Abstract

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 in the sparse case. It is also shown that those limit probabilities are well-behaved with respect to some parameters that represent edge densities. An application of these results to the problem of random Boolean satisfiability is presented afterwards. We show that there is no first order property of k-CNF formulas that implies unsatisfiability and holds for almost all typical unsatisfiable formulas when the number of clauses is linear.

Introduction

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

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

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

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

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

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

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

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

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

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

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

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

1 Preliminaries

1.1 General notation

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

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

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

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

Given two real functions over the natural numbers $f,g : \mathbb{N} \to \mathbb{R}$ 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. If $g(n) \neq 0$ for sufficiently large values of n then we will write $f \simeq g$ if $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$.

1.2 Logical preliminaries

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

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

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

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

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

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

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

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

1.3 Structures as multi-hypergraphs

For the rest of the article consider fixed:

- Positive natural numbers t, $\bar{a} = a_1, \dots, a_t$.
- A relational vocabulary $\sigma = \{\bar{R}\}$, with $\bar{R} = R_1, \dots, R_t$ such that a_i is the arity of R_i.
- Groups $\overline{\Phi} = \Phi_1, \dots, \Phi_t$ such that each Φ_i is consists of permutations on $[a_i]$ with the usual composition as its operation.
- Sets $\bar{P} = P_1, \dots, P_t$ satisfying $P_i \subseteq \binom{[a_i]}{2}$

We define the class $\mathscr{C}^{\sigma}_{\overline{\Phi}, \bar{P}}$ as the class of σ-structures that satisfy the following axioms:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

1.4 Ehrenfeucht-Fraisse Games

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

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

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

The following is satisfied:

Theorem 1.1 (Ehrenfeut, 9)**.** *Let G_1 and G_2 be hypergraphs. Then Duplicator wins $\textsc{Ehr}_k(G_1,G_2)$ if and only if $G_1 =_k G_2$.*

Now consider $\bar{v}$, and $\bar{u}$ lists of vertices of the same length, l, from G_1 and G_2 respectively. We define the k round Ehrenfeucht-Fraisse game on G_1 and G_2 with initial position given by $\bar{v}$ and $\bar{u}$, denoted by $\textsc{Ehr}_k(G_1,\bar{v},G_2,\bar{u})$, the same way as $\textsc{Ehr}_k(G_1,G_2)$, but in this case the game has l extra rounds at the beginning where the vertices in $\bar{v}$ and $\bar{u}$ are played successively. After this, k more rounds are played normally.

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

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

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

1.5 Some winning strategies for Duplicator

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

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

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

Now suppose that $X \subseteq V(G_1)$ and $Y \subseteq V(G_2)$ can be partitioned into lists $X = \overline{v_1} \cup \cdots \cup \overline{v_a}$ and $Y = \overline{u_1} \cup \cdots \cup \overline{u_b}$ such that $N(\overline{v_i};r)$'s, and the $N(\overline{u_i};r)$'s, are connected and disjoint. We say that $N(X;r)$ and $N(Y;r)$ are k-agreeable, or that they have k-agreeable neighborhoods, if any $\overline{w}$ among the $\overline{v_i}$'s or among the $\overline{u_i}$'s satisfies:

- The number of $\overline{v_i}$'s and the number of $\overline{u_i}$'s satisfying that "$\overline{v_i}$ (resp. $\overline{u_i}$) and $\overline{w}$ have k-similar r-neighborhoods" are the same or are both greater or equal than k.

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

Theorem 1.2. *Set $r = (3^k-1)/2$. Let G_1, G_2 be hypergraphs, and suppose there exist sets $X \subseteq V(G_1)$, $Y \subseteq V(G_2)$ with the following properties:*

(1) $N(X;r)$ and $N(Y;r)$ are k-agreeable.

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

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

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

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

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

Observation 1.2. *Let H_1, H_2 be hypergraphs and $\bar{v}$, $\bar{u}$, be lists of vertices from $V(H_1)$ and $V(H_2)$ respectively. Suppose Duplicator wins $d\text{EHR}_k(H_1,\bar{v},H_2,\bar{u})$. Let $v' \in V(H_1), u' \in V(H_2)$ be vertices played in the first round of an instance of the game where Duplicator is following a winning strategy. Then Duplicator also wins $d\text{EHR}_{k-1}(H_1,\overline{v_2},H_2,\overline{u_2})$, where $\overline{v_2} := \bar{v},v'$ and $\overline{u_2} := \bar{u},u'$.*

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

Proof. Using observation 1.2 we get that Duplicator wins

$$d\text{EHR}_k(N^{G_1}(\bar{v};3r+1),\overline{v_2},N^{G_2}(\bar{u};3r+1),\overline{u_2})$$

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

$$d\text{EHR}_k(N^{H_1}(\overline{v_2};r),\overline{v_2},N^{H_2}(\overline{u_2};r),\overline{u_2}).$$

Because of this if we prove $N^{G_1}(\overline{v_2};r) = N^{H_1}(\overline{v_2};r)$ and $N^{G_2}(\overline{u_2};r) = N^{H_2}(\overline{u_2};r)$, then we are finished. Let $z \in N^{G_1}(v';r)$. Then $d(z,\bar{v}) \le d(z,v') + d(v',\bar{v}) = 3r+1$. In consequence, $N^{G_1}(v;r) \subset H_1$. Thus, $N^{G_1}(\overline{v_2};r) \subseteq H_1$, and $N^{G_1}(\overline{v_2};r) = N^{H_1}(\overline{v_2};r)$. Analogously we obtain $N^{G_2}(\overline{u_2};r) = N^{H_2}(\overline{u_2};r)$, as we wanted. □

Now we are in conditions to prove theorem 1.2.

Proof of theorem 1.2. Define $r_0 = 0$ and $r_i = 3r_{i-1}+1$ for $i > 0$. Let us denote by w_i and z_i the vertices played in G_1 and G_2 respectively during the i-th round of $\text{EHR}_k(G_1,G_2)$.

Let $\overline{v_1},\ldots,\overline{v_a}$ and $\overline{u_1},\ldots,\overline{u_b}$ be lists forming partitions of X and Y respectively, and assume they are as in the definition of k-agreeability. Set

$$\mathscr{X}[0] = \{\overline{v_1},\ldots,\overline{v_a}\}, \quad \mathscr{Y}[0] = \{\overline{u_1},\ldots,\overline{u_b}\}.$$

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

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

Case 1: The vertex w_s satisfies $d(w_s,\bar{v}) > 2r+1$ for all $\bar{v} \in \mathscr{X}[s-1]$. Then Duplicator can find a vertex z_s in G_2 such that $d(z_s,\bar{u}) > 2r+1$ for all $\bar{u} \in \mathscr{Y}[s-1]$ satisfying that w_s and z_s have $(k-s)$-similar r-neighborhoods. To form $\mathscr{X}[s]$ and $\mathscr{Y}[s]$, add to $\mathscr{X}[s-1]$ and $\mathscr{Y}[s-1]$ the lists consisting of only w_s and only z_s respectively, and mark them with the number s.

Case 2: The vertex w_s satisfies $d(w_s,\bar{v}) \le 2r+1$ for a unique $\bar{v} \in \mathscr{X}[s-1]$,and $\bar{v}$ is marked. In this case, find the list $\bar{u} \in \mathscr{Y}[s-1]$ with the same mark. Duplicator then can chose $z_s \in N(\bar{u},2r+1)$ in response to w_s according to a winning strategy for

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

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

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

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

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

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

Case 1: Notice that the lists in $\mathscr{Y}[s-1]$ only contain the vertices previously played in G_2 and the ones from Y. Thus, assumption (3) of the theorem, (or assumption (2) in the symmetric case where Spoiler plays in G_2) assures us that Duplicator can always find such z_s sufficiently far away from all the other lists. In this case, the only new marked lists in $\mathscr{X}[s]$ and $\mathscr{Y}[s]$ are the ones consisting of w_s and z_s respectively. By assumption w_s and z_s have $(k-s)$-similar r_{k-s}-neighborhoods.

Case 2: Notice that by the induction hypothesis $\bar{v}$ and $\bar{u}$ have $(k-s+1)$-similar r_{s-k+1}-neighborhoods, and in consequence a winning strategy for Duplicator exists. Using lemma 1.1 we obtain that the extended lists $\bar{v}, w_s$ and $\bar{u}, z_s$ have $(k-s)$-similar r_{s-k}-neighborhoods.

Case 3: This case is analogous to the previous one. The definition of k-agreeability implies that there is such an unmarked list $\bar{u}$ available. Using lemma 1.1 we obtain that the extended lists $\bar{v}, w_s$ and $\bar{u}, z_s$ have $(k-s)$-similar r_{s-k}-neighborhoods.

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

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

1.6 The random model

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

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

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

1.7 Outline of the proof

We show now an outline of the proof.

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

(1) a.a.s the rank k type of any two graphs in the same class coincide,
(2) a.a.s. any random graph belongs to some of them, and
(3) the limit probability of random graph belonging to any of them is an expression in Θ.

After this is archived the theorem follows easily. Indeed, let ϕ be a sentence in the first order language $\mathscr{L}$ of graphs whose quantifier rank is k, and denote by G a random graph in $G(n, \beta/n)$.

The objective of next sections will be to define the classes $C_1, \ldots, C_{n_k}$ and to show that they satisfy properties (1), (2) and (3). Later we will prove a stronger result, so we will allow ourselves to just sketch some of the proofs during this chapter.

$\square$

References

[1] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

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

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

[4] Edgar N Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.

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

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

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

[8] H-D Ebbinghaus, Jörg Flum, and Wolfgang Thomas. *Mathematical logic*. Springer Science & Business Media, 2013.

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

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