Probabilities of first order sentences on sparse random relational structures: An application to definability on random CNF formulas

Lázaro Alberto Larrauri
Universitat Politècnica de Catalunya
lazaro.alberto.larrauri@upc.edu

Abstract
We extend the convergence law for sparse random graphs proven by Lynch to arbitrary relational languages. 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 when the expected number of tuples satisfying each relation in $\sigma$ is linear. It is also shown that these limit probabilities are well-behaved with respect to several parameters that represent the density of tuples in each relation $R$ in the vocabulary $\sigma$. An application of these results to the problem of random Boolean satisfiability is presented. We show that in a random $k$-CNF formula on $n$ variables, where each possible clause occurs with probability $\sim c/n^{k-1}$, independently any first order property of $k$-CNF formulas that implies unsatisfiability does almost surely not hold as $n$ tends to infinity.

Keywords: random hypergraphs, convergence law, random SAT, asymptotic probability, unsatisfiability certificate.
Introduction

We say that a sequence of random structures \( \{G_n\}_n \) satisfies a limit law with respect to some logical language \( L \) if for every property \( P \) expressible in \( L \) the probability that \( G_n \) satisfies \( P \) tends to some limit as \( n \to \infty \). If that limit takes only the values zero and one then we say that \( \{G_n\}_n \) satisfies a zero-one law with respect to \( L \).

Convergence and zero-one laws have been extensively studied on the binomial graph \( G(n,p) \). The seminal theorem on this topic, due to Fagin [7] and Glebskii et al. [9] independently, concerns general relational structures. When applied to graphs it states that if \( p \) is fixed, then \( G(n,p) \) satisfies a zero-one law with respect to the first order (FO) language of graphs.

This zero-one law was later extended by Shelah and Spencer in [12]. There it is proven, among other results, that if \( p := p(n) \) is a decreasing function of the form \( n^{-\alpha} \) and \( \alpha > 0 \) is irrational, then \( G(n,p(n)) \) obeys a zero-one law with respect to FO logic. Moreover, it is also proven that if \( \alpha \in (0,1) \) is rational then \( G(n,p(n)) \) does not obey a convergence law.

This was further studied by Lynch in [10], where it is shown that in the case where the expected number of edges is linear, i.e. when \( p(n) \sim \beta/n \) for some \( \beta > 0 \), then \( G(n,p(n)) \) satisfies a limit law with respect to FO logic. The following is a restatement of the main result in that article.

**Theorem** (Lynch, 1992). Let \( p(n) \sim \beta/n \). For every FO sentence \( \phi \), the function \( F_\phi : (0,\infty) \to [0,1] \) given by

\[
F_\phi(\beta) = \lim_{n \to \infty} \Pr(G(n,p(n)) \text{ satisfies } \phi)
\]

is well defined and is given by an expression with parameter \( \beta \) built using rational constants, addition, multiplication and exponentiation with base \( e \).

A relevant aspect of this result is that the limit probability of any FO property in \( G(n,p(n)) \) when \( p(n) \sim \beta/n \) varies analytically with \( \beta \). A consequence of this is that FO logic cannot “capture” sudden changes in the structure of \( G(n,p(n)) \).

It was left open at the end of [10] whether the convergence law obeyed by \( G(n,p(n)) \) in the range \( p(n) \sim \beta/n \) could be generalized to other random models of relational structures that contain relations of arity greater than 2. A result in this direction was obtained in [11], among other zero-one and convergence laws. They consider the random model of \( d \)-uniform hypergraphs \( G^d(n,p) \) where each \( d \)-edge is added to a set of \( n \) labeled vertices independently with probability \( p \). It is shown that when \( p(n) \sim \beta/n^{d-1} \), i.e. when the expected number of edges is linear, \( G^d(n,p(n)) \) obeys a convergence law with respect to the FO language of \( d \)-uniform hypergraphs. With little additional work it can be shown that in these conditions the limit probability of any FO property of \( G^d(n,p(n)) \) varies analytically with \( \beta \). We extend this result to arbitrary relational structures on whose relations we can impose symmetry and anti-reflexivity constraints (Theorem 1.3).

This generalization is motivated by an application to the problem of random SAT. We continue the study started by Atserias in [1] with respect to the definability in first order logic of certificates for unsatisfiability that hold for typical unsatisfiable formulas. A random model for 3-CNF formulas where each possible clause over \( n \) variables is added independently with probability \( p \) is considered there. In this model the expected number
of clauses $m$ is $\Theta(n^3p)$ as $n$ grows. The main result of that article states the following: (1) if $m = \Theta(n^{2-\alpha})$ for an irrational number $\alpha > 0$, then no FO property of 3-CNF formulas that implies unsatisfiability holds asymptotically almost surely (a.a.s.) for unsatisfiable formulas, and (2) if $m = \Theta(n^{2+\alpha})$ for $\alpha > 0$, then there exists some FO property that implies unsatisfiability and holds a.a.s. for unsatisfiable formulas.

The second part of the statement is the simpler one to prove: it can be shown that when $m = \Theta(n^{2+\alpha})$ for some $\alpha > 0$ the random 3-CNF formula a.a.s. contains some fixed unsatisfiable subformula (which depends on the choice of $\alpha$). This is clearly expressible in FO logic, so (2) follows. The proof of (1) is more involved and, in fact, shows something stronger: if $m = \Theta(n^{2-\alpha})$ for $\alpha > 0$ irrational, then all FO properties that imply unsatisfiability a.a.s. do not hold. This proof employs techniques based in those used by Shelah and Spencer in [12] to prove that $G(n, p)$ satisfies a zero-one law with respect to FO logic when $p$ is an irrational power of $n$.

Since the techniques used to prove (2) rely on the fact that $\alpha$ is irrational, the study of the range $m = \Theta(n)$ (that is, $m = \Theta(n^{2-\alpha})$ with $\alpha = 1$), was left open. This range is of special interest because it is where the phase transition from almost sure satisfiability to almost sure unsatisfiability takes place. It was shown in [3] that a random $k$-CNF formula with $m$ clauses over $n$ variables satisfying that $m \sim cn$ is a.a.s satisfiable for all sufficiently small values of $c$ and is a.a.s unsatisfiable for all sufficiently large values of $c$.

The possibility of studying FO definability of certificates for unsatisfiability in random $l$-CNF formulas with a linear expected number of clauses using a generalization of Lynch theorem was suggested by Atserias. This application is discussed in Section 5. We give a brief overview of it here. Let $F(l, n, p)$ be a random model of $l$-CNF formulas where each $l$-clause over $n$ variables is chosen independently with probability $p$. Let $F_n^l(\beta)$ denote a random formula in $F(l, n, p)$ where $p := p(n) \sim \beta/n^{l-1}$. Suppose that every FO property of $l$-CNF formulas has a well defined asymptotic probability in $F_n^l(\beta)$ for any $\beta > 0$. Further suppose that these asymptotic probabilities vary analytically with $\beta$. Then any FO property that implies unsatisfiability a.a.s does not hold in $F_n^l(\beta)$ for $\beta > 0$. Indeed, let $P$ be one such FO property. One can find a value $\beta_0 > 0$ satisfying that a.a.s $F_n^l(\beta)$ is satisfiable when $0 < \beta < \beta_0$. As a consequence $P$ a.a.s does not hold in $F_n^l(\beta)$ when $0 < \beta < \beta_0$. Since the asymptotic probability of $P$ varies analytically with $\beta$ and it vanishes in the non-empty interval $(0, \beta_0)$, because of the Principle of analytical continuation it must be true that a.a.s $P$ does not hold in $F_n^l(\beta)$ for all $\beta > 0$.

1 Preliminaries

1.1 General notation

Given a positive natural number $n$, we write $[n]$ to denote the set $1, 2, \ldots, n$. Given numbers, $n, m \in \mathbb{N}$ with $m \leq n$ we denote by $(n)_m$ the $m$-th falling factorial of $n$. Given a set $S$ and a natural number $k \in \mathbb{N}$ we use $\binom{S}{k}$ to denote the set of subsets of $S$ of size $k$. Given a set $S$ and $n \leq |S|$, 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 $(S)_* := \bigcup_{n \leq |S|} (S)_n$.

We use the convention that over-lined variables, like $\overline{x}$, denote ordered tuples of arbitrary length. Given an ordered tuple $\overline{x}$ we define $\text{len}(\overline{x})$ as its length. Given a tuple $\overline{x}$ and an element $x$ the expression $x \in \overline{x}$ means that $x$ appears as some coordinate in
Theorem 1.1. Let \( \pi \) be a positive integer variable over the same probability space. Let \( X \) denote conditional expectation of \( X \). Then the shortening some expressions in this situation, given a random variable \( X \) establishes that if \( \lambda \) means surely (a.a.s.) if \( \Pr(\lambda) = 0 \) for \( \lambda \) large enough we write \( \lambda \equiv \lambda \) by this action. Given \( \Phi \), the quotient of \( S^a/\Phi \) by \( [x_1, \ldots, x_a] \) or \( [\pi] \). Thus, for \( g \in \Phi \), by definition \([x_1, \ldots, x_a] = [x_{g(1)}, \ldots, x_{g(a)}] \). The notations \( \pi \) and \( (x_1, \ldots, x_a) \) represent ordered tuples while \( [\pi] \) and \( [x_1, \ldots, x_a] \) denote ordered tuples modulo the action of some arbitrary group of permutations. Which group it is will depend on the ambient set where \( [x_1, \ldots, x_a] \) belongs and it should either be clear from context or not be relevant.

Given real functions over the natural numbers \( f, g : \mathbb{N} \to \mathbb{R} \) the expressions \( f = O(g) \), \( f = o(g) \) and \( f = \Theta(g) \) have their usual meaning. If \( g(n) \neq 0 \) for \( n \) large enough we write \( f \sim g \) if \( \lim_{n \to \infty} \frac{f(n)}{g(n)} = 1 \).

1.2 Probabilistic preliminaries

We assume familiarity with basic probability theory. We denote by \( \text{Poiss}_\lambda(n) \) the discrete probability mass function of a random Poisson variable with mean \( \lambda \). That is, \( \text{Poiss}_\lambda(n) = e^{-\lambda \frac{n}{n!}} \). We define \( \text{Poiss}_\lambda(\geq n) = 1 - \sum_{i=0}^{n-1} \text{Poiss}_\lambda(i) \).

Given some sequence of events \( \{A_n\}_n \) we say that \( A_n \) is satisfied asymptotically almost surely (a.a.s.) if \( \Pr(A_n) \) tends to \( 1 \) as \( n \to \infty \). Given a sequence of random variables \( \{X_n\}_n \), the first moment method is an application of Markov’s inequality that establishes that if \( \text{E}[X_n] \) tends to zero as \( n \to \infty \) then a.a.s \( X_n = 0 \).

If \( A, B \) are events we may write the conditioned probability \( \Pr(A \mid B) \) as \( \Pr_B(A) \) to shorten some expressions. In this situation, given a random variable \( X \) we put \( \text{E}_B[X] \) to denote conditional expectation of \( X \) given the event \( B \).

Our main tool for proving the convergence in distribution to Poisson variables is the next result, which can be found in [2, Theorem 1.23].

**Theorem 1.1.** Let \( l \in \mathbb{N} \). For each \( n \in \mathbb{N} \), let \( X_{n,1}, \ldots, X_{n,l} \) be non-negative random integer variables over the same probability space. Let \( \lambda_1, \ldots, \lambda_l \) be real numbers. Suppose for any \( r_1, \ldots, r_l \in \mathbb{N} \)

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

Then the \( X_{n,1}, \ldots, X_{n,l} \) converge in distribution to independent Poisson variables with means \( \lambda_1, \ldots, \lambda_l \) respectively.

We use the following observation in order to compute the binomial moments of our random variables.

**Observation 1.1.** Let \( X_1, \ldots, X_l \) be non-negative random integer variables over the same probability space. Let \( r_1, \ldots, r_l \in \mathbb{N} \). Suppose each \( X_i \) is the sum of indicator random variables (i.e. variables that only take the values 0 and 1) \( X_i = \sum_{j=1}^{n_i} Y_{i,j} \). Define
$\Omega := \prod_{i=1}^l \binom{[n]}{r_i}$. 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

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

### 1.3 Logical preliminaries

We assume familiarity with first order logic (FO). We follow the convention that first order logic contains the equality symbol. Given a vocabulary $\sigma$ we denote by $FO[\sigma]$ the set of first order formulas of vocabulary $\sigma$. Given a relation symbol $R \in \sigma$ we denote by $ar(R)$ the arity of $R$. Given a formula $\phi \in FO[\sigma]$ we use the notation $\phi(\bar{y})$ to express that $\bar{y}$ is a tuple of (different) variables which contains all free variables in $\phi$ and none of its bounded variables, but it may contain variables which do 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 $\phi$, written as $qr(\phi)$, is the maximum number of nested quantifiers in $\phi$. We call **edge sentence** to any consistent open formula that contains no occurrence of the equality symbol ‘$=$’.

### 1.4 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 $P_R \subseteq \binom{[ar(R)]}{2}$ for all $R \in \sigma$.

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

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

  $$\forall \bar{x} := x_1, \ldots, x_{ar(R)} \ (R(\bar{x}) \iff R(g\bar{x}))$$

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

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

Structures in $\mathcal{C}$ generalize the usual notion of a hypergraph in the sense that they contain multiple “adjacency” relations with arbitrary symmetry and anti-reflexivity axioms.

We use the usual graph theory nomenclature and notation with some minor changes. In the scope of this article **hypergraphs** are structures in $\mathcal{C}$. Given a hypergraph $G$ its **vertex set** $V(G)$ is its universe.

In order to define the edge sets of $G$ we need the following auxiliary definition
Definition 1.1. Let $V$ be a set, and let $R \in \sigma$. We define the set of possible edges over $V$ given by $R$ as

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

where

$$X = \big\{[v_1, \ldots, v_{ar(R)}] \mid v_1, \ldots, v_{ar(R)} \in V, \text{ and } v_i = v_j \text{ for some } \{i, j\} \in P_R\big\}.$$

We call edges to the elements of $E_R[V]$ and we say that the sort of an edge $e \in E_R[V]$ is $R$. In the case where $V = [n]$ we write simply $E_R[n]$ instead of $E_R[[n]]$.

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 $P_R$.

Let $G$ be a hypergraph with vertex set is $V$ and let $R \in \sigma$ be a relation. We define the edge set of $G$ given by $R$, denoted by $E_R(G)$, as the set of edges $[\overline{v}] \in E_R[V]$ such that $\overline{v} \in R^G$. We define the total edge set of $G$ as the set $E(G) := \bigcup_{R \in \sigma} E_R(G)$.

An hypergraph $G$ is connected if for any two vertices $v, u \in V(G)$ there is a sequence of edges $e_1, \ldots, e_m \in E(G)$ 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 $\text{ex}(G) \geq -1$ for any connected hypergraph.

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

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

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 $d^G(u, v) = \infty$. When $G$ is understood or not relevant we simply write $d$ instead of $d^G$. Equivalently, the distance $d$ coincides with the usual one defined over the Gaifman graph of the structure $G$. The diameter of a hypergraph is the maximum distance between any pair of vertices. We extend naturally the distance $d$ to sets and tuples of vertices, as usual. Given a vertex/set/tuple $X$ and a number $r \in \mathbb{N}$ we define the neighborhood $N^G(X; r)$, or simply $N(X; r)$ when $G$ is not relevant, as the set of vertices $v$ such that $d^G(X, v) \leq r$. 


A connected hypergraph $G$ is a **path** between two of its vertices $v, u \in V(G)$ if $G$ does not contain any connected proper sub-hypergraph containing both $v, u$. A connected hypergraph $G$ is a **tree** if $\text{ex}(G) = -1$ and **dense** if $\text{ex}(G) > 0$. An hypergraph is called **$r$-sparse** if it does not contain any dense sub-hypergraph $H$ such that $\text{diam}(H) \leq r$. A connected hypergraph $G$ with $\text{ex}(G) \geq 0$ is called **saturated** if for any non-empty proper sub-hypergraph $H \subset G$ it holds $\text{ex}(H) < \text{ex}(G)$. A connected hypergraph $G$ with $\text{ex}(G) = 0$ is called a **unicycle**. A saturated unicycle is called a **cycle**. We say that an edge $e := [\overline{v}]$ contains a **loop** if some vertex $v$ appears in $\overline{v}$ more than once.

A **rooted tree** $(T, v)$ is a tree $T$ with a distinguished vertex $v \in V(T)$ called its **root**.

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

**Isomorphisms** between hypergraphs are defined as isomorphisms between relational structures. Isomorphisms between $\Sigma$-hypergraphs are just isomorphisms between the underlying hypergraphs that also preserve their colorings. In both cases we denote the isomorphism relation by $\cong$. Given a hypergraph $H$, resp. a $\Sigma$-hypergraph $(H, \chi)$, an **automorphism** of $H$, resp. $(H, \chi)$, is an isomorphism from $H$, resp. $(H, \chi)$, to itself. We denote by $\text{aut}(H)$, resp. $\text{aut}(H, \chi)$, the number of such automorphisms.

Let $H$ be a hypergraph and let $V$ be a set. We define the set of **copies of $H$ over $V$**, denoted as $\text{ Copies}(H, V)$, as the set of hypergraphs $H'$ such that $V(H') \subset V$ and $H \cong H'$. Let $\chi$ be a $\Sigma$-coloring of $H$. Analogously, we define the set $\text{ Copies}((H, \chi), V)$ as the set of $\Sigma$-hypergraphs $(H', \chi')$ satisfying $V(H') \subset V$ and $(H, \chi) \cong (H', \chi')$. Let $\mathbb{H}$ be an isomorphism class of $\Sigma$-hypergraphs. Then the set $\text{ Copies}(\mathbb{H}, V)$ 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 $\text{ Copies}(\mathbb{H}, V; (v, s))$ as the set of $\Sigma$-hypergraphs $(H', \chi') \in \text{ Copies}(\mathbb{H}, V)$ that satisfy $v \in V(H')$ as well as $\chi'(v) = s$.

Given $\mathbb{H}$ an isomorphism class of hypergraphs or $\Sigma$-hypergraphs, we define expressions such as $\text{ex}(\mathbb{H})$, $\text{aut}(\mathbb{H})$, $|V(\mathbb{H})|$, $|E(\mathbb{H})|$ or $\text{ Copies}(\mathbb{H}, V)$ via representatives of $\mathbb{H}$.

### 1.5 Ehrenfeucht-Fraisse Games

We assume familiarity with Ehrenfeucht-Fraisse (EF) games. An introduction to the subject can be found for instance in [1][Section 2], for example. Given hypergraphs $H_1$ and $H_2$ we denote the $k$-round EF game played on $H_1$ and $H_2$ by $\text{EHR}_k(H_1; H_2)$. The following is satisfied:

**Theorem 1.2** (Ehrenfeucht, [1]). 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 $q_f(\phi) \leq k$.

Given lists $\overline{v} \in V(H_1)^*$, and $\overline{u} \in V(H_2)^*$ of the same length, we denote the $k$-round Ehrenfeucht-Fraisse game on $H_1$ and $H_2$ with initial position given by $\overline{v}$ and $\overline{u}$ by $\text{EHR}_k(H_1, \overline{v}; H_2, \overline{u})$.

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: For any
i, j ∈ [k], \(d^{H_1}(v_i, v_j) = d^{H_2}(u_i, u_j)\), where \(v_s\) and \(u_s\) denote the vertex played on \(H_1\), resp. \(H_2\) in the \(s\)-th round of the game. 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 \(dEHR_k(H_1, \overline{v}; H_2, \overline{u})\) analogously to \(EHR_k(H_1, \overline{v}; H_2, \overline{u})\).

### 1.6 The random model

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

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

Equivalently, this is the probability space obtained by assigning to each edge \(e \in E_R[n]\) probability \(p_R\) independently for each \(R \in \sigma\).

As in the case of Lynch theorem, we are interested in the ”sparse regime” of \(G^\sigma(n, \{p_R\}_{R})\), were the expected number of edges of each sort is linear. This is achieved when for each \(R \in \sigma\) it holds \(p_R(n) \sim \beta_R/n^\alpha(R) - 1\) for some \(\beta_R > 0\). We write \(G_n(\{\beta_R\}_{R})\) to denote a random sample of \(G^\sigma(n, \{p_R\}_{R})\) when the probabilities \(p_R\) satisfy \(p_R(n) \sim \beta_R/n^\alpha(R) - 1\). When the choice of \(\{\beta\}_{R}\) is not relevant we write \(G_n\) instead of \(G_n(\{\beta_R\}_{R})\).

### 1.7 Main definitions

Our main definition follow closely the ones in [10] adapted to the context of hypergraphs.

**Definition 1.2.** Let \(H\) be a connected hypergraph. Then \(H\) contains a unique maximal saturated sub-hypergraph \(H'\) satisfying satisfies \(\text{ex}(H') = \text{ex}(H)\) if \(\text{ex}(H) \geq 0\), and \(H' = \emptyset\) otherwise. Given \(\overline{v} \in V(H)^*\) we define \(\text{Center}(H, \overline{v})\) as the minimal connected sub-hypergraph in \(H\) that contains both \(H'\) and the vertices in \(\overline{v}\). If \(H\) is not connected we define \(\text{Center}(H, \overline{v})\), as the union of \(\text{Center}(H^{\prime}, \overline{v})\) for all connected components \(H^{\prime} \subset H\), where \(\overline{v} \in V(H)^*\) contains exactly the vertices in \(\overline{v}\) belonging to \(V(H^{\prime})\). When \(\overline{v}\) is empty we simply write \(\text{Center}(H)\).

**Definition 1.3.** Let \(H\) be a hypergraph, \(\overline{v} \in V(H)^*\) and \(r \in \mathbb{N}\). Let \(X\) be the set of vertices \(v \in V(H)\) that either belong to \(\overline{v}\) or belong to some saturated sub-hypergraph of \(H\) with diameter at most \(2r + 1\). We define \(\text{Core}(H, \overline{v}; r)\) as \(N(X; r)\). If \(\overline{v}\) is empty we write \(\text{Core}(H; r)\). We say that \(H\) is \(r\)-simple if all connected components of \(\text{Core}(H; r)\) are unicycles.

**Definition 1.4.** Let \(H\) be a hypergraph, let \(\overline{v} \in V(H)^*\) and let \(v \in H\) be such that \(d(\text{Center}(H, \overline{v}), v) < \infty\). Let \(X \subset V(H)\) be the set

\[
X := \{ u \in V(H) \mid d(\text{Center}(H, \overline{v}), u) = d(\text{Center}(H, \overline{v}), v) + d(v, u) \}.
\]

Then we define \(\text{Tr}(H, \overline{v}; v)\) as the tree \(H[X]\) with \(v\) as a root. That is, \(\text{Tr}(H, \overline{v}; v)\) is the tree formed of all vertices whose only path to \(\text{Center}(H, \overline{v})\) contains \(v\). One can easily check that \(H[X]\) is indeed a tree: if it were not then it would contain some saturated sub-hypergraph, leading to a contradiction. Given \(r \in \mathbb{N}\) we define \(\text{Tr}(H, \overline{v}; v; r)\) as \(\text{Tr}(\text{Core}(H, \overline{v}; r), \overline{v}; v)\). In the case that \(\overline{v}\) is the empty list we write simply \(\text{Tr}(H; v)\) or \(\text{Tr}(H; v; r)\).
For any $k \in \mathbb{N}$ we define an equivalence relation over rooted trees which generalizes both the relation of "$k$-morphism" as defined in [10], and the notion of "$(k, r)$-values" defined in [11].

**Definition 1.5.** 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 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 $\sim_k$ classes of trees with radius at most $r - 1$. Let $\rho$ be an special symbol called the root symbol. Set $\hat{\Sigma}_{k,r-1} := \Sigma_{k,r-1} \cup \{\rho\}$. Then a $(k, r)$-pattern is isomorphism class of $\hat{\Sigma}_{k,r-1}$-hypergraphs $(e, \tau)$ that consist of only one edge with no loops and no isolated vertices, and satisfy $\tau(v) = \rho$ for exactly one vertex $v \in V(e)$. We denote by $P(k, r)$ the set of $(k, r)$-patterns.

Given a rooted tree $(T, v)$ of radius $r$ we define its canonical $k$-coloring as the map $\tau_{k,(T,v)}^k : V(T) \rightarrow \hat{\Sigma}_{k,r-1}$ satisfying that $\tau_{k,(T,v)}^k(u)$ is the $\sim_k$ class of $\text{Tr}(T, u; v)$ for any $u \neq v$, and $\tau_{k,(T,v)}^k(v) = \tau$.

Let $T_1$ and $T_2$ be rooted trees of radius $r$. We say that $(T_1, v_1) \sim_k (T_2, v_2)$ if for any pattern $\epsilon \in P(k, r)$ the "quantity of initial edges $e_1 \in E(T_1)$ such that $(e, \tau_{(T_1,v_1)}^k) \in \epsilon$" and the "quantity of initial edges $e_2 \in E(T_2)$ such that $(e, \tau_{(T_2,v_2)}^k) \in \epsilon$" are equal or are both greater than $k - 1$.

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

**Observation 1.2.** Let $T$ be a $\sim_k$ class of rooted trees with radii at most $r$. Then there is a partition $E_1^T, E_2^T$ of $P(k, r)$ and natural numbers $a_\epsilon < k$ for each $\epsilon \in E_1^T$ that depends only on $T$ such that a rooted tree $(T, v)$ belongs to $T$ if and only if the following hold: (1) For any pattern $\epsilon \in E_1^T$ there are at least $k$ initial edges $e \in E(T)$ such that $(e, \tau_{(T,v)}^k) \in \epsilon$, and (2) for any pattern $\epsilon \in E_2^T$ there are exactly $a_\epsilon$ initial edges $e \in E(T)$ such that $(e, \tau_{(T,v)}^k) \in \epsilon$.

From this characterization of the $\sim_k$ relation it follows, by induction over $r$, that the quantity of $\sim_k$ classes of trees with radii at most $r$ is finite, for any $r \in \mathbb{N}$.

**Definition 1.6.** Let $k \in \mathbb{N}$. Given a non-tree connected hypergraph $H$, we define its canonical $k$-coloring $\tau_H^k$ as the one that assigns to each vertex $v \in V(H)$ the $\sim_k$ class of the tree $\text{Tr}(H, v)$. Let $H_1$ and $H_2$ be connected hypergraphs which are not trees. Set $H_1' := \text{Center}(H_1)$ and $H_2' := \text{Center}(H_2)$. We say that $H_1$ and $H_2$ are $k$-equivalent, written as $H_1 \sim_k H_2$, if $(H_1', \tau_{H_1'}^k) \simeq (H_2', \tau_{H_2'}^k)$.

**Definition 1.7.** Let $k, r \in \mathbb{N}$ and let $H_1$ and $H_2$ be hypergraphs. Let $H_1' := \text{Core}(H_1; r)$ and $H_2' := \text{Core}(H_2; r)$. 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 $\mathbf{H}$ "the number of connected components in $H_1'$ that belong to $\mathbf{H}$" and "the number of connected components in $H_2'$ that belong to $\mathbf{H}$" are the same or are both greater than $k - 1$.
Definition 1.8. Let \( k, r \in \mathbb{N} \) and let \( \Sigma_{(k, r)} \) be the set of \( \sim_k \) classes of rooted trees with radii at most \( r \). Then a \((k, r)\)-cycle is an isomorphism class of \( \Sigma_{(k, r)} \) -hypergraphs \((H, \tau)\) that are cycles of diameter at most \( 2r + 1 \). We denote by \( C(k, r) \) the set of \((k, r)\)-cycles.

Observation 1.3. Let \( k, r \in \mathbb{N} \) and let \( \mathbf{O} \) be a \( \approx_{k, r} \) class of \( r \)-simple hypergraphs. Then there is a partition \( U^0, U^1 \) of \( C(k, r) \) and natural numbers \( a_\omega < k \) for each \( \omega \in U^0 \) that depend only on \( \mathbf{O} \) such that a \( r \)-simple hypergraph \( G \) belongs to \( \mathbf{O} \) if and only if it holds that (1) for any \( \omega \in U^0 \) there are at least \( k \) connected components \( H \subseteq \text{Core}(G; r) \) whose cycle \( H' = \text{Center}(H) \) satisfies \( (H', \tau^k_H) \in \omega \), and (2) for any \( \omega \in U^1 \) there are exactly \( a_\omega \) connected components \( H \subseteq \text{Core}(G; r) \) whose cycle \( H' = \text{Center}(H) \) satisfies that \( (H', \tau^k_H) \in \omega \).

Definition 1.9. Let \( H \) be a hypergraph and let \( k, r \in \mathbb{N} \). Let \( X \subseteq V(H) \) 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 \( r' \leq r \), vertices \( v_1, \ldots, v_k \) and \( \sim_k \) class \( \mathbf{T} \) of trees with radius at most \( r' \) there exists a vertex \( v \in V(H) \) such that \( d(v, X) > 2r' + 1 \), \( d(v, v_i) > 2r' + 1 \) for all \( v_i \) and \( T := N(v; r') \) is a tree satisfying \( (T, v) \in \mathbf{T} \).

1.8 Main result and outline of the proof

Our goal is to prove the following theorem.

Theorem 1.3. Let \( \phi \) be a sentence in \( FO[\sigma] \). Then the function \( F_\phi : (0, \infty)^{|\sigma|} \rightarrow \mathbb{R} \) given by

\[
\{\beta_R\}_{R \in \sigma} \mapsto \lim_{n \to \infty} \Pr(G_n(\{\beta_R\}_R) \models \phi)
\]

is well defined and analytic.

In fact we prove something stronger. We show that the limit in last theorem is given by an expression with parameters \( \{\beta_R\}_R \) built using rational constants, sums, products and exponentiation with base \( e \). We do so by giving a family of expressions which contains the ones that define limit probabilities of \( FO \) properties in \( G_n(\{\beta\}_R) \).

The main arguments are similar to the ones in the proof of [10, Theorem 2.1], adapted to fit our context. As in that article the proof is divided into two parts: a model theoretic part and a probabilistic part. The main result of the first part is the following

Theorem 2.4. Let \( k \in \mathbb{N} \) and let \( H_1, H_2 \) be hypergraphs. Set \( r := (3k - 1)/2 \). Suppose that both \( H_1 \) and \( H_2 \) are \((k, r)\)-rich and \( H_1 \approx_{k, r} H_2 \). Then Duplicator wins \( \text{EHR}_k(H_1, H_2) \)

With regards to the second part, the “landscape” of \( G_n \) can be described similarly to the one of \( G(n, c/n) \) as in [13]: A.a.s for any fixed radius \( r \) all neighborhoods \( N(v; r) \) in \( G_n \) are trees or unicycles, so cycles in \( G_n \) are far apart. One can find arbitrarily many copies of any fixed tree, while the expected number of copies of any fixed cycle is finite. The main probabilistic results are the following:

Theorem 3.2. Let \( r \in \mathbb{N} \). Then a.a.s \( G_n \) is \( r \)-simple.

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

Theorem 3.5. Let \( k, r \in \mathbb{N} \). Let \( \mathbf{O} \) be a \( \approx_{k, r} \) class of \( r \)-simple hypergraphs. Then

\[
\lim_{n \to \infty} \Pr(G_n(\{\beta_R\}_{R \in \sigma}) \in \mathbf{O})
\]

exists and is an analytic expression in \( \{\beta_R\}_{R \in \sigma} \).
A sketch of the proof of Theorem 3.5 using these results as follows. Let $\Phi \in FO[\sigma]$ be a sentence and let $k := qr(\Phi)$, $r := (3^k - 1)/2$. Because of Theorems 2.3 and 3.4 it holds that for any $\approx_{k,r}$ class $O$

$$\lim_{n \to \infty} \Pr (G_n \models \Phi \mid G_n \in O) = 0 \text{ or } 1.$$ 

This together with Theorem 3.2 and the fact that there is a finite number of $\approx_{k,r}$-classes of $r$-simple hypergraphs imply that $\lim_{n \to \infty} \Pr (G_n \models \Phi)$ equals a finite sum of limits of the form $\lim_{n \to \infty} \Pr (G_n \in O)$, where $O$ is some $\approx_{k,r}$-class of $r$-simple hypergraphs. Finally, using Theorem 3.5 we get that $\lim_{n \to \infty} \Pr (G_n \models \Phi)$ exists and is an analytic expression in $\{\beta_R\}_R$, as we wanted.

2 Model theoretic results

2.1 Winning strategies for Duplicator

During this section $H_1$ and $H_2$ stand for hypergraphs and $V_1 := V(H_1)$, $V_2 := V(H_2)$.

**Definition 2.1.** Let $\bar{v} \in V_1^r$, $\bar{w} \in V_2^r$ be tuples of the same length. We write $(H_1, \bar{v}) \simeq_{k,r} (H_2, \bar{w})$, if Duplicator wins $dEHR_k (N(\bar{v}; r), \bar{v}; N(\bar{w}; r), \bar{w})$. Given $X \subseteq V_1$ and $Y \subseteq V_2$ we write $(H_1, X) \simeq_{k,r} (H_2, Y)$, if we can order $X$, resp. $Y$, to form lists $\bar{v}$, resp. $\bar{w}$, such that $(H_1, \bar{v}) \simeq_{k,r} (H_2, \bar{w})$. Given $X \subseteq V_1$, $Y \subseteq V_2$ and tuples of the same length $\bar{v} \in V_1^r$ and $\bar{w} \in V_2^r$ we write $(H_1, (X, \bar{v})) \simeq_{k,r} (H_2, (Y, \bar{w}))$, if $X$ and $Y$ can be ordered to form lists $\bar{v}$, resp. $\bar{w}$ such that $(H_1, \bar{v} - \bar{v}) \simeq_{k,r} (H_2, \bar{v} - \bar{w})$.

**Definition 2.2.** Fix $r \in \mathbb{N}$. Suppose $X \subseteq V_1$ and $Y \subseteq V_2$ can be partitioned into sets $X = X_1 \cup \cdots \cup X_6$ and $Y = Y_1 \cup \cdots \cup Y_6$ such that all $N(X_i; r)$ and $N(Y_i; r)$ are connected and disjoint. We write $(H_1, X) \cong_{k,r} (H_2, Y)$, if for any set $Z \subseteq V_6$ with $\delta \in \{1, 2\}$, among the $X_i$ or the $Y_i$ it is satisfied that “the number of $X_i$ such that $(H_\delta, Z) \simeq_{k,r} (H_1, X_i)$” and “the number of $Y_i$ 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 strengthening of [14] Theorem 2.6.7], is the following.

**Theorem 2.1.** Let $k \in \mathbb{N}$. Set $r := (3^k - 1)/2$. 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 $\bar{v} \in (V_2)^{k-1}$ be a tuple of vertices. Then there exists $u \in V_2$ such that $d(Y, u) > 2r' + 1$, $d(V, u) > 2r' + 1$ and $(H_1, v) \simeq_{k,r} (H_2, u)$.

Then Duplicator wins $EHR_k (H_1; H_2)$.
In order to prove this theorem we need to make two observations and prove a previous lemma.

**Observation 2.1.** Let \( k \in \mathbb{N} \) and let \( \pi \in V(H_1)^* \), \( \pi \in V(H_2)^* \) be of equal length. Suppose Duplicator wins \( dEHR_k(H_1, 3r + 1) \). Then, for any \( r \in \mathbb{N} \), \( (H_1, \pi) \simeq_{k,r} (H_2, \pi) \).

**Observation 2.2.** Let \( k \in \mathbb{N} \) and let \( \pi \in V(H_1)^* \), \( \pi \in V(H_2)^* \) be of equal length. Suppose Duplicator wins \( dEHR_k(H_1, 3r + 1) \). Let \( v \in V(H_1) \), \( u \in V(H_2) \) be the vertices played in the first round of an instance of the game where Duplicator is following a winning strategy. Then Duplicator also wins \( dEHR_{k-1}(H_1, \pi; H_2, \pi) \), where \( \pi := \pi^* u \).

**Lemma 2.1.** Let \( k, r \in \mathbb{N} \). Let \( \pi \in V_1^* \) and \( \pi \in V_2^* \) be of equal length. \( (H_1, \pi) \simeq_{k,3r+1} (H_2, \pi) \). Let \( v \in V_1 \) and \( u \in V_2 \) be vertices played in the first round of an instance of

\[
dEHR_k(N(\pi; 3r + 1), \pi; N(\pi; 3r + 1), \pi)
\]

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

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

\[
dEHR_{k-1}(N(\pi; 3r + 1), \pi; N(\pi; 3r + 1), \pi)
\]

as well. Call \( H_1' = N(\pi; 3r + 1) \), \( H_2' = N(\pi; 3r + 1) \). Then by Observation 2.2 Duplicator wins

\[
dEHR_{k-1}(N^{H_1}(\pi; r), \pi; N^{H_2}(\pi; r), \pi)
\]

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

**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 \( \simeq_{k,r} \). Let \( r_0 := (3^k - 1)/2 \) and \( r_i := (r_{i-1} - 1)/3 \) for each \( 1 \leq i \leq k \). Let \( v_i \) and \( v_i' \) be the vertices played in \( H_1 \) and \( H_2 \) respectively during the \( i \)-th round of \( EHR_k(H_1, H_2) \). We show a winning strategy for Duplicator in \( EHR_k(H_1, H_2) \). For each \( 0 \leq i \leq k \), 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 \( \pi(T) \in V_\delta^* \) 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 \( \pi(S) := v_i \) and \( \pi(T) := v_i' \).
- Given two sets \( S \subset V_1 \), \( T \subset V_2 \) that were previously marked during the same round, append \( v_i \) and \( v_i' \) to \( \pi(S) \) and \( \pi(T) \) respectively.

We show that Duplicator can play in such a way that at the end round the following are satisfied:
(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 $\overline{v}(T)$.

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

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

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

$$dEHR_{k-i} \left( N(\overline{v}^1; r_i), \overline{v}^1; N(\overline{v}^2; r_i), \overline{v}^2 \right),$$

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

The game $dEHR_k(H_1; H_2)$ 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 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 $S \subset V_1$ we have $d(S \cup \overline{v}, v_i^1) \leq 2r_i + 1$. In this case let $T \subset V_2$ be the set in $H_2$ marked in the same round as $T$. By hypothesis

$$\left( H_1, (S, \overline{v}(S)) \right) \simeq_{k-i+1, 3r_i+1} \left( H_2, (T, \overline{v}(T)) \right).$$

Then, by definition, for some orderings $\overline{w}, \overline{z}$ of the vertices in $S$ and $T$ respectively it holds that Duplicator wins

$$dEHR_{k-i+1} \left( N(\overline{w}; \overline{v}(S); 3r_i+1), \overline{w}; N(\overline{z}; \overline{v}(T); 3r_i+1), \overline{z} \right).$$

Thus Duplicator can choose $v_i^2 \in V_2$ according to the winning strategy in that game. After this Duplicator sets $\overline{v}(S) := \overline{v}(S) \setminus v_i^1$, and $\overline{v}(T) := \overline{v}(T) \setminus v_i^2$. Notice that because of Lemma 2.1 now

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

- For all marked sets $S \subset V_1$ it holds $d(S \cup \overline{v}(S), v_i^1) > 2r_i + 1$, but there is a unique $S$ among $X_1, \ldots, X_a$ 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 $Y_1, \ldots, Y_b$ such that

$$\left( H_1, S \right) \simeq_{k-i+1, 3r_i+1} \left( H_2, T \right).$$

Thus, by definition, for some orderings $\overline{w}, \overline{z}$ of the vertices in $S$ and $T$ respectively, Duplicator wins

$$dEHR_{k-i+1} \left( N(\overline{w}; 3r_i+1), \overline{w}; N(\overline{z}; 3r_i+1), \overline{z} \right).$$
Then Duplicator can choose \( v^2_i \in V_2 \) according to a winning strategy for this game. After this Duplicator marks both \( S \) and \( T \) and sets \( \tau(S) := v^1_1 \), and \( \tau(T) := v^2_1 \). Notice that because of Lemma 2.1 now

\[(H_1, (S, \tau(S))) \simeq_{k-i,r_i} (H_2, (T, \tau(T))).\]

- For all marked sets \( S \subset V_1 \) we have \( d(S \cup \tau(S), v^1_1) > 2r_i + 1 \), and for all sets \( S \) among \( X_1, \ldots, X_n \) it also holds \( d(S, v^1_1) > 2r_i + 1 \). In this case from condition (2) of the statement it follows that Duplicator can choose \( v^2_i \in V_2 \) such that (A) \( d(T \cup \tau(T), v^2_i) > 2r_i + 1 \) for all marked sets \( T \subset V_2 \), (B) \( d(T, v^2_i) > 2r_i + 1 \) for all sets \( T \) among \( Y_1, \ldots, Y_n \), and (C) \( (H_1, v^1_1) \simeq_{k-i,r_i} (H_2, v^2_i) \). After this Duplicator marks both \( S = \{v^1_1\} \) and \( T = \{v^2_i\} \) and sets \( \tau(S) := v^1_1 \) and \( \tau(T) := v^2_i \).

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.

\[\square\]

### 2.2 k-Equivalent trees

We want prove the following.

**Theorem 2.2.** Let \( k \in \mathbb{N} \). 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 \( \text{Tr}(T, v; e) \) as the induced tree \( T[X] \) on the set \( X := \{v\} \cup \{u \in V(T) \mid d(v, u) = 1 + d(e, u)\} \), with \( v \) as the root. In other words, \( \text{Tr}(T, v; e) \) is the tree consisting of \( v \) and all the vertices in \( T \) whose only path to \( v \) contains \( e \).

**Lemma 2.2.** Let \( k \in \mathbb{N} \) and fix \( r > 0 \). Suppose 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 \( \tau^k_{(T_1, v_1)} \) and \( \tau^k_{(T_2, v_2)} \) be colorings over \( T_1 \) and \( T_2 \) as in Definition 1.5. Let \( e_1 \) and \( e_2 \) be initial edges of \( T_1 \) and \( T_2 \) respectively satisfying \( (e_1, \tau^k_{(T_1, v_1)}) \simeq (e_2, \tau^k_{(T_2, v_2)}) \). Name \( T'_1 \) := \( \text{Tr}(T_1, v_1; e_1) \) and \( T'_2 := \text{Tr}(T_2, v_2; e_2) \). 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 an isomorphism \( f : V(e_1) \to V(e_2) \) between \( (e_1, \tau^k_{(T_1, v_1)}) \) and \( (e_2, \tau^k_{(T_2, v_2)}) \). Suppose in the \( i \)-th round of the game Spoiler plays on \( T'_i \). The other case is symmetric. If Spoiler plays \( v_1 \) then Duplicator chooses \( v_2 \). Otherwise, Spoiler plays a vertex \( v \) that belongs to some \( \text{Tr}(T'_1, v_1; u) \) for a unique \( u \in V(e_1) \) different from the root \( v_1 \). Set \( T''_1 := \text{Tr}(T'_1, v_1; u) \) and \( T''_2 := \text{Tr}(T'_2, v_2; f(u)) \) Then, as \( \tau^k_{(T'_1, v_1)}(u) = \tau^k_{(T'_2, v_2)}(f(u)) \), we obtain \( (T''_1, u) \sim_k (T''_2, f(u)) \). As both these trees have radii at most \( r \), by assumption Duplicator has a winning strategy in \( d\text{EHR}_k(T''_1, u; T''_2, f(u)) \) and they can follow it considering the previous plays in \( T''_1 \) and \( T''_2 \).

**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 smaller values of \( r \). Let \( \tau^k_{(T_1; v_1)} \) and \( \tau^k_{(T_2; v_2)} \) be the colorings over \( T_1 \) and \( T_2 \) as in Definition 1.5. We 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 in the \( i \)-th round Spoiler plays \( T_1 \). 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 \( (e_1, \tau^k_{(T_1; v_1)}) \sim (e_2, \tau^k_{(T_2; v_2)}) \). 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 \( (e_1, \tau^k_{(T_1; v_1)}) \sim (e_2, \tau^k_{(T_2; v_2)}) \). 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 \( T'_2 \).

2.3 \( k \)-Equivalent hypergraphs

**Theorem 2.3.** Let \( H_1 \) and \( H_2 \) be non-tree connected hypergraphs satisfying \( H_1 \sim_k H_2 \). Set \( H'_1 := \text{Center}(H_1) \) and \( H'_2 := \text{Center}(H_2) \). Let \( \tau^k_{H_1}, \tau^k_{H_2} \) be as in Definition 1.6. Let \( f \) be an isomorphism between \( (H'_1, \tau^k_{H_1}) \) and \( (H'_2, \tau^k_{H_2}) \). Let \( \overline{\tau} \) be an ordering of the vertices of \( H'_1 \) and let \( \overline{\tau} := f(\overline{\tau}) \) be the corresponding ordering of the vertices of \( H'_2 \). Then Duplicator wins \( d\text{EHR}_k(H'_1, \overline{\tau}; H'_2, \overline{\tau}) \).

**Proof.** The winning strategy for Duplicator is as follows. Suppose 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 \( \text{Tr}(H_1; u) \) for a unique \( u \in H'_1 \). Set \( T_1 := \text{Tr}(H_1; u) \) and \( T_2 := \text{Tr}(H_2; f(u)) \). 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 \( d\text{EHR}_k(T_1, u; T_2, f(u)) \), and they can follow it taking into account the previous moves made in \( T_1 \) and \( T_2 \), if any. 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.

2.4 Main result

**Lemma 2.3.** Let \( k, r \in \mathbb{N} \) 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 a saturated sub-hypergraph of diameter at most \( 2r + 1 \). Then \( (H_1, X) \approx_{k,r} (H_2, Y) \) in the sense of Definition 2.2.

**Proof.** Let \( X_1, \ldots, X_6 \) and \( Y_1, \ldots, Y_6 \) be partitions of \( X \) and \( Y \) such that each \( N(X_i; r) \) and \( N(Y_i; r) \) is a connected component of \( \text{Core}(H_1; r) \), resp. \( \text{Core}(H_2; r) \). Because of Theorem 2.3 \( N(X_i; r) \sim_k N(Y_i; r) \) implies \( (H_1, X_i) \sim_{k,r} (H_2, Y_j) \) in the sense of Definition 2.1. The result follows now from the definition of \( H_1 \approx_{k,r} H_2 \).
Theorem 2.4. Let \( k \in \mathbb{N} \), and set \( r := \frac{(3^k - 1)}{2} \). Let \( H_1, H_2 \) be hypergraphs. Suppose that both \( H_1 \) and \( H_2 \) are \((k,r)\)-rich and \( H_1 \approx_{k,r} H_2 \). Then Duplicator wins 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 \( Y \subset V(H_2) \) defined as before. The hypothesis of \((k,r)\)-richness on both \( H_1, H_2 \) ensures that condition (2) in the statement of Theorem 2.1 holds. \( \square \)

3 Probabilistic results

3.1 Almost all hypergraphs are simple

**Lemma 3.1.** Let \( H \) be a hypergraph, and let \( X_n \) be the random variable equal to the number of copies of \( H \) in \( G_n \). Then \( \mathbb{E}[X_n] = \Theta(n^{-\text{ex}(H)}) \).

**Proof.** We have
\[
\mathbb{E}[X_n] = \sum_{H' \in \text{Copies}(H,[n])} \Pr(H' \subset G_n).
\]
We also have that \( |\text{Copies}(H,[n])| = \frac{(n)!}{\text{aut}(H)} \). Also, for any \( H' \in \text{Copies}(H,[n]) \) it holds that
\[
\Pr(H' \subset G_n) \sim n^{\text{ex}(H)} \prod_{R \in \sigma} \left( \frac{\beta_R}{n^{\text{ar}(R)-1}} \right)^{|E_R(H)|}.
\]
Substituting in the first equation we get
\[
\mathbb{E}[X_n] \sim \frac{(n)!}{\text{aut}(H)} \prod_{R \in \sigma} \left( \frac{\beta_R}{n^{\text{ar}(R)-1}} \right)^{|E_R(H)|} \sim n^{-\text{ex}(H)} \prod_{R \in \sigma} \frac{\beta_R^{E_R(H)}}{\text{aut}(H)}.
\]

**Lemma 3.2.** Let \( H \) be a hypergraph such that \( \text{ex}(H) > 0 \). Then a.a.s there are no copies of \( H \) in \( G_n \).

**Proof.** Because of the previous lemma \( \mathbb{E}[\# \text{ copies of } H \text{ in } G_n] \xrightarrow{n \to \infty} 0 \). An application of the first moment method yields the desired result. \( \square \)

**Lemma 3.3.** Let \( H \) be a hypergraph. Let \( \overline{v} \in (\mathbb{N})_n \) be a list of vertices with \( \text{len}(\overline{v}) \leq |V(H)| \). 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 \( \mathbb{E}[X_n] = \Theta(n^{-\text{ex}(H)-\text{len}(\overline{v})}) \).

**Proof.** The number of hypergraphs \( H' \in \text{Copies}(H,[n]) \) that contain all vertices in \( \overline{v} \) is asymptotically \( \sim n^{|V(H)|-\text{len}(\overline{v})} \) for some constant \( C \). Then,
\[
\mathbb{E}[X_n] \sim C n^{|V(H)|-\text{len}(\overline{v})} \prod_{R \in \tau} \left( \frac{\beta_R}{n^{\text{ar}(R)-1}} \right)^{e_R(n)} = n^{-\text{ex}(H)-\text{len}(\overline{v})} C \prod_{R \in \tau} (\beta_R)^{e_R(n)}.
\]

Given a 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 $H \subseteq G$ be a connected sub-hypergraph with $\text{ex}(H) < \text{ex}(G)$. Then there is a connected sub-hypergraph $H' \subseteq G$ satisfying $H \subseteq H'$, $\text{ex}(H) < \text{ex}(H')$ and that $|E(H')| \leq |E(H)| + 2r + 1$.

**Proof.** Suppose there is some edge $e \in E(G) \setminus E(H)$ with and $\text{ex}(e) \geq 0$. 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 $\text{ex}(e) = -1$. 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 a hypergraph $G'$ with $\text{ex}(G') < \text{ex}(G)$. Let $e$ be the edge that was cut last. Then $V(G') \cap V(e) = \text{ex}(G) - \text{ex}(G') + 1 \geq 2$. 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 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 $l := \max_{R \in \sigma} ar(R)$ the 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 $|H| \leq l(4r + 2)$.

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

**Proof.** Because of the 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,

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

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

As a corollary we obtain the needed result.

Theorem 3.2. Let $r \in \mathbb{N}$. Then a.a.s $G_n$ is $r$-simple.

**Proof.** If some connected component of $\text{Core}(G_n; r)$ is not a cycle then either $G_n$ contains a dense hypergraph of diameter at most $4r + 1$, or $G_n$ contains two cycles of diameter at most $2r + 1$ that are at 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 $6r + 3$. Hence 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 $v \in (\mathbb{N})$, and let $r \in \mathbb{N}$. Then a.a.s, for all vertices $v \in v$ the neighborhoods $N(v; r)$ are all trees and they are all disjoint.
Definition 3.2. Let \( \lambda \) and \( \mu \) be a.a.s the \( N(v; r) \) are disjoint trees. But if any of the \( N(v; r) \) was an unicycle then in \( G_n \) there would exist a path \( P \) of length at most \( 2r + 1 \) joining some vertex \( v \in v \) with a cycle \( C \) of diameter at most \( 2r + 1 \). Using Lemma 3.3 and the fact that there is a finite number of possible \( \simeq \) classes for \( P \cup C \), we obtain that a.a.s no such \( P \) and \( C \) exist. In consequence all the \( N(v; r) \) are disjoint trees as we wanted to prove. \( \square \)

Lemma 3.7. Let \( \bar{v} \subset N \) be a finite set of fixed vertices and let \( \pi(\bar{v}) \) be an edge sentence such that \( \text{len}(\bar{v}) = \text{len}(\bar{v}) \). Define \( G'_n = G_n \setminus E[\bar{v}] \) (i.e. \( G_n \) minus all the edges induced on \( \bar{v} \)). Fix \( r \in N \). Then a.a.s for all vertices \( v \in \bar{v} \) the neighborhoods \( N^{G'_n}(v; r) \) are disjoint trees.

Proof. Let \( A_n \) be the event that the \( N^{G'_n}(v; r) \) are disjoint trees. Notice that \( A_n \) does not concern the possible edges induced over \( \bar{v} \). Because edges are independent in our random model, we have that \( \Pr (A_n | \pi(\bar{v})) = \Pr (A_n) \). Now the result follows from Lemma 3.6 using that \( G'_n \subset G_n \).

\( \square \)

3.2 Probabilities of trees

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 \( b \in N \), and \( \bar{\lambda} \in \Lambda \), the expression \( (\beta_R/b) \prod_{\lambda \in \bar{\lambda}} \lambda \) belongs to \( M \), (3) for any \( \mu \in M \) and any \( n \in N \) both \( \text{Pois}_n \) and \( \text{Pois}_\mu(\geq n) \) are in \( \Lambda \), and (4) for any \( \lambda_1, \lambda_2 \in \Lambda \), the product \( \lambda_1 \lambda_2 \) belongs to \( \Lambda \) as well.

Definition 3.2. Let \( r \in N \) and let \( T \) be a \( \sim_r \) class of trees with radius at most \( r \). Let \( v \in N \) be an arbitrary vertex. We define \( \Pr[r, T] \) as the limit

\[
\lim_{n \to \infty} \Pr (Tr(G_n, v; v; r) \in T).
\]

Note that the definition of \( \Pr[r, T] \) does not depend on the choice of \( v \). The goal of this section is to show that \( \Pr[r, T] \) exists and is an expression with parameters \( \{\beta_R\}_{R \in \sigma} \) belonging to \( \Lambda \) for any choice of \( r \) and \( T \).

Theorem 3.3. Fix \( r \in N \). Let \( k \in N \) The following hold:

(1) Let \( T \) be a \( k \)-equivalence class of trees with radii at most \( r \). Then \( \Pr[r, T] \) exists, is positive for all choices of \( \{\beta_R\}_{R \in (0, \infty]} \), and is an expression in \( \Lambda \).

(2) Let \( \bar{v} \in (N)_\ast \), and let \( \pi(\bar{v}) \in FO[\sigma] \) be a consistent edge sentence such that \( \text{len}(\bar{v}) = \text{len}(\bar{v}) \). Let \( \bar{v} \in (N)_\ast \) be vertices contained in \( \bar{v} \). For each \( v \in \bar{v} \) let \( T_v \) be a \( k \)-equivalence class of trees with radii at most \( r \). Then

\[
\lim_{n \to \infty} \Pr \left( \bigwedge_{v \in \bar{v}} Tr (G_n, \bar{v}; v; r) \in T_v \mid \pi(\bar{v}) \right) = \prod_{v \in \bar{v}} \Pr[r, T_v].
\]

We devote the rest of this section to proving this theorem. The proof is by induction on \( r \). Recall that all trees with radius zero are \( k \)-equivalent. Thus, the limits appearing in conditions (1) and (2) are both equal to 1 in the case \( r = 0 \).
Lemma 3.8. Conditions (1) and (2) of Theorem 3.3 are satisfied for \( r = 0 \).

Definition 3.3. Let \( k \in \mathbb{N} \) and \( r > 0 \). Suppose that Theorem 3.3 holds for \( r - 1 \). Given a \((k, r)\)-pattern \( \epsilon \) we define the expressions \( \lambda_{r, \epsilon} \) and \( \mu_{r, \epsilon} \) as follows. Let \((e, \tau)\) be a representative of \( \epsilon \) whose root is \( v \). Then for all vertices \( u \in V(e) \) such that \( u \neq v \) it holds that \( \tau(u) \) is a \( \sim_k \) class of trees with radius at most \( r \) and we can set

\[
\lambda_{r, \epsilon} := \prod_{u \in V(e)} \Pr[r - 1, \tau(u)], \quad \text{and} \quad \mu_{r, \epsilon} = \frac{\beta_R(\epsilon)}{\aut(\epsilon)} \lambda_{r, \epsilon}.
\]

Clearly the definitions of \( \lambda_{r, \epsilon} \) and \( \mu_{r, \epsilon} \) are independent of the chosen representative. By hypothesis it holds that \( \mu_{r, \epsilon} \) is positive for all values of \( \{\beta_R\}_{R \in \sigma} \in (0, \infty)^{[\sigma]} \) and it is an expression belonging to \( M \).

Lemma 3.9. Let \( k \in \mathbb{N} \), \( r > 0 \) and \( \bar{\nu} \in (\mathbb{N})_* \). Let \( \pi(\bar{\nu}) \in FO[\sigma] \) be a consistent edge sentence such that \( \text{len}(\pi(\bar{\nu})) = \text{len}(\bar{\nu}) \). Let \( \pi \in (\mathbb{N})_* \) be vertices contained in \( \bar{\nu} \). For each \( v \in \pi \) set \( T_{n,v} := \text{Tr}(G_n, \bar{\nu}; v; r) \). Given a pattern \( \epsilon \in P(k, r) \) and \( v \in \pi \) we define the random variable \( X_{n,v,\epsilon} \) as the number of initial edges \( e \in E(T_{n,v}) \) such that \((e, \tau^k_{(T_{n,v}, v)}) \in \epsilon \). Suppose that Theorem 3.3 holds for \( r - 1 \). Then the conditional distributions of the variables \( X_{n,v,\epsilon} \) given \( \pi(\bar{\nu}) \) converge to independent Poisson distributions whose respective mean values are given by the \( \mu_{r, \epsilon} \).

Proof. To avoid excessively complex notation we prove only the case where \( \bar{\nu} \) consists of a single vertex \( v \). The general case is proven using the same arguments. Set \( T_n := T_{n,v} \) and \( X_{n,\epsilon} := X_{n,v,\epsilon} \) for all \( \epsilon \in P(k, r) \). By Theorem 1.1 in order to prove the result it is enough to show that for any choice of natural numbers \( \{b_{\epsilon}\}_{\epsilon \in P(k, r)} \) it holds that

\[
\lim_{n \to \infty} E_{\pi(\bar{\nu})} \left[ \prod_{\epsilon \in P(k, r)} \left( \frac{X_{n,\epsilon}}{b_{\epsilon}} \right) \right] = \prod_{\epsilon \in P(k, r)} \frac{(\mu_{r, \epsilon})^{b_{\epsilon}}}{b_{\epsilon}!}.
\]

Consider the numbers \( \{b_{\epsilon}\}_{\epsilon \in P(k, r)} \) fixed. For each \( n \in \mathbb{N} \) define

\[
\Omega_n := \left\{ \{E_{\epsilon}\}_{\epsilon \in P(k, r)} \mid \forall \epsilon \in P(k, r) \quad E_{\epsilon} \subset \text{Copies}(\epsilon, [n], (v, \rho)), \quad |E_{\epsilon}| = b_{\epsilon} \right\}.
\]

Informally, elements of \( \Omega_n \) represent choices of \( b_{\epsilon} \) possible initial edges of \( T_n \) whose \( k \)-pattern is \( \epsilon \) for all \((k, r)\)-patterns \( \epsilon \). Using Observation 1.1 we obtain

\[
E_{\pi(\bar{\nu})} \left[ \prod_{\epsilon \in P(k, r)} \left( \frac{X_{n,\epsilon}}{b_{\epsilon}} \right) \right] = \sum_{\{E_{\epsilon}\} \in \Omega_n} \Pr_{\pi(\bar{\nu})} \left( \bigwedge_{\epsilon \in P(k, r)} \left( \bigwedge_{(e, \tau) \in E_{\epsilon}} \left( e \in E(T_n) \bigwedge_{u \in V(e) \atop u \neq v} \text{Tr}(T_n, v; u) \in \tau(u) \right) \right) \right).
\]

We say that a choice \( \{E_{\epsilon}\} \in \Omega_n \) is disjoint if the edges \( (e, \tau) \in \bigcup_{\epsilon \in P(k, r)} E_{\epsilon} \) satisfy that no vertex \( w \in \bar{\nu} \) other than \( v \) belongs to any of those edges and each vertex \( w \in [n] \setminus \{v\} \) belongs to at most one of those edges. For each \( n \in \mathbb{N} \) let \( \Omega'_n \subset \Omega_n \) be the set of disjoint elements in \( \Omega_n \) and set \( \Omega'_n = \bigcup_{n \in \mathbb{N}} \Omega'_n \). If for some \( \{E_{\epsilon}\} \in \Omega_n \) we have that \( e \in E(T_n) \) for all \((e, \tau) \in \bigcup_{\epsilon \in P(k, r)} E_{\epsilon} \) then \( \{E_{\epsilon}\} \) is necessarily disjoint. This is because \( T_n \) is a tree and the only vertex in \( \bar{\nu} \) that belongs to \( T_n \) is \( v \) by definition. Thus, in the last sum it suffices to consider only the disjoint \( \{E_{\epsilon}\} \). Because of the symmetry of the random model the
probabilities in that sum are the same for all disjoint choices of \( \{E_e\}_e \). Hence, if we fix \( \{E_e\}_e \in \Omega_n^\epsilon \) we obtain

\[
E_{\pi(\overline{\pi})} \left[ \prod_{e \in P(k,r)} \left( \frac{X_n,e}{b_e} \right) \right] = |\Omega_n'| |\Omega_n| \Pr_{\pi(\overline{\pi})} \left( \bigwedge_{e \in P(k,r)} \left( e \in E(T_n) \bigwedge_{u \in V(e)} \bigwedge_{u \neq v} Tr(T_n, v; u) \in \tau(u) \right) \right).
\]

(2)

Set \( N := \sum_{e \in P(k,r)} (|e| - 1)b_e \). Counting vertices and automorphisms we get that

\[
|\Omega_n'| = (n - \text{len}(\overline{\pi})) N \prod_{e \in P(k,r)} \frac{1}{b_e!} \left( \frac{1}{\text{aut}(e)} \right)^{b_e}.
\]

(3)

Let \( \overline{w} \in (\mathbb{N})_* \) be a list containing exactly the vertices \( u \in V(e) \) for all \( e \in \bigcup_{e \in P(k,r)} E_e \).

Clearly, the event

\[
\bigwedge_{e \in P(k,r)} \left( e \in E(G_n) \right)
\]

can be described via an edge sentence whose variables are interpreted as vertices in \( \overline{w} \).

Let \( \psi(\overline{\pi}) \) be one of such edge sentences. This event is independent of \( \pi(\overline{\pi}) \) because edges are independent in \( G_n \).

Thus, a simple computation yields

\[
\Pr_{\pi(\overline{\pi})} \left( \bigwedge_{e \in P(k,r)} \left( e \in E(G_n) \right) \right) = \prod_{e \in P(k,r)} \left( \frac{\beta_{R(e)}}{n^{\text{ar}(R(e)) - 1}} \right)^{b_e} = \frac{1}{n^N} \prod_{e \in P(k,r)} \beta_{R(e)}^{b_e}.
\]

Because of Lemma 3.7 a.a.s if \( e \in E(G_n) \) and \( v \in V(e) \), then \( e \in E(T_n) \). Thus,

\[
\Pr_{\pi(\overline{\pi})} \left( \bigwedge_{e \in P(k,r)} \left( e \in E(T_n) \bigwedge_{u \in V(e)} \bigwedge_{u \neq v} Tr(T_n, u) \in \tau(u) \right) \right) \sim \left( \frac{1}{n^N} \prod_{e \in P(k,r)} \beta_{R(e)}^{b_e} \right) \Pr_{\pi(\overline{\pi}) \wedge \psi(\overline{\pi})} \left( \bigwedge_{e \in P(k,r)} \bigwedge_{u \in V(e)} \bigwedge_{u \neq v} Tr(T_n, u) \in \tau(u) \right).
\]

(4)

The trees \( Tr(T_n; u) \) in the last probability coincide with \( Tr(G_n, \overline{w}; \overline{w} u; r - 1) \) for all \( u \).

As a consequence, using the hypothesis that Theorem 3.3 holds for \( r - 1 \), we obtain

\[
\Pr_{\pi(\overline{\pi}) \wedge \psi(\overline{\pi})} \left( \bigwedge_{e \in P(k,r)} \bigwedge_{u \in V(e)} Tr(T_n, u) \in \tau(u) \right) \sim \prod_{e \in P(k,r)} (\lambda_{r,e})^{b_e}.
\]
Proof. Let \( G \) be a concatenation of all the \( \pi \) be as in the statement of (2). Using the previous lemma we obtain that the last expression equals the following expression:

\[
\left( \prod_{r \in E_1^k} \text{Poiss}_{\mu_r, r} (\geq k) \right) \left( \prod_{r \in E_2^k} \text{Poiss}_{\mu_r, r} (a_r) \right).
\]

Using the definition of the \( \mu_r, \epsilon \) we obtain that the last expression belongs to \( \Lambda \) as we wanted to prove. Furthermore, as the \( \mu_r, \epsilon \) are positive, this expression is also positive for all values of \( \{\beta_r\}_{r \in \epsilon} \in (0, \infty)^{\epsilon} \). Now we proceed to prove condition (2). Let \( \pi, \bar{\pi}, \{T_v\}_{v \in \bar{\pi}} \) and \( \pi(\bar{\pi}) \) be as in the statement of (2). Using the previous lemma we obtain that the events \( Tr(G_n, \bar{\pi}; v; \epsilon) \in \bar{T}_v \) for all \( v \in \bar{\pi} \) are asymptotically independent and are also independent of \( \pi(\bar{\pi}) \). Then the desired result follows from condition (1).

\[\square\]

3.3 Almost all graphs are \((k,r)\)-rich

Theorem 3.4. Let \( k, r \in \mathbb{N} \). 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 \( r \). Let \( m > k \). For each \( T \in \Sigma \) let \( \nu(T) \in [N]_m \) be tuples satisfying that all the \( \nu(T) \) are disjoint. Let \( \bar{w} \in (\mathbb{N}) \) be a concatenation of all the \( \nu(T) \). For each \( T \in \Sigma \) define \( X_{n, T} \) as the number of vertices \( v \in \nu(T) \) such that \( Tr(G_n, \bar{w}; v; r) \in T \). Because of Theorem 3.3 the \( \sim_k \) types of the trees \( Tr(G_n, \bar{w}; v; r) \) for all \( v \in \bar{w} \) are asymptotically independent and given any \( v \in \bar{w} \) and \( T \) it holds that \( \Pr(Tr(G_n, \bar{w}; v; r) \in T) \) tends to \( \Pr(r, T) \) as \( n \) goes to infinity. Hence, the variables \( X_{n, T} \) converge in distribution to independent binomial
variables whose respective parameters are \( m \) and \( \Pr[r, T] \). That is, given natural numbers \( 0 \leq l_T \leq m \) for all \( T \in \Sigma \),

\[
\lim_{n \to \infty} \Pr \left( \bigwedge_{T \in \Sigma} X_{n,T} = l_T \right) = \prod_{T \in \Sigma} \left( \frac{m}{l_T} \right)^{l_T} \Pr[r, T]^{|l_T|} (1 - \Pr[r, T])^{m-l_T}. 
\]

Fix \( \delta > 0 \) such that \( \delta < \Pr[r, T] \) for all \( T \in \Sigma \) and fix \( \epsilon > 0 \) arbitrarily small. Because of the Law of large numbers, if \( m \) is large enough

\[
\lim_{n \to \infty} \Pr \left( \left| X_{n,T} / m - \Pr[r, T] \right| \geq \delta \right) \leq \epsilon \quad \text{for all } T \in \Sigma. 
\]  

Also, for \( m \) large enough we have

\[
\Pr[r, T] > k / m + \delta \quad \text{for all } T \in \Sigma. 
\]  

Suppose that \( m \) is large enough for both Equations (5) and (6) to hold. Then

\[
\lim_{n \to \infty} \Pr \left( X_{n,T} < k \right) \leq \epsilon \quad \text{for all } T \in \Sigma
\]

We define \( A_n \) as the event that for any \( v \in \overline{w} \) we have \( N(v; r) \cap \text{Core}(G_n; r) = \emptyset \) (in particular this implies that \( N(v; r) \) is a tree), and for any two \( v_1, v_2 \in \overline{w} \) it is satisfied that \( d^\sigma_G(v_1, v_2) \geq 2r+1 \). If \( A_n \) holds then for all \( v \in \overline{w} \) we have that \( N(v; r) = Tr(G_n, \overline{w}; v; r) \) and the \( N(v; r) \) are disjoint trees. Thus, if both \( A_n \) holds and \( X_{n,T} \geq k \) for all \( T \) then \( G_n \) is \((k, r)\)-rich. Because of Lemma 3.6 a.a.s \( A_n \) holds, and we obtain

\[
\lim_{n \to \infty} \Pr \left( G_n \text{ is not } (k, r)\text{-rich} \right) \leq \lim_{n \to \infty} \Pr \left( A_n \land \left( \bigvee X_{n,T} < k \right) \right) = \lim_{n \to \infty} \Pr \left( \bigvee X_{n,T} < k \right) \leq \epsilon |\Sigma|. 
\]

As \( \epsilon \) can be arbitrarily small given a suitable choice of \( m \) we obtain that necessarily a.a.s \( G_n \) is \((k, r)\)-rich, as was to be proved.

3.4 Probabilities of cycles

**Definition 3.4.** We define \( \Gamma \) and \( \Upsilon \) as the minimal families of expressions with arguments \( \{ \beta_R \}_{R \in \sigma} \) that satisfy the following conditions: (1) given natural numbers \( a_R \) for each \( R \in \sigma \), a positive number \( b \in \mathbb{N} \) and a \( \lambda \in \Lambda \), the expression \( \lambda^b \prod_{R \in \sigma} \beta_R^{a_R} \) belongs to \( \Gamma \), (2) given a \( \gamma \in \Gamma \) and a \( a \in \mathbb{N} \), the expressions \( \text{Poiss}_\gamma(a) \) and \( \text{Poiss}_\gamma(\geq a) \) both belong to \( \Upsilon \), and (3) if \( v_1, v_2 \in \Upsilon \) then \( v_1 v_2 \in \Upsilon \) as well.

**Definition 3.5.** Let \( k, r \in \mathbb{N} \) and \( O \in C(k, r) \). Let \((H, \tau)\) be a representative of \( O \). We define \( \lambda_{r,O} \) and \( \gamma_{r,O} \) in the following way:

\[
\lambda_{r,O} := \prod_{v \in V(H)} \Pr[r, \tau(v)], \quad \text{and} \quad \gamma_{r,O} := \prod_{R \in \sigma} \frac{\beta_R^{\text{aut}(H)}}{\text{aut}(H, \tau)} \lambda_{r,O}. 
\]

Clearly the definitions of \( \lambda_{r,O} \) and \( \gamma_{r,O} \) are independent of the chosen representative and the expression \( \gamma_{r,O} \) belongs to \( \Gamma \).
Lemma 3.11. Let \( k, r \in \mathbb{N} \). For any \( O \in C(k, r) \) let \( X_{n,O} \) be the random variable equal to the number of connected components \( H \) of \( \text{Core}(G_n; \tau'_n) \) such that \( H' := \text{Center}(H) \) satisfies that \((H', \tau'_n) \in O. \) Then the \( X_{n,O} \) converge in distribution to independent Poisson variables whose respective expected values are given by the \( \gamma_{r,O}. \)

Proof. The proof is similar to the one of Lemma 3.9. By Theorem 1.1, to prove the result is enough to show that for any natural numbers \( \{b_O\}_{O \in C(k, r)} \) it holds

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

(7)

For each \( n \in \mathbb{N} \) we define

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

Given a cycle \( H \) such that \( V(H) \subseteq [n] \) we say that \( H \subset G_n \) if \( H = \text{Center}(H') \) for some connected component \( H' \) of \( \text{Core}(G_n; r) \). Using observation Observation 1.1 we obtain

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

We call a choice \( \{F_O\}_{O \in \Omega_n} \) disjoint if no vertex \( v \in [n] \) belongs to two cycles \((H, \tau) \in \cup O F_O. \) Define \( \Omega'_n \) as the set of disjoint elements in \( \Omega_n \) and set \( \Omega'_n := \cup_{n \in \mathbb{N}} \Omega'_n. \) If for some \( \{F_O\}_{O \in \Omega_n} \) it holds that \( H \subset G_n \) for all \((H, \tau) \in \cup O F_O \) then necessarily \( \{F_O\}_{O \in \Omega_n} \) is disjoint. Indeed, suppose the opposite. Then for some \((H_1, \tau_1), (H_2, \tau_2) \in \cup O F_O \) it holds that \( V(H_1) \cap V(H_2) \neq \emptyset. \) Then both \( H_1 \) and \( H_2 \) belong to the same connected component \( H \) of \( \text{Core}(G_n; r) \) and thus \( H_1 \cup H_2 \subset \text{Center}(H). \) As a consequence neither \( H_1 \subset G_n \) or \( H_2 \subset G_n \) hold. \((H_1, \tau_1), (H_2, \tau_2) \in \cup O \in C(k, r) \) \( F_O. \) Hence in the last sum it suffices to consider disjoint choices \( \{F_O\}_{O \in \Omega_n}. \) Because of the symmetry of the random model the probability in that sum is the same for all disjoint choices of \( \{F_O\}_{O \in \Omega_n}. \) In consequence, if we fix \( \{F_O\}_{O \in \Omega_n} \) we obtain

\[
E \left[ \prod_{O \in C(k, r)} \left( \frac{X_{n,O}}{b_O} \right) \right] = |\Omega'_n| \Pr \left( \bigwedge_{O \in C(k, r)} \left( H \subset G_n \bigwedge_{v \in V(H)} Tr(G_n, v; r) \in \tau(v) \right) \right).
\]

(8)

Set \( N := \sum_{O \in C(k, r)} |O|b_O. \) We have that

\[
|\Omega'_n| = \frac{(n)_N}{\prod_{O \in C(k, r)} b_O \! \varnothing (O)^{b_O}}.
\]

(9)

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

\[
\bigwedge_{O \in C(k, r)} \left( H \subset G_n \right)
\]

...
Because of Theorem 3.2, a.a.s if some cycle \( H \) of diameter at most \( 2r+1 \) satisfies \( H \subset G_n \), then \( H \subset G_n \). Hence,

\[
\Pr \left( \bigwedge_{(O \in C(k,r) \setminus (H,\tau) \in F_O)} H \subset G_n \wedge \bigwedge_{v \in V(H)} Tr(G_n, v; r) \in \tau(v) \right) \sim \frac{1}{n^N} \prod_{O \in C(k,r)} \left( \prod_{R \in \sigma} \beta_{R}^{[E_R(O)]} \right)^{b_{\sigma}} \Pr_{\phi(\overline{\tau})} \left( \bigwedge_{(O \in C(k,r) \setminus (H,\tau) \in F_O)} \bigwedge_{v \in V(H)} Tr(G_n, v; r) \in \tau(v) \right).
\]

As all the vertices \( v \in \overline{\tau} \) belong to \( \text{Core}(G_n; r) \), the trees \( Tr(G_n; v; r) \) in the last probability coincide with \( Tr(G_n, \overline{\tau}; v; r) \). By Theorem 3.3 we have that

\[
\Pr_{\phi(\overline{\tau})} \left( \bigwedge_{(O \in C(k,r) \setminus (H,\tau) \in F_O)} \bigwedge_{v \in V(H)} Tr(G_n, v; r) \in \tau(v) \right) \sim \prod_{O \in C(k,r)} (\lambda_{r,O})^{b_{\sigma}}.
\]

Combining this with Equations (8) to (10) we obtain

\[
E \left[ \prod_{O \in C(k,r)} \left( X_{n,O} \right)^{b_{\sigma}} \right] \sim \left( \frac{n^N}{n^N} \right) \prod_{O \in C(k,r)} \frac{1}{b_{\sigma}!} \left( \frac{\lambda_{r,O} \prod_{R \in \sigma} \beta_{R}^{[E_R(O)]}}{\text{aut}(O)} \right) \sim \prod_{O \in C(k,r)} \frac{(\gamma_{r,O})^{b_{\sigma}}}{b_{\sigma}!}.
\]

This proves Equation (7) and the statement. \( \square \)

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

**Proof.** For each \( O \in C(k,r) \) let \( X_{n,O} \) be as in the previous lemma. Let \( U_O^1, U_O^2 \) and \( \{a_O\}_{O \in U_O^2} \) be as in Observation 1.3. Let \( A_n \) be the event that \( G_n \) is \( r \)-simple. Then

\[
\lim_{n \to \infty} \Pr (G_n \in O) = \lim_{n \to \infty} \Pr \left( A_n \wedge \left( \bigwedge_{O \in U_O^1} X_{n,O} \geq k \right) \wedge \left( \bigwedge_{O \in U_O^2} X_{n,O} = a_O \right) \right).
\]

Because of Theorem 3.2 a.a.s \( A_n \) holds. Thus, using the last lemma the previous limit equals the following expression

\[
\left( \prod_{O \in C_1} \text{Poiss}_{\gamma_{r,O}}(\geq k) \right) \left( \prod_{O \in C_2} \text{Poiss}_{\gamma_{r,O}}(a_O) \right).
\]

As all the \( \gamma_{r,O} \) belong to \( \Gamma \), this last expression belongs to \( \Upsilon \) and 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
\[
\{ \beta_R \}_{R \in \sigma} \mapsto \lim_{n \to \infty} \Pr (G_n (\{ \beta_R \}_R) \models \phi)
\]
is well defined and it is given by a finite sum of expressions in \( \Upsilon \).

**Proof.** Let \( k \) be the quantifier rank of \( \phi \) and let \( r = 3^k \). Let \( G_n := G_n (\{ \beta_R \}_{R \in \sigma}) \) and let \( \Sigma \) be the set of \((k, r)\)-agreeability classes of \( r \)-simple hypergraphs. Because of Theorem 3.2 a.a.s \( G_n \) is \((k, r)\)-rich. This together with Theorem 2.4 implies that for any \( O \in \Sigma \)
\[
\lim_{n \to \infty} \Pr (G_n \models \phi \mid G_n \in O) = 0 \text{ or } 1.
\]
Let \( \Sigma' \subset \Sigma \) be the set of classes \( O \) for which last limit equals 1. Then
\[
\lim_{n \to \infty} \Pr (G_n \models \phi \mid G_n \in O) = \sum_{O \in \Sigma'} \lim_{n \to \infty} \Pr (G_n \in O).
\]
Because of Theorem 3.3 we know that each of the limits inside the last sum exists and is given by an expression that belongs to \( \Upsilon \). As a consequence the theorem follows. \( \square \)

5 Application to random SAT

We define a binomial model of random CNF formulas, in analogy with the one in [3], but the generality in Theorem 1.3 allows for many variants.

**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 **non-tautological** 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 \( l \in \mathbb{N} \) a **l-CNF formula** is a set of non-tautological clauses that contain exactly \( l \) literals. We say that a formula \( F \) on the variables \( x_1, \ldots, x_n \) is **satisfiable** if there is an assignment \( f : \{x_1, \ldots, x_n\} \to \{0, 1\} \) that satisfies all clauses in \( F \).

Given \( n, l \in \mathbb{N} \) and a real number \( 0 \leq p \leq 1 \) we define the random model \( F(l, n, p) \) as the discrete probability space that assigns to each \( l \)-CNF formula \( F \) on the variables \( \{x_i\}_{i \in [n]} \) the probability
\[
\Pr (F) = p^{|F|}(1-p)^{2^l(n)^{|F|}},
\]
where \( |F| \) is the number of clauses in \( F \). Equivalently, a random formula in \( F(l, n, p) \) is obtained by choosing each of the \( 2^l(n)^{|F|} \) non-tautological clauses of size \( l \) on the variables
\{x_i\}_i with probability \(p\) independently. When \(p\) is a function of \(n\) satisfying \(p(n) \sim \frac{\beta}{n^{l-1}}\) we denote by \(F^n_l(\beta)\) a random sample of \(F(l, n, p(n))\).

We consider \(l\)-CNF formulas, as defined above, as relational structures with a language \(\sigma\) consisting of \(l + 1\) relation symbols \(R_0, \ldots, R_l\) of arity \(l\). We do that in such a way that the expression \(R_j(x_1, \ldots, x_n)\) means that our formula contains the clause consisting of \(\neg x_{i_1}, \ldots, \neg x_{i_l}\) and \(x_{i_l+1}, \ldots, x_{i_n}\). The relations \(R_1, \ldots, R_l\) satisfy the following axioms: (1) given \(0 \leq j \leq l\) and variables \(y_1, \ldots, y_l\) the fact that \(R_j(y_1, \ldots, y_l)\) holds is invariant under any permutation of the variables \(y_1, \ldots, y_l\) or \(y_{j+1}, \ldots, y_l\), and (2) for any \(0 \leq j \leq l\) and any variables \(y_1, \ldots, y_l\) it holds that \(R_j(y_1, \ldots, y_l)\) only if all the \(y_i\) are different. Call \(C\) to the family of \(\sigma\)-structures satisfying the last two axioms. The language \(\sigma\) and the family \(C\) satisfy the conditions in Section 1.4. The random model \(F_l(n, p)\) coincides with the model \(G(n, \{p_R\}_R)\) of random \(C\)-hypergraphs described in Section 1.6 when all the \(p_R\) are equal. As a particular case of Theorem 1.3 we obtain the following result.

**Theorem 5.1.** Let \(l > 1\) be a natural number. Then for each sentence \(\Phi \in FO[\sigma]\) it is satisfied that the map \(f_\Phi : (0, \infty) \rightarrow \mathbb{R}\) given by

\[
\beta \mapsto \lim_{n \to \infty} \Pr \left( F^n_l(\beta) \models \Phi \right)
\]

is well defined and analytic.

The following is a well known result regarding random CNF formulas.

**Theorem 5.2.** Let \(l \geq 2\) be a natural number, and let \(c \in (0, \infty)\) be an arbitrary real number. Let \(m : \mathbb{N} \rightarrow \mathbb{N}\) be such that \(m(n) \sim cn\). For each \(n\) let \(C_{n,1}, \ldots, C_{n,m(n)}\) be clauses chosen uniformly at random independently among the \(2^l(n)\) non-tautological clauses of size \(l\) over the variables \(x_1, \ldots, x_n\). For each \(n\), let \(UNSAT_n\) denote the event that there is no assignment of the variables \(x_1, \ldots, x_n\) that satisfies all clauses \(C_{n,1}, \ldots, C_{n,m(n)}\). Then there are two real constants \(0 < c_1 < c_2\), such that a.a.s \(UNSAT_n\) does not hold if \(c < c_1\), and a.a.s \(UNSAT_n\) holds if \(c > c_2\).

The existence of \(c_1\) is proven in [3, Theorem 1]. The fact that \(c_2\) exists follows from a direct application of the first moment method and is also shown for instance in [3, 8, 4]. We want to show that an analogous “phase transition” also happens in \(F(l, n, p)\) when \(p \sim \frac{\beta}{n^{l-1}}\). 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 \(m : \mathbb{N} \rightarrow \mathbb{N}\) satisfy \(m(n) \sim cn\). For each \(n\) let \(F_{n,m(n)}\) be a random formula chosen uniformly at random among all sets of \(m(n)\) non-tautological clauses of size \(l\) over the variables \(x_1, \ldots, x_n\). Then there are two real positive constants \(0 < c_1 < c_2\) such that a.a.s \(F_{n,m(n)}\) is satisfiable if \(c < c_1\), and a.a.s \(F_{n,m(n)}\) is unsatisfiable if \(c > c_2\).

**Proof.** For each \(n\) let \(C_{n,1}, \ldots, C_{n,m(n)}\) and \(UNSAT_n\) be as in the previous theorem. One can consider \(F_{n,m(n)}\) to be the result of selecting clauses \(C_{n,1}, \ldots, C_{n,m(n)}\) uniformly at random independently among all possible clauses, given the fact that no two clauses \(C_{n,i}, C_{n,j}\) are equal. Hence,

\[
\Pr \left( F_{n,m(n)} \text{ is unsatisfiable } \right) = \Pr \left( UNSAT_n \mid \text{all the } C_{n,j} \text{ are different} \right).
\]

An application of the first moment method yields that for \(l \geq 3\) a.a.s the number of unordered pairs \(\{i, j\}\) such that \(C_{n,i} = C_{n,j}\) is equal to zero. In the case of \(l = 2\),
an application of Theorem 5.1 proves that the number of such pairs \( \{i, j\} \) converges in distribution to a Poisson variable. In either case all the \( C_{n,i} \) are different with positive asymptotic probability. Thus the constants \( c_1 \) and \( c_2 \) from the previous theorem satisfy our statement.

Let \( F_{n,m(n)} \) be as in last result. Note that because of 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 \( F(l, n, p(n)) \) given that the number of clauses is \( m(n) \). Using this observation we can prove the following.

**Theorem 5.3.** Let \( l > 1 \). Then there are real positive values \( \beta_1 < \beta_2 \) such that a.a.s \( F_n^1(\beta) \) is satisfiable for \( 0 < \beta < \beta_1 \) and a.a.s \( F_n^1(\beta) \) is unsatisfiable and for \( \beta > \beta_2 \).

**Proof.** For each \( n \in \mathbb{N} \) let \( X_n(\beta) \) be the random variable equal to the number of clauses in \( F_n^1(\beta) \). We have that \( E[X_n(\beta)] \sim \frac{\beta^2 l^2}{2} n \). Let \( c_1, c_2 \) be as in last corollary. Define \( \beta_1 := \frac{c_1 l^2}{2} \) and \( \beta_2 := \frac{c_2 l^2}{2} \). Fix \( \beta \in \mathbb{R} \) satisfying \( 0 < \beta < \beta_1 \). Let \( \epsilon > 0 \) be a real number such that \( \frac{\beta^2 l^2}{2} + \epsilon < c_1 \). For each \( n \in \mathbb{N} \) set \( \delta_1(n) := \left[ \left( \frac{\beta^2 l^2}{2} - \epsilon \right) n \right] \) and \( \delta_2(n) := \left[ \left( \frac{\beta^2 l^2}{2} + \epsilon \right) n \right] \).

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

\[
\Pr \left( F_n^1(\beta) \text{ is unsatisfiable } \right) \sim \int_{\delta_1(n)}^{\delta_2(n)} \Pr \left( F_n^1(\beta) \text{ is unsatisfiable } \mid X_n(\beta) = m \right) dp_n(m).
\]

Note that the property of being unsatisfiable is monotonous. As a consequence,

\[
\int_{\delta_1(n)}^{\delta_2(n)} \Pr \left( F_n^1(\beta) \text{ is unsatisfiable } \mid X_n(\beta) = m \right) dp_n(m) \leq \Pr \left( F_n^1(\beta) \text{ is unsatisfiable } \mid X_n(\beta) = \delta_2(n) \right) \Pr(\delta_1(n) \leq X_n(\beta) \leq \delta_2(n)).
\]

Because of the Law of large numbers,

\[
\lim_{n \to \infty} \Pr(\delta_1(n) \leq X_n(\beta) \leq \delta_2(n)) = 1.
\]

As \( \delta_2(n) < c_2 n \), because of the previous corollary

\[
\lim_{n \to \infty} \Pr \left( F_n^1(\beta) \text{ is unsatisfiable } \mid X_n(\beta) = \delta_2(n) \right) = 0.
\]

Combining the previous equations we obtain that for any \( \beta < \beta_1 \) it holds that \( F_n^1(\beta) \) a.a.s is satisfiable, as it was to be proven. Showing that for any \( \beta > \beta_2 \), a.a.s \( F_n^1(\beta) \) is unsatisfiable is analogous.

A direct consequence of the last theorem, due to A. Atserias (personal communication, July, 2019), is the following

**Theorem 5.4.** Let \( l > 1 \) be a natural number. Let \( \Phi \in FO[\sigma] \) be a first order sentence that implies unsatisfiability. Then for all \( \beta > 0 \) a.a.s \( F_n^1(\beta) \) does not satisfy \( \Phi \).

**Proof.** Let \( \beta_1 \) and \( \beta_2 \) be as in Theorem 5.3. As \( \Phi \) implies unsatisfiability \( \Pr \left( F_n^1(\beta) \models \Phi \right) \leq \Pr \left( F_n^1(\beta) \text{ is unsatisfiable } \right) \). Thus, by Theorem 5.3 we get that for all \( \beta \in (0, \beta_1] \)

\[
\lim_{n \to \infty} \Pr \left( F_n^1(\beta) \models \Phi \right) = 0.
\]

By Theorem 5.1 last limit varies analytically with \( \beta \). It vanishes in the proper interval \((0, \beta_1] \) then by the Principle of analytic continuation it has to vanish in the whole \((0, \infty) \), and the result holds.
Acknowledgments

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

I would like to thank both my supervisor Marc Noy and Albert Atserias for suggesting the topic of this research. I am grateful to M. Noy for introducing me to the topic of zero-one laws and for helpful discussions on the subject. I am also thankful to A. Atserias for his insight on random SAT problems and for suggesting the proof of Theorem 5.1. The feedback given by both of them has been very helpful.

References

[1] A. Atserias. Definability on a random 3-cnf formula. In 20th Annual IEEE Symposium on Logic in Computer Science (LICS'05), pages 458–466. IEEE, 2005.

[2] B. Bollobás and B. Béla. Random graphs. Number 73. Cambridge university press, 2001.

[3] V. Chvátal and B. 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.

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

[5] H.-D. Ebbinghaus and J. Flum. Finite model theory. Springer Science & Business Media, 2005.

[6] A. Ehrenfeucht. An application of games to the completeness problem for formalized theories. Fund. Math, 49(129-141):13, 1961.

[7] R. Fagin. Probabilities on finite models 1. The Journal of Symbolic Logic, 41(1):50–58, 1976.

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

[9] Y. V. Glebskii, D. I. Kogan, M. Liogon’kil, and V. Talanov. Range and degree of realizability of formulas in the restricted predicate calculus. Cybernetics and Systems Analysis, 5(2):142–154, 1969.

[10] J. F. Lynch. Probabilities of sentences about very sparse random graphs. Random Structures & Algorithms, 3(1):33–53, 1992.

[11] N. C. Saldanha and M. Telles. Spaces of completions of elementary theories and convergence laws for random hypergraphs. arXiv preprint arXiv:1602.06537, 2016.

[12] S. Shelah and J. Spencer. Zero-one laws for sparse random graphs. Journal of the American Mathematical Society, 1(1):97–115, 1988.
[13] S. Shelah and J. Spencer. Can you feel the double jump? *Random Structures & Algorithms*, 5(1):191–204, 1994.

[14] J. Spencer. *The strange logic of random graphs*, volume 22. Springer Science & Business Media, 2013.