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 those limit probabilities are well-behaved with respect to some parameters that represent the density of tuples satisfying R for each relation R in the vocabulary σ. An application of these results to the problem of random Boolean satisfiability is presented afterwards. We show that in a random k-CNF formula over 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.

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 any property P expressible in L the probability that G_n satisfies P tends to some limit as n goes to ∞. 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 with regards to random binomial graphs $G(n,p)$. The seminal theorem on this topic, due to Fagin [1] and Glebskii et al. [2] independently, concerns general relational structures. When applied to graphs it states that if p is a fixed probability 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 [3]. There it is proven, among other results, that if $p := p(n)$ was a decreasing function on n of the form $n^{-\alpha}$, and $\alpha \in (0,1)$ 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 extended by Lynch in [4], where it is shown that in the case where the expected number of edges is linear, when $p(n) \sim \beta/n$ for some $\beta \in (0,\infty)$, then $G(n,p(n))$ satisfies a limit law with respect to FO logic. The following is a restatement of the main theorem in that article.

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

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

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 in this conditions no FO property can "capture" sudden changes in the structure of $G(n,p(n))$.

It was left open at the end of [4] whether the convergence law obeyed by $G(n,\beta/n)$ could be generalized for other random models of relational structures that contain relations of degree greater than 2. A result in this regard was obtained in [5], among other zero-one and convergence laws. There it is considered $G^d(n,p)$ the random model of d-uniform hypergraphs where each d-edge is added to a set of n labeled vertices with probability p independently. It is shown that when $p(n) \sim \beta/n^{d-1}$ -i.e., when the expected number of edges is linear- then $G^d(n,p(n))$ obeys a convergence law with respect of the FO language of d-uniform hypergraphs. With little additional work it can be shown that in this conditions the limit probability of any FO property of $G^d(n,p(n))$ varies analytically with β.

We aim to extend this last result to arbitrary relational structures to whose relations we can impose symmetry and anti-reflexivity constraints. This generalization is motivated by an application to the problem of random satisfiability. We continue the study started by Atserias in [6] with regards to the definability in first order logic of certificates for unsatisfiability that hold for typical unsatisfiable formulas. There it is considered a random model for 3-CNF formulas where each possible clause over n variables is added with probability p independently. This way the expected number of clauses is $\Theta(n^3p)$ as n grows. The main result of that article states that when the expected number of clauses is $\Theta(n^{2+\alpha})$ for any arbitrary $\alpha > 0$ then there exists some FO property of 3-CNF formulas that implies unsatisfiability and asymptotically almost surely (a.a.s) holds for unsatisfiable formulas. Furthermore, if the expected number of clauses is $\Theta(n^{2-\alpha})$ for any irrational $\alpha > 0$ then no FO property that implies unsatisfiability holds a.a.s for unsatisfiable formulas.

1 Preliminaries

1.1 General notation

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

Given 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 will use $\binom{S}{k}$ to denote the set of subsets of S whose size is k.Given a set S and a number $n \in \mathbb{N}$ with $n \leq |S|$ 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 will use the convention that over-lined variables , like $\overline{x}$, denote ordered tuples of arbitrary length. Given an ordered tuple $\overline{x}$ we define the number $len(\overline{x})$ 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 $f : X \to Y$ between two sets X,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.

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

The notations $\overline{x}$ and $(x_1,\dots,x_a)$ will be reserved to ordered tuples while $[\overline{x}]$ and $[x_1,\dots,x_a]$ will denote ordered tuples modulo the action of some arbitrary group of permutations. Which group is this will depend on the ambient set where $[x_1,\dots,x_a]$ belongs and it should either be clear from context or not be relevant.

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

1.2 Logical preliminaries

We assume familiarity with first order logic (FO). We follow the convention that first order logic contains the equality symbol. Given a vocabulary σ we will denote by $FO[\sigma]$ the set of first order formulas of vocabulary σ. We define the set of **free variables** of a formula as usual. Given a relation symbol $R \in \sigma$ we denote by $ar(R)$ the arity of R. Given a formula $\phi \in FO[\sigma]$ we will use the notation $\phi(\overline{y})$ to express that $\overline{y}$ is a tuple of (different) variables which contains all free variables in ϕ and none of its bounded variables, but it may contain variables which do not appear in ϕ. Formulas with no free variables are called **sentences** and formulas with no quantifiers are called **open formulas**.

1.3 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 that for all $R \in \sigma$, $P_R \subseteq \binom{[ar(R)]}{2}$

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

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

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

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

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

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

We will use the usual graph theory nomenclature and notation with some minor changes. In the scope of this article we will call **hypergraphs** to structures in $\mathscr{C}$. Given an hypergraph G we will call its **vertex set**, denoted by $V(G)$, to its universe.

In order to define the edge sets of G we need the following auxiliary definition

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

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

where

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

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

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

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

Clearly an 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.

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

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

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

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

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

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

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

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

That is, the **distance** between v and u is the minimum number of edges necessary to connect v and u. If such number does not exist we define $d^G(u,v) = \infty$. When G is understood or not relevant we will usually simply denote the distance by 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 an hypergraph is the maximum distance between any two of its 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 $ex(G) = -1$ and **dense** if $ex(G) > 0$. An 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 $ex(G) \geq 0$ is called **saturated** if for any non-empty proper sub-hypergraph $H \subset G$ it holds $ex(H) < ex(G)$. A connected hypergraph G with $ex(G) = 0$ is called **unicycle**. A saturated unicycle is called a **cycle**.

A **rooted tree** (T,v) is a tree T with a distinguished vertex $v \in V(T)$ called its **root**. We will usually omit the root when it is not relevant and write just T instead of (T,v). The **initial edges** of a rooted tree (T,v) are the edges in T that contain v. 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 an hypergraph and $\chi : V(H) \to \Sigma$ is a map called **Σ-coloring** of H.

Isomorphisms between hypergraphs are defined the same 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 an 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 will denote by $\mathrm{aut}(H)$, resp. $\mathrm{aut}(H,\chi)$ the number of such automorphisms.

Let H be an 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. As before, we define the set $Copies\left((H,\chi),\, V\right)$ 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 $ex(\mathbb{H})$, $\mathrm{aut}(\mathbb{H})$, $|V(\mathbb{H})|$, $|E(\mathbb{H})|$ or $Copies(\mathbb{H},V)$ via representatives of $\mathbb{H}$.

1.4 Ehrenfeucht-Fraisse Games

We assume familiarity with Ehrenfeucht-Fraisse (EF) games. An introduction to the subject can be found in [7, 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 $\mathrm{EHR}_k(H_1;H_2)$. The following is satisfied:

Theorem 1.1 (Ehrenfeut, 8)**.** Let H_1 and H_2 be hypergraphs. Then Duplicator wins $\mathrm{EHR}_k(H_1;H_2)$ if and only if H_1 and H_2 satisfy the same sentences $\phi \in FO[\sigma]$ with $qr(\phi) \leq k$.

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

We also define the k-round distance Ehrenfeucht-Fraisse game on H_1 and H_2, denoted by $d\mathrm{EHR}_k(H_1;H_2)$, the same way as $\mathrm{EHR}_k(H_1;H_2)$, 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 $d\mathrm{EHR}_k(H_1,\overline{v};H_2,\overline{u})$ analogously to $\mathrm{EHR}_k(H_1,\overline{v};H_2,\overline{u})$.

1.5 The random model

For each $R \in \sigma$ let p_R be a real number between zero and one. The random model $G^{\mathscr{C}}\left(n,\{p_R\}_{R\in\sigma}\right)$ 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's theorem, we are interested in the "sparse regime" of $G^{\mathscr{C}}(n,\overline{p})$, were the expected number of edges each sort is linear. This is achieved when each of the p_R's are of the form $\beta_R/n^{ar(R)-1}$ for some positive real numbers $\{\beta_R\}_{R\in\sigma}$. We will denote a random sample of $G^{\mathscr{C}}\left(n,\{p_R\}_{R\in\sigma}\right)$ by $G_n\left(\{\beta_R\}_{R\in\sigma}\right)$ when the probabilities p_R satisfy $p_R(n) \sim \beta_R/n^{ar(R)-1}$ for all $R \in \sigma$. When the choice of $\{\beta_R\}_{R\in\sigma}$ is not relevant we will write G_n instead of $G_n\left(\{\beta_R\}_{R\in\sigma}\right)$.

Our goal is to prove the following theorem:

Theorem 1.2. Let ϕ be a sentence in $FO[\sigma]$. Then the function $F_\phi : (0,\infty)^{|\sigma|} \to \mathbb{R}$ given by

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

is well defined and analytic.

1.6 Main definitions

Definition 1.2. Let H be a connected hypergraph. Then it holds that H contains a unique maximal saturated sub-hypergraph H' and it satisfies $ex(H') = ex(H)$ if $ex(H) \geq 0$ and $H' = \emptyset$ if H is a tree. 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 1.3. Let H be an 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 1.4. Let H be an hypergraph, let $\overline{v} \in V(H)^*$ and let $v \in H$ be such that $d(Center(H, \overline{v}), v) < \infty$. Let $X \subset V(H)$ be the set

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

Then we define $\mathrm{Tr}\left(H, \overline{v};\, v\right)$ as the tree $H[X]$ to which we assign v as a root. That is, $\mathrm{Tr}\left(H, \overline{v};\, v\right)$ 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 was not then it would contain some saturated sub-hypergraph, leading to 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 will write simply $\mathrm{Tr}(H;\, v)$ or $\mathrm{Tr}(H;\, v;\, r)$.

Definition 1.5. Fix a natural number k. We define the ***k*-equivalence** relation over rooted trees, written as $\sim_k$, by induction over their radii as follows:

- Any two trees with radius zero are k-equivalent. Notice that those trees consist only of one vertex: their respective roots.

- Let $r > 0$. Suppose the k-equivalence relation has been defined for rooted trees with radius at most $r-1$. Let $\Sigma_{k,r-1}$ be the set consisting of the $\sim_k$ classes of trees with radius at most $r-1$. Let ρ be an special symbol called the **root symbol**. Set $\widehat{\Sigma}_{k,r-1} := \Sigma_{k,r-1} \cup \{\rho\}$. Then we call a (k,r)-**pattern** to an isomorphism class of $\widehat{\Sigma}_{k,r-1}$-hypergraphs (e, τ) that consist of only one loop-free edge and no isolated vertices, and satisfy $\tau(v) = \rho$ for exactly one vertex $v \in V(e)$. We will denote by $P(k,r)$ the set of (k,r)-patterns.

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

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

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

Observation 1.1. 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 $a_\varepsilon < k$ for each $\varepsilon \in E^2_{\mathbf{T}}$ that only depend on

$\mathbf{T}$ such that any rooted tree (T,v) belongs to $\mathbf{T}$ if and only if the following hold: (1) For any pattern $\varepsilon \in E^1_{\mathbf{T}}$ there are at least k initial edges $e \in E(T)$ such that $(e, \tau_{(T,v)}) \in \varepsilon$, and (2) for any pattern $\varepsilon \in E^2_{\mathbf{T}}$ there are exactly a_ε initial edges $e \in E(T)$ such that $(e, \tau_{(T,v)}) \in \varepsilon$.

Observation 1.2. Using last characterization of $\sim_k$ classes it is easy to show that for any $r \in \mathbb{N}$ the quantity of $\sim_k$ classes of trees with radii at most r is finite. We proceed by induction. For $r = 0$ there is only one $\sim_k$ class. Now let $r > 0$ and suppose the statement holds for $r-1$. Then the number of (k,r)-patterns is finite and so is the number of $\sim_k$ classes of trees with radii at most r.

Definition 1.6. Given H be a non-tree connected hypergraph, we define its **canonical coloring** τ_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 which are not trees. Set $H_1' := Center(H_1)$ and $H_2' := Center(H_2)$. We say that H_1 and H_2 are k-equivalent, written as $H_1 \sim_k H_2$, if $(H_1', \tau_{H_1}) \simeq (H_2', \tau_{H_2})$

Definition 1.7. Let H_1 and H_2 be hypergraphs and let $r \in \mathbb{N}$. Let $H_1' := Core(H_1; r)$ and $H_2' := Core(H_2; r)$. We say that H_1 and H_2 are (k,r)-agreeable, written as $H_1 \approx_{k,r} H_2$ if for any $\sim_k$ class $\mathscr{H}$ "the number of connected components in H_1' that belong to $\mathscr{H}$" and "the number of connected components in H_2' that belong to $\mathscr{H}$" are the same or are both greater than $k-1$.

Definition 1.8. Let $\Sigma_{(k,r-1)}$ be the set of $\sim_k$ classes of rooted trees with radii at most $r-1$. Then we call a (k,r)-**cycle** to an isomorphism class of $\Sigma_{(k,r-1)}$-hypergraphs (H,τ) that are cycles of diameter at most $2r+1$. We will denote by $C(k,r)$ the set of (k,r)-cycles.

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

Definition 1.9. Let H be an hypergraph and let $r \in \mathbb{N}$. Let $X \subset V(H)$ be the set of vertices in H belonging to some saturated sub-hypergraph of diameter at most $2r+1$. We say that H is (k,r)-**rich** if for any $r' \leq r$, any vertices $v_1, \dots, v_k$ and any $\sim_k$ class $\mathbf{T}$ of trees with radius at most r' it holds that there exists a vertex $v \in V(H)$ such that $d(v,X) > 2r'+1$, $d(v,v_i) > 2r'+1$ for all v_i's and $T := N(v;r')$ is a tree satisfying $(T,v) \in \mathbf{T}$.

1.7 Outline of the proof

We show now an outline of the proof of theorem 1.2.

The arguments are similar to the ones in the proof of [4, Theorem 2.1], adapted to fit our context. As in that article the proof is divided in two parts: a model theoretic part and a probabilistic part. The main result of the first part is the following

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

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 [9]: A.a.s for any fixed radius r all neighborhoods $N(v;r)$ in G_n are trees or unicycles, so cycles in G_n are far apart. One can find arbitrarily many copies of any fixed tree, while the expected number of copies of any fixed cycle is finite. The main probabilistic results are the following:

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

Theorem 3.5. *Let $k,r \in \mathbb{N}$. Then a.a.s G_n is (k,r)-rich.*

Theorem 3.6. *Let $k,r \in \mathbb{N}$. Let $\mathscr{O}$ be a $\approx_{k,r}$ class of r-simple hypergraphs. Then*

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

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

An sketch of the proof of theorem 1.2 using these results is the following. Let $\Phi \in FO[\sigma]$ be a sentence and let $k := qr(\Phi)$, $r := (3^k-1)/2$. Because of theorem 2.4 and theorem 3.5 it holds that for any $\approx_{k,r}$ class $\mathscr{O}$

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

This together with theorem 3.2 and the fact that there are a finite number of $\approx_{k,r}$-classes of r-simple hypergraphs imply that $\lim_{n\to\infty} \Pr(G_n \models \Phi)$ equals a finite sum of limits of the form $\lim_{n\to\infty} \Pr(G_n \in \mathscr{O})$ where $\mathscr{O}$ is some $\approx_{k,r}$-class of r-simple hypergraphs. Finally, using theorem 3.6 we get that $\lim_{n\to\infty} \Pr(G_n \models \Phi)$ exists and is an analytic expression in $\{\beta_R\}_{R\in\sigma}$, as we wanted.

2 Model theoretic results

2.1 Winning strategies for Duplicator

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

Definition 2.1. Let $\overline{v} \in V_1^*, \overline{u} \in V_2^*$ be tuples of the same length. We write $(H_1,\overline{v}) \simeq_{k,r} (H_2,\overline{u})$, if Duplicator wins $d\mathrm{EHR}_k\left(N(\overline{v};r),\overline{v}; N(\overline{u};r),\overline{u}\right)$. 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 2.2. Fix $r \in \mathbb{N}$. Suppose $X \subseteq V_1$ and $Y \subseteq V_2$ can be partitioned into sets $X = X_1 \cup \cdots \cup X_a$ and $Y = Y_1 \cup \cdots \cup Y_b$ such that $N(X_i;r)$'s, and the $N(Y_i;r)$'s, are connected and disjoint. We 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's or the Y_i's it is satisfied that "the number of X_i's such that $(H_\delta,Z) \simeq_{k,r} (H_1,X_i)$" and "the number of Y_i's such that $(H_\delta,Z) \simeq_{k,r} (H_2,Y_i)$" are both equal or are both greater than $k-1$.

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

Theorem 2.1. Set $r = (3^k-1)/2$. Suppose there exist sets $X \subseteq V_1$, $Y \subseteq V_2$ with the following properties:

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

(2) • Let $r' \leq r$. Let $v \in V_1$ be a vertex such that $d(X,v) > 2r'+1$. Let $\overline{u} \in (V_2)^{k-1}$ 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 2.1. 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 2.2. 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 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 2.1. Let $\overline{v} \in V_1^*$ and $\overline{u} \in V_2^*$ be of equal length. Let $r \in \mathbb{N}$ be greater than zero. Suppose $(H_1,\overline{v}) \simeq_{k,3r+1} (H_2,\overline{u})$. Let $v \in V_1$ and $u \in V_2$ be vertices played in the first round of an instance of

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

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

Proof. Using observation 2.2 we get that Duplicator wins

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

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

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

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 $d(z,\overline{v}) \leq d(z,v') + d(v',\overline{v}) = 3r+1$. In consequence, $N^{H_1}(v;r) \subset H_1'$. Thus, $N^{H_1}(\overline{v_2};r) \subseteq H_1'$, and $N^{H_1}(\overline{v_2};r) = N^{H_1'}(\overline{v_2};r)$. Analogously we obtain $N^{H_2}(\overline{u_2};r) = N^{H_2'}(\overline{u_2};r)$, as we wanted. □

Proof of theorem 2.1. Let $X_1,\ldots,X_a$ and $Y_1,\ldots,Y_b$ be partitions of X and Y respectively as in the definition of $\cong_{k,r}$. Define $r_0 = (3^k-1)/2$ and $r_i = (r_{i-1}-1)/3$ for each $1 \leq i \leq k$. 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 $\text{EHR}_k(H_1,H_2)$. We show a winning strategy for Duplicator in $\text{EHR}_k(H_1;H_2)$. For each $0 \leq i \leq k$, Duplicator will keep track of some marked sets of vertices $T \subset V_1$, $S \subset V_2$. For $\delta = 1,2$ each marked set $T \subset V_\delta$ will have associated a tuple of vertices $\overline{v}(T) \in V_\delta^*$ 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 a way such that at the end round the following are satisfied:

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

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

(iii)
   - Let $S \subset V_1$ be a marked set. Then for any different marked $S' \subset V_1$ of any different S' among $X_1,\dots,X_a$ 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

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

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

The game $d\mathrm{EHR}_k(H_1; H_2)$ proceeds as follows. Clearly properties (i) to (iv) hold at the beginning of the game. Suppose 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 $S \subset V_1$ it holds that $d(S \cup \overline{v}, \quad v_i^1) \leq 2r_i + 1$. In this case let $T \subset V_2$ be the set in H_2 marked in the same round as T. By hypothesis

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

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

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

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

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

- 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 it holds that 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 2.1 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 $d(S \cup \overline{v}(S), \quad v_i^1) > 2r_i + 1$, and for all sets S among $X_1, \dots, X_a$ it also holds $d(S, v_i^1) > 2r_i + 1$. In this case from condition (2) of the statement follows that Duplicator can choose $v_i^2 \in V_2$ such that (A) $d(T \cup \overline{v}(T), \quad v_i^2) > 2r_i + 1$ for all marked sets $T \subset V_2$, (B) $d(T, v_i^2) > 2r_i + 1$ for all sets T among $Y_1, \dots, Y_b$, and (C) $(H_1, v_i^1) \simeq_{k-i, r_i} (H_2, v_i^2)$. After this Duplicator marks both $S = \{v_i^1\}$ and $T = \{v_i^2\}$ and sets $\overline{v}(S) := v_i^1$, and $\overline{v}(T) := v_i^2$.

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

□

2.2 k-Equivalence relation

2.2.1 k-Equivalent trees

We want prove the following.

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

Before proceeding with the proof we need an auxiliary result. Let (T, v) be a rooted tree and e an initial edge of T. We define $\text{Tr}(T, v;\ e)$ as the induced tree $T[X]$ on the set $X := \{v\} \cup \{\, u \in V(T) \,|\, d(v, u) = 1 + d(e, u) \,\}$, to which we assign v as the root. In other words, $\text{Tr}\,(T, v;\ e)$ is the tree formed of v and all the vertices in T whose only path to v contain e.

Lemma 2.2. Fix $r > 0$. Suppose theorem 2.2 holds for rooted trees with radii at most r. Let (T_1, v_1) and (T_2, v_2) be rooted trees with radius $r + 1$. Let $\tau_{(T_1, v_1)}$ and $\tau_{(T_2, v_2)}$ be colorings over T_1 and T_2 as in the definition of k-equivalence. Let e_1 and e_2 be initial edges of T_1 and T_2 respectively satisfying $(e_1, \tau_{(T_1, v_1)}) \simeq (e_2, \tau_{(T_2, v_2)})$. Name $T_1' := \text{Tr}(T_1, v_1;\ e_1)$ 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 $f : V(e_1) \to V(e_2)$ an isomorphism between $(e_1, \tau_{(T_1, v_1)})$ and $(e_2, \tau_{(T_2, v_2)})$. Suppose 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 $\mathrm{Tr}(T_1', v_1;\, u)$ for a unique $u \in V(e_1)$ different from the root v_1. Set $T_1'' := \mathrm{Tr}(T_1', v_1;\, u)$ and $T_2'' := \mathrm{Tr}(T_2', v_2;\, f(u))$ Then, as $\tau_{(T_1,v_1)}(u) = \tau_{(T_2,v_2)}(f(u))$, we obtain $(T_1'', u) \sim_k (T_2'', f(u))$. As both these trees have radii at most r, by assumption Duplicator has a winning strategy in $d\mathrm{EHR}_k(T_1'', u;\quad T_2'', f(u))$ and they can follow it considering the previous plays in T_1'' and T_2''. □

Proof of theorem 2.2.

Notice that, as $(T_1, v_1) \sim_k (T_2, v_2)$, both T_1 and T_2 have the same radius r. We prove the result by induction on r. If $r = 0$ then both T_1 and T_2 consist of only one vertex and we are done. Now let $r > 0$ and assume that the statement is true for all lesser values of r. Let $\tau_{(T_1,v_1)}$ and $\tau_{(T_2,v_2)}$ be the colorings over T_1 and T_2 as in the definition of $\sim_k$. We show that there is a winning strategy for Duplicator in $d\mathrm{EHR}_k(T_1, v_1;\, T_2, v_2)$. 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 $\mathrm{Tr}(T_1, v_1;\, e_1)$ 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 $\left(e_1, \tau_{(T_1,v_1)}\right) \simeq \left(e_2, \tau_{(T_2,v_2)}\right)$. Mark e_2 as well. Set $T_1' := \mathrm{Tr}(T_1, v_1;\, e_1)$ and $T_2' := \mathrm{Tr}(T_2, v_2;\, e_2)$ Because of lemma 2.2, Duplicator has a winning strategy in $d\mathrm{EHR}k(T_1', v_1;\, T_2', v_2)$ 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 $\left(e_1, \tau_{(T_1,v_1)}\right) \simeq \left(e_2, \tau_{(T_2,v_2)}\right)$. Again, because of lemma 2.2, Duplicator has a winning strategy in $d\mathrm{EHR}k(T_1', v_1;\, T_2', v_2)$ and can continue playing according to it taking into account the plays made previously in T_1' and T_2'.

□

2.2.2 k-Equivalent hypergraphs

Theorem 2.3. Let H_1 and H_2 be non-tree connected hypergraphs satisfying $H_1 \sim_k H_2$. Set $H_1' := Center(H_1)$ and $H_2' := Center(H_2)$. Let f be an isomorphism between (H_1', τ_{H_1}) and (H_2', τ_{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 $d\mathrm{EHR}_k(H_1', \overline{v};\, H_2', \overline{u})$.

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 $\mathrm{Tr}(H_1;\, u)$ for a unique $u \in H_1'$. Set $T_1 := \mathrm{Tr}(H_1;\, u)$ and $T_2 := \mathrm{Tr}(H_2;\, f(u))$. By hypothesis $(T_1, u) \sim_k (T_2, f(u))$. Then because of theorem 2.2 we have that Duplicator has a winning strategy in $d\mathrm{EHR}_k(T_1, u;\, T_2, f(u))$, and they can follow it taking into account the previous plays made in T_1 and T_2, if any. In particular, if Spoiler has chosen u then Duplicator will necessarily choose $f(u)$. One can easily check that distances are preserved following this strategy. □

2.3 r-Cores

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

Proof. Let $X_1, \ldots, X_a$ 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. $Core(H_2; r)$. Using the definition of $H_1 \approx_{k,r} H_2$ as well as the fact that because of theorem 2.3 $N(X_i; r) \sim_k N(Y_j; r)$ implies $(H_1, X_i) \simeq_{k,r} (H_2, Y_j)$ the result follows. $\square$

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

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

3 Probabilistic results

3.1 Almost all hypergraphs are simple

Lemma 3.1. Let H be an hypergraph, and let X_n be the random variable that counts the copies of H in G_n. Then $\mathrm{E}\left[X_n\right] = \Theta(n^{-ex(H)})$.

Proof. It holds

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

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

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

Substituting in the first equation we get

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

$\square$

Lemma 3.2. Let H be an hypergraph such that $ex(H) > 0$. Then a.a.s there are no copies of H in G_n.

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

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

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

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

□

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

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

Proof. Suppose there is some edge $e \in E(G) \setminus E(H)$ with and $ex(e) \geq 0$. Let P be a path of length at most r joining H and e in G. Then $H' := H \cup P \cup e$ satisfies the conditions of the statement. Otherwise, all edges $e \in E(G) \setminus E(H)$ satisfy $ex(e) = -1$. In this case we successively cut edges e from G such that $d(e,H)$ is the maximum possible (notice that this always yields a connected hypergraph) until we obtain an hypergraph G' with $ex(G') < ex(G)$. Let e be the edge that was cut last. Then $V(G') \cap V(e) = ex(G) - ex(G') + 1 \geq 2$. Let $v_1, v_2 \in V(G') \cap V(e)$, and let P_1, P_2 be paths of length at most r that join H with v_1 and v_2 respectively in G'. Then the hypergraph $H' := H \cup e \cup P^1 \cup P^2$ satisfies the conditions in the statement. □

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

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

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

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

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

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

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

As a corollary we obtain the needed result.

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

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

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

Proof. An application of the first moment method together with lemma 3.3 and the fact that there are a finite number of $\simeq$ classes of paths whose length is at most $2r+1$ implies that a.a.s the $N(v;r)$'s are disjoint. Also, because of theorem 3.1 a.a.s the $N(v;r)$'s are either trees or unicycles. But if any of the $N(v;r)$'s was an unicycle then that would mean that in G_n there exists a path P of length at most $2r+1$ joining some vertex $v \in \overline{v}$ with a cycle C of diameter at most $2r+1$. Using lemma 3.3 again as well as the fact that there are 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 $N(v;r)$'s are disjoint trees as we wanted to prove. $\square$

Lemma 3.7. Let $\overline{v} \subset \mathbb{N}*$ be a finite set of fixed vertices and let $\pi(\overline{x})$ be an edge sentence such that $len(\overline{x}) = len(\overline{v})$. Define $G'_n = G_n \setminus E[\overline{v}]$ (i.e. G_n minus all the edges induced over $\overline{v}$). Fix $R \in \mathbb{N}$. Let A_n be the event that G'_n contains a path of length at most R between any two vertices $u, w \in \overline{v}$. Let B_n be the event that G'_n contains a cycle of diameter at most R at distance at most R to some vertex $u \in \overline{v}$. Then

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

Proof. Notice that the events A_n and B_n do not concern the possible edges induced over $\overline{v}$. In consequence, because edges are independent in our random model, $\Pr(A_n \,|\, \pi(\overline{v})) = \Pr(A_n)$ and $\Pr(B_n \,|\, \pi(\overline{v})) = \Pr(B_n)$. Now both limits from the statement follow from lemma 3.6 using that $G'_n \subset G_n$. $\square$

3.2 Convergence to Poisson variables

Our main tool for computing probabilities in the following sections will be the next result, which can be found in [11, Theorem 1.23].

Theorem 3.3. Fix $k \in \mathbb{N}$. 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 $r_1, \ldots, r_l \in \mathbb{N}$

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

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

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

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

3.3 Probabilities of trees

Definition 3.1. 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 3.2. 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

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

Notice 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 $\{\beta_R\}_{R\in\sigma}$, belongs to Λ for any choice of r and $\mathbf{T}$.

Theorem 3.4. Fix $r\in\mathbb{N}$. The following are satisfied:

(1) Let $\mathbf{T}$ be a k-equivalence class for trees with radii at most r. Then $\Pr[r,\mathbf{T}]$ exists, is positive for all choices of $\{\beta_R\}_{R\in\varepsilon}\in(0,\infty)^{|\sigma|}$, 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 sentence such that $len(\overline{x}) = len(\overline{u})$. Let $\overline{v}\in(\mathbb{N})_*$ be vertices contained in $\overline{u}$. For each $v\in\overline{v}$ let $\mathbf{T}_v$ be a k-equivalence class of trees with radii at most r. Then

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

We will devote the rest of this section to proving this theorem. The proof will be done 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 $r=0$. In consequence we obtain the following.

Lemma 3.8. Conditions (1) and (2) of theorem 3.4 are satisfied for $r=0$.

Definition 3.3. Let $r \in \mathbb{N}$ be a positive number and suppose that theorem 3.4 holds for $r-1$. Given a (k,r)-pattern ε we define the expressions $\lambda_{r,\varepsilon}$ and $\mu_{r,\varepsilon}$ in the following way. 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

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

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

Lemma 3.9. Let $r > 0$ and $\overline{u} \in (\mathbb{N})_*$. Let $\pi(\overline{x}) \in FO[\sigma]$ be a consistent edge sentence such that $len(\overline{x}) = len(\overline{u})$. Let $\overline{v} \in (\mathbb{N})_*$ be vertices contained in $\overline{u}$. For each $v \in \overline{v}$ set $T_{n,v} := \operatorname{Tr}(G_n, \overline{u};\, v;\, r)$. Given a pattern $\varepsilon \in P(k,r)$ and $v \in \overline{v}$ we define the random variable $X_{n,v,\varepsilon}$ as the number of initial edges $e \in E(T_{n,v})$ such that $(e, \tau_{(T_{n,v},v)}) \in \varepsilon$. Suppose that theorem 3.4 holds for $r-1$. Then the conditional distributions of the variables $X_{n,v,\varepsilon}$'s given $\pi(\overline{u})$ converge to independent Poisson distributions whose respective mean values are given by the $\mu_{r,\varepsilon}$'s.

Proof. To avoid excessively complex notation we will show only the case where $\overline{v}$ consists only of one vertex v. The general case is proven using the same arguments. Set $T_n := T_{n,v}$ and $X_{n,\varepsilon} := X_{n,v,\varepsilon}$ for all $\varepsilon \in P(k,r)$. By theorem 3.3, in order to prove the result it is enough to show that for any choice of natural numbers $\{b_\varepsilon\}_{\varepsilon\in P(k,r)}$ it holds

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

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

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

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 3.1 we obtain

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

We say that a choice $\{E_\varepsilon\}_\varepsilon \in \Omega_n$ is **disjoint** if the edges $(e,\tau) \in \bigcup_{\varepsilon\in P(k,r)} E_\varepsilon$ satisfy that no vertex $w \in \overline{u}$ other than v belong 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 $\{E_\varepsilon\}_\varepsilon \in \Omega_n$ it holds that $e \in E(T_n)$ for all $(e,\tau) \in \bigcup_{\varepsilon\in P(k,r)} E_\varepsilon$ then $\{E_\varepsilon\}_\varepsilon$ 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 last sum it suffices to consider only the disjoint $\{E_\varepsilon\}_\varepsilon$. Because of the symmetry of the random model the probability in that sum is the same for all disjoint choices of $\{E_\varepsilon\}_\varepsilon$. In consequence, if we fix $\{E_\varepsilon\}_\varepsilon \in \Omega'_{\mathbb{N}}$ we obtain

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

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

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

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

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

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

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

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

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

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

The trees $Tr(T_n;u)$ in last probability coincide with $Tr(G_n,\overline{u}^\frown\overline{w};u;r-1)$ for all u's. In consequence, using the hypothesis that theorem 3.4 holds for $r-1$ we obtain

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

Joining this with eq. (2), eq. (3) and eq. (4) we obtain

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

This proves eq. (11) and the statement. □

Next lemma completes the proof of theorem 3.4.

Lemma 3.10. Let $r\in\mathbb{N}$, $r>0$. Suppose that theorem 3.4 holds for $r-1$. Then it also holds for r.

Proof. We start showing condition (1) of theorem 3.4. 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 $\varepsilon \in P(k,r)$ let $X_{n,\varepsilon}$ be the random variable that counts the number of initial edges in T_n whose pattern is ε. Let $E^1_{\mathbf{T}}, E^2_{\mathbf{T}}, \{a_\varepsilon\}_\varepsilon$ be as in observation 1.1. Then

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

And using the previous lemma we obtain that last limit equals the following expression.

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

Using the definition of the $\mu_{r,\varepsilon}$'s we obtain that last expression belongs to Λ as we wanted to prove. Furthermore, as the $\mu_{r,\varepsilon}$'s are positive, this expression is also positive for all values of $\{\beta_R\}_{R\in\sigma} \in (0,\infty)^{|\sigma|}$. 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 $\pi(\overline{u})$. In consequence the desired result follows from condition (1). $\square$

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

Theorem 3.5. Let $r \in \mathbb{N}$. 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 $\overline{v}(\mathbf{T})$'s are disjoint. Let $\overline{w} \in (\mathbb{N})_*$ be a concatenation of all the $\overline{v}(\mathbf{T})$'s. 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 3.4 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 $\mathbf{T} \in \Sigma$, it is satisfied that

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

Fix $\delta > 0$ such that $\delta < \Pr[r,\mathbf{T}]$ for all $\mathbf{T} \in \Sigma$ and fix $\varepsilon > 0$ arbitrarily small. Because of the Law Of Large Numbers if m is big enough it holds that

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

Also, for m big enough we have that

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

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

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

We define the event 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 $N(v;r)$'s 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 3.6 a.a.s A_n holds, and we obtain

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

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 it was to be proved. □

3.5 Probabilities of cycles

Definition 3.4. 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 $R \in \sigma$, any positive number $b \in \mathbb{N}$ and any $\lambda \in \Lambda$ the expression $\frac{\lambda}{b} \cdot \prod_{R\in\sigma} \beta_R^{a_R}$ belongs to Γ, (2) given any $\gamma \in \Gamma$ and any $a \in \mathbb{N}$, 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 $\upsilon_1 \cdot \upsilon_2 \in \Upsilon$ as well.

Definition 3.5. Let $r \in \mathbb{N}$ 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:

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

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 3.11. Let $r \in \mathbb{N}$. For any $O \in C(k,r)$ let $X_{n,O}$ be the random variable that counts the connected components H of $Core(G_n;r)$ such that $H' := Center(H)$ satisfies that $(H',\tau_H) \in O$. Then the variables $X_{n,O}$'s converge in distribution to independent Poisson variables whose respective mean values are given by the $\gamma_{r,O}$'s.

Proof. The proof is similar to the one of lemma 3.9. By theorem 3.3 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

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

Given a cycle H such that $V(H) \subseteq [n]$ we say that $H \sqsubset G_n$ if $H = Center(H')$ for some connected component H' of $Core(G_n;r)$. Using observation observation 3.1 we obtain

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

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 $Core(G_n;r)$ and thus $H_1 \cup H_2 \subset Center(H)$. In consequence neither $H_1 \sqsubset G_n$ or $H_2 \sqsubset G_n$ 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

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

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

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

Let $\bar{v} \in (\mathbb{N})_*$ 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 sentence concerning the vertices in $\bar{v}$. Let $\varphi(\bar{x})$ be one of such sentences. We have that

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

Because of theorem 3.2 a.a.s if some cycle H of diameter at most $2r+1$ satisfies $H \subset G_n$ then $H \sqsubset G_n$. In consequence:

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

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

As all the vertices $v \in \bar{v}$ belong to $Core(G_n;r)$, the trees $Tr(G_n;v;r)$ in last probability coincide with $Tr(G_n,\bar{v};v;r)$. Hence, by theorem 3.4 we have that

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

Joining this with eq. (8), eq. (9) and eq. (10) we obtain

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

This proves eq. (7) and the statement. □

Theorem 3.6. Let $r\in\mathbb{N}$ and let $\mathscr{O}$ be a simple (k,r)-agreeability class of hypergraphs. Then $\lim_{n\to\infty}\Pr(G_n\in\mathscr{O})$ 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 $U^1_{\mathscr{O}},U^2_{\mathscr{O}}$ and $\{a_O\}_{O\in U^2_{\mathscr{O}}}$ be as in observation 1.3. Let A_n be the event that G_n is r-simple. Then

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

Because of theorem 3.2, a.a.s A_n holds. Thus, using last lemma the previous limit equals the following expression

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

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

4 Proof of the main theorem

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

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

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

Proof. Let k be the quantifier rank of ϕ and let $r=3^k$. Let $G_n:=G_n\left(\{\beta_R\}_{R\in\sigma}\right)$. Let Σ be the set of (k,r)-agreeability classes of r-simple hypergraphs. Because of theorem 3.2 a.a.s G_n is r-simple. In consequence

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

Because the set Σ is finite, sum and we can exchange last summation and limit symbols. Because theorem 3.5 a.a.s G_n is (k,r)-rich. This together with theorem 2.4 implies that for $\mathscr{O}\in\Sigma$

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

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

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

Because of theorem 3.6 we know that each of the limits inside last sum exists and is given by an expression that belongs to Υ. As a consequence the theorem follows. □

5 Application to random SAT

We define a binomial model of random CNF formulas, in analogy with the one in [12], but the generality in theorem 1.2 allows for many modifications.

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

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

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

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

We consider l-CNF formulas, as defined above, as relational structures with a language σ consisting of $l+1$ relation symbols $R_0, \ldots, R_l$ of arity l. We do that in such a way that the expression $R_j(x_{i_1}, \ldots, x_{i_l})$ means that our formula contains the clause consisting of $\neg x_{i_1}, \ldots, \neg x_{i_j}$ and $x_{i_{j+1}}, \ldots x_{i_l}$. The relations $R_1, \ldots, R_l$ satisfy the following axioms: (1) for any $0 \le j \le l$ and for any variables $y_1, \ldots, y_l$ the fact that $R_j(y_1, \ldots, y_l)$ holds is invariant under any permutation of the variables $y_1, \ldots, y_j$ or $y_{j+1}, \ldots, y_l$, and (2) f or any $0 \le j \le l$ and any variables $y_1, \ldots, y_l$ it holds that $R_j(y_1, \ldots, y_l)$ only if all the y_i's are different. Call $\mathscr{C}$ to the family of σ-structures satisfying these last two axioms. The language σ and the family $\mathscr{C}$ satisfy the conditions in section 1.3. The random model $F_l(n,p)$ coincides with the model $G(n, \{p_R\}_R)$ of random $\mathscr{C}$-hypergraphs described in section 1.5 when all the p_R's are equal. As a particular case of theorem 1.2 we obtain the following result.

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

$$\beta \mapsto \lim_{n \to \infty} \Pr(F_n(\beta) \models \Phi)$$

is well defined and analytic.

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

Theorem 5.2. Let $l \ge 2$ be a natural number, and let $c \in (0, \infty)$ be an arbitrary real number. Let $m : \mathbb{N} \to \mathbb{N}$ be a map such that $m(n) = (c + o(1))n$. For each n let $C_{n,1}, \ldots, C_{n,m(n)}$ be clauses chosen uniformly at random independently among the $2^l \binom{n}{l}$ ordinary clauses of size l over the variables $x_1, \ldots, x_n$. For each n, let $UNSAT_n$ denote the event that there is no assignment of the variables $x_1, \ldots, x_n$ that satisfies all clauses $C_{n,1}, \ldots, C_{n,m(n)}$. Then there are two real constants $0 < c_1 < c_2$, independent of 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 [12, Theorem 1]. The fact that c_2 exists follows from a direct application of the first order method and is also shown in [12], as well as [13], [14], [15] and possibly others. 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 5.1. Let $l \geq 2$ be a natural number. Let $c \in (0,\infty)$ be an arbitrary real number and let $m : \mathbb{N} \to \mathbb{N}$ satisfy $m(n) = (c+o(1))n$. For each $n \in \mathbb{N}$ let $F_{n,m(n)}$ be a random formula chosen uniformly at random among all the sets of $m(n)$ ordinary clauses of size l over the variables $x_1,\dots,x_n$. Then there are two real positive constants $0 < c_1 < c_2$ such that 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. In consequence

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

An application of the first order method yields that for $l > 3$ a.a.s the number of unordered pairs $\{i,j\}$ such that $C_{n,i} = C_{n,j}$ is zero. For the case of $l = 2$ an application of the factorial moments method proves that the number of such pairs $\{i,j\}$ converges in distribution to a Poisson variable. In either case all the $C_{n,i}$'s are different with positive asymptotic probability. In consequence the constants c_1 and c_2 from the previous theorem satisfy our statement. $\square$

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

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

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

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 it holds

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

Notice that the property of being unsatisfiable is monotonous. In consequence,

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

Because of the Law of Large Numbers it holds

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

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

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

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

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

Theorem 5.4. Let $l > 1$ be a natural number. For each $n \in \mathbb{N}$ let $F_n(\beta)$ be a random formula from $F(l,n,\beta/n^{l-1})$. Let $\Phi \in FO[\sigma]$ be a first order sentence that implies unsatisfiability. Then for all $\beta \in (0,\infty)$ a.a.s $F_n(\beta)$ does not satisfy Φ.

Proof. Let β_1 and β_2 be as in theorem 5.3. As Φ implies unsatisfiability it holds $\Pr(F_n(\beta) \models \Phi) \leq \Pr(F_n(\beta) \text{ is unsatisfiable })$. Thus, using theorem 5.3 we get that for all $\beta \in (0,\beta_1]$

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

Because theorem 5.1 last limit varies analytically with β, so if it vanishes in a proper interval $(0,\beta_1]$ then it has to vanish in the whole $(0,\infty)$ by the Principle of Analytic Continuation, and the result holds. □

Conclusions

Acknowledgments

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

Albert and Marc Noy

References

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

[2] Yu V Glebskii, Do I Kogan, MI Liogon'kiI, and VA Talanov. Range and degree of realizability of formulas in the restricted predicate calculus. *Cybernetics and Systems Analysis*, 5(2):142–154, 1969.

[3] Saharon Shelah and Joel Spencer. Zero-one laws for sparse random graphs. *Journal of the American Mathematical Society*, 1(1):97–115, 1988.

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

[5] Nicolau C Saldanha and Márcio Telles. Spaces of completions of elementary theories and convergence laws for random hypergraphs. *arXiv preprint arXiv:1602.06537*, 2016.

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

[7] Heinz-Dieter Ebbinghaus and Jörg Flum. *Finite model theory*. Springer Science & Business Media, 2005.

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

[9] Saharon Shelah and Joel Spencer. Can you feel the double jump? *Random Structures & Algorithms*, 5(1):191–204, 1994.

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

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

[12] Vašek Chvátal and Bruce Reed. Mick gets some (the odds are on his side)(satisfiability). In *Proceedings., 33rd Annual Symposium on Foundations of Computer Science*, pages 620–627. IEEE, 1992.

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

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

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