n,v;r) \in \tau(v) \right)\right).$$

Consider the subset $\Omega'_n \subset \Omega_n$ that contains the elements $(F_O)_{O\in C(k,r)} \in \Omega_n$ such that there exists some vertex $v \in [n]$ contained in two graphs $(H_1,\tau_1),(H_2,\tau_2) \in \bigcup_{O\in C(k,r)} F_O$. We want to argue that

$$\lim_{n\to\infty} \sum_{(F_O)_{O\in C(k,r)}\in\Omega'_n} \Pr\left( \bigwedge_{\substack{O\in C(k,r)\\ (H,\tau)\in F_O}} \left( H \subset G_n \bigwedge_{v\in V(H)} Tr(G_n,v;r) \in \tau(v) \right)\right) = 0. \tag{7}$$

Given an element $(F_O)_{O\in C(k,r)} \in \Omega_n$ we define the hypergraph $G\Big((F_O)_{O\in C(k,r)}\Big)$ as follows:

$$G\Big((F_O)_{O\in C(k,r)}\Big) := \bigcup_{H\in F} H,$$

where

$$F := \left\{ H \;\middle|\; (H,\tau) \in \bigcup_{O\in C(k,r)} F_O \right\}.$$

That is, $G\Big((F_O)_{O\in C(k,r)}\Big)$ is the union of all hypergraphs chosen in $(F_O)_{O\in C(k,r)}$. Then, for all $(F_O)_{O\in C(k,r)} \in \Omega_n$ it is satisfied

$$\Pr\left( \bigwedge_{\substack{O\in C(k,r)\\(H,\tau)\in F_O}} \left( H\subset G_n \bigwedge_{v\in V(H)} Tr(G_n,v;r)\in\tau(v) \right)\right) \leq \Pr\left( \bigwedge_{\substack{O\in C(k,r)\\(H,\tau)\in F_O}} H\subset G_n \right) =$$
$$\Pr\left( G\Big((F_O)_{O\in C(k,r)}\Big) \subset G_n \right).$$

Let

$$t = \sum_{O\in C(k,r)} |V(O)|\cdot b_O.$$

Then $V\left(G\Big((F_O)_{O\in C(k,r)}\Big)\right) \leq t$ for any $(F_O)_{O\in C(k,r)} \in \Omega_n$.

Consider the following facts

(1) If $(F_O)_{O\in C(k,r)} \in \Omega'_n$ then $G\big((F_O)_{O\in C(k,r)}\big)$ is dense.

(2) Given an hypergraph $H$ with $V(H)\subset\mathbb{N}$, the number of elements $(F_O)_{O\in C(k,r)} \in \Omega'_{\mathbb{N}}$ such that $H = G\big((F_O)_{O\in C(k,r)}\big)$ is finite and it is the same for all $H'\simeq H$ with $V(H')\subset\mathbb{N}$.

(3) There is a finite amount of unlabeled dense hypergraphs with size bounded by $t$.

Then it follows that

$$\sum_{(F_O)_{O\in C(k,r)}\in\Omega'_n} \Pr\left( G\Big((F_O)_{O\in C(k,r)}\Big) \subset G_n \right)$$
$$= \mathrm{O}\left(\mathrm{E}\left[\#\text{ of dense subgraphs in } G_n \text{ with size bounded by } t\right]\right).$$

And this, together with lemma 3.2 proves eq. (7).

For all $n$ define $\Omega''_n = \Omega_n \setminus \Omega'_n$. That is, $\Omega''_n$ contains the elements $(F_O)_{O\in C(k,r)}$ in $\Omega_n$ such that all vertices $v\in[n]$ belong to at most one hypergraph $(H,\tau)\in\bigcup_{O\in C(k,r)} F_O$. We also define $\Omega''_{\mathbb{N}}$. Because of eq. (7) we have

$$\mathrm{E}\left[\prod_{O\in C(k,r)} \binom{X_{n,O}}{b_O}\right] = \sum_{(F_O)_{O\in C(k,r)}\in\Omega''_n} \Pr\left( \bigwedge_{\substack{O\in C(k,r)\\(H,\tau)\in F_O}} \left( H\subset G_n \bigwedge_{v\in V(H)} Tr(G_n,v;r)\in\tau(v) \right)\right) + o(1).$$

Because of the symmetry of the model the probability inside of last sum is the same for all elements $(F_O)_{O\in C(k,r)} \in \Omega''_n$. Also, counting all different vertices and automorphisms we obtain that

$$|\Omega''_n| = \frac{(n)_{\sum_{O\in C(k,r)}|V(O)|\cdot b_O}}{\prod_{O\in C(k,r)} b_O! \cdot \mathrm{aut}(O)^{b_O}}.$$

Fix $(F_O)_{O\in C(k,r)} \in \Omega''_{\mathbb{N}}$. Then

$$\lim_{n\to\infty} \mathrm{E}\left[\prod_{O\in C(k,r)} \binom{X_{n,O}}{b_O}\right] =$$

$$\lim_{n\to\infty} \frac{(n)_{\sum_{O\in C(k,r)}|V(O)|\cdot b_O}}{\prod_{O\in C(k,r)} b_O! \cdot \mathrm{aut}(O)^{b_O}} \cdot \Pr\left(\bigwedge_{\substack{O\in C(k,r)\\(H,\tau)\in F_O}} \left(H\subset G_n \bigwedge_{v\in V(H)} Tr(G_n,v;r)\in\tau(v)\right)\right).$$

It holds that the probability in last expression equals

$$\prod_{O\in C(k,r)} \left(\frac{\prod_{R\in\sigma}\beta_R^{|E_R(O)|}}{n^{|V(O)|}}\right)^{b_O} \cdot \Pr\left(\bigwedge_{\substack{O\in C(k,r)\\(H,\tau)\in F_O\\v\in V(H)}} Tr(G_n,v;r)\in\tau(v) \,\middle|\, \bigwedge_{\substack{O\in C(k,r)\\(H,\tau)\in F_O}} H\subset G_n\right).$$

Let $\overline{v}\in(\mathbb{N})_*$ be a list that contains exactly the vertices in $G\big((F_O)_{O\in C(k,r)}\big)$. Then the event

$$A_n := \bigwedge_{\substack{O\in C(k,r)\\(H,\tau)\in F_O}} H\subset G_n$$

can be written as an edge sentence concerning the vertices in $\overline{w}$. Also, if $A_n$ holds then all vertices in $\overline{w}$ belong to $Core(G_n;r)$. Thus, for all $v\in\overline{v}$, $Tr(G_n,v;r) = Tr(G_n,\overline{w};r)$ and using theorem 3.3 we obtain

$$\Pr\left(\bigwedge_{\substack{O\in C(k,r)\\(H,\tau)\in F_O\\v\in V(H)}} Tr(G_n,v;r)\in\tau(v) \,\middle|\, \bigwedge_{\substack{O\in C(k,r)\\(H,\tau)\in F_O}} H\subset G_n\right) \sim \prod_{O\in C(k,r)} (\lambda_O)^{b_O}.$$

Joining everything together we obtain

$$\lim_{n\to\infty} \mathrm{E}\left[\prod_{O\in C(k,r)} \binom{X_{n,O}}{b_O}\right] =$$

$$\lim_{n\to\infty} \frac{(n)_{\sum_{O\in C(k,r)}|V(O)|\cdot b_O}}{\prod_{O\in C(k,r)} b_O! \cdot \mathrm{aut}(O)^{b_O}} \cdot \prod_{O\in C(k,r)} \left(\frac{\lambda_O\cdot\prod_{R\in\sigma}\beta_R^{|E_R(O)|}}{n^{|V(O)|}}\right)^{b_O} =$$

$$\prod_{O\in C(k,r)} \frac{1}{b_O!}\left(\frac{\lambda_O\cdot\prod_{R\in\sigma}\beta_R^{|E_R(O)|}}{\mathrm{aut}(O)}\right)^{b_O} = \prod_{O\in C(k,r)} \frac{(\gamma_O)^{b_O}}{b_O!},$$

as we wanted. With this, because of theorem 3.2, it is proven that when $n$ tends to infinity the $X_{n,O}$'s are asymptotically distributed like independent Poisson variables with the $\gamma_O$'s as their respective means.

Fix a $(k,r)$-agreeability class of $r$-simple hypergraphs $\mathscr{O}$. Because observation 2.5 it holds that there is a partition $C_1, C_2 \subset C(k,r)$, $C_1 \cup C_2 = C(k,r)$ and natural numbers $a_O \leq k-1$ for any $O \in C_2$ such that $C_1, C_2, (a_O)_{O \in C_2}$ depend only on $\mathscr{O}$ and

$$\lim_{n\to\infty} \Pr\big(G_n \in \mathscr{O}\big) =$$
$$\lim_{n\to\infty} \Pr\left( G_n \text{ is } r\text{-sparse } \wedge \left( \bigwedge_{O\in C_1} X_{n,O} \geq k \right) \wedge \left( \bigwedge_{O\in C_1} X_{n,O} = a_O. \right)\right).$$

Because of theorem 3.1, last limit equals

$$\lim_{n\to\infty} \Pr\left(\left( \bigwedge_{O\in C_1} X_{n,O} \geq k \right) \wedge \left( \bigwedge_{O\in C_1} X_{n,O} = a_O \right)\right) =$$
$$\left(\prod_{O\in C_1} \mathrm{Poiss}_{\geq k}(\gamma_O)\right) \cdot \left(\prod_{O\in C_2} \mathrm{Poiss}_{a_O}(\gamma_O)\right).$$

This last expression belongs to $\Upsilon$, so the theorem is proven. $\square$

# 4 Proof of the main theorem

**Theorem 4.1.** *Let $\phi \in FO[\sigma]$. Then the function $F_\phi : [O,\infty)^{|\sigma|} \to [0,1]$ given by*

$$\{\beta_R\}_{R\in\sigma} \mapsto \lim_{n\to\infty} \Pr\left(G_n\left(\{\beta_R\}_{R\in\sigma}\right) \models \phi\right)$$

*is well defined and it is given by a finite sum of expressions in $\Theta$.*

*Proof.* Let $k$ be the quantifier rank of $\phi$ and let $r = 3^k$. Let $G_n := G_n\left(\{\beta_R\}_{R\in\sigma}\right)$. Using corollary 3.1 we obtain

$$\lim_{n\to\infty} \Pr\left(G_n \models \phi\right) = \lim_{n\to\infty} \Pr\left( G_n \models \phi \,\middle|\, G_n \text{ is } r\text{-sparse}\right).$$

Let $\Sigma$ be the set of $(k,r)$-agreeability classes of $(k,r)$-simple hypergraphs. Then

$$\lim_{n\to\infty} \Pr\left(G_n \models \phi\right) = \lim_{n\to\infty} \sum_{\mathscr{O}\in\Sigma} \Pr\left(G_n \in \mathscr{O}\right) \cdot \Pr\left( G_n \models \phi \,\middle|\, G_n \in \mathscr{O}\right). \tag{8}$$

Notice that, because the set $\Sigma$ is finite, this is the limit of a finite sum and we can exchange summation and limit. Also, using theorem 3.4, we obtain that for any $\mathscr{O} \in \Sigma$ it holds

$$\lim_{n\to\infty} \Pr\left( G_n \models \phi \,\middle|\, G_n \in \mathscr{O}\right) = \lim_{n\to\infty} \Pr\left( G_n \models \phi \,\middle|\, G_n \in \mathscr{O} \text{ and } G_n \text{ is } (k,r)\text{-rich}\right).$$

Because theorem 2.4 we have that given any two hypergraphs $H_1$ and $H_2$ such that $H_1$ and $H_2$ are $(k,r)$-agreeable and they are both $(k,r)$-rich then they both satisfy the same first order sentences with quantifier rank at most $k$. Then the LHS of last equation always equals either zero or one. Let $\Sigma' \subset \Sigma$ be the set of classes $\mathscr{O}$ for which last limit equals one. Then

$$\lim_{n\to\infty} \Pr\left(G_n \models \phi\right) = \sum_{\mathscr{O}\in\Sigma'} \lim_{n\to\infty} \Pr\left(G_n \in \mathscr{O}\right).$$

Because of theorem 3.5 we know that each of the limits inside last sum exists and is given by an expression that belongs to $\Theta$. As a consequence the theorem follows. $\square$

# 5 Application to random SAT

We will define a binomial model of random CNF formulas, in analogy with the one in [7], but clearly the generality in theorem 1.1 allows for many modifications.

**Definition 5.1.** Given a variable $x$ both expressions $x$ and $\neg x$ are called **literals**. A **clause** is a set of literals. A clause $C$ is called **ordinary** if no variable $x$ satisfies that both $x$ and $\neg x$ belong to $C$. An **assignment** over a set of variables $X$ is a map $f$ that assigns 0 or 1 to each variable of $X$. A clause $C$ is **satisfied** by an assignment $f$ if either there is some variable $x$ such that $x \in C$ and $f(x) = 1$ or there is some variable $x$ such that $\neg x \in C$ and $f(x) = 0$. Given a natural number $l \in \mathbb{N}$ a $l$-**CNF formula** is a set of ordinary clauses that contain exactly $l$ literals. We say that a formula $F$ over the variables $x_1, \dots, x_n$ is **satisfiable** if there is an assignment $f : \{x_1, \dots, x_n\} \to \{0,1\}$ that satisfies all clauses for any clause $C \in F$.

Given $n, l \in \mathbb{N}$ and a real number $0 \le p \le 1$ we define the random model $F(l,n,p)$ as the discrete probability space that assigns to each $l$-CNF formula $F$ formed of clauses over the variables $\{x_i\}_{i\in[n]}$ the probability

$$\Pr\big(F\big) = p^{|F|} \cdot (1-p)^{2^l \binom{n}{l} - |F|}.$$

Equivalently, a random formula in $F(l,n,p)$ is obtained by choosing each one of the $2^l \binom{n}{l}$ normal clauses of size $l$ over the variables $\{x_i\}_{i\in[n]}$ with probability $p$ independently.

We can model $l$-CNF formulas, as we have defined them, as relational structures with a language $\sigma$ consisting of $l+1$ relation symbols $R_0, \dots, R_l$ with arity $l$. We do that in such a way that the expression $R_j(x_{i_1}, \dots, x_{i_l})$ means that our formula contains the clause consisting of $\neg x_{i_1}, \dots, \neg x_{i_j}$ and $x_{i_{j+1}}, \dots x_{i_l}$. In consequence we need $R_1, \dots, R_l$ to satisfy the following additional axioms:

- For each $0 \le j \le l$ and for any variables $y_1, \dots, y_j, y_{j+1}, \dots, y_l$ it holds that $R_j(y_1, \dots y_j, y_{j+1}, \dots, y_l)$ if and only if it still holds after applying any permutations on the variables $y_1, \dots, y_j$ and the variables $y_{j+1}, \dots, y_l$.
- For each $0 \le j \le l$ and for any variables $y_1, \dots, y_l$ it holds that $R_j(y_1, \dots, y_l)$ only if all the $y_i$'s are different.

Call $\mathscr{C}$ to the family of $\sigma$-structures satisfying these last two axioms. Then, there is a clear correspondence between $l$-CNF formulas over the variables $\{x_i\}_{i\in[n]}$ and the structures in $\mathscr{C}$ whose universe is $\{x_i\}_{i\in[n]}$.

The language $\sigma$ and the family $\mathscr{C}$ satisfy the conditions in section 1.3. The random model $F_l(n,p)$ coincides with the model $G(n, \{p_R\}_{R\in\sigma})$ of random $\mathscr{C}$-hypergraphs described in section 1.6 when all the $p_R$'s are equal. In consequence the following theorem holds

**Theorem 5.1.** *Let $l > 1$ be a natural number. For each $n \in \mathbb{N}$ let $F_n(\beta)$ be a random formula from $F(l,n,\beta/n^{l-1})$. Then for each sentence $\Phi \in FO[\sigma]$ it is satisfied that the map $f_\Phi : [0,\infty) \to \mathbb{R}$ given by*

$$\beta \mapsto \Pr\big(F_n(\beta) \models \Phi\big)$$

*is well defined and analytic.*

A different model of random CNF formulas is studied in [7]. There formulas are viewed as families of non necessary different clauses rather than sets. They define a random formula with $m$ clauses of size $l$ over $n$ variables as a sequence of independent random clauses $C_1,\dots,C_m$ where each $C_i$ is chosen uniformly at random among the $2^l\binom{n}{l}$ ordinary clauses of size $l$ over $n$ variables. The following holds

**Theorem 5.2.** *Let $l\geq 2$ be a natural number, and let $c\in[0,\infty)$ be an arbitrary real number. Let $m:\mathbb{N}\to\mathbb{N}$ be a map such that $m(n)=(c+o(1))n$. For each $n$ let $C_{n,1},\dots,C_{n,m(n)}$ be clauses chosen uniformly at random independently among the $2^l\binom{n}{l}$ ordinary clauses of size $l$ over the variables $x_1,\dots,x_n$. For each $n$, let $UNSAT_n$ denote the event that there is no assignment of the variables $x_1,\dots,x_n$ that satisfies all clauses $C_{n,1},\dots,C_{n,m(n)}$. Then there are two real constants $0<c_1<c_2$, independent from such that*

$$\lim_{n\to\infty}\Pr\big(UNSAT_n\big)=0$$

*if $c<c_1$, and*

$$\lim_{n\to\infty}\Pr\big(UNSAT_n\big)=1$$

*if $c>c_2$.*

The existence of $c_1$ is proven in theorem 1 of [7]. The existence of the other constant $c_2$ follows from a simple application of the first order method that also appears in [7], as well as [12], [13], [14] and possibly others. We want to show that this "phase transition" also happens in our model $F(l,n,p)$ when $p\sim\beta/n^{l-1}$. We start by showing the following

**Corollary 5.1.** *Let $l\geq 2$ be a natural number. Let $c\in[0,\infty)$ be an arbitrary real number and let $m:\mathbb{N}\to\mathbb{N}$ satisfy $m(n)=(c+o(1))n$. For each $n\in\mathbb{N}$ let $F_{n,m(n)}$ be a random formula chosen uniformly at random among all the sets of $m(n)$ ordinary clauses of size $l$ over the variables $x_1,\dots,x_n$. Then there are two real positive constants $0<c_1<c_2$ such that*

$$\lim_{n\to\infty}\Pr\big(F_{n,m(n)}\text{ is unsatisfiable }\big)=0$$

*if $c<c_1$, and*

$$\lim_{n\to\infty}\Pr\big(F_{n,m(n)}\text{ is unsatisfiable }\big)=1$$

*if $c>c_2$.*

*Proof.* One can consider $F_{n,m(n)}$ to be the result of 'selecting clauses $C_{n,1},\dots,C_{n,m(n)}$ uniformly at random independently among all possible clauses' as in the previous theorem, but only in the case that 'no two clauses $C_{n,i},C_{n,j}$ are equal'. In consequence

$$\Pr\big(F_{n,m(n)}\text{ is unsatisfiable }\big)=\Pr\big(UNSAT_n\,\big|\text{ all the }C_{n,i}\text{'s are different }\big),$$

where the event $UNSAT_n$ is defined as in the previous theorem. An application of the first order method yields that for $l>3$ a.a.s the number of unordered pairs $\{i,j\}$ such that $C_{n,i}=C_{n,j}$ is zero. For the case of $l=2$ an application of the factorial moments method proves that the number of such pairs $\{i,j\}$ converges in distribution to a Poisson variable of positive mean. In either case we have

$$\lim_{n\to\infty}\Pr\big(\text{ all the }C_{n,i}\text{'s are different }\big)>0.$$

In consequence the constants $c_1$ and $c_2$ from the previous theorem satisfy the statement of this corollary. □

---

Let $F_{n,m(n)}$ be as in last theorem. Notice that because the symmetry in the random model $F(l,n,p(n))$ one can consider $F_{n,m(n)}$ to be a random sample of the space $F(l,n,p(n))$ conditioned to the event that the number of clauses is $m(n)$. Using this observation we can prove the following result:

**Theorem 5.3.** *Let $l \geq 2$ be a natural number. For each $n \in \mathbb{N}$ let $F_n(\beta)$ be a random formula from $F(l,n,\beta/n^{l-1})$. Then there are real positive values $\beta_1 < \beta_2$ such that for any $0 < \beta < \beta$ it holds*

$$\lim_{n\to\infty} \Pr\big(F_n(\beta) \textit{ is unsatisfiable }\big) = 0,$$

*and for any $\beta > \beta_2$ it holds*

$$\lim_{n\to\infty} \Pr\big(F_n(\beta) \textit{ is unsatisfiable }\big) = 1.$$

*Proof.* For each $n \in \mathbb{N}$ let $X_n(\beta)$ be the random variable that counts the clauses in $F_n(\beta)$. It is satisfied that $\mathrm{E}[X_n(\beta)] \sim \frac{\beta \cdot 2^l}{l!} n$. Let $c_1, c_2$ be as in last corollary. Define $\beta_1 := \frac{c_1 \cdot l!}{2^l}$ and $\beta_2 := \frac{c_2 \cdot l!}{2^l}$. Fix $\beta \in \mathbb{R}$ satisfying $0 < \beta < \beta_1$. Let $\varepsilon > 0$ be a real number such that $\frac{\beta \cdot 2^l}{l!} + \varepsilon < c_1$. For each $n \in \mathbb{N}$ set $\delta_1(n) := \left\lfloor \left(\frac{\beta \cdot 2^l}{l!} - \varepsilon\right) n \right\rfloor$ and $\delta_2(m) := \left\lfloor \left(\frac{\beta \cdot 2^l}{l!} + \varepsilon\right) n \right\rfloor$. Because of the Central Limit Theorem it holds

$$\lim_{n\to\infty} \Pr\left(\delta_1(n) \leq X_n(\beta) \leq \delta_2(n)\right) = 1. \tag{9}$$

Denote by $dp$ the probability density function of the variable $X_n(\beta)$. That is $dp(m) = \Pr(X_n(\beta) = m)$. Then, because of the previous equation it holds

$$\Pr\left(F_n(\beta) \text{ is unsatisfiable }\right) \sim \int_{\delta_1(n)}^{\delta_2(n)} dp(m) \cdot \Pr\left(F_n(\beta) \text{ is unsatisfiable } \middle| X_n(\beta) = m\right).$$

Notice that the property of being unsatisfiable is monotonous. That is, for any two natural numbers $m_1 < m_2$ it holds

$$\Pr\left(F_n(\beta) \text{ is unsatisfiable } \middle| X_n(\beta) = m_2\right) < \Pr\left(F_n(\beta) \text{ is unsatisfiable } \middle| X_n(\beta) = m_2\right).$$

In consequence,

$$\int_{\delta_1(n)}^{\delta_2(n)} dp(m) \cdot \Pr\left(F_n(\beta) \text{ is unsatisfiable } \middle| X_n(\beta) = m\right) \leq$$
$$\Pr\left(F_n(\beta) \text{ is unsatisfiable } \middle| X_n(\beta) = \delta_2(n)\right) \cdot \Pr\big(\delta_1(n) \leq X_n(\beta) \leq \delta_2(n)\big).$$

And because of eq. (9),

$$\Pr\left(F_n(\beta) \text{ is unsatisfiable } \middle| X_n(\beta) = \delta_2(n)\right) \cdot \Pr\left(\delta_1(n) \leq X_n(\beta) \leq \delta_2(n)\right) \sim$$
$$\Pr\left(F_n(\beta) \text{ is unsatisfiable } \middle| X_n(\beta) = \delta_2(n)\right).$$

Finally, as $\delta_2(n) < c_2 n$, because of the previous corollary

$$\lim_{n\to\infty} \Pr\left(F_n(\beta) \text{ is unsatisfiable } \middle| X_n(\beta) = \delta_2(n)\right) = 0.$$

Thus, joining everything, we have proven that for any $\beta < \beta_1$, it holds that $F_n(\beta)$ a.a.s is satisfiable, as we wanted. Showing that for any $\beta > \beta_2$, a.a.s $F_n(\beta)$ is unsatisfiable is analogous. We fix $\varepsilon > 0$ such that $\frac{\beta \cdot 2^l}{l!} - \varepsilon > c_2$. We define $\delta_1(n)$ and $\delta_2(n)$ as before. Then similarly to before using the Central Limit Theorem and the fact that the property of being unsatisfiable is monotonous one can prove the bound

$$\lim_{n\to\infty} \Pr\left(F_n(\beta) \text{ is unsatisfiable }\right) \geq \lim_{n\to\infty} \Pr\left(F_n(\beta) \text{ is unsatisfiable } \middle| X_n(\beta) = \delta_1(n)\right).$$

And using the previous corollary we obtain that last limit equals one, and the result follows.

$\square$

---

A direct consequence of last theorem, due to Albert Atserias, is the following

**Theorem 5.4.** *Let $l > 1$ be a natural number. For each $n \in \mathbb{N}$ let $F_n(\beta)$ be a random formula from $F(l, n, \beta/n^{l-1})$. Let $\Phi \in FO[\sigma]$ be a first order sentence that implies unsatisfiability. Then for all $\beta \in [0, \infty)$ it holds*

$$\lim_{n\to\infty} \Pr\big(F_n(\beta) \models \Phi\big) = 0.$$

*Proof.* Let $\beta_1$ and $\beta_2$ be as in theorem 5.3. As $\Phi$ implies unsatisfiability it holds $\Pr\big(F_n(\beta) \models \Phi\big) \leq \Pr\big(F_n(\beta) \text{ is unsatisfiable }\big)$. Thus, using theorem 5.3 we get that for all $\beta \in [0, \beta_1]$

$$\lim_{n\to\infty} \Pr\big(F_n(\beta) \models \Phi\big) = 0.$$

Because theorem 5.1 last limit varies analytically with $\beta$, so if it vanishes in a proper interval $[0, \beta_1]$ then it has to vanish in the whole $[0, \infty)$ by the principle of analytic continuation, and the result holds. $\square$

# Conclusions

# Acknowledgments

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement ERC-2014-CoG 648276 AUTAR).

Albert and Marc Noy

# 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.

[11] Béla Bollobás and Bollobás Béla. *Random graphs*. Number 73. Cambridge university press, 2001.

[12] John Franco and Marvin Paull. Probabilistic analysis of the davis putnam procedure for solving the satisfiability problem. *Discrete Applied Mathematics*, 5(1):77–87, 1983.

[13] Vašek Chvátal and Endre Szemerédi. Many hard examples for resolution. *Journal of the ACM (JACM)*, 35(4):759–768, 1988.

[14] J-C Simon, Jacques Carlier, Olivier Dubois, O Moulines, and J-L LIONS. Etude statistique de l'existence de solutions de problèmes sat, application aux systèmes-experts. *Comptes rendus de l'Académie des sciences. Série 1, Mathématique*, 302(7):283–286, 1986.