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, Dept. Ciències de la Computació, Jordi Girona 29 08034 Barcelona, Catalonia, Spain.*
E-mail: 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 σ and a first-order theory T for σ composed of symmetry and anti-reflexivity axioms. We define a binomial random model of finite σ-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 σ 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 σ. 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

1 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 ϕ, the function $F_\phi : (0,\infty) \to [0,1]$ given by

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

Vol. 31, No. 2, © The Author(s) 2020. Published by Oxford University Press.

Advance Access Publication on 27 October 2020 doi:10.1093/logcom/exaa061

is well defined and is given by an expression with parameter β 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 β. 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 β. 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 FO logic of certificates for unsatisfiability that hold for typical unsatisfiable formulas. A conjunctive normal form formula, shortened to CNF formula, is a Boolean formula consisting in a conjunction of one or more clauses, where a clause is a disjunction of one or more literals. 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 α). 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 on 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 α is irrational, the study of the range $m = \Theta(n)$ (i.e. $m = \Theta(n^{2-\alpha})$ 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 $F^l_n(\beta)$ 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 $F^l_n(\beta)$ for any $\beta > 0$. Further suppose that these asymptotic probabilities vary analytically with β. Then any FO property that implies unsatisfiability a.a.s. does not hold in $F^l_n(\beta)$ for $\beta > 0$. Indeed, let P be one such FO property. One can find a value $\beta_0 > 0$ satisfying that a.a.s. $F^l_n(\beta)$ is satisfiable when $0 < \beta < \beta_0$. As a consequence P a.a.s. does not hold in $F^l_n(\beta)$ when $0 < \beta < \beta_0$. Since the asymptotic probability of P varies analytically with β 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 $F^l_n(\beta)$ for all $\beta > 0$.

2 Preliminaries

2.1 General notation

Given a positive natural number n, we write $[n]$ to denote the set $1, 2, \dots, n$. 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 $\overline{x}$ define $\mathrm{len}(\overline{x})$ 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 $\overline{x}$. Given a map $f : X \to Y$ and an ordered tuple $\overline{x} := (x_1, \dots, x_a) \in X^*$, we define $f(\overline{x}) \in Y^*$ as the tuple $(f(x_1), \dots, f(x_a))$. Given two tuples $\overline{x}, \overline{y}$, we write $\overline{x}^\frown\overline{y}$ to denote their concatenation. Given a set S and elements x_s for each $s \in S$, we write $\{x_s\}_{s \in S}$, or just $\{x_s\}_s$ when S is understood, to denote the tuple indexed by S which contains the element x_s at the position given by s.

Let S be a set, a a positive natural number and Φ a group of permutations over $[a]$. Then Φ acts naturally on S^a in the following way: given $g \in \Phi$ and $\overline{x} := (x_1, \dots, x_a) \in S^a$, let $g\overline{x} := (x_{g(1)}, \dots, x_{g(a)})$. We denote by S^a/Φ the quotient of S^a by this action. Given $\overline{x} := (x_1, \dots, x_a) \in S^a$, we denote its equivalence class in S^a/Φ by $[x_1, \dots, x_a]$ or $[\overline{x}]$. Thus, for $g \in \Phi$, by definition $[x_1, \dots, x_a] = [x_{g(1)}, \dots, x_{g(a)}]$.

The notations $\overline{x}$ and $(x_1, \dots, x_a)$ represent ordered tuples while $[\overline{x}]$ and $[x_1, \dots, x_a]$ denote ordered tuples modulo the action of some arbitrary group of permutations. Which group it is will depend on the ambient set where $[x_1, \dots, x_a]$ 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 $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$.

2.2 Probabilistic preliminaries

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

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 $\mathrm{E}[X_n]$ 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 $\mathrm{E}_B[X]$ 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 2.1
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 $r_1, \ldots, r_l \in \mathbb{N}$

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

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 2.2
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) $X_i = \sum_{j=1}^{a_i} Y_{i,j}$. Define $\binom{\Omega := \prod_{i=1}^{l}}{[a_i] r_i}$. 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

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

2.3 Logical preliminaries

We assume familiarity with FO logic. We follow the convention that FO logic contains the equality symbol. Given a vocabulary σ, we denote by $FO[\sigma]$ the set of FO formulas of vocabulary σ. 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 that contains all free variables in ϕ and none of its bounded variables, but it may contain variables that do not appear in ϕ. Formulas with no free variables are called **sentences** and formulas with no quantifiers are called **open formulas**. The **quantifier rank** of a formula ϕ, written as $\mathrm{qr}(\phi)$, is the maximum number of nested quantifiers in ϕ. We call an **edge formula** any consistent open formula that contains no occurrence of the equality symbol '='.

2.4 Structures as multi-hypergraphs

For the rest of the article consider fixed:

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

We define $\mathcal{C}$ as the class of σ-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 $\{i,j\} \in P_R$

$$\forall \bar{x} \left((x_i = x_j) \implies \neg R(\bar{x})\right).$$

Families of structures defined in the same fashion as $\mathcal{C}$ generalize the notion of uniform hypergraphs (with edges of size at least 2) in the following sense: those hypergraphs can be regarded as relational structures whose only relation, the one given by their edges, is completely symmetric and anti-reflexive, while structures in $\mathcal{C}$ contain multiple relations with arbitrary symmetry and anti-reflexivity restrictions.

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 2.3
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

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

In other words, the set $E_R[V]$ is the set of all the '$ar(R)$-tuples of elements in V modulo the permutations in ϕ_R' excluding those that contain some repetition of elements in the positions given by P_R.

We call elements of $E_R[V]$ **edges**. In the case where $V = [n]$, we write simply $E_R[n]$ instead of $E_R[[n]]$. To avoid the ambiguity that arises when an edge e belongs to two sets $E_{R_1}[V], E_{R_2}[V]$, where R_1, R_2 are relations of the same arity, we introduce the convention that each edge has a unique sort and the sort of an edge $e \in E_R[V]$ is R. This way, each edge has its own identity, determined by their underlying tuple of vertices and their sort. For example, in the previous situation instead of having an edge $e \in E_{R_1}[V] \cap E_{R_2}[V]$ we have two different ones $e_1 \in E_{R_1}[V], e_2 \in E_{R_2}[V]$ (because their sorts, R_1 and R_2, are different) even if they have the same underlying tuples of vertices and the same symmetries.

Let G be a hypergraph, $V := V(G)$, 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 $[\bar{v}] \in E_R[V]$ such that $\bar{v} \in R^G$. We define **the total edge set of** G as the set $E(G) := \cup_{R\in\sigma} E_R(G)$. Given an edge, $e \in E(G)$ we denote by $V(e)$ the set of all vertices that participate in e.

Clearly, a hypergraph G is completely given by its vertex set $V(G)$ and its edge set $E(G)$. Notice that edges $e \in E(G)$ are sorted according to the relation they represent. The **size** of G, written as $|G|$, is its number of vertices.

Given two hypergraphs H and G, we say that H is a **sub-hypergraph** of G, written 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 denote by $G[U]$ the **hypergraph induced by** G **on** U. That is, $G[U]$ is a hypergraph $H = (V(H), \{E(H)_R\}_{R\in\sigma})$ such that $V(H) = U$ and for any $R \in \sigma$ an edge $e \in E_R(G)$ belongs to $E_R(H)$ if and only if $V(e) \subset U$.

We define the **excess** $\mathrm{ex}(G)$ of a hypergraph G as the number

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

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

A 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 $\mathrm{ex}(G) \geq -1$ for any connected hypergraph.

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

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

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 $d^G(u, v) = \infty$. 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 $\mathrm{ex}(G) = -1$ and **dense** if $\mathrm{ex}(G) > 0$. A hypergraph is called r**-sparse** if it does not contain any dense sub-hypergraph H such that $diam(H) \leq r$. A connected hypergraph G with $\mathrm{ex}(G) \geq 0$ is called **saturated** if for any non-empty proper sub-hypergraph $H \subset G$, it holds $\mathrm{ex}(H) < \mathrm{ex}(G)$. A connected hypergraph G with $\mathrm{ex}(G) = 0$ 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**. We 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. We define the radius of a rooted tree as the maximum distance between its root and any other vertex.

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

Isomorphisms between hypergraphs are defined as isomorphisms between relational structures. Isomorphisms between Σ-hypergraphs are just isomorphisms between the underlying hypergraphs that also preserve their colorings. In both cases, we denote the isomorphism relation by $\simeq$. Given a hypergraph H, resp. a Σ-hypergraph (H, χ), an **automorphism** of H, resp. (H, χ), is an isomorphism from H, resp. (H, χ), to itself. We denote by $\mathrm{aut}(H)$, resp. $\mathrm{aut}(H, \chi)$, 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 $Copies(H, V)$, as the set of hypergraphs H' such that $V(H') \subset V$ and $H \simeq H'$. Let χ be a Σ-coloring of H. Analogously, we define the set $Copies\,((H, \chi),\ V)$ as the set of Σ-hypergraphs (H', χ') satisfying $V(H') \subset V$ and $(H, \chi) \simeq (H', \chi')$. Let $\mathbb{H}$ be an isomorphism class of Σ-hypergraphs. Then the set $Copies(\mathbb{H}, V)$ is defined as the set of Σ-hypergraphs (H', χ') 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* $(\mathbb{H}, V;\ (v, s))$ as the set of Σ-hypergraphs $(H', \chi') \in Copies(\mathbb{H}, V)$ that satisfy $v \in V(H')$ as well as $\chi'(v) = s$.

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

2.5 Ehrenfeucht–Fraisse games

We assume familiarity with Ehrenfeucht–Fraisse (EF) games. An introduction to the subject can be found for instance in [5, 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 2.4 (Ehrenfeucht, 6).
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 $\text{qr}(\phi) \le k$.

Given lists $\overline{v} \in V(H_1)^*$, and $\overline{u} \in V(H_2)^*$ of the same length, we denote the k round EF 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 EF 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 $i, j \in [k]$, $d^{H_1}(v_i, v_j) = d^{H_2}(u_i, u_j)$, 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 $d\text{EHR}_k(H_1, \overline{v}; H_2, \overline{u})$ analogously to $\text{EHR}_k(H_1, \overline{v}; H_2, \overline{u})$.

2.6 The random model

For each $R \in \sigma$, let p_R be a real number between zero and one. The random model $G^{\mathcal{C}}\left(n, \{p_R\}_{R \in \sigma}\right)$ 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 for each $R \in \sigma$.

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

2.7 Main definitions

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

DEFINITION 2.5
Let H be a connected hypergraph. Then H contains a unique maximal saturated sub-hypergraph H' satisfying $\text{ex}(H') = \text{ex}(H)$ if $\text{ex}(H) \ge 0$, and $H' = \emptyset$ otherwise. Given $\overline{v} \in V(H)^*$, we define *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 $Center(H, \overline{v})$ as the union of $Center(H'', \overline{u})$ for all connected components $H'' \subset H$, where $\overline{u} \in V(H)^*$ contains exactly the vertices in $\overline{v}$ belonging to $V(H'')$. When $\overline{v}$ is empty, we simply write $Center(H)$.

DEFINITION 2.6
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 $Core(H, \overline{v}; r)$ as $N(X; r)$. If $\overline{v}$ is empty, we write $Core(H; r)$. We say that H is r**-simple** if all connected components of $Core(H; r)$ are unicycles.

DEFINITION 2.7
Let H be a hypergraph, let $\overline{v} \in V(H)^*$ and let $v \in V(H)$ be such that $d(Center(H, \overline{v}),\ v) < \infty$. Let $X \subset V(H)$ be the set

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

Then we define Tr $(H, \overline{v};\ v)$ as the tree $H[X]$ with v as a root. That is, Tr $(H, \overline{v};\ v)$ is the tree formed of all vertices whose only path to $Center(H, \overline{v})$ 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 $\mathrm{Tr}(H, \overline{v};\ v;\ r)$ as $\mathrm{Tr}(Core(H, \overline{v};\ r), \overline{v};\ v)$. In the case that $\overline{v}$ is the empty list, we write simply $\mathrm{Tr}(H;\ v)$ or $\mathrm{Tr}(H;\ v;\ r)$.

For any $k \in \mathbb{N}$, we define an equivalence relation over rooted trees that generalizes both the relation of 'k-morphism' as defined in [10], and the notion of '(k, r)-values' defined in [11].

DEFINITION 2.8
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 ρ be an special symbol called the **root symbol**. Set $\widehat{\Sigma}_{k,r-1} := \Sigma_{k,r-1} \cup \{\rho\}$. Then a (k, r)-**pattern** is isomorphism class of $\widehat{\Sigma}_{k,r-1}$-hypergraphs (e, τ) 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 $\tau^k_{(T,v)} : V(T) \to \widehat{\Sigma}_{k,r-1}$ satisfying that $\tau^k_{(T,v)}(u)$ is the $\sim_k$ class of $\mathrm{Tr}(T, u;\ v)$ for any $u \neq v$, and $\tau^k_{(T,v)}(v) = \tau$.
- 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 $(e, \tau^k_{(T_\delta, v_\delta)}) \in \epsilon$' and the 'quantity of initial edges $e_2 \in E(T_2)$ such that $(e, \tau^k_{(T_\delta, v_\delta)}) \in \epsilon$ ' 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 2.9
Let $\mathbf{T}$ be a $\sim_k$ class of rooted trees with radii at most r. Then there is a partition $E^1_{\mathbf{T}}, E^2_{\mathbf{T}}$ of $P(k,r)$ and natural numbers $a_\epsilon < k$ for each $\epsilon \in E^2_{\mathbf{T}}$ that depends only on $\mathbf{T}$ such that a rooted tree (T,v) belongs to $\mathbf{T}$ if and only if the following hold: (1) for any pattern $\epsilon \in E^1_{\mathbf{T}}$, there are at least k initial edges $e \in E(T)$ such that $(e, \tau^k_{(T,v)}) \in \epsilon$, and (2) for any pattern $\epsilon \in E^2_{\mathbf{T}}$, there are exactly a_ϵ initial edges $e \in E(T)$ such that $(e, \tau^k_{(T,v)}) \in \epsilon$.

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 2.10
Let $k \in \mathbb{N}$. Given a non-tree connected hypergraph H, we define its **canonical k-coloring** τ^k_H as the one that assigns to each vertex $v \in V(H)$ the $\sim_k$ class of the tree $\mathrm{Tr}(H,v)$. Let H_1 and H_2 be connected hypergraphs that are not trees. Set $H'_1 := \mathit{Center}(H_1)$ and $H'_2 := \mathit{Center}(H_2)$. We say that H_1 and H_2 are k-equivalent, written as $H_1 \sim_k H_2$, if $(H'_1, \tau^k_{H_1}) \simeq (H'_2, \tau^k_{H_2})$.

DEFINITION 2.11
Let $k, r \in \mathbb{N}$ and let H_1 and H_2 be hypergraphs. Let $H'_1 := \mathit{Core}(H_1; r)$ and $H'_2 := \mathit{Core}(H_2; r)$. 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 2.12
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,τ) that are cycles of diameter at most $2r+1$. We denote by $C(k,r)$ the set of (k,r)-cycles.

OBSERVATION 2.13
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 $U^1_{\mathbf{O}}, U^2_{\mathbf{O}}$ of $C(k,r)$ and natural numbers $a_\omega < k$ for each $\omega \in U^2_{\mathbf{O}}$ 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 $\omega \in U^1_{\mathbf{O}}$, there are at least k connected components $H \subset \mathit{Core}(G; r)$ whose cycle $H' = \mathit{Center}(H)$ satisfies that $(H', \tau^k_H) \in \omega$, and (2) for any $\omega \in U^2_{\mathbf{O}}$, there are exactly a_ω connected components $H \subset \mathit{Core}(G; r)$ whose cycle $H' = \mathit{Center}(H)$ satisfies that $(H', \tau^k_H) \in \omega$.

DEFINITION 2.14
Let H be a hypergraph and let $k, r \in \mathbb{N}$. 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 $r' \le r$, vertices $v_1, \dots, v_k$ 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 $(T,v) \in \mathbf{T}$.

2.8 Main result and outline of the proof

Our goal is to prove the following theorem.

THEOREM 2.15
Let ϕ be a sentence in $FO[\sigma]$. Then the function
$F_\phi : (0,\infty)^{|\sigma|} \to \mathbb{R}$ given by

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

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 that 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 2.16
Let $k \in \mathbb{N}$ and let H_1, H_2 be hypergraphs. Set $r := (3^k - 1)/2$. Suppose that both H_1 and H_2 are (k, r)-rich and $H_1 \approx_{k,r} H_2$. Then Duplicator wins $\mathrm{EHR}_k(H_1, H_2)$.

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 2.17
Let $r \in \mathbb{N}$. Then a.a.s. G_n is r-simple.

THEOREM 2.18
Let $k, r \in \mathbb{N}$. Then a.a.s. G_n is (k, r)-rich.

THEOREM 2.19
Let $k, r \in \mathbb{N}$. Let $\mathbf{O}$ be a $\approx_{k,r}$ class of r-simple hypergraphs. Then

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

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

A sketch of the proof of Theorem 2.15 using these results as follows. Let $\Phi \in FO[\sigma]$ be a sentence and let $k := \mathrm{qr}(\Phi)$, $r := (3^k - 1)/2$. Because of Theorems 2.16 and 2.18, it holds that for any $\approx_{k,r}$ class $\mathbf{O}$

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

This together with Theorem 4.7 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 $\lim_{n\to\infty} \Pr(G_n \in \mathbf{O})$, where $\mathbf{O}$ is some $\approx_{k,r}$-class of r-simple hypergraphs. Finally, using

Theorem 2.19 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.

3 Model theoretic results

3.1 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 3.1
Let $\overline{v} \in V_1^*, \overline{u} \in V_2^*$ be tuples of the same length. We write $(H_1, \overline{v}) \simeq_{k,r} (H_2, \overline{u})$, if Duplicator wins $d\text{EHR}_k\,(N(\overline{v}; r), \overline{v};\ N(\overline{u}; r), \overline{u})$. 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 $\overline{v}$, resp. $\overline{u}$, such that $(H_1, \overline{v}) \simeq_{k,r} (H_2, \overline{u})$. Given $X \in V_1$, $Y \in V_2$ and tuples of the same length $\overline{v} \in V_1^*$ and $\overline{u} \in V_2^*$, we write $(H_1, (X, \overline{v})) \simeq_{k,r} (H_2, (Y, \overline{u}))$, if X and Y can be ordered to form lists $\overline{w}$, resp. $\overline{z}$ such that $(H_1, \overline{w}^\frown\overline{v}) \simeq_{k,r} (H_2, \overline{z}^\frown\overline{u})$.

DEFINITION 3.2
Fix $r \in \mathbb{N}$. Suppose $X \subseteq V_1$ and $Y \subseteq V_2$ can be partitioned into sets $X = X_1 \cup \cdots \cup X_a$ and $Y = Y_1 \cup \cdots \cup Y_b$ 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 $Z \subset V_\delta$, 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 3.3
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) $(H_1, X) \cong_{k,r} (H_2, Y)$.

- 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 $\text{EHR}_k\,(H_1; H_2)$.

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

OBSERVATION 3.4
Let $k \in \mathbb{N}$ and let $\overline{v} \in V(H_1)^*$, $\overline{u} \in V(H_2)^*$ be of equal length. Suppose Duplicator wins $d\text{EHR}_k(H_1, \overline{v};\ H_2, \overline{u})$. Then, for any $r \in \mathbb{N}$, $(H_1, \overline{v}) \simeq_{k,r} (H_2, \overline{u})$.

OBSERVATION 3.5
Let $k \in \mathbb{N}$ and let $\overline{v} \in V(H_1)^*$, $\overline{u} \in V(H_2)^*$ be of equal length. Suppose Duplicator wins $d\text{EHR}_k(H_1, \overline{v};\ H_2, \overline{u})$. 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 $d\text{EHR}_{k-1}(H_1, \overline{v_2};\ H_2, \overline{u_2})$, where $\overline{v_2} := \overline{v}^\frown v$ and $\overline{u_2} := \overline{u}^\frown u$.

LEMMA 3.6
Let $k, r \in \mathbb{N}$. Let $\overline{v} \in V_1^*$ and $\overline{u} \in V_2^*$ be of equal length. $(H_1, \overline{v}) \simeq_{k,3r+1} (H_2, \overline{u})$. Let $v \in V_1$ and $u \in V_2$ be vertices played in the first round of an instance of

$$d\textsc{Ehr}_k\left(N(\overline{v}; 3r+1), \overline{v}; \quad N(\overline{u}; 3r+1), \overline{u}\right),$$

where Duplicator is following a winning strategy. Further, suppose that $d(\overline{v}, v) \le 2r+1$ (and in consequence $d(\overline{u}, u) \le 2r+1$ 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 3.5, we get that Duplicator wins

$$d\textsc{Ehr}_{k-1}\left(N(\overline{v}; 3r+1), \overline{v_2}; \quad N(\overline{u}; 3r+1), \overline{u_2}\right)$$

as well. Call $H_1' = N(\overline{v}; 3r+1)$, $H_2' = N(\overline{u}; 3r+1)$. Then by Observation 3.5 Duplicator wins

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

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 $d(z, \overline{v}) \le d(z, v') + d(v', \overline{v}) = 3r+1$. As a 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. □

PROOF OF THEOREM 3.3. Let $X_1, \dots, X_a$ and $Y_1, \dots, Y_b$ be partitions of X and Y, respectively, as in the definition of $\cong_{k,r}$. Let $r_0 := (3^k - 1)/2$ and $r_i := (r_{i-1} - 1)/3$ for each $1 \le i \le k$. Let v_i^1 and v_i^2 be the vertices played in H_1 and H_2, respectively, during the i-th round of $\textsc{Ehr}_k(H_1, H_2)$. We show a winning strategy for Duplicator in $\textsc{Ehr}_k(H_1; H_2)$. 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_δ 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.

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 $\overline{v}(T)$.

– Let $S \subset V_1$ be a marked set. Then for any different marked $S' \subset V_1$ of any different S' among $X_1, \dots, X_a$, 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 $Y_1, \dots, Y_b$, 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

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

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

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

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 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

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

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

  $$d\text{EHR}_{k-i+1}\left(N(\overline{w}^\frown \overline{v}(S); 3r_i + 1), \overline{w}^\frown \overline{v}(S); \quad N(\overline{z}^\frown \overline{v}(T); 3r_i + 1), \overline{z}^\frown \overline{v}(T) \right).$$

  Thus, Duplicator can choose $v_i^2 \in V_2$ according to the winning strategy in that game. After this, Duplicator sets $\overline{v}(S) := \overline{v}(S)^\frown v_i^1$, and $\overline{v}(T) := \overline{v}(T)^\frown v_i^2$. Notice that because of Lemma 3.6 now

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

- 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, Duplicator wins

  $$d\text{EHR}_{k-i+1}\left(N(\overline{w}; 3r_i + 1), \overline{w}; \quad N(\overline{z}; 3r_i + 1), \overline{z} \right).$$

  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 3.6 now

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

- For all marked sets $S \subset V_1$, we have $d(S \cup \overline{v}(S), v_i^1) > 2r_i + 1$, 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 it follows that Duplicator can choose $v_i^2 \in V_2$ such that (A) $d(T \cup \overline{v}(T), v_i^2) > 2r_i + 1$ 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 Observations 3.4 and 3.5. □

3.2 k-Equivalent trees

We want prove the following.

THEOREM 3.7
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 3.8
Let $k \in \mathbb{N}$ and fix $r > 0$. Suppose Theorem 3.7 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 2.8. 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'_1. 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 $T''_1 := \text{Tr}\left(T'_1, v_1;\ u\right)$ and $T''_2 := \text{Tr}\left(T'_2, v_2;\ f(u)\right)$. Then, as $\tau^k_{(T_1,v_1)}(u) = \tau^k_{(T_2,v_2)}(f(u))$, we obtain $\left(T''_1, u\right) \sim_k \left(T''_2, f(u)\right)$. As both these trees have radii at most r, by assumption Duplicator has a winning strategy in $d\text{EHR}_k\left(T''_1, u;\quad T''_2, f(u)\right)$ and they can follow it considering the previous plays in T''_1 and T''_2. □

PROOF OF THEOREM. 3.7 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 2.8. 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 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 $\left(e_1, \tau^k_{(T_1,v_1)}\right) \simeq \left(e_2, \tau^k_{(T_2,v_2)}\right)$. 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 3.8, 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 $\left(e_1, \tau^k_{(T_1,v_1)}\right) \simeq \left(e_2, \tau^k_{(T_2,v_2)}\right)$. Again, because of Lemma 3.8, 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. □

3.3 k-Equivalent hypergraphs

THEOREM 3.9
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 $\tau^k_{H_1}, \tau^k_{H_2}$ be as in Definition 2.10. Let f be an isomorphism between $(H_1', \tau^k_{H_1})$ and $(H_2', \tau^k_{H_2})$. Let $\overline{v}$ be an ordering of the vertices of H_1', and let $\overline{u} := f(\overline{v})$ be the corresponding ordering of the vertices of H_2'. Then Duplicator wins $d\text{EHR}_k\left(H_1', \overline{v};\ H_2', \overline{u}\right)$.

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 3.7, 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. ☐

3.4 Main result

LEMMA 3.10
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 $(H_1, X) \cong_{k,r} (H_2, Y)$ in the sense of Definition 3.2.

PROOF. Let $X_1, \ldots, X_a$ and $Y_1, \ldots, Y_b$ 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. $Core(H_2; r)$. Because of Theorem 3.9 $N(X_i; r) \sim_k N(Y_j; r)$ implies $(H_1, X_i) \simeq_{k,r} (H_2, Y_j)$ in the sense of Definition 3.1. The result follows now from the definition of $H_1 \approx_{k,r} H_2$. ☐

THEOREM 3.11
Let $k \in \mathbb{N}$, and set $r := (3^k - 1)/2$. 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 $\text{EHR}_k(H_1, H_2)$.

PROOF. Because of the previous lemma, we can apply Theorem 3.3 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 3.3 holds. ☐

4 Probabilistic results

4.1 Almost all hypergraphs are simple

LEMMA 4.1
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 $\text{E}\big[X_n\big] = \Theta(n^{-\text{ex}(H)})$.

PROOF. We have

$$\text{E}\big[X_n\big] = \sum_{H' \in Copies(H,[n])} \Pr\left(H' \subset G_n\right).$$

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

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

Substituting in the first equation, we get

$$\mathrm{E}\big[X_n\big] \sim \frac{(n)_{|H|}}{\mathrm{aut}(H)}\prod_{R\in\sigma}\left(\frac{\beta_R}{n^{ar(R)-1}}\right)^{|E_R(H)|} \sim n^{-\mathrm{ex}(H)}\frac{\prod_{R\in\sigma}\beta_R^{|E_R H|}}{\mathrm{aut}(H)}.$$

□

LEMMA 4.2
Let H be a hypergraph such that $\mathrm{ex}(H) > 0$. Then a.a.s. there are no copies of H in G_n.

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

LEMMA 4.3
Let H be a hypergraph. Let $\overline{v} \in (\mathbb{N})_*$ be a list of vertices with $\mathrm{len}(\overline{v}) \leq |V(H)|$. 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 $\mathrm{E}\big[X_n\big] = \Theta(n^{-\mathrm{ex}(H)-\mathrm{len}(\overline{v})})$.

PROOF. The number of hypergraphs $H' \in Copies(H,[n])$ that contain all vertices in $\overline{v}$ is asymptotically $\sim n^{|V(H)|-\mathrm{len}(\overline{v})}$ for some constant C. Then

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

□

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 4.4
Let G be a dense hypergraph with diameter at most r, and let $H \subset G$ be a connected sub-hypergraph with $\mathrm{ex}(H) < \mathrm{ex}(G)$. Then there is a connected sub-hypergraph $H' \subset G$ satisfying $H \subset H'$, $\mathrm{ex}(H) < \mathrm{ex}(H')$ and that $|E(H')| \leq |E(H)| + 2r + 1$.

PROOF. Suppose there is some edge $e \in E(G) \setminus E(H)$ with and $\mathrm{ex}(e) \geq 0$. 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 $\mathrm{ex}(e) = -1$. 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 $\mathrm{ex}(G') < \mathrm{ex}(G)$. Let e be the edge that was cut last. Then $V(G') \cap V(e) = \mathrm{ex}(G) - \mathrm{ex}(G') + 1 \geq 2$. 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 4.5
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 4.6
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,

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

Because of Lemma 4.2, 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 $\sim$ classes of dense hypergraphs whose size bounded by R, we deduce that the right hand side of the last inequality tends to 1. □

As a corollary, we obtain the needed result.

THEOREM 4.7
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 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 4.8
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 $N(v; r)$ are all trees and they are all disjoint.

PROOF. An application of the first moment method together with Lemma 4.3 and the fact that there is a finite number of $\simeq$ classes of paths whose length is at most $2r+1$, implies that a.a.s. the $N(v; r)$ are disjoint. Also, because of Theorem 4.6 a.a.s the $N(v; r)$ are either trees or unicycles. But if any of the $N(v; r)$ was a unicycle, then in G_n there would exist 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 4.3 again, as well as 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. □

LEMMA 4.9
Let $\overline{v} \subset \mathbb{N}*$ be a finite set of fixed vertices, and let $\pi(\overline{x})$ be an edge formula such that $\text{len}(\overline{x}) = \text{len}(\overline{v})$. Define $G'_n = G_n \setminus E[\overline{v}]$ (i.e. G_n minus all the edges induced on $\overline{v}$). Fix $r \in \mathbb{N}$. Then a.a.s. for all vertices $v \in \overline{v}$ 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 $\overline{v}$. Because edges are independent in our random model, we have that $\Pr(A_n \mid \pi(\overline{v})) = \Pr(A_n)$. Now the result follows from Lemma 4.8 using that $G'_n \subset G_n$. □

4.2 Probabilities of trees

DEFINITION 4.10
We define Λ 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 Λ; and **(4)** for any $\lambda_1, \lambda_2 \in \Lambda$, the product $\lambda_1\lambda_2$ belongs to Λ as well.

DEFINITION 4.11
Let $r \in \mathbb{N}$ and let $\mathbf{T}$ be a $\sim_k$ class of trees with radius at most r. Let $v \in \mathbb{N}$ be an arbitrary vertex. We define $\Pr[r, \mathbf{T}]$ as the limit

$$\lim_{n\to\infty} \Pr\left(Tr(G_n,\ v;\ v;\ r) \in \mathbf{T}\right).$$

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

THEOREM 4.12
Fix $r \in \mathbb{N}$. Let $k \in \mathbb{N}$. The following hold:

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

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

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

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 4.13
Conditions (1) and (2) of Theorem 4.12 are satisfied for $r = 0$.

DEFINITION 4.14
Let $k \in \mathbb{N}$ and $r > 0$. Suppose that Theorem 4.12 holds for $r - 1$. Given a (k, r)-pattern ϵ, we define the expressions $\lambda_{r,\epsilon}$ and $\mu_{r,\epsilon}$ as follows. Let (e, τ) be a representative of ϵ 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

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

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 4.15
Let $k \in \mathbb{N}$, $r > 0$ and $\overline{u} \in (\mathbb{N})_*$. Let $\pi(\overline{x}) \in FO[\sigma]$ be a consistent edge formula such that $\text{len}(\overline{x}) = \text{len}(\overline{u})$. Let $\overline{v} \in (\mathbb{N})_*$ be vertices contained in $\overline{u}$. For each $v \in \overline{v}$ set $T_{n,v} := \text{Tr}\,(G_n, \overline{u};\; v;\; r)$. Given a pattern $\epsilon \in P(k,r)$ and $v \in \overline{v}$, 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 4.12 holds for $r-1$. Then the conditional distributions of the variables $X_{n,v,\epsilon}$ given $\pi(\overline{u})$ 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 $\overline{v}$ 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 2.1, 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

$$\lim_{n\to\infty} \mathrm{E}_{\pi(\overline{u})}\left[\prod_{\epsilon\in P(k,r)} \binom{X_{n,\epsilon}}{b_\epsilon}\right] = \prod_{\epsilon\in P(k,r)} \frac{(\mu_{r,\epsilon})^{b_\epsilon}}{b_\epsilon!}. \tag{1}$$

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

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

Informally, elements of Ω_n represent choices of b_ϵ possible initial edges of T_n whose k- pattern is ϵ for all (k,r)-patterns ϵ. Using Observation 2.2, we obtain

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

We say that a choice $\{E_\epsilon\}_\epsilon \in \Omega_n$ is **disjoint** if the edges $(e,\tau) \in \bigcup_{\epsilon\in P(k,r)} E_\epsilon$ satisfy that no vertex $w \in \overline{u}$ 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 Ω_n and set $\Omega'_{\mathbb{N}} = \cup_{n\in\mathbb{N}}\Omega'_n$. If for some $\{E_\epsilon\}_\epsilon \in \Omega_n$, we have that $e \in E(T_n)$ for all $(e,\tau) \in \bigcup_{\epsilon\in P(k,r)} E_\epsilon$ then $\{E_\epsilon\}_\epsilon$ is necessarily disjoint. This is because T_n is a tree and the only vertex in $\overline{u}$ that belongs to T_n is v by definition. Thus, in the last sum it suffices to consider only the disjoint $\{E_\epsilon\}_\epsilon$. Because of the symmetry of the random model, the probabilities in that sum are the same for all disjoint choices of $\{E_\epsilon\}_\epsilon$. Hence, if we fix $\{E_\epsilon\}_\epsilon \in \Omega'_{\mathbb{N}}$ we obtain

$$\mathrm{E}_{\pi(\overline{u})}\left[\prod_{\epsilon\in P(k,r)} \binom{X_{n,\epsilon}}{b_\epsilon}\right] = |\Omega'_n|\mathrm{Pr}_{\pi(\overline{u})}\left(\bigwedge_{\substack{\epsilon\in P(k,r)\\(e,\tau)\in E_\epsilon}}\left(e \in E(T_n) \bigwedge_{\substack{u\in V(e)\\u\neq v}} Tr(T_n, v; u) \in \tau(u)\right)\right). \tag{2}$$

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

$$|\Omega'_n| = (n - \text{len}(\overline{u}))_N \prod_{\epsilon\in P(k,r)} \frac{1}{b_\epsilon!}\left(\frac{1}{\text{aut}(\epsilon)}\right)^{b_\epsilon}. \tag{3}$$

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

$$\bigwedge_{\substack{\epsilon \in P(k,r) \\ (e,\tau) \in E_\epsilon}} e \in E(G_n)$$

can be described via an edge formula whose variables are interpreted as vertices in $\overline{w}$. Let $\psi(\overline{x})$ be one of such edge formulas. This event is independent of $\pi(\overline{u})$ because edges are independent in G_n. Thus, a simple computation yields

$$\Pr_{\pi(\overline{u})}\left(\bigwedge_{\substack{\epsilon \in P(k,r) \\ (e,\tau) \in E_\epsilon}} e \in E(G_n)\right) = \prod_{\epsilon \in P(k,r)} \left(\frac{\beta_{R(\epsilon)}}{n^{ar(R(\epsilon)-1)}}\right)^{b_\epsilon} = \frac{1}{n^N} \prod_{\epsilon \in P(k,r)} \beta_{R(\epsilon)}^{b_\epsilon}.$$

Because of Lemma 4.9 a.a.s. if $e \in E(G_n)$ and $v \in V(e)$, then $e \in E(T_n)$. Thus,

$$\Pr_{\pi(\overline{u})}\left(\bigwedge_{\substack{\epsilon \in P(k,r) \\ (e,\tau) \in E_\epsilon}} \left(e \in E(T_n) \bigwedge_{\substack{u \in V(e) \\ u \neq v}} Tr(T_n;\, u) \in \tau(u)\right)\right) \tag{4}$$

$$\sim \left(\frac{1}{n^N} \prod_{\epsilon \in P(k,r)} \beta_{R(\epsilon)}^{b_\epsilon}\right) \Pr_{\pi(\overline{u}) \wedge \psi(\overline{w})}\left(\bigwedge_{\substack{\epsilon \in P(k,r) \\ (e,\tau) \in E_\epsilon}} \bigwedge_{\substack{u \in V(e) \\ u \neq v}} Tr(T_n;\, u) \in \tau(u)\right).$$

The trees $Tr(T_n; u)$ in the last probability coincide with $Tr(G_n, \overline{u}^\frown\overline{w};\, u;\, r-1)$ for all u. As a consequence, using the hypothesis that Theorem 4.12 holds for $r-1$, we obtain

$$\Pr_{\pi(\overline{u}) \wedge \psi(\overline{w})}\left(\bigwedge_{\substack{\epsilon \in P(k,r) \\ (e,\tau) \in E_\epsilon}} \bigwedge_{\substack{u \in V(e) \\ u \neq v}} Tr(T_n;\, u) \in \tau(u)\right) \sim \prod_{\epsilon \in P(k,r)} (\lambda_{r,\epsilon})^{b_\epsilon}.$$

Combining this with Equations (2)–(4), we obtain

$$\mathrm{E}_{\pi(\overline{u})}\left[\prod_{\epsilon \in P(k,r)} \binom{X_{n,\epsilon}}{b_\epsilon}\right] \sim \frac{(n - \mathrm{len}(\overline{u}))_N}{n^N} \prod_{\epsilon \in P(k,r)} \frac{1}{b_\epsilon!} \left(\frac{\beta_{R(\epsilon)}\lambda_{r,\epsilon}}{\mathrm{aut}(\epsilon)}\right)^{b_\epsilon} \sim \prod_{\epsilon \in P(k,r)} \frac{(\mu_{r,\epsilon})^{b_\epsilon}}{b_\epsilon!}.$$

This proves (11) and the statement. □

Next lemma completes the proof of Theorem 4.12.

LEMMA 4.16
Let $r > 0$. Suppose that Theorem 4.12 holds for $r-1$. Then it also holds for r.

PROOF. Fix $k \in \mathbb{N}$. We start showing condition (1) of Theorem 4.12. Fix $\mathbf{T}$ a $\sim_k$ class of trees with radius at most r. Fix a vertex $v \in \mathbb{N}$ as well. Set $T_n := \mathrm{Tr}(G_n, v;\, v;\, r)$. For each $\epsilon \in P(k,r)$, let $X_{n,\epsilon}$ be the random variable that counts the number of initial edges in T_n whose pattern is ϵ.

Let $E^1_{\mathbf{T}}, E^2_{\mathbf{T}}, \{a_\epsilon\}_\epsilon$ be as in Observation 2.2. Then

$$\Pr[r,\mathbf{T}] = \lim_{n\to\infty} \Pr(T_n \in \mathbf{T}) = \lim_{n\to\infty} \Pr\left(\left(\bigwedge_{\epsilon\in E^1_{\mathbf{T}}} X_{n,\epsilon} \geq k\right) \wedge \left(\bigwedge_{\epsilon\in E^2_{\mathbf{T}}} X_{n,\epsilon} = a_\epsilon\right)\right).$$

Using the previous lemma, we obtain that the last limit equals the following expression:

$$\left(\prod_{\epsilon\in E^1_{\mathbf{T}}} \mathrm{Poiss}_{\mu_{r,\epsilon}}(\geq k)\right)\left(\prod_{\epsilon\in E^2_{\mathbf{T}}} \mathrm{Poiss}_{\mu_{r,\epsilon}}(a_\epsilon)\right).$$

Using the definition of the $\mu_{r,\epsilon}$, we obtain that the last expression belongs to Λ as we wanted to prove. Furthermore, as the $\mu_{r,\epsilon}$ are positive, this expression is also positive for all values of $\{\beta_R\}_{R\in\sigma} \in (0,\infty)^{|\sigma|}$. Now we proceed to prove condition (2). Let $\overline{u}, \overline{v}, \{\mathbf{T}_v\}_{v\in\overline{v}}$ and $\pi(\overline{x})$ be as in the statement of (2). Using the previous lemma, we obtain that the events $Tr(G_n, \overline{u};\, v;\, r) \in \mathbf{T}_v$ for all $v \in \overline{v}$ are asymptotically independent and are also independent of $\pi(\overline{u})$. Then the desired result follows from condition (1). □

4.3 Almost all graphs are (k,r)-rich

THEOREM 4.17
Let $k, r \in \mathbb{N}$. Then a.a.s. G_n is (k,r)-rich.

PROOF. Let Σ be the set of all $\sim_k$ classes of rooted trees with radii at most r. Let $m > k$. For each $\mathbf{T} \in \Sigma$, let $\overline{v}(\mathbf{T}) \in (\mathbb{N})_m$ be tuples satisfying that all the $\overline{v}(\mathbf{T})$ are disjoint. Let $\overline{w} \in (\mathbb{N})_*$ be a concatenation of all the $\overline{v}(\mathbf{T})$. For each $\mathbf{T} \in \Sigma$, define $X_{n,\mathbf{T}}$ as the number of vertices $v \in \overline{v}(\mathbf{T})$ such that $Tr(G_n, \overline{w};\, v;\, r) \in \mathbf{T}$. Because of Theorem 4.12, the $\sim_k$ types of the trees $Tr(G_n, \overline{w};\, v;\, r)$ for all $v \in \overline{w}$ are asymptotically independent and given any $v \in \overline{w}$ and $\mathbf{T}$ it holds that $\Pr(Tr(G_n, \overline{w};\, v;\, r) \in \mathbf{T})$ tends to $\Pr[r,\mathbf{T}]$ as n goes to infinity. Hence, the variables $X_{n,\mathbf{T}}$ converge in distribution to independent binomial variables whose respective parameters are m and $\Pr[r,\mathbf{T}]$. That is, given natural numbers $0 \leq l_{\mathbf{T}} \leq m$ for all $\mathbf{T} \in \Sigma$,

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

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

$$\lim_{n\to\infty} \Pr\left(\left|X_{n,\mathbf{T}}/m - \Pr[r,\mathbf{T}]\right| \geq \delta\right) \leq \epsilon \quad \text{for all} \mathbf{T} \in \Sigma. \tag{5}$$

Also, for m large enough, we have

$$\Pr[r,\mathbf{T}] > k/m + \delta \quad \text{for all} \mathbf{T} \in \Sigma. \tag{6}$$

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

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

We define A_n as the event that for any $v \in \overline{w}$, we have $N(v;r) \cap Core(G_n;r) = \emptyset$ (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 $d^{G_n}(v_1,v_2) > 2r+1$. 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 $X_{n,\mathbb{T}} \geq k$ for all $\mathbb{T}$ then G_n is (k, r)-rich. Because of Lemma 4.8 a.a.s. A_n holds, and we obtain

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

As ϵ 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. □

4.4 Probabilities of cycles

DEFINITION 4.18
We define Γ and Υ 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 $\frac{\lambda}{b}\prod_{R\in\sigma}\beta_R^{a_R}$ belongs to Γ; (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 Υ; and (3) if $\upsilon_1, \upsilon_2 \in \Upsilon$ then $\upsilon_1\upsilon_2 \in \Upsilon$ as well.

DEFINITION 4.19
Let $k, r \in \mathbb{N}$ and $O \in C(k, r)$. Let (H, τ) be a representative of O. We define $\lambda_{r,O}$ and $\gamma_{r,O}$ in the following way:

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

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 Γ.

LEMMA 4.20
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 $Core(G_n; r)$ such that $H' := Center(H)$ satisfies that $(H', \tau_H^k) \in O$. 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 4.15. By Theorem 2.1, to prove the result is enough to show that for any natural numbers $\{b_O\}_{O\in C(k,r)}$ it holds

$$\lim_{n\to\infty} \mathrm{E}\left[\prod_{O\in C(k,r)}\binom{X_{n,O}}{b_O}\right] = \prod_{O\in C(k,r)}\frac{(\gamma_{r,O})^{b_O}}{b_O!}. \tag{7}$$

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

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

Given a cycle H such that $V(H) \subseteq [n]$, we say that $H \sqsubset G_n$ if $H = \mathit{Center}(H')$ for some connected component H' of $\mathit{Core}(G_n;\, r)$. Using observation Observation 2.2, we obtain

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

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 Ω'_n as the set of disjoint elements in Ω_n and set $\Omega'_{\mathbb{N}} := \cup_{n\in\mathbb{N}} \Omega'_n$. If for some $\{F_O\}_O \in \Omega_n$ it holds that $H \sqsubset G_n$ for all $(H, \tau) \in \cup_O F_O$ then necessarily $\{F_O\}_O$ 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 $\mathit{Core}(G_n;\, r)$ and thus $H_1 \cup H_2 \subset \mathit{Center}(H)$. As a consequence, neither $H_1 \sqsubset G_n$ nor $H_2 \sqsubset G_n$ hold. $(H_1, \tau_1), (H_2, \tau_2) \in \bigcup_{O\in C(k,r)} F_O$. Hence, in the last sum it suffices to consider disjoint choices $\{F_O\}_O$. 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 consequence, if we fix $\{F_O\}_O \in \Omega'_{\mathbb{N}}$ we obtain

$$\mathrm{E}\left[\prod_{O\in C(k,r)} \binom{X_{n,O}}{b_O}\right] = |\Omega'_n| \Pr\left(\bigwedge_{\substack{O\in C(k,r)\\ (H,\tau)\in F_O}} \left(H \sqsubset G_n \bigwedge_{v\in V(H)} \mathit{Tr}(G_n, v; r) \in \tau(v) \right)\right). \tag{8}$$

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

$$|\Omega'_n| = \frac{(n)_N}{\prod_{O\in C(k,r)} b_O!\, \mathrm{aut}(O)^{b_O}}. \tag{9}$$

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

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

can be written as an edge formula concerning the vertices in $\overline{v}$. Let $\varphi(\overline{x})$ be one of such sentences. We have that

$$\Pr\left(\bigwedge_{\substack{O\in C(k,r)\\ (H,\tau)\in F_O}} H \subset G_n \right) = \prod_{O\in C(k,r)} \left(\frac{\prod_{R\in\sigma} \beta_R^{|E_R(O)|}}{n^{|O|}} \right)^{b_O} = \frac{1}{n^N} \prod_{O\in C(k,r)} \left(\prod_{R\in\sigma} \beta_R^{|E_R(O)|} \right)^{b_O}.$$

Because of Theorem 4.7 a.a.s. if some cycle H of diameter at most $2r+1$ satisfies $H \subset G_n$, then $H \sqsubset G_n$. Hence,

$$\Pr\left(\bigwedge_{\substack{O\in C(k,r)\\(H,\tau)\in F_O}}\left(H\sqsubset G_n \bigwedge_{v\in V(H)} Tr(G_n,v;r)\in\tau(v)\right)\right)$$

$$\sim \frac{1}{n^N}\prod_{O\in C(k,r)}\left(\prod_{R\in\sigma}\beta_R^{|E_R(O)|}\right)^{b_O}\Pr_{\varphi(\overline{v})}\left(\bigwedge_{\substack{O\in C(k,r)\\(H,\tau)\in F_O}}\bigwedge_{v\in V(H)} Tr(G_n,v;r)\in\tau(v)\right). \tag{10}$$

As all the vertices $v\in\overline{v}$ belong to $Core(G_n;\, r)$, the trees $Tr(G_n;\, v;\, r)$ in the last probability coincide with $Tr(G_n,\, \overline{v};\, v;\, r)$. By Theorem 4.12, we have that

$$\Pr_{\varphi(\overline{v})}\left(\bigwedge_{\substack{O\in C(k,r)\\(H,\tau)\in F_O}}\bigwedge_{v\in V(H)} Tr(G_n,v;r)\in\tau(v)\right)\sim\prod_{O\in C(k,r)}(\lambda_{r,O})^{b_O}.$$

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

$$\mathrm{E}\left[\prod_{O\in C(k,r)}\binom{X_{n,O}}{b_O}\right]\sim\frac{(n)_N}{n^N}\prod_{O\in C(k,r)}\frac{1}{b_O!}\left(\frac{\lambda_{r,O}\prod_{R\in\sigma}\beta_R^{|E_R(O)|}}{\mathrm{aut}(O)}\right)\sim\prod_{O\in C(k,r)}\frac{(\gamma_{r,O})^{b_O}}{b_O!}.$$

This proves (7) and the statement. □

THEOREM 4.21

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

PROOF. For each $O\in C(k,r)$, let $X_{n,O}$ be as in the previous lemma. Let $U^1_{\mathbf{O}}, U^2_{\mathbf{O}}$ and $\{a_O\}_{O\in U^2_{\mathbf{O}}}$ be as in Observation 2.13. Let A_n be the event that G_n is r-simple. Then

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

Because of Theorem 4.7, a.a.s A_n holds. Thus, using the last lemma the previous limit equals the following expression:

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

As all the $\gamma_{r,O}$ belong to Γ, this last expression belongs to Υ and the theorem is proven. □

5 Proof of the main theorem

THEOREM 5.1
Let $\phi \in FO[\sigma]$. Then the function $F_\phi : [O, \infty)^{|\sigma|} \to [0, 1]$ given by

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

is well defined and it is given by a finite sum of expressions in Υ.

PROOF. Let k be the quantifier rank of ϕ and let $r = 3^k$. Let $G_n := G_n\left(\{\beta_R\}_{R\in\sigma}\right)$ and let Σ be the set of (k, r)-agreeability classes of r-simple hypergraphs. Because of Theorem 4.7 a.a.s. G_n is r-simple. Thus,

$$\lim_{n\to\infty} \Pr\left(G_n \models \phi\right) = \lim_{n\to\infty} \sum_{\mathbf{O}\in\Sigma} \Pr\left(G_n \in \mathbf{O}\right) \Pr\left(G_n \models \phi \,\middle|\, G_n \in \mathbf{O}\right). \tag{11}$$

Because the set Σ is finite, we can exchange the summation and the limit. By Theorem 4.17 a.a.s. G_n is (k, r)-rich. This together with Theorem 3.11 implies that for any $\mathbf{O} \in \Sigma$

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

Let $\Sigma' \subset \Sigma$ be the set of classes $\mathbf{O}$ for which last limit equals 1. Then

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

Because of Theorem 4.21, we know that each of the limits inside the last sum exists and is given by an expression that belongs to Υ. As a consequence, the theorem follows. □

6 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 2.15 allows for many variants.

DEFINITION 6.1
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 $0 \le p \le 1$, we define the random model $F(l, n, p)$ as the discrete probability space that assigns to each l-CNF formula F on the variables $\{x_i\}_{i\in[n]}$ the probability

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

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 $\binom{2^l}{nl}$ non-tautological clauses of size l on the variables $\{x_i\}_i$ with probability p

independently. When p is a function of n satisfying $p(n) \sim \beta/n^{l-1}$, 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 σ consisting of $l+1$ relation symbols $R_0,\dots,R_l$ of 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 $x_{i_{j+1}},\dots x_{i_l}$. The relations $R_1,\dots,R_l$ satisfy the following axioms: (1) given $0\leq j\leq l$ and variables $y_1,\dots,y_l$ the fact that $R_j(y_1,\dots,y_l)$ holds is invariant under any permutation of the variables $y_1,\dots,y_j$ or $y_{j+1},\dots,y_l$, and (2) for any $0\leq j\leq l$ and any variables $y_1,\dots,y_l$ it holds that $R_j(y_1,\dots,y_l)$ only if all the y_i are different. Call $\mathcal{C}$ the family of σ-structures satisfying the last two axioms. The language σ and the family $\mathcal{C}$ satisfy the conditions in Section 2.4. The random model $F_l(n,p)$ coincides with the model $G(n,\{p_R\}_R)$ of random $\mathcal{C}$-hypergraphs described in Section 2.6 when all the p_R are equal. As a particular case of Theorem 2.15, we obtain the following result.

THEOREM 6.2
Let $l>1$ be a natural number. Then for each sentence $\Phi\in FO[\sigma]$, it is satisfied that the map $f_\Phi:(0,\infty)\to\mathbb{R}$ 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 6.3
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 such that $m(n)\sim cn$. For each n, let $C_{n,1},\dots,C_{n,m(n)}$ be clauses chosen uniformly at random independently among the $\binom{2l}{nl}$ non-tautological 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 $0<c_1<c_2$, 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, 4, 8]. We want to show that an analogous 'phase transition' also happens in $F(l,n,p)$ when $p\sim\beta/n^{l-1}$. We start by showing the following.

COROLLARY 6.4
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 $m(n)\sim cn$. For each $n\in\mathbb{N}$, 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 $x_1,\dots,x_n$. 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 $n\in\mathbb{N}$, let $C_{n,1},\dots,C_{n,m(n)}$ and $UNSAT_n$ be as in the previous theorem. One can consider $F_{n,m(n)}$ to be the result of selecting clauses $C_{n,1},\dots,C_{n,m(n)}$ uniformly at random independently among all possible clauses, given the fact that no two clauses $C_{n,i},C_{n,j}$ are equal. Hence,

$$\Pr\left(F_{n,m(n)}\text{ is unsatisfiable}\right)=\Pr\left(UNSAT_n\,\middle|\,\text{all the }C_{n,i}\text{ are different}\right).$$

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 1.1 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 6.5
Let $l > 1$. Then there are real positive values $\beta_1 < \beta_2$ such that a.a.s. $F_n^l(\beta)$ is satisfiable for $0 < \beta < \beta_1$, and a.a.s. $F_n^l(\beta)$ 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 $F_n^l(\beta)$. We have that $\mathrm{E}[X_n(\beta)] \sim \frac{\beta 2^l}{l!} n$. Let c_1, c_2 be as in last corollary. Define $\beta_1 := \frac{c_1 l!}{2^l}$ and $\beta_2 := \frac{c_2 l!}{2^l}$. Fix $\beta \in \mathbb{R}$ satisfying $0 < \beta < \beta_1$. Let $\epsilon > 0$ be a real number such that $\frac{\beta 2^l}{l!} + \epsilon < c_1$. For each $n \in \mathbb{N}$ set $\delta_1(n) := \lfloor * \rfloor \left(\frac{\beta 2^l}{l!} - \epsilon \right) n$ and $\delta_2(n) := \lfloor * \rfloor \left(\frac{\beta 2^l}{l!} + \epsilon \right) n$.

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,

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

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

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

Because of the law of large numbers,

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

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

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

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

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

THEOREM 6.6
Let $l > 1$ be a natural number. Let $\Phi \in FO[\sigma]$ be a FO sentence that implies unsatisfiability. Then for all $\beta > 0$ a.a.s. $F_n^l(\beta)$ does not satisfy Φ.

PROOF. Let β_1 and β_2 be as in Theorem 6.5. As Φ implies unsatisfiability $\Pr\left(F_n^l(\beta) \models \Phi\right) \leq \Pr\left(F_n^l(\beta)\text{is unsatisfiable}\right)$. Thus, by Theorem 6.5, we get that for all $\beta \in (0, \beta_1)$

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

By Theorem 6.2, last limit varies analytically with β. It vanishes in the proper interval $(0, \beta_1)$ (recall that $\beta_1 > 0$), then by the principle of analytic continuation, it has to vanish in the whole $(0, \infty)$, and the result holds. □

Concluding remarks

The question of whether our main result Theorem 2.15 could be generalized for structures containing unary relations was raised by a reviewer. We claim that this question has positive answer.

Let $\mathcal{C}$ be a family of hypergraphs defined as in Section 2.4 whose vocabulary is σ_2, and let σ_1 be a finite set of unary relation symbols. We consider the family $\widehat{\mathcal{C}}$ whose elements are of the form $\widehat{G} = (V, E, \{R^G\}_{R\in\sigma_1})$, where (V, E) is a hypergraph in $\mathcal{C}$ and for each $R \in \sigma_1$ the set R^G is a unary relation over V. Put σ for the vocabulary $\sigma_1 \cup \sigma_2$. The language $FO[\sigma]$ is then interpreted naturally over the structures in $\widehat{\mathcal{C}}$. We define the random model $G^{\widehat{\mathcal{C}}}(n, \{p_R\}_{R\in\sigma})$ as the discrete probability space where we obtain a structure $(V, E, \{R^G\}_{R\in\sigma_1}) \in \widehat{\mathcal{C}}$ by taking first a random hypergraph $(V, E) \in G^{\mathcal{C}}(n, \{p_R\}_{R\in\sigma_2})$ and then setting $v \in R^G$ with probability p_R for all $v \in V$ and all $R \in \sigma_1$ independently.

Let $\beta_R \in (0, 1)$ for each $R \in \sigma_1$, and let $\beta_R \in (0, \infty)$ for each $R \in \sigma_2$. We put $\widehat{G}_n(\{\beta_R\}_R)$ for a random sample of $G^{\widehat{\mathcal{C}}}(n, \{p_R\}_R)$ where each probability p_R satisfies $p_R(n) \sim \beta_R/n^{ar(R)-1}$ In this context, the following can be proven:

THEOREM 6.7
Let ϕ be a sentence in $FO[\sigma]$. Then the function
$F_\phi : (0, 1)^{|\sigma_1|} \times (0, \infty)^{|\sigma_2|} \to \mathbb{R}$ given by

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

is well defined and analytic.

To prove this theorem, one can proceed as in our proof of Theorem 2.15. Now vertices are ‘decorated’ by unary relations that have to be taken into account when defining the $\sim_k$ classes of trees and hypergraphs. With regards to the probabilistic part of the proof, in this setting the probability that a vertex satisfies a given unary relation is asymptotically constant, and hence the computations remain very similar.

Acknowledgements

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 6.2. 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)*, pp. 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*, pp. 620–627. IEEE, 1992.

[4] V. Chvátal and E. Szemerédi. Many hard examples for resolution. *Journal of the ACM*, **35**, 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. *Fundamenta Mathematicae*, **49**, 13, 1961.

[7] R. Fagin. Probabilities on finite models 1. *The Journal of Symbolic Logic*, **41**, 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**, 77–87, 1983.

[9] Y. V. Glebskii, D. I. Kogan, M. I. Liogon'kii and V. A. Talanov. Range and degree of realizability of formulas in the restricted predicate calculus. *Cybernetics and Systems Analysis*, **5**, 142–154, 1969.

[10] J. F. Lynch. Probabilities of sentences about very sparse random graphs. *Random Structures and Algorithms*, **3**, 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**, 97–115, 1988.

[13] S. Shelah and J. Spencer. Can you feel the double jump? *Random Structures and Algorithms*, **5**, 191–204, 1994.

[14] J. Spencer. *The Strange Logic of Random Graphs*, vol. 22. Springer Science & Business Media, 2013.

Received 11 June 2020