MASTER'S THESIS

# Master of Science in Advanced Mathematics and Mathematical Engineering

**Title:** First Order Logic of Sparse Random Hyper-graphs

**Author:** Lázaro Alberto Larrauri Borroto

**Advisor:** Marc Noy Serrano

**Department:** Applied Mathematics

**Academic year**: 2018 - 2019



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

**Abstract**

We give a generalization the results from Lynch in [1] on the convergence law for sparse random graphs to sparse random hypergraphs.

# Contents

# Introduction

This work is the thesis presented for the Master in Applied Mathematics and Mathematical Engineering of the Universitat Politcnica de Catalunya in June 2019.

This text belongs to the study of the asymptotic properties of random structures. This is an active area of mathematics that has gained recent popularity due to the increasing need for methods that allow for analysis of large random structures. These structures appear in a wide array of “real-life” problems, ranging from the study of social networks to biology, and they are arguably interesting in their own right.

The objective of this work, proposed by Marc Noy and Albert Atserias, was to generalize the results obtained in [1] to more general random structures. This was motivated by applications to the study of the satisfability of random CNF formulas, but those are out of the scope of this work.

Going into further detail, in [1] the asymptotic behavior under first order logic of the Erds-Rnyi binomial model $\mathcal{G}(n,p)$, where $p$ is taken as a decreasing function on $n$ of the form $\beta \cdot n^{-1}$, is studied. The first order properties of graphs are the ones expressible in terms of quantification over vertices and Boolean combinations of statements of the form “$x$ adjacent to $y$” and “$x$ equal to $y$”. The main theorem in [1] states that the limit probability that a graph in $\mathcal{G}(n,\beta/n)$ satisfies any given first-order property exists and has good properties with respect to $\beta$.

The original goal was to give a generalization for the case of uniform hypergraphs but the techniques used for that case also allowed for a generalization for hypergraphs with multiple edge sets of different sizes.

# 1 Logic of Sparse Random Graphs

We give a brief overview of [1]. The asymptotic behavior of graphs in $G(n, \beta/n)$ is as follows

- The number of cycles of length $3, 4 \ldots, r$ are asymptotically distributed like independent Poisson variables.
- Small cycles are a.a.s (asymptotically almost surely) far away.
- Fixed vertices are a.a.s far away.
- The ball of a given radius centered in fixed vertex is a.a.s a tree. Any tree occurs with a positive probability.

This way, asymptotically, trees of any kind occur arbitrarily often in a typical random graph. Given the locality of first order logic up to a fixed quantifier rank this is tells us that the only way to distinguish large random graphs in $G(n, \beta/n)$ under first order logic (up to a fixed quantifier rank) is by looking at the neighborhoods of their small cycles.

It is given a finite and exhaustive classification of the uni-cyclic graphs with less than a given radius. After that it is shown that if two graphs $G_1$ and $G_2$ contain the same number of uni-cycles of each type or they contain a large enough number of those then they are a.a.s equivalent under first order logic up to some quantifier rank $k$. This is done by giving a winning strategy for duplicator in the Ehrenfeucht Fraisse game of $k$ rounds played on $G_1$ and $G_2$.

Finally it is shown that the quantities of uni-cycles of each type in a random graph are asymptotically distributed like independent Poisson variables whose means are given by some other nested "Poisson expressions" that correspond to the trees hanging from each uni-cycle.

# 2 Random Hyper-graphs

Work has been done to generalize zero-one laws and limit laws of graphs to the setting of random hyper-graphs. The results in [2] from Shelah and Spencer have been partially generalized for random $c$-uniform hypergraphs in [3], and a some limit laws, including one for $p(n) = (n^{1-c})$ have been obtained in [4] using approaches different from the ones appearing here.

# 3 Basic definitions and conventions.

## 3.1 General relational structures.

Given a natural number $n$, we will use the notation $[n] := \{1, \ldots, n\}$. We will denote by $S_n$ the symmetric group on $[n]$, and by $\Delta_n$ the diagonal set $\{(a, a) \in [n]^2\}$.

Given a set $X$, then $S_n$ acts on $X^n$ in a natural way. That is, given $g \in S_n$ and $(x_1, \ldots, x_n)$ one can define

$$g \cdot (x_1, \ldots, x_n) = (y_1, \ldots, y_n),$$

where $y_{g(i)} = x_i$ for all $1 \leq i \leq n$.

Given $\Phi$ a subgroup of $S_n$ we will denote by $X^n/\Phi$ the orbit set associated to the action of $\Phi$ over $X^n$.

We will use the notation $[x_1, \ldots, x_n]$ to refer to the equivalence class of the $n$-tuple $(x_1, \ldots, x_n)$ in any sort of quotient $X^n/\Phi$. That is, while the notation $(x_1, \ldots, x_n)$ will be reserved to ordered $n$-tuples, $[x_1, \ldots, x_n]$ will denote an ordered $n$-tuple modulo the action of some arbitrary group of permutations. Which group is this will depend solely on the ambient set where $[x_1, \ldots, x_n]$ is considered.

**Definition 3.1.** Let $n, a \in \mathbb{N}$, with $a \geq 2$, let $\Phi$ be a subgroup of $S_a$, and let $A$ be a subset

$$A \subseteq [a]^2 \setminus \Delta_a.$$

The total edge set $\mathcal{H}_{(a,\Phi,A)}(n)$ of size $a$, symmetry group $\Phi$ and restrictions $A$ on $n$ elements is the set:

$$\mathcal{H}_{(a,\Phi,A)}(n) = ([n]^a/\Phi) \ \setminus \{ [x_1, \ldots, x_a] \in [n]^a/\Phi \mid x_i = x_j \text{ for some } (i,j) \in A\}$$

**Definition 3.2.** An (hyper)-graph $([n], H_1, \ldots, H_c)$ with edge colors $1, \ldots, c$, sizes $a_1, \ldots, a_c$, symmetry groups $\Phi_1, \ldots, \Phi_c$ and restrictions $A_1, \ldots, A_c$ consists of

- The set $[n]$ for some natural number $n$.
- For $i = 1, \ldots, c$, a colored edge set $H_i \subseteq \mathcal{H}_{(a_i,\Phi_i,A_i)}(n)$ whose elements have color $i$.

**Definition 3.3.** Let $p = (p_1, \ldots, p_c)$, where all $p_i$'s are real numbers between 0 and 1. The random model $HG(n,p)$ with edge colors $1, \ldots, c$, sizes $a_1, \ldots, a_c$, symmetry groups $\Phi_1, \ldots, \Phi_c$ and restrictions $A_1, \ldots, A_c$, is the one that assigns to each graph $G = ([n], H_1, \ldots, H_c)$ probability

$$\Pr(G) = \prod_{i=1}^{c} p_i^{|H_i|}(1-p_i)^{|\mathcal{H}_{(a_i,\Phi_i,A_i)}(n)|-|H_i|}.$$

Equivalently, this is the probability space obtained by assigning to each colored edge $e \in \mathcal{H}_{(a_i,\Phi_i,A_i)}(n)$ probability $p_i$ independently.

For the rest of the work we will consider fixed

- the total number of colors $c$,
- the sizes $a_1, \ldots, a_c$,
- the symmetry group $\Phi_1, \ldots, \Phi_c$ and,
- the restrictions $A_1, \ldots, A_c$.

When we say "graph" from now on what we will mean is "hiper-graph with edge colors $1, \dots, c$, sizes $a_1, \dots, a_c$, symmetry groups $\Phi_1, \dots, \Phi_c$ and restrictions $A_1, \dots, A_c$".

Given a graph $G = ([n], H_1, \dots, H_c)$ we will denote by $H_i(G)$ the edge set $H_i$, and by $V(G)$ the vertex set $[n]$. Also, we will write $H(G)$ to denote the disjoint union of colored sets $\cup_{i=1}^{c} H_i$. This way, an edge $e \in H(G)$ with color $i$ is an element $[x_1, \dots, x_{a_i}] \in H_i(G)$, and the $x_i$'s are vertices belonging to $V(G)$.

Given a set of vertices, $X \subseteq V(G)$, we will denote the by $G[X]$ the induced sub-graph on $X$.

As usual, we will sometimes treat edges as sets of vertices rather than "tuples modulo the action of some permutation group". This way, expressions like $e_1 \cap e_2$ for $e_1, e_2 \in H(G)$ will make sense and mean "the set of vertices that occupy some place in $e_1$ and in $e_2$".

Some other times we will treat edges $e \in H(G)$ as sub-graphs of $G$ in the natural way. That is, the subgraph denoted by $e$ is the one whose vertex set is $e$-i.e., the vertices in $e$- and whose only edge is $e$. This way, when we have some edges $e_1, \dots, e_l \in H(G)$ it will make sense to talk about the subgraph $\cup_{i=1}^{l} e_i$, which is the graph whose vertex set is the set of vertices belonging to any of the $e_i$'s, and whose edges are exactly the $e_i$'s. In spite of these abuses of notation the "type" of any "term" involving edges should be derivable from the context.

Another usual abuse of notation we will make is to sometimes treat graphs as their underlying vertex sets. Hence, expressions defined for sets of vertices will also be defined for graphs.

## 3.2 The First Order Language

From now on when we talk about "first order formulas" will be referring to formulas in the first order relational language $\mathcal{L}$ with relations $R_1, ..., R_c$ of arities $a_1, \dots, a_c$ respectively.

Graphs are $\mathcal{L}$-structures in an evident way. The universe of a graph $G$ is $V(G)$, and for each $1 \leq i \leq c$ and any $x_1, \dots, x_{a_i}$ we say

$$R_i(x_1, \dots, x_{a_i}) \text{ if } [x_1, \dots, x_{a_i}] \in H_i(G).$$

That is, variables in $L$ are interpreted as vertices in $G$ and relations are interpreted as edges. By definition, all graphs $G$ satisfy the formulas

- Symmetry formulas:

  $$S_g := \big(R_i(x_1 \dots, x_{a_i}) \iff R_i(x_{g(1)} \dots, x_{g(a_i)})\big),$$

  where $g$ is an element from $\Phi_i$.

- Anti-reflexivity formulas:

  $$AR_{i,(j,l)} := \big(R_i(x_1 \dots, x_{a_i}) \implies \neg(x_j = x_l)\big),$$

  where $(j, l) \in A_i$.

## 4 The Theorem

From now on we will adopt the following two conventions:

- We will always work in the random model $HG(n, p(n))$, where $p(n) = p(n, \beta)$ is defined as the tuple $(\beta_1 n^{1-a_1}, \dots, \beta_c n^{1-a_c})$, and $\beta = (\beta_1, \dots, \beta_c)$. The symbols $\beta_1, \dots, \beta_c$ denote non-negative real variables, but for the most part we will treat them as non-negative positive real constants.
- The first order language we will always refer to is the one defined in last section, $\mathcal{L}$.

We will denote by $\text{Poiss}_\lambda$ the probability function of the Poisson distribution with mean $\lambda$. That is, the one given by $\text{Poiss}_\lambda(n) = e^{-\lambda}\lambda^n/n!$ for any $n \in \mathbb{N}$. Also, we define $\text{Poiss}_\lambda(\leq n)$ and $\text{Poiss}_\lambda(> n)$ as $\sum_{i=0}^{n} \text{Poiss}_\lambda(n)$ and $1 - \text{Poiss}_\lambda(\leq n)$ respectively. Notice that for a fixed $n$, both $\text{Poiss}_\lambda(\leq n)$ and $\text{Poiss}_\lambda(> n)$ can be considered real functions of parameter $\lambda$.

We define $\Lambda$ and $M$ as the minimal families of expressions with arguments $\beta_1, \dots, \beta_c$ that satisfy the conditions:

- $1 \in \Lambda$
- For any $\mu \in M$ and any $n \in \mathbb{N}$ both $\text{Poiss}_\mu(n)$ and $\text{Poiss}_\mu(\geq n)$ are in $\Lambda$
- For any $\lambda_1, \lambda_2 \in \Lambda$, the product $\lambda_1\lambda_2$ belongs to $\Lambda$ as well.
- For any $b, i \in \mathbb{N}$ with $1 \leq i \leq c$, $b > 0$, and $\lambda_1, \dots, \lambda_{a_i - 1} \in \Lambda$, the expression $\frac{\beta_i \prod_{j=1}^{a_i-1} \lambda_j}{b}$ belongs to $M$.

We also define another families $\widehat{\Theta}$ and $\Theta$ of expressions with arguments $\beta_1, \dots, \beta_c$ as the minimal ones satisfying:

- For any $l, s, b \in \mathbb{N}$ with $b > 0$ and any non-necessarily-different $\lambda_1, \dots, \lambda_l \in \Lambda$, $\alpha_1, \dots, \alpha_s \in \{\beta_1, \dots \beta_c\}$, the expression
$$\frac{\left(\prod_{i=1}^{l} \lambda_i\right)\left(\prod_{i=1}^{s} \alpha_i\right)}{b}$$
lies in $\widehat{\Theta}$.
- For any $u \in \widehat{\Theta}$ and any $n \in \mathbb{N}$ both $\text{Poiss}_u(n)$ and $\text{Poiss}_u(\geq n)$ are in $\Theta$.
- For any $\theta_1, \theta_2 \in \Theta$ the product $\theta_1\theta_2$ belongs to $\Theta$ as well.

For any first order sentence $\varphi$ we will use the notation
$$\text{Pr}_n(\varphi) := \text{Pr}(HG(n, p(n)) \models \varphi).$$

The rest of the work will be devoted to prove the following theorem:

**Theorem 4.1.** *Let $\beta = (\beta_1, \dots, \beta_c)$, and let $\varphi$ be a F.O sentence in $\mathcal{L}$. Then the function*
$$\mathfrak{F}(\beta) := \lim_{n\to\infty} \text{Pr}(\, HG(n, p(\beta, n)) \models \varphi\,)$$
*is well defined for all values of $\beta$ and it is a finite sum of expressions in $\Theta$.*

# 5 Sub-critical, Critical and Super-critical Graphs.

We define a distance over arbitrary graphs $G$. For each $x, y \in V(G)$,

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

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

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

As usual, when $X = \{x\}$ we will omit the brackets and write $d(x,Y)$ instead of $d(\{x\},Y)$, for example.

**Definition 5.1.** A path between $x$ and $y$ is minimal graph among the connected graphs containing both $x$ and $y$.

**Proposition 5.1.** *A path $P$ between $x$ and $y$ in a graph $G$ is a union of edges $e_1, \ldots, e_l \in H(G)$ such that*

- *$x$ only belongs to $e_1$ and $y$ only belongs to $e_l$.*
- *For any $1 \leq j < i \leq l$, $e_i$ intersects $e_j$ if and only if $i = j+1$.*

*Sketch of the proof.* Order the edges of the path in a way that (1) $x$ belongs to the first and $y$ to the last, and (2) each edge intersects the previous one and the next one. Such ordering exists because a path is connected. If an edge intersects any other edge than its previous one or next one then we can remove some edge of the path, contradicting the minimality condition. □

**Definition 5.2.** The likelihood $L(G)$ of an hypergraph $G = (V, H_1, \ldots, H_c)$ is the number

$$\left( |V(G)| - \sum_{i=1}^{c} |H_i|(a_i - 1) \right).$$

An hypergraph is $L$-balanced if it contains no subgraph with less likelihood than itself.

Given a graph $G$, and an edge $e \in H(G)$ of color $i$, the operation of "cutting" the edge $e$ is the one where we remove $e$ from $H_i$ and afterwards we also remove the isolated vertices from the resulting graph.

**Proposition 5.2.** *Any connected graph $G$ is $L$-balanced.*

*Proof.* Suppose $G$ is non-empty. The proof is by induction on the number of edges in $G$.

If $G$ has zero edges it is an isolated vertex so the statement is true.

Suppose that $G$ has $m > 0$ edges. Choose a vertex $x \in V(G)$ and an edge $e \in H(G)$ such that the distance from the $x$ to $e$ is maximum. The subgraph $F$ obtained from cutting $e$ must be connected, so by the induction hypothesis $F$ is L-balanced. The original graph $G$ was connected, so $e$ must intersect $F$ in at least one vertex, and

$$L(G) \leq L(F). \tag{1}$$

Suppose that $G$ contains a sub-graph $G_2$ with $L(G_2) < L(G)$. Then, as $F$ is $L$-balanced and 1 holds, $e \in H(G_2)$. Call $F_2$ to the result of cutting $e$ in $G_2$. Then $F_2$ is a subgraph of $F$ and one can check

$$L(F_2) - L(G_2) \leq L(F) - L(G).$$

In consequence, as $L(F) - L(F_2)$ is non-positive, so is $L(G) - L(G_2)$, arriving at a contradiction. □

**Corollary 5.1.** *A (non-empty) connected graph $G$ cannot have more likelihood than* 1*.*

*Proof.* A because of the previous proposition $G$ is $L$-balanced. If $G$ is non-empty it contains some vertex, and vertices have likelihood 1. □

**Definition 5.3.** We will call sub-critical, critical and super-critical graphs to $L$-balanced graphs with likelihood greater than zero, zero, and less than zero respectively.

**Definition 5.4.** A cluster is a connected graph $G$ with non-positive likelihood such that all of its subgraphs $T$ have greater likelihood than itself.

We will call unicycles to connected critical graphs, and cycles to minimal unicycles. In particular, cycles are clusters.

**Proposition 5.3.** *Any unicycle contains exactly one cycle.*

*Sketch of the proof.* Suppose the unicycle $G$ contains two different cycles $F_1, F_2$.

If $F_1$ and $F_2$ have nonempty intersection then $F_1 \cup F_2$ has negative likelihood.

Otherwise, consider a path $P$ between two vertices $x \in V(F_1)$ and $y \in V(F_2)$. The union $F_1 \cup F_2 \cup P$ has negative likelihood. □

**Proposition 5.4.** *A cycle $G$ is either:*

*(1) An edge $e$ where exactly one vertex appears exactly twice.*

*(2) A path $P$, with $L(P) = 1$ between two vertices $x$ and $y$, together with an edge $e$ that intersects $P$ exactly in $x$ and $y$.*

*Sketch of the proof.*

If $G$ contains only one edge then (1) holds necessarily.

Otherwise, $G$ cannot contain an edge that intersects the rest of the graph only in one vertex, because cutting it would yield a smaller connected graph with likelihood 0.

This way, by double counting one can obtain that each edge intersects the rest of $G$ exactly in two vertices. Choose an edge $e \in H(G)$ and call $F$ to the graph obtained by cutting $e$ in $G$. The intersection $e \cap F$ contains exactly two vertices, $x, y$. The graph $F$ must be both

- Connected. Otherwise one of its connected components would have likelihood 0 or less.
- A path between $x$ and $y$. Otherwise it contains a path $P$ between $x$ and $y$ and $P \cup E$ has likelihood 0.

Thus (2) holds, as we wanted. □

Given a graph $G$, its diameter $\operatorname{diam}(G)$ will be the maximum distance between any two of its vertices:

$$\operatorname{diam}(G) = \max_{x,y \in V(G)} d(x, y).$$

**Corollary 5.2.** *For any $r \in \mathbb{N}$ there is a finite number of cycles with diameter at most $r$.*

*Sketch of the proof.* Let $G$ be a cycle with $\operatorname{diam}(G) = r$, and let $x, y \in V(G)$ be such that $d(x, y) = r$. Let $e$ be an edge containing $y$ and not $x$. If such edge does not exists then $G$ is composed of at most two edges. Otherwise consider $F$ the graph resulting from cutting $e$ in $G$. One can check that $F$ must be a path of size at most $2r$, and there are a finite number of those. In consequence we can represent all cycles of diameter $r$ either as an edge, as an union of two edges, or as an union of a path with size at most $r$ with an edge. In particular $|V(G)| \leq 2r + 2a$, where $a$ is the maximum of $a_1, \ldots, a_c$. □

**Proposition 5.5.** *A cluster is a connected graph $G$ where each edge $e \in H(G)$ satisfies at least one of the following conditions:*

*(1) The edge $e$ intersects the union of all the other edges in $H(G)$ in at least two vertices.*

*(2) The edge $e$ contains some vertex at least twice.*

*Proof.* Suppose that $e \in H(G)$ satisfies neither (1) nor (2). Then the graph $G'$ obtained from cutting $e$ in $G$ is a subgraph of $G$ with

$$L(G') = L(G).$$

In consequence $G$ is not a cluster. □

**Definition 5.5.** A tree is a connected graph with likelihood 1.

**Proposition 5.6.** *There is a unique path between any two vertices of a tree.*

*Sketch of the proof.* Let $T$ be a tree. Suppose there are two vertices $x, y \in V(T)$ with two different paths, $P_1, P_2$, between them. Then the graph $P_1 \cup P_2$ would have non positive likelihood, contradicting proposition 5.2. □

**Proposition 5.7.** *A tree $T$ is a graph obtained from successively adding edges to a single vertex in such a way that any new edge intersects the current graph in exactly one vertex.*

*Sketch of the proof.* Choose a vertex $x \in V(T)$ and an edge $e \in H(T)$ such that the distance between the two is the maximum one. Call $T_2$ to the graph obtained by cutting $e$ in $T$. The graph $T_2$ is a tree and intersects $e$ in exactly one vertex. One can continue this process until reaching a single edge. Repeating the process backwards now, adding the removed edges successively, yields the result. □

# 6 Agreeability classes. Winning Strategy For Duplicator

We define the set of constants symbols as $Const := \{\mathfrak{c}_i\}_{i\in\mathbb{N},\, i>0}$. For any $n \in \mathbb{N}$, $n > 0$, let $Const_n$ be the set $\{\mathfrak{c}_1, \dots, \mathfrak{c}_n\}$. A co-labeling of a graph $G$ is a map $\nu : Const_n \to V(G)$, where for some $n > 0$. Given $\mathfrak{c}_i \in Const_n$, we will say that the vertex $\nu(\mathfrak{c}_i)$ is labeled $\mathfrak{c}_i$. Equivalently, we can denote a labeling $\nu$ as a tuple $(\mathfrak{c}_1[x_1], \dots, \mathfrak{c}_m[x_m])$ where each $x_j$ in $V(G)$ is labeled $\mathfrak{c}_j$.

**Definition 6.1.** A graph with constants $G(\mathfrak{c}_1, \cdots, \mathfrak{c}_m)$ is graph $G$ together with a co-labeling $(\mathfrak{c}_1[x_1], \dots, \mathfrak{c}_m[x_m])$.

To keep our notation compact we will often drop he $x_i$'s and say $G(\mathfrak{c}_1, \dots, \mathfrak{c}_m)$ and sometimes we will even omit the $(\mathfrak{c}_1, \dots, \mathfrak{c}_m)$ and denote by $G$ the graph with constants when the co-labeling is not relevant.

We will often identify constants $\mathfrak{c}_i$ with their labeled vertices $\nu(\mathfrak{c}_i)$. This should not lead to confusion. However, note that for two different indices $i, j$, $\nu(\mathfrak{c}_i)$ and $\nu(\mathfrak{c}_j)$ may be the same vertex, but the constant symbols $\mathfrak{c}_i$ and $\mathfrak{c}_j$ are different.

**Definition 6.2.** The center of a connected graph with constants $G(\mathfrak{c}_1, \cdots, \mathfrak{c}_m)$ is its minimal connected subgraph containing all the constants and clusters. If $G(\mathfrak{c}_1, \cdots, \mathfrak{c}_m)$ is not connected then its center is the union of the centers of its connected components.

Given $x \in V(G)$ and $X \subseteq V(G)$ we define the neighbourhood graphs $N(x;r)$ and $N(X;r)$ as

$$N(x;r) = G[S], \quad \text{where } S = \{\, y \in V(G) \,|\, d(x,y) \le r \,\},$$

$$N(X;r) = G[S], \quad \text{where } S = \{\, y \in V(G) \,|\, d(X,y) \le r \,\},$$

for each $r \in \mathbb{N}$.

**Definition 6.3.** A rooted tree $(T, x)$ is a tree $T$ with a distinguished vertex $x$ called root. The radius of the tree is the maximum distance between its root and any of its vertices. The initial edges of $(T, x)$ are the edges in $T$ containing the root.

We will often omit the root and denote by $T$ the whole rooted tree $(T, x)$.

In a rooted tree $(T, x)$ all the edges can be rooted in a canonical way. The root of an edge $e$ is the vertex $y \in e$ such that $d(x, y) = d(x, e)$.

**Definition 6.4.** Let $G(\mathfrak{c}_1, \cdots, \mathfrak{c}_m)$ be a graph with constants and $X, Y \subset V(G)$. Then $Tree(x, G)$ is the rooted tree $(G[V_x], x)$, where

$$V_x = \{\, y \in V \mid d(Center(G), y) = d(Center(G), x) + d(x, y) \,\}.$$

Informally, $Tree(x, G)$ is the graph composed of all the vertices that have to pass through $x$ in order to reach $Center(G)$.

**Remark 6.1.** The graph $Tree(x, G)$ is indeed a tree.

When $(T, x)$ is a rooted tree, we will use the notation

$$Tree(y, T) := Tree(y, T(\mathfrak{c}_1[x])).$$

That is, $Tree(y, T)$ is the tree that "hangs from $y$" in $T$.

When the "ambient graph" is clear we will write $Tree(x)$ instead of $Tree(x, G)$.

**Definition 6.5.** The $k$-morphism equivalence relation $\overset{k}{\simeq}$ between rooted trees of the same radius, and the $k$-type of an edge in a rooted tree, are defined inductively as follows:

- If $T_1$ and $T_2$ are radius 0 rooted trees, then they consist only of their roots and $T_1 \overset{k}{\simeq} T_2$.
- The radius of a $k$-morphism class of rooted trees $C$ is the radius that all trees in $C$ have.
- Let $T$ be a rooted tree of radius $r$. The $k$-type of an edge

  $$e = [x_1, \ldots, x_{j-1}, x, x_j, \ldots, x_{a_d-1}] \in H(T)$$

  with color $d$, where $x$ is the root of $e$, is the colored "tuple modulo the action of $\Phi_d$"

  $$E = [C_1, \ldots, C_{j-1}, \mathfrak{r}, C_j, \ldots, C_{a_d-1}],$$

  with color $d$, where

  - $\mathfrak{r}$ is the "root symbol",
  - $C_i$ is the $k$-morphism class of the tree $Tree(x_i, T)$, for $1 \leq i \leq a_d - 1$.

  The radius of $E$ is the maximum radius of any of the $C_i$'s.

- If $T_1$ and $T_2$ are trees of radius $r > 0$ and roots $x$ and $y$, then $T_1 \overset{k}{\simeq} T_2$ means that for any $k$-type of edges $E$ with radius less than $r$ at least one of the following is satisfied:

  - The number of initial edges in $T_1$ and $T_2$ of $k$-type $E$ is the same.
  - Both $T_1$ and $T_2$ contain no less than $k + 1$ initial edges of $k$-type $E$.

**Remark 6.2.** If $(T, x)$ is a rooted tree with radius $r$, any initial edge $e \in H(T)$ has ($k$-type of) radius less than $r$.

When working with graphs $G_1(\mathfrak{c}_1, \cdots, \mathfrak{c}_m)$ and $G_2(\mathfrak{c}_1, \cdots, \mathfrak{c}_m)$ with the same constants we will use super-indices to distinguish between the constants $\mathfrak{c}_i \in V(G_1)$ and $\mathfrak{c}_i \in V(G_2)$. We will call them $\mathfrak{c}_i^1$ and $\mathfrak{c}_i^2$ respectively.

**Definition 6.6.** Let $G_1(\mathfrak{c}_1, \cdots, \mathfrak{c}_m)$ and $G_2(\mathfrak{c}_1, \cdots, \mathfrak{c}_m)$ be graphs with the same constants. Then we will say that they are $k$-morphic, denoted by $G_1(\mathfrak{c}_1, \cdots, \mathfrak{c}_m) \overset{k}{\simeq} G_2(\mathfrak{c}_1, \cdots, \mathfrak{c}_m)$, if there is an isomorphism $f : Center(G_1) \to Center(G_2)$ such that

- $f(c_i^1) = c_i^2$ for all constants, and
- $Tree(x, G_1) \overset{k}{\simeq} Tree(f(x), G_2)$ for all $x \in V(Center(G_1))$.

Notice that the relation symbol $\overset{k}{\simeq}$ is "overloaded". It is used to denote both the $k$-morphism relation between rooted trees and the $k$-morphism relation between graphs with constants. We give now another additional meaning to the symbol $\overset{k}{\simeq}$. Let $T_1$ and $T_2$ be rooted trees and let $e_1 \in H(T_1)$, $e_2 \in H(T_2)$ be edges. We will write $e_1 \overset{k}{\simeq} e_2$ to denote that $e_1$ and $e_2$ have the same $k$-type.

**Remark 6.3.** The $k$-type of an edge $e$ in a rooted tree $T_1$ does not only depend on the edge $e$. It also contains some non-local information about $T_1$. Namely, it contains the $\overset{k}{\simeq}$ classes of the trees "hanging" from $e$. If $T_2$ is a sub-tree of $T_1$ also containing $e$, the $k$-type of $e$ in $T_2$ may differ from the $k$-type of $e$ in $T_1$. However we will usually refer to "the $k$-type of $e$" instead of "the $k$-type of $e$ in $T_1$" when the "ambient tree" is clear from the context.

Let $T$ be a rooted tree of radius $r$, and let $E$ be a $k$-type of edges with radius less than $r$ for some $k \in \mathbb{N}$. We will denote by $\langle T, E\rangle$ the number of initial edges in $T$ of $k$-type $E$.

Let $C$ be a $k$-morphism class of rooted trees with radius $r$. Consider a representative $T \in C$. We define the number $\langle C, E\rangle$ as the maximum between $\langle T, E\rangle$ and $k+1$. One can check that $\langle C, E\rangle$ does not depend on the choice of $T$.

With this new notation, if $C$ is a $k$-morphism class of rooted trees with radius $r$ and $T$ is a rooted tree of radius $r$, it is satisfied that $T \in C$ if and only if for each $k$-type $E$ of edges with radius less than $r$ either

$$\langle C, E\rangle = \langle T, E\rangle, \quad \text{or} \quad \langle C, E\rangle, \langle T, E\rangle \geq k+1.$$

**Lemma 6.1.** *Let $(T_1, x)$ and $(T_2, y)$ be rooted trees such that for some $k > 0$, $T_1 \overset{k}{\simeq} T_2$. Then, for any $0 \leq j < k$, it is satisfied $T_1 \overset{j}{\simeq} T_2$.*

*Proof.* Fix $j < k$. The proof is by induction on the radius of $T_1$ and $T_2$.

- If they have radius 0 they are $j$-morphic by definition.

- Suppose now that they have radius $r$. Let $e_1$ be an initial edge in $T_1$ and $e_2$ be an initial edge in $T_2$ such that $e_1 \overset{k}{\simeq} e_2$. Then, because of the definition of $k$-type, $e_1$ and $e_2$ have the same color $l$ and we can write $e_1 = [x_1, \dots, x_{a_l}]$, $e_2 = [y_1, \dots, y_{a_l}]$ for some vertices such that for all $i = 1, \dots, a_l$, either
  - $Tree(x_i) \overset{k}{\simeq} Tree(y_i)$, and because of the induction hypothesis $Tree(x_i) \overset{j}{\simeq} Tree(y_i)$.
  - Or both $x_i = x$ and $y_i = y$ are the roots of $T_1$ and $T_2$.

  In consequence $e_1 \overset{j}{\simeq} e_2$. Then, the $\overset{k}{\simeq}$ relation implies the $\overset{j}{\simeq}$ relation on edges with radius less than $r$. Thus, for any $j$-type of edges $E$ there are $k$-types of edges $E_1, \dots, E_m$ such that for all rooted trees $T$ of radius $r$

$$\langle T, E \rangle = \sum_{i=1}^{m} \langle T, E_i \rangle.$$

  In particular

$$\langle T_g, E \rangle = \sum_{i=1}^{m} \langle T_g, E_i \rangle \quad , \text{for } g = 1, 2.$$

  Both of the above sums coincide if all the numbers $\langle T_1, E_i \rangle$'s are not greater than $k$ or both sums are greater than $k$ otherwise. As $j \leq k$ this implies

$$\text{Either } \langle T_1, E \rangle = \langle T_2, E \rangle \text{ or } \langle T_1, E \rangle, \langle T_2, C \rangle \geq j + 1,$$

  and $T_1 \overset{j}{\simeq} T_2$.

□

**Corollary 6.1.** *Let $G_1(\mathfrak{c}_1, \dots, \mathfrak{c}_t)$, and $G_2(c_1, \dots, c_t)$ be graphs with constants such that $G_1(\mathfrak{c}_1, \dots, \mathfrak{c}_t) \overset{k}{\simeq} G_2(c_1, \dots, c_t)$.*
*Then, for any $j < k$, $G_1(\mathfrak{c}_1, \dots, \mathfrak{c}_t) \overset{j}{\simeq} G_2(c_1, \dots, c_t)$.*

**Remark 6.4.** Given a rooted tree $(T, x)$, the neighborhood $N(x; r)$ for any $r \in \mathbb{N}$ together with the root $x$ is a rooted tree.

**Lemma 6.2.** *Let $(T_1, x)$ and $(T_2, y)$ be rooted trees such that for some $k > 0$ $T_1 \overset{k}{\simeq} T_2$. Then, for any $r \geq 0$,*

$$(N(x; r), x) \overset{k}{\simeq} (N(y; r), y).$$

*Proof.* The proof is by induction on the radius of $T_1$ and $T_2$.

- If they have at most radius 0 the statement is vacuously true.
- Suppose now that they have radius $s$. Let $e_1$ be an edge with root $x$ in $N(x; r)$ and $e_2$ be an edge of root $y$ in $N(y; r)$ such that $e_1 \overset{k}{\simeq} e_2$ when both edges are considered in $T_1$ and $T_2$ respectively. We want to prove that $e_1 \overset{k}{\simeq} e_2$ as well when they are considered as edges in $N(x; r)$ and $N(y; r)$. By definition we can write $e_1 = [x_1, \dots, x_a]$ and $e_2 = [y_1, \dots, y_a]$ for some vertices such that for all $i = 1, \dots, a$ either

- $Tree(x_i) \overset{k}{\simeq} Tree(y_i)$, and because of the induction hypothesis

$$Tree(x_i) \cap N(x_i; r-a) \overset{k}{\simeq} Tree(y_i) \cap N(y_i; r-a).$$

- Or $x_i = x$ and $y_i = y$.

In consequence $e_1 \overset{k}{\simeq} e_2$ as edges in $N(x;r)$ and $N(y;r)$. This implies that for any $k$-type $E$ edges of trees with radius less than $r$ there are $k$-types of edges classes $E_1, \ldots, E_l$ of rooted edges such that for all trees $(T_1, x)$,$(T_2, y)$ with radius $s$:

$$\langle N(x;r), E\rangle = \sum_{i=1}^{l} \langle T_1, E_i\rangle,$$

and the same for $N(y;r)$ and $T_2$.

From this follows that either

$$\langle N(x;r), E\rangle = \langle N(y;r), E\rangle \quad \text{or} \quad \langle N(x;r), E\rangle, \langle N(y;r), E\rangle \geq k+1.$$

Thus $N(x;r) \overset{k}{\simeq} N(y;r)$.

□

**Lemma 6.3.** *Let $(T_1, x), (T_1', x), (T_2, y)$ and $(T_2', x)$ be rooted trees satisfying $T_1 \overset{k}{\simeq} T_2$, and $T_1' \overset{k}{\simeq} T_2'$ for some $k \geq 0$ and $V(T_1) \cap V(T_1') = x$, $V(T_2) \cap V(T_2') = y$. Then $T_1 \cup T_1' \overset{k}{\simeq} T_2 \cup T_2'$.*

*Proof.* Let $E$ be a $k$-morphism class of rooted edges. Then

$$\langle T_1 \cup T_1', E\rangle = \langle T_1, E\rangle + \langle T_1', E\rangle, \text{and}$$

$$\langle T_2 \cup T_1', E\rangle = \langle T_2, E\rangle + \langle T_2', E\rangle.$$

And it follows that either

$$\langle T_1 \cup T_1', E\rangle = \langle T_2 \cup T_2', E\rangle,$$

or both quantities are greater than $k$. □

Let $T$ be a tree and $T'$ be a sub-tree of $T$. For any $y \in V(T')$ we will denote by $Span(y, T', T)$ the rooted tree $(T[Y], y)$, where

$$Y := \{\, x \in V(T) \mid d(T', x) = d(x, y) \,\}.$$

That is, $Span(y, T', T)$ is the tree consisting of the vertices in $T$ that "need to pass through $y$ in order to reach $T'$". We also define the graph $Span(T', T)$ as the union

$$\bigcup_{y \in V(T')} Span(y, T', T).$$

**Lemma 6.4.** *Let $(T_1, x)$, $(T_2, y)$ be rooted trees and let $(T_1', x)$ and $(T_2', y)$ be subtrees of $T_1$ and $T_2$ respectively. Let $k \geq 0$. If there is an isomorphism $f : T_1' \to T_2'$ such that $f(x) = y$ and for all $x_1 \in V(T_1')$*

$$Span(x_1, T_1', T_1) \overset{k}{\simeq} Span(f(x_1), T_2', T_2),$$

*then $T_1 \overset{k}{\simeq} T_2$.*

*Proof.* The proof is, again, by induction on the radius of $T_1'$ and $T_2'$.

- If $T_1'$ and $T_2'$ have radius zero, then they consist only of $x$ and $y$ respectively.
- Suppose that $T_1'$ and $T_2'$ have at most $r$. Let $x_1 \neq x$ be a vertex in an edge $e$ rooted at $x$ such that $e \in H(T_1')$. Then the four graphs $Tree(x_1, T_g)$ and $Tree(x_1, T_g')$ for $g = 1, 2$ satisfy the hypothesis of the lemma. The radius of $Tree(x_1, T_1')$ is strictly less than $r$, so because of the induction hypothesis $Tree(x_1, T_1) \overset{k}{\simeq} Tree(f(x_1), T_2)$. This happens for any vertex $x_1$ in $e$ different from the root $x$. In consequence $e \overset{k}{\simeq} f(e)$ and using the previous lemma successively we get $T_1 \overset{k}{\simeq} T_2$.

□

**Definition 6.7.** Let $G(\mathfrak{c}_1, \cdots, \mathfrak{c}_m)$ be a graph with constants. We define $Core(G, r)$ to be $N(X, r)$ where $X$ is the union of constants and clusters of diameter at most $r$.

**Theorem 6.1.** *Let $G_1(\mathfrak{c}_1, \ldots, \mathfrak{c}_t)$, and $G_2(\mathfrak{c}_1, \ldots, \mathfrak{c}_t)$ be graphs with constants such that for some $r \in N$*

$$Core(G_1(\mathfrak{c}_1, \ldots, \mathfrak{c}_t); r) \overset{k}{\simeq} Core(G_2(\mathfrak{c}_1, \ldots, \mathfrak{c}_t); r)$$

*through $f$. Then, for any $s < r$*

$$Core(G_1(\mathfrak{c}_1, \ldots, \mathfrak{c}_t); s) \overset{k}{\simeq} Core(G_2(\mathfrak{c}_1, \ldots, \mathfrak{c}_t); s).$$

*Proof.* Fix $s < r$. Let us introduce some notation

$$F_i := Core(G_i(\mathfrak{c}_1, \ldots, \mathfrak{c}_t); r) \quad \text{for } i = 1, 2.$$

$$F_i' := Core(G_i(\mathfrak{c}_1, \ldots, \mathfrak{c}_t); s) \quad \text{for } i = 1, 2.$$

One can check $F_i' \subseteq F_i$ for $i = 1, 2$, and that the isomorphism $f : Center(F_1) \to Center(F_2)$ restricts to one between $Center(F_1')$ and $Center(F_2')$. Let $v \in Center(F_1')$. Let $T = Tree(v, F_1') \cap Center(F_1)$. It is not hard to see that $T$ is connected and in consequence is a tree. The following identity holds:

$$Tree(v, F_1') = (Tree(v, F_1) \cap N(v; s)) \cup Span(T, Tree(v, F_1')) \tag{2}$$

And analogously for $f(v), f(T), F_2$ and $F_2'$.

Also, for any $w \in V(T)$

$$Span(w, T, Tree(v, F_1')) = Tree(w, F_1) \cap N(w; s - d(Center(F_1'), w)).$$

So, by lemma 6.2

$$Span(w, T, Tree(v, F_1')) \overset{k}{\simeq} Span(f(w), f(T), Tree(f(v), F_2')).$$

In consequence, by lemma 6.4,

$$Span(T, Tree(v, F_1')) \overset{k}{\simeq} Span(f(T), Tree(f(v), F_2')). \tag{3}$$

Again, by lemma 6.2

$$(Tree(v, F_1) \cap N(v; s)) \overset{k}{\simeq} (Tree(f(v), F_1) \cap N(f(v); s)). \tag{4}$$

Finally, using that

$$(Tree(v, F_1) \cap N(v; s)) \cup Span(T, Tree(v, F_1')) = v$$

and

$$(Tree(f(v), F_2) \cap N(f(v); s)) \cup Span(f(T), Tree(f(v), F_2')) = f(v)$$

together with eq. (2), eq. (3), eq. (4) and lemma 6.3 we get:

$$Tree(v, F_1') \overset{k}{\simeq} Tree(f(v), F_2').$$

Hence, $F_1' \overset{k}{\simeq} F_2'$, as desired. □

Let $(T, x)$ be a rooted tree and let $e \in H(T)$ be an initial edge. We will denote by $T \setminus e$ the tree $T[X]$, where

$$X = \{\, y \in V(T) \,|\, \forall z \in e, z \neq x : \; y \notin V(Tree(z, T)) \,\}.$$

In other words, $T \setminus e$ is the result of removing from $T$ the edge $e$ and all the trees that "hang" from $e$.

**Lemma 6.5.** *Let $(T_1, x)$, $(T_2, y)$ be rooted trees such that for some $k \geq 0$ $T_1 \overset{k}{\simeq} T_2$. Let $e_1$ and $e_2$ be initial edges of $T_1$ and $T_2$ such that $e_1 \overset{k}{\simeq} e_2$. Then $T_1 \setminus e_1 \overset{k-1}{\simeq} T_2 \setminus e_2$.*

*Sketch of the proof.* For any $k$-morphism class $E$ of rooted edges clearly either

$$\langle T_1 \setminus e_1, E \rangle = \langle T_2 \setminus e_2, E \rangle,$$

or both quantities are greater than $k - 1$. Now, after an induction process analogous to the one in lemma 6.1 the result follows. □

**Lemma 6.6.** *Let $(T_1, x_1)$, $(T_2, x_2)$ be rooted trees such that for some $k \geq 0$ $T_1 \overset{k}{\simeq} T_2$. For $i = 1, 2$, given a vertex $v \in V(T_i)$ let us denote by $P(v)$ the unique path between $v$ and $x_i$. Then for any $v \in V(T_1)$ there is a vertex $w \in V(T_2)$ and an isomorphism $f : P(v) \to P(w)$ such that*

*(1) $f(x_1) = x_2$ and $f(v) = w$.*

*(2) For any edge $e \in E(P(v))$, $e \overset{k}{\simeq} f(e)$.*

*(3) For any vertex $y \in V(P(v))$, $Tree(y, T_1) \overset{k}{\simeq} Tree(f(y), T_2)$.*

*Proof.* The proof is by induction on $d(x_1, v)$.

- If $d(x_1, v) = 0$ then $x_1 = v$ and the statement is true taking $w = x_2$.
- Suppose now that $d(x_1, v) = r$. Then, by proposition 5.1 one can write the path $P(v)$ as a succession of edges $e_1, e_2, \dots, e_s$, where $v \in e_s$. Let $v'$ be the root of $e_s$, and let $b$ be the color of $e_s$. Then $d(x_1, v') = r - a_b + 1$. Thus, by the induction hypothesis there exists $w'$ such that there is an isomorphism $f : P(v') \to P(w')$ with the required properties. In particular, $Tree(v', T_1) \overset{k}{\simeq} Tree(w', T_2)$, so there is an edge $e'$ rooted at $w'$ with the same $k$-type as $e_s$. Then one can write

$$e_s = [v_1, \dots, v_a], \quad e' = [w_1, \dots, w_a],$$

in a such a way that for some $i$, $v_i = v'$ and $w_i = w'$ and for all $j \neq i$ $Tree(v_j, T_1) \overset{k}{\simeq} Tree(w_j, T_2)$. Let $j$ be such that $v = v_j$. Then we can take $w = w_j$, and extend the isomorphism $f$ to one between $P(v)$ and $P(w') \cup e'$ in the natural way.

□

**Theorem 6.2.** *Let $G_1(\mathfrak{c}_1, \dots, \mathfrak{c}_t)$, and $G_2(c_1, \dots, c_t)$ be graphs with constants such that*

$$Core(G_1(\mathfrak{c}_1, \dots, \mathfrak{c}_t); r) \overset{k}{\simeq} Core(G_2(c_1, \dots, c_t); r)$$

*by means of $f$. Then, for any vertex $x \in Core(G_1(\mathfrak{c}_1, \dots, \mathfrak{c}_t); r)$ there is a vertex $y \in Core(G_2(\mathfrak{c}_1, \dots, \mathfrak{c}_t); r)$ such that*

$$Core(G_1(\mathfrak{c}_1, \dots, \mathfrak{c}_t, \mathfrak{c}_{t+1}[v_1]); r) \overset{k-1}{\simeq} Core(G_2(c_1, \dots, c_t, \mathfrak{c}_{t+1}[v_2]); r)$$

*Proof.* Let us introduce some notation:

$$F_i := Core(G_i(\mathfrak{c}_1, \dots, \mathfrak{c}_t); r) \quad \text{for } i = 1, 2.$$

$$F_1' := Core(G_1(\mathfrak{c}_1, \dots, \mathfrak{c}_t, \mathfrak{c}_{t+1}[v_1]); r).$$

The vertex $v_1$ belongs to $Tree(x_1, F_1)$ for a unique $x_1 \in Center(F_1)$. By the previous lemma there exist a vertex $v_2$ in $Tree(f(x_1), F_2)$ such that the path $P(v_1)$, joining $v_1$ and $x_1$, is isomorphic to the path $P(v_2)$, joining $v_2$ and $f(x_1)$, through an isomorphism $f' : P(v_1) \to P(v_2)$ satisfying properties (1), (2) and (3). Let

$$F_2' := Core(G_1(\mathfrak{c}_1, \dots, \mathfrak{c}_t, \mathfrak{c}_{t+1}[v_2]); r).$$

We are going to show that $F_1' \overset{k-1}{\simeq} F_2'$. Clearly $Center(F_i') = Center(F_i) \cup P(v_1)$ for $i = 1, 2$ , so we can glue $f$ and $f'$ into an isomorphism $g : Center(F_1') \to Center(F_2')$. Let $w_1 \in V(F_1')$. We have to show that

$$Tree(w_1, F_1') \overset{k-1}{\simeq} Tree(g(w_1), F_2')$$

and the result will be proven. The following two cases may occur:

- $Tree(w_1, F_1)$ contains no edges in $P(v_1)$. In this case $Tree(w_1, F_1') = Tree(w_1, F_1)$. Thus, $Tree(w_1, F_1') \overset{k}{\simeq} Tree(g(w_1), F_1')$ and $Tree(w_1, F_1') \overset{k-1}{\simeq} Tree(g(w_1), F_1')$ in consequence by lemma 6.1.
- $Tree(w_1, F_1)$ contains edges from $P(v_1)$. In this case, $Tree(w_1, F_1)$ contains exactly one initial edge $e_1$ in $P(v_1)$, and

  $$Tree(w_1, F_1') = Tree(w_1, F_1) \setminus e_1.$$

  One can check that $Tree(g(w_1), F_2)$ contains exactly one edge in $P(v_2)$, namely $g(e_1)$, and

  $$Tree(g(w_1), F_2') = Tree(g(w_1), F_2) \setminus g(e_1).$$

  We had $Tree(w_1, F_1) \overset{k}{\simeq} Tree(g(w_1), F_2)$, so by lemma 6.5

  $$Tree(w_1, F_1') \overset{k-1}{\simeq} Tree(g(w_1), F_2').$$

□

**Definition 6.8.** Let $G_1, G_2$ be graphs with constants. We say that $G_1$ and $G_2$ are $k$-agreeable if for each $\overset{k}{\simeq}$ class $U$ of connected graphs either

- $G_1$ and $G_2$ have the same number of connected components of type $U$, or
- Both $G_1$ and $G_2$ have no less than $k$ connected components of type $U$.

Given a graph with constants $G$ and a $\overset{k}{\simeq}$ class of connected graphs $U$, we define the number $\langle G, U \rangle$ as the number of connected components of $G$ belonging to $U$.

If $O$ is a $k$-agreeability class and $G \in O$ is a representative of it, we define the number $\langle O, U \rangle$ to be the maximum between $\langle G, U \rangle$ and $k+1$.

**Definition 6.9.** A graph $G$ is $(i,k,r)$-rich for some $i,k,r \in \mathbb{N}$, if for any rooted tree $T$ of radius at most $r$ there are $x_1, \dots, x_i \in V(G)$ such that

- The $d(x_l, x_s) \geq 4r$ if $l \neq s$.
- The $N(x_l; r)$'s do not intersect $Core(G, r)$
- $(N(x_l; r), x_l) \overset{k}{\simeq} T$.

The following is a corollary of lemma 6.1 and lemma 6.2.

**Lemma 6.7.** *Let $i,k,r \in \mathbb{N}$ be all positive numbers, and let $G(\mathfrak{c}_1, \cdots, \mathfrak{c}_m)$ be a $(i,k,r)$-rich graph with constants. Then, for any $x \in V(G)$, the graph $G(\mathfrak{c}_1, \cdots, \mathfrak{c}_m, \mathfrak{c}_{m+1}(x))$ is $(i-1, j, s)$-rich, where $j, s$ are arbitrary natural numbers satisfying $j \leq k$, and $s \leq r$.*

For reference about Ehrenfeucht Fraisse games see [5] and [6].

**Theorem 6.3.** *Let $G_1, G_2$ be $(k,k,3^k)$-rich graphs such that $Core(G_1; 3^k)$ is $k$-agreeable with $Core(G_2; 3^k)$. Then $G_1$ and $G_2$ have the same rank $k$ type.*

*Proof.* We will prove that Duplicator has a winning strategy in the $E.F$ game on $G_1$ and $G_2$ with $k$ rounds. We will show, by induction on $i$, that Duplicator can play in such a way that in the $i$-th round

$$Core(G_1(\mathfrak{c}_1[x_1], \dots, \mathfrak{c}_i[x_i]); 3^{k-i})$$

is $(k-i)$-agreeable with

$$Core(G_2(\mathfrak{c}_1[y_1], \dots, \mathfrak{c}_i[y_i]); 3^{k-i}),$$

where for each $1 \leq j \leq i$, the vertices $x_j$ and $y_j$ are the ones chosen in the $j$-th round of the game in the graphs $G_1$ and $G_2$ respectively. After this the theorem follows because $Core(G_1(\mathfrak{c}_1[x_1], \dots, c_i[x_k]); 1)$ being 0-agreeable with $Core(G_2(\mathfrak{c}_1[y_1], \dots, c_i[y_k]); 1)$ implies that the map given by $x_i \mapsto y_i$ defines a partial isomorphism.

- For $i = 0$ the statement is true by hypothesis.
- Assume now that $Core(G_1(\mathfrak{c}_1[x_1], \dots, \mathfrak{c}_i[x_{i-1}]); 3^{k-i+1})$ is $(k-i+1)$-agreeable with $Core(G_2(\mathfrak{c}_1[y_1], \dots, \mathfrak{c}_i[y_{i-1}]); 3^{k-i+1})$. Without loss of generality we can suppose that Spoiler chooses a vertex $x_i$ in $G_1$ in the $i$-th round. We have two cases:

  **Case 1.** $N(x_i; 3^{k-i})$ is contained in $Core(G_1(\mathfrak{c}_1[x_1], \dots, \mathfrak{c}_i[x_{i-1}]); 3^{k-i+1})$. Let $F_1$ be the connected component of $Core(G_1(\mathfrak{c}_1[x_1], \dots, \mathfrak{c}_i[x_{i-1}]); 3^{k-i+1})$ containing $N(x_i; 3^{k-i})$. Then there is a connected component $F_2$ in $Core(G_2(\mathfrak{c}_1[y_1], \dots, \mathfrak{c}_i[y_{i-1}]); 3^{k-i+1})$ such that $F_1 \overset{k-i+1}{\simeq} F_2$. Applying theorem 6.2 and theorem 6.1 successively Duplicator can choose $y_i \in V(F_2)$ such that

  $$Core(F_1(\mathfrak{c}_i[x_i]); 3^{k-1}) \overset{k-i}{\simeq} Core(F_2(\mathfrak{c}_i[y_i]); 3^{k-1}).$$

  Using theorem 6.1 and counting the connected components in each $k-i$-morphism class one can check that now $Core(G_1(\mathfrak{c}_1[x_1], \dots, \mathfrak{c}_i[x_i]); 3^{k-i})$ is $(k-i)$-agreeable with $Core(G_2(\mathfrak{c}_1[y_1], \dots, \mathfrak{c}_i[y_i]); 3^{k-i})$.

  **Case 2.** $N(x_i; 3^{k-i})$ is contained in $Core(G_1(c_1[x_1], \dots, c_i[x_{i-1}]); 3^{k-i+1})$. Then $N(x_i; 3^{k-i})$ is disjoint from $Core(G_1(\mathfrak{c}_1[x_1], \dots, \mathfrak{c}_i[x_{i-1}]); 3^{k-i})$ and in particular, $N(x_i; 3^{k-i})$ is a tree. As $G_2$ was originally $k, k, 3^k$-rich, using lemma 6.7 we obtain that $G_2(\mathfrak{c}_1[y_1], \dots, \mathfrak{c}_i[y_{i-1}])$ is $(k-i+1, k-i+1, 3^{k-i+1})$-rich. Hence, Duplicator can choose $y_i$ in $G_2$ such that $N(y_i; 3^{k-i})$ is disjoint from $Core(G_2(\mathfrak{c}_1[y_1], \dots, \mathfrak{c}_i[y_{i-1}]); 3^{k-i})$ and the tree $N(x_i; 3^{k-i})$ rooted at $x_i$ is $(k-i)$-morphic to the tree $N(y_i; 3^{k-i})$ rooted at $y_i$. Counting connected components of each type we can conclude that $Core(G_1(\mathfrak{c}_1[x_1], \dots, \mathfrak{c}_i[x_i]); 3^{k-i})$ is $(k-i)$-agreeable with $Core(G_2(\mathfrak{c}_1[y_1], \dots, \mathfrak{c}_i[y_i]); 3^{k-i})$.

□

# 7 Probabilistic results

## 7.1 Almost All Graphs are Simple

**Definition 7.1.** A graph $G$ is $r$-simple if all connected components of $Core(G; r)$ are unicycles.

A $k$-agreeability class $O$ is called simple if all the graphs belonging to it are disjoint unions of unicycles.

**Proposition 7.1.** *Let $F$ be a cluster with $L(F) < 0$, and let $X_n$ be the random variable that counts the number of times that $H$ appears as a subgraph of $HG(n, p(n))$. Then*

$$\lim_{n\to\infty} \Pr(X_n > 0) = 0.$$

*Proof.* Let $v = |V(F)|$ and $h_i = |H_i(F)|$ for $i = 1, \ldots, l$. Chose a ordering of the vertices in $H$. For any ordered sequence of vertices $S = (x_1, \ldots, x_v)$, let $X_{n,S}$ be the indicator variable that equals 1 if $F$ is a subgraph of $G[S]$ (in a way that respects the ordering) and 0 otherwise. Clearly $X_n$ is the sum of all the $X_{n,S}$'s, so

$$E(X_n) = \frac{n(n-1)\cdots(n-v+1)}{b} \prod_{i=1}^{l} \Big(\frac{c_i}{n^{a_1-1}}\Big)^{h_i},$$

where $b$ is the cardinality of $F$'s group of isomorphisms. Then, for some constant $A$,

$$\lim_{n\to\infty} E(X_n) \leq \lim_{n\to\infty} A \cdot n^{L(H)},$$

and using that $L(E) < 0$, this limit is zero. Using the first moment method the result follows. □

**Lemma 7.1.** *Let $G$ be a critical graph of radius $r$. Then $G$ contains a critical subgraph with size no greater than $(a+2)(r+1)+2a$, where $a$ is the largest edge size in $G$.*

*Proof.* Choose $x \in V(G)$. Successively remove from $G$ edges $e$ such that $d(x, e)$ is maximum until the resulting graph $G'$ has likelihood no less than 0. We have two cases:

- $L(G') = 1$. Let $e = [x_1, \ldots, x_b]$ be the last removed edge and $e \cap G' = \{x_{i_1}, \ldots, x_{i_d}\}$. For any $j = 1, \ldots, d$ choose $P_j$ a path of size no greater than $r+1$ joining $x$ and $x_{i_j}$ in $G'$. Then $P_1 \cup \ldots P_d \cup e$ is a critical subgraph of $G$ of size less than $a(r+1)+a < (a+2)(r+1)+2a$.
- $L(G') = 0$. Let $e_1 = [x_1, \ldots, x_{b_1}]$ be the last removed edge. Continue removing the edges of $G'$ that are at maximum distance from $x$ until you obtain $G''$ with $L(G'') = 1$. Let $e_2 = [y_1, \ldots, y_{b_1}]$ be the last removed edge. As before, let $e_1 \cap G' = \{x_{i_1}, \ldots, x_{i_d}\}$ and for $j = 1, \ldots, d$ let $P_j$ a path of size no greater than $r+1$ joining $x$ and $x_{i_j}$ in $G'$ Then $e_2 \cup G'' = \{y_{i_1}, y_{i_2}\}$. Let $Q_1, Q_2$ be paths size no greater than $r+1$ from $x$ to $y_{i_1}$ and $y_{i_2}$ in $G''$. Then $Q_1 \cup Q_2 \cup e_2$ is a graph of likelihood 0 and size less than $2r+2+a$, and $Q_1 \cup Q_2 \cup P_1 \cup \cdots \cup P_d \cup e_1 \cup e_2$ is a critical graph with size less than $(2+a)(r+1)+2a$

□

**Corollary 7.1.** *Let $A_n$ be the event that $HG(n, p(n))$ contains critical subgraph with diameter no greater than $r$. Then*

$$\lim_{n\to\infty} \Pr(A_n) = 0.$$

*Proof.* If a random graph contains $G$ such critical graph, then by the previous lemma it has to contain a critical graph of size less than some constant $M$. The number of critical graphs of such size is finite and by proposition 7.1 the probability that any one of those appears as a subgraph of $G$ is asymptotically zero. $\square$

**Corollary 7.2.** *For any $r$,*

$$\lim_{n\to\infty} \Pr(G(n, \beta_1/n^{a_1-1}, \ldots, \beta_l/n^{a_l-1}) \text{ is } r\text{-simple}) = 1.$$

*Proof.* One can check that if $G$ contains no super-critical subgraphs of diameter at most $4r$ then $G$ is $r$-simple. $\square$

## 7.2 Probabilities of Trees. Almost All Graphs are Rich.

For any formula $\phi$ with free variables $x_1, \ldots, x_l$, we define $\Pr_n(\phi(x_1, \ldots, x_l)) = \sum_{G \models \phi(a_1, \ldots, a_l)} \Pr_n(G)$, where $a_1, \ldots, a_l$ are fixed **different** natural numbers in $[n]$. If $\phi$ and $\sigma$ are formulas with possibly some free variables, then we define $\Pr_n(\phi \,|\sigma) = \Pr_n(\phi \wedge \sigma)/\Pr_n(\sigma)$.

We introduce some notation now. For any numbers $l, r \in \mathbb{N}$ we denote by $\phi_r(x_1, \ldots, x_l)$ the formula with free variables $x_1, \ldots, x_l$ that is satisfied if for any $y$ there is a unique (minimal) path from $y$ to the set of $x_i$'s.

**Lemma 7.2.** *Let $\sigma$ be a consistent open formula (i.e., a formula with no quantifiers) with free variables $x_1, \ldots, x_l$. Then, for any $r \in \mathbb{N}$*

$$\lim_{n\to\infty} \Pr_n(\phi_r(x_1, \ldots x_l) \,|\sigma) = 1.$$

*Proof.* We will see that

$$\lim_{n\to\infty} \Pr_n(\neq \phi_r(x_1, \ldots, x_l) \,|\sigma) = 0.$$

Fix $n$ and $x_1, \ldots, x_l \in [n]$. Notice that without loss of generality we can assume that $\sigma$ is Boolean combination of atomic formulas of the form $R_i(y_1, \ldots, y_{a_i})$, where the $y_i$'s are among the free variables $x_1, \ldots, x_l$. In particular $G \models \sigma(x_1, \ldots, x_l)$ and $G \models \phi_r(x_1, \ldots, x_l)$ are independent events, because $\phi_r$ depends only on the edges not in $x_1, \ldots, x_l$. Thus, $\Pr_n(\neg\phi_r \,|\sigma) = \Pr_n(\neg\phi_r)$. Consider $G$ a random graph in $G(n, p(n))$. If $\phi_r$ is not satisfied in $G$ then there exists a $y$ different from the $x_i$'s and two paths $P_1, P_2$ between $y$ and the set of $x_i$'s with $V(P_i) \leq r + 1$. We have two cases

- $P_1$ and $P_2$ contain some vertex $z \neq y$ in their intersection. Then, using that $L(P_1), L(P_2) \leq 1$ and counting we get that $P_1 \cup P_2$ is a critical graph of size no greater than $2r + 1$.
- The union $P_1 \cup P_2$ is a path between some $x_{i_1}, x_{i_2}$ of size no greater than $2r+1$ that contains vertices different from the $x_j$'s. Let us denote by $A$ the event that such a path exists. One can check that $\Pr_n(A) \leq C/n$ for some fixed $C$.

Thus, if we denote by $B$ the event that $G$ contains some super-critical subgraph of size no greater than $2r+1$, by the union bound:

$$\Pr_n(\neg\phi_r) \leq \Pr_n(B) + \Pr_n(A) \overset{n\to\infty}{\longrightarrow} 0.$$

$\square$

We give without proof (see Chapter 8, [7]) the following theorem, which will be our main technical tool for the rest of the work.

**Theorem 7.1.** *(Multivatiate Brun's Sieve) Let $r \in \mathbb{N}$ and for each $i = 1, \ldots, r$ let $\{X_i(n)\}_{n\in\mathbb{N}}$ be a succession of random variables such that for each $n \in \mathbb{N}$, $X_i(n)$ is a sum of random indicator variables (i.e. variables taking only values 0 and 1). Let $\lambda_1, \ldots, \lambda_r$ be real numbers. If for each $r$-tuple of natural numbers $b_1, \ldots, b_r$ is satisfied*

$$\lim_{n\to\infty} E\left[\prod_{i=1}^{r}\binom{X_i}{b_i}\right] = \prod_{i=1}^{r}\frac{\lambda_i^{b_i}}{b_i!},$$

*then the random variable$(X_1, \ldots, X_r)$ converges in distribution to a tuple of independent Poisson variables with means $\lambda_1 \ldots \lambda_r$. That is,*

$$\forall x_1, \ldots, x_r \in \mathbb{N}: \quad \lim_{n\to\infty} \Pr(\wedge_{i=1}^{r} X_i = x_i) = \prod_{i=1}^{r} \mathrm{Poiss}_{\lambda_i}(x_i).$$

**Remark 7.1.** Let $X$ be a random variable sum of indicator variables $Y_1, \ldots, Y_s$. For each $i \in \mathbb{N}$, let $X_i$ be the random variable

$$X_i = |\{(j_1, \ldots, j_l) \,|\, j_1 < \cdots < j_i,\, Y_{j_1} = \cdots = Y_{j_i} = 1\}|.$$

That is, $X_i$ counts the unordered $i$-tuples of $Y_j$'s that take value 1. Then it is not difficult to check that

$$\binom{X}{i} = X_i, \text{ for all } i \in \mathbb{N}.$$

Let $x_1, \ldots, x_l$ be vertices of a random graph $G$. For each $y \in V(G)$ we abbreviate by $T_r(y; x_1, \ldots, x_l)$ the rooted tree

$$Tree(y, Core(G(c_1[x_1], \ldots, c_l[x_l]); r)).$$

**Theorem 7.2.** *Let $k \in N$. For all $r \in N$ and any $k$-morphism class of trees with radius at most $r$ there exist expressions $\lambda(k, C, r) \in \Lambda$ such that:*

*for any consistent open formula $\sigma(x_1, \ldots, x_l)$, $r, s \in \mathbb{N}$ $s \leq l$ and any $k$-morphism classes $C_1, \ldots, C_s$ of trees with radii at most $r$ it is satisfied*

$$\lim_{n\to\infty} \Pr_n\Big(\bigwedge_{i=1}^{s} Tree_r(x_i; x_1, \ldots, x_l) \in C_i \,|\, \sigma(x_1, \ldots, x_l)\,\Big) = \prod_{i=1}^{s} \lambda(k, C_i, r).$$

*Proof.* Consider $k$ fixed. The proof is by induction on $r$.

For $r = 0$ there is only one class of $k$-morphic trees and the probability in the statement is always 1 for all $n$. Thus, taking $C$ the $k$-morphism class of the isolated vertex, one can define $\lambda(k, C, 0) = 1$.

Fix $r > 0$ and assume that the statement is true for all lesser values of $r$.

Let $\mathcal{E}$ be the set of $k$-types of edges $E$ of radius at most $r-1$. For each $E \in \mathcal{E}$ pick a representative $(C_{E,1}, \dots, C_{E,j-1}, r, C_{(E,j)} \dots, C_{E,a_E-1})$, and denote by $j_E$ the index of the root in that representative. Denote by $a_E$ and $c_E$ the arity and color of $E$, and denote by $\psi_E$ the subgroup of the symmetry group of $\phi_{c_E}$ that fixes the chosen representative of $E$. Consider, for each $i = 1, \dots, s$ and each the random variables

$$X_{i,E}(n) = \text{number of initial edges of type } E \text{ in } T_r(x_i; x_1, \dots, x_l).$$

Given $e = [y_1, \dots, y_{j_E-1}, x_i, y_{j_E}, \dots, y_{a_E-1}] \in \mathcal{H}_{c_E}(n)$ we can define the indicator random variable $X_{i,E,e}$ that takes value 1 if the following are all satisfied

- $e \in H_{c_E}$,
- $e$ belongs to $T_r(x_i; x_1, \dots, x_l)$, and
- the $k$-type of $e$ is $E$.

One can check that for fixed $i, E$

$$X_{i,E}(n) = \sum_{e=[y_1,\dots,y_{j_E-1},x_i,y_{j_E},\dots,y_{a_E-1}]\in\mathcal{H}_{c_E}(n)} X_{i,E,e}(n).$$

Thus we can apply the multivariate Brun's Sieve to the variables $X_{i,E}$.

Let $(b_{(i,E)})_{\substack{i=1,\dots,s\\ E\in\mathcal{E}}}$ be natural numbers. We want to compute

$$\lim_{n\to\infty} E\left[\prod_{i=0}^{s}\prod_{E\in\mathcal{E}}\binom{X_{i,E}(n)}{b_{(i,E}}\,\middle|\,\sigma\right].$$

Define by $\Omega$ the set

$$\begin{aligned}\Omega := \{(i,E,b,j) \,|\, i,b,j \in \mathbb{N},\ E \in \mathcal{E},\\ 1 \le i \le s,\ b = 1 \le b \le b_{(i,E},\ 1 \le j \le a_E - 1\}.\end{aligned}$$

And let $\hat{\Omega}$ be the projection of $\Omega$ onto its first three coordinates. That is,

$$\begin{aligned}\widehat{\Omega} := \{(i,E,b) \,|\, i,b \in \mathbb{N},\ E \in \mathcal{E},\\ 1 \le i \le s,\ b = 1 \le b \le b_{(i,E}\}.\end{aligned}$$

Denote by $X$ be the set $\{x_1, \dots, x_l\}$. Choose a function $y : \Omega \to [n] \setminus X$. Informally, $y()$ represents a choice of edges in $G$. We say that $y()$ satisfies the property $P$ if for any fixed $1 \le i \le s$, $E \in \mathcal{E}$ and $1 \le b_1 < b_2 \le b_{(i,E}$, and $t = 1, 2$, the tuples

$$[y(i,E,b_t,1), \dots, y(i,E,b_t,j_E-1), x_i, y(i,E,b_t,j_E), \dots y(i,E,b_t,a_E-1)]$$

represent different elements in $\mathcal{H}_{c_E}(n)$. In other words, $y()$ is a choice of different edges.

Define the event $A(y)$ as

$$\bigwedge_{(i,E,b,j)\in\Omega} y(i,E,b,j) \in T_r(x_i; x_1,\dots,x_l).$$

Define also the event $B(y)$ as

$$\bigwedge_{\omega=(i,E,b)\in\widehat{\Omega}} [y(\omega,1),\dots,y(\omega,j_E-1),\, x_i\,, y(\omega,j_E),\dots,y(\omega,a_E-1)] \in H_{c_E}.$$

Finally, let $T(y)$ be the event that

$$\bigwedge_{(i,E,b,j)\in\Omega} T_{(r-a_E+1)}(y(i,E,b,j); x_1\dots,x_l) \in C_{(E,j)}.$$

That is,

- $A(y)$ is the event that for any fixed $(i,E,b)$ the vertices $y(i,E,b,j)$ belong to the tree of $x_i$,
- $B(y)$ is the event that for any fixed $(i,E,b)$ the vertices $y(i,E,b,j)$ together with $x_i$ form an edge in $H_{c_E}$ when ordered in a particular way, and
- $T(y)$ is the event that the tree hanging from each vertex $y(i,E,b,j)$ belongs to the particular $k$-morphism class given by the edge type $E$ and the position $j$.

Then it is satisfied

$$\begin{aligned}
&E\left[\prod_{i=0}^{s}\prod_{E\in\mathcal{E}}\binom{X_{i,E}}{b_{(i,E)}}\,\middle|\,\sigma\right]=\\
&=\prod_{i=1}^{s}\prod_{E\in\mathcal{E}}\left(\frac{1}{|\psi_E|^{b_{(i,E}}\cdot b_{(i,E)}}\right)\cdot\sum_{\substack{y:\Omega\to[n]\setminus X\\ y\text{ satisfies }P}}\Pr(T(y)\,\wedge\,A(y)\wedge\,B(y)\,|\,\sigma\,).
\end{aligned}\tag{5}$$

Notice that $A(y)$ implies that $y$ is injective. Indeed, if a vertex $v$ belongs to two edges incident to some $x_i$ then both edges cannot belong to the tree of $x_i$ because they would form a cycle (or a super-critical graph). Also, if $v$ belongs to the edges $e_1, e_2$ incident to $x_i$ and $x_j$ respectively then it cannot happen that $e_1$ is in the tree of $x_i$ and $e_2$ is in the tree of $x_j$ at the same time, because $e_1\cup e_2$ would belong to the center of $Core(G(c_i[x_i], c_j[x_j]); r)$. In consequence we only need to take in consideration injective $y$'s in last equation. Also, by the symmetry of the random model, the probability written in that equation is equal for all injective $y$'s. Hence we have

$$\begin{aligned}
&\prod_{i=1}^{s}\prod_{E\in\mathcal{E}}\left(\frac{1}{|\psi_E|^{b_{(i,E}}\cdot b_{(i,E)}}\right)\cdot\sum_{\substack{y:\Omega\to[n]\setminus X\\ y\text{ injective}}}\Pr(T(y)\,\wedge\,A(y)\wedge\,B(y)\,|\,\sigma\,)=\\
&\prod_{i=1}^{s}\prod_{E\in\mathcal{E}}\left(\frac{1}{|\psi_E|^{b_{(i,E}}\cdot b_{(i,E)}}\right)\cdot\Pr(T(z)\,\wedge\,A(z)\wedge\,B(z)\,|\,\sigma\,)\cdot\sum_{\substack{y:\Omega\to[n]\setminus X\\ y\text{ injective}}}1,
\end{aligned}\tag{6}$$

where $z$ is an arbitrary injective map from $\Omega$ to $[n] \setminus X$.

We can write

$$\Pr(T(z) \wedge A(z) \wedge B(z) \mid \sigma) = \Pr(T(z) \wedge A(z) \mid B(z) \wedge \sigma) \cdot \Pr(B(z) \mid \sigma)$$

Notice that $\phi_r(x_1, \ldots, x_l) \wedge B(z)$ implies $A(z) \wedge B(z)$, and in consequence the following chain of inequalities holds

$$\begin{aligned}\Pr(T(z) \mid B(z) \wedge \sigma) &\geq \Pr(T(z) \wedge A(z) \mid B(z) \wedge \sigma) \geq \\ &\geq \Pr(T(z) \wedge \phi_r(x_1, \ldots, x_l) \mid B(z) \wedge \sigma).\end{aligned}$$

But using lemma 7.2 we get

$$\lim_{n\to\infty} \Pr_n(T(z) \wedge \phi_r(x_1, \ldots, x_l) \mid B(z) \wedge \sigma) = \lim_{n\to\infty} \Pr_n(T(z) \mid B(z) \wedge \sigma),$$

so

$$\lim_{n\to\infty} \Pr_n(T(z) \wedge A(z) \mid B(z) \wedge \sigma) = \lim_{n\to\infty} \Pr_n(T(z) \mid B(z) \wedge \sigma). \tag{7}$$

Notice that $B(z)$ can be written in terms of a purely relational open formula with free variables the $y(i, E, b, j)$'s. Thus by the induction hypothesis we have

$$\lim_{n\to\infty} \Pr_n(T(z) \mid B(z) \wedge \sigma) = \Gamma,$$

where

$$\Gamma := \prod_{\substack{1\leq i\leq s\\ E\in\mathcal{E}\\ 1\leq j\leq a_E-1}} (\lambda(k, C_{(E,j)}, r - a(E) + 1))^{b_{(i,E)}}, \tag{8}$$

where $a(E)$ is the size of $E$. In particular $\Gamma$ is different from 0. Hence, last term in eq. (6) is equal to

$$\lim_{n\to\infty} \prod_{i=1}^{s} \prod_{E\in\mathcal{E}} \left( \frac{1}{|\psi_E|^{b_{(i,E}} \cdot b_{(i,E)}} \right) \cdot \Gamma \cdot \Pr_m(B(z) \mid \sigma) \cdot \sum_{\substack{y:\Omega\to[n]\setminus X\\ y \text{ injective}}} 1. \tag{9}$$

Also,

$$\sum_{\substack{y:\Omega\to[n]\setminus X\\ y \text{ injective}}} 1 = |[n] \setminus X| \cdot (|[n] \setminus X|) \cdots (|[n] \setminus X| - |\Omega| + 1),$$

and using that $X$ and $\Omega$ are constant in size,

$$\sum_{\substack{y:\Omega\to[n]\setminus X\\ y \text{ injective}}} 1 \simeq n^{|\Omega|}, \tag{10}$$

where $f(n) \simeq g(n)$ means that $\lim_{n\to\infty} f(n)/g(n) = 1$. Finally, as $\sigma$ only affects the edges between the $x_i$'s , $B(z)$ and $\sigma$ are independent. Hence,

$$\Pr_n(B(z) \mid \sigma) = \prod_{i=1}^{s} \prod_{E\in\mathcal{E}} \left( \frac{\beta_{c_E}}{n^{(a_E-1)}} \right)^{b_{(i,E}},$$

and using that

$$n^{|\Omega|} = \prod_{i=1}^{s} \prod_{E\in\mathcal{E}} (n^{(a_E-1)})^{b_{(i,E)}}$$

and eq. (10) we obtain

$$\Pr_n(\, B(z) \,|\, \sigma\,) \cdot \sum_{\substack{y:\Omega\to[n]\setminus X\\ y \text{ injective}}} 1 \simeq \prod_{E\in\mathcal{E}} \beta_{c_E}^{b_{(i,E}}. \tag{11}$$

In consequence, using eq. (11) and eq. (8) in eq. (9) we get

$$\lim_{n\to\infty} E\left[\prod_{i=0}^{s}\prod_{E\in\mathcal{E}} \binom{X_{i,E}}{b_{(i,E}} \,\middle|\, \sigma\right] = \prod_{i=0}^{s}\prod_{E\in\mathcal{E}}\left[\left(\frac{\beta_{c_E}\prod_{j=1}^{a_E-1}\lambda(k, C_{(E,j)}, r-a(E)+1)}{|\psi_E|}\right)^{b_{(i,E}} \frac{1}{b_{(i,E)}}\right].$$

And using the multivariate Brun's Sieve we get that for each choice of natural numbers $\{b_{(i,E}\}_{\substack{1\leq i\leq s\\ E\in\mathcal{E}}}$ it is satisfied

$$\lim_{n\to\infty}\Pr_n\left(\bigwedge_{i=0}^{s}\bigwedge_{E\in\mathcal{E}} X_{i,E} = b_{(i,E)} \,\middle|\, \sigma\right) = \prod_{i=0}^{s}\prod_{E\in\mathcal{E}} \mathrm{Poiss}_{\mu(k,E,r-a_E+1)}(b_{(i,E)}),$$

where we define

$$\mu(k, E, r-a_E+1) = \frac{\beta_{c_E}\prod_{j=1}^{a_E-1}\lambda(k, C_{(E,j)}, r-a(E)+1)}{|\psi_E|}.$$

Notice that $\mu(k, E, r-a_E+1)$ is an expression in $M$. The $k$-morphism class of $T_r(x_j; x_1, \ldots, x_l)$ depends exclusively on the number, up to $k+1$, of its initial edges of each type. More explicitly

$$\begin{aligned}&T_r(x_j; x_1, \ldots, x_l) \in C \iff\\ &\bigwedge_{E\in\mathcal{E}} \big(X_{j,E} = \langle C, E\rangle \text{ if } \langle C, E\rangle \leq k\big), \text{ or } \big(X_{j,E} \geq k+1 \text{ otherwise}\big).\end{aligned}$$

In consequence

$$\begin{aligned}&\lim_{n\to\infty}\Pr_n(\bigwedge_{i=1}^{s} Tree_r(x_i; x_1, \ldots, x_l) \in C_i \,|\, \sigma\,) =\\ &= \prod_{i=0}^{s}\left[\left(\prod_{\substack{E\in\mathcal{E}\\ \langle C_i,E\rangle<k+1}} \mathrm{Poiss}_{\mu(k,E,r-a_E+1)}(\langle C_i, E\rangle)\right)\left(\prod_{\substack{E\in\mathcal{E}\\ \langle C_i,E\rangle\geq k+1}} \mathrm{Poiss}_{\mu(k,E,r-a_E+1)}(\geq (k+1))\right)\right] =\\ &= \prod_{i=0}^{s}\lambda(k, C_i, r),\end{aligned}$$

where we define

$$\lambda(k, C_i, r) = \left(\prod_{\substack{E\in\mathcal{E}\\ \langle C_i,E\rangle<k+1}} \mathrm{Poiss}_{\mu(k,E,r-a_E+1)}(\langle C_i, E\rangle)\right)\left(\prod_{\substack{E\in\mathcal{E}\\ \langle C_i,E\rangle\geq k+1}} \mathrm{Poiss}_{\mu(k,E,r-a_E+1)}(\geq (k+1))\right).$$

Notice that $\lambda(k, C_i, r)$ belongs to $\Lambda$, and its definition depends only on the previously defined $\lambda$'s for lesser values of $r$, and on the choice of $C_i, r$, as we wanted.

□

**Corollary 7.3.** *For any* $i, k, r \in \mathbb{N}$

$$\lim_{n\to\infty} \Pr_n(\,G \textit{ is } i,k,r\text{- } \textit{rich}\,) = 1.$$

*Proof.* Fix $i, k, r$, and let $\mathcal{C}$ be the set of all $k$-morphism classes of trees with radius at most $r$. For any $s \in \mathbb{N}$ and $C \in \mathcal{C}$ with $s \geq i$ we define the event $A_{s,C}$ as:

$$\exists x_1, \dots, x_i \in [s] : \quad T_r(x_j; 1, dots, s) \in C \quad \forall 1 \leq j \leq i.$$

One can check that for all $s$,

$$\phi_{3r}(1, \dots, s) \bigwedge_{C\in\mathcal{C}} A_{s,C} \implies G \text{ is } i,k,r\text{- rich}\,.$$

This is because $\phi_{3r}(1, \dots, s)$ means that the vertices $x \in [s]$ are "far from each other and far from the cycles of radius $r$". Thus

$$\lim_{n\to\infty} \Pr_n\left(\phi_{3r}(1, \dots, s) \bigwedge_{C\in\mathcal{C}} A_{s,C}\right) \geq \lim_{n\to\infty} \Pr_n(\,G \text{ is } i,k,r\text{- rich}\,). \tag{12}$$

But also, because of lemma 7.2

$$\lim_{n\to\infty} \Pr_n\left(\phi_{3r}(1, \dots, s) \bigwedge_{C\in\mathcal{C}} A_{s,C}\right) = \lim_{n\to\infty} \Pr_n\left(\bigwedge_{C\in\mathcal{C}} A_{s,C}\right). \tag{13}$$

Using the intersection bound, we get

$$Pr\left(\bigwedge_{C\in\mathcal{C}} A_{s,C}\right) \geq 1 - \sum_{C\in\mathcal{C}} \Pr(\neq A_{s,C}).$$

Define, for each $s \geq i$, $C \in \mathcal{C}$ the random variable

$$X_{s,C} = |\{\, y \in [s] \,|\, T_r(y; 1, \dots, s) \in C \,\}|.$$

Because of last theorem $X_{s,C}$ converges in distribution to a binomial variable whose defining probability is $\lambda(k, C, r)$. Notice that the event $A_{s,C}$ is precisely the event $X_{s,C} \geq i$. Thus:

$$\lim_{n\to\infty} \Pr_n\left(\bigwedge_{C\in\mathcal{C}} A_{s,C}\right) \geq \lim_{n\to\infty} 1 - \sum_{C\in\mathcal{C}} \Pr_n(\neq A_{s,C}) =$$
$$= \lim_{n\to\infty} 1 - \sum_{C\in\mathcal{C}} \Pr_n(X_{s,C} < i) = 1 - \sum_{j<i} \binom{s}{j} \lambda(k, C, r)^j (1 - \lambda(k, C, r))^{(s-j)}.$$

This expression goes to 1 as $s$ goes to infinity because $\lambda(k, C, r)$ is positive. This means

$$\lim_{s\to\infty} \lim_{n\to\infty} \Pr_n\left(\bigwedge_{C\in\mathcal{C}} A_{s,C}\right) = 1.$$

Finally using eq. (12) and eq. (13) we get that

$$\lim_{n\to\infty} \Pr_n(\,G \text{ is } i,k,r\text{- rich}\,) \geq 1$$

and the result is proven. $\square$

## 7.3 Probabilities of Unicycles.

**Theorem 7.3.** *Let $k \in \mathbb{N}$. Let $O$ be a simple $k$-agreeability class. Then it is satisfied*

$$\lim_{n\to\infty} \Pr_n(Core(G;r) \in O) = \theta,$$

*for some $\theta \in \Theta$.*

*Proof.* This is an easier version of theorem 7.2.

Let $\mathcal{U}$ be the set of all $k$-morphism classes of unicycles with radius at most $r$. For each class $U \in \mathcal{U}$ choose a representative $rep(U) \in U$, and let $cycle(U)$ be the cycle in $rep(U)$, whose number of vertex will be denoted by $n_U$. Choose an ordering $x_{(U,1)}, \dots, x_{(U,n_U)}$ of the vertices in $cycle(U)$ and for each $1 \le i \le n_U$ denote by $C_{(U,i)}$ the $k$-morphism class of $Tree(x_i, rep(U))$. One can consider $cycle(U)$ to be a vertex colored graph where the color assigned to each vertex $x_i$ is $C_{(U,i)}$. Isomorphisms of the colored cycle $cycle(U)$ induce permutations of $[n_U]$ via this ordering. Let us denote by $\psi_U$ that group of permutations.

For any $U \in \mathcal{U}$ we define the random variable

$$X_U(n) = \text{number of connected components of } Core(G;r) \text{ in } U,$$

and for each element $g \in V(G)^{n_U}/\psi_U$ we define the indicator variable $X_{U,g}(n)$ that equals 1 if

- $g = [x_1, \dots, x_{n_U}]$, for some vertices $x_1, \dots, x_{n_U}$ such that the map $f : G[X] \to cycle(U)$ ,where $X = \{x_1, \dots x_{n_U}\}$, defined by $x_i \mapsto x_{(U,i)}$ is an isomorphism.
- $N(X;r)$ is a connected component of $Core(G;r)$.
- $N(X;r) \overset{k}{\simeq} rep(U)$ via $f$. In particular this means that $Tree(x_i, G)$ belongs to $C_{(U,i)}$.

In other words, $X_{U,g}$ indicates if there is a graph in the $k$-morphism class $U$ embedded in $G$ in a particular way represented by $g$. One can check that for all $U \in \mathcal{U}$

$$X_U(n) = \sum_{g \in [n]/\psi_U} X_{U,g}(n),$$

so we can apply the multivariate Brun's Sieve to the $X_U$'s.

Let $(b_U)_{U\in\mathcal{U}}$ be fixed natural numbers. We are interested in obtaining

$$\lim_{n\to\infty} E\left[\prod_{U\in\mathcal{U}} \binom{X_U(n)}{b_U}\right].$$

Let $\Omega$ be the set defined as

$$\Omega := \{\, (U,b,i) \,|\, U \in \mathcal{U},\ b,i \in \mathbb{N},\ 1 \le b \le b_U,\ 1 \le i \le n_U \,\}$$

and let $\widehat{\Omega}$ be the projection of $\Omega$ onto its two first coordinates. That is,

$$\widehat{\Omega} := \{\, (U,b) \,|\, U \in \mathcal{U},\ b \in \mathbb{N},\ 1 \le b \le b_U \,\}.$$

Let $y : \Omega \to [n]$ be a map. Informally, $y()$ represents a choice of embeddings of graphs in with the appropriate $k$-morphism classes. We will say that $y()$ satisfies the property $P$ if for any fixed $U \in \mathcal{U}$ and $1 \le b_1 < b_2 \le b_U$ the tuples

$$[y(U, b_1, 1), \dots, y(U, b_1, n_U)], \text{ and } [y(U, b_2, 1), \dots, y(U, b_2, n_U)]$$

represent different elements in $[n]^{n_U}/\psi_U$. That is, $y()$ is a choice of different embeddings.

Define for any $(U, b) \in \widehat{\Omega}$ the set $Y(U, b) = \{\, y(U, b, i) \,|\, 1 \le i \le n_U \,\}$.

We define the following events for a given $y : \Omega \to [n]$.

- Let $A(y)$ be the event that for each $(U, b) \in \widehat{\Omega}$, the map $f_{U,b} : \mathit{cycle}(U) \to G[Y(U, b)]$ given by $x_{(U,i)} \to y(U, b, i)$ is an embedding.
- Let $B(y)$ be the event that $\mathit{Center}(N(Y(U, b); r)$ is the image of $f_{U,b}$, for each $(U, b) \in \widehat{\Omega}$.
- Let $T(y)$ be the event that

$$\bigwedge_{(U,b,i)\in\Omega} T_r(y(U, b, i); Y) \in C_{(U,i)},$$

  where $Y$ denotes set of vertices in the image of $y$.

Then,

$$E\left[\prod_{U\in\mathcal{U}} \binom{X_U(n)}{b_U}\right] = \prod_{U\in\mathcal{U}} \frac{1}{|\psi_U|^{b_U} b_U!} \cdot \sum_{\substack{y:\Omega\to[n]\\ y \text{ satisfies } P}} \Pr(A(y) \wedge B(y) \wedge T(y)).$$

Property $P$,together with events $A(y)$ and $B(y)$ imply that $y$ is injective, so we can consider only such $y$'s in last equation. Again, by the symmetry of the random model the probability appearing there is the same for all injective $y$'s. Hence,

$$\begin{aligned} &\prod_{U\in\mathcal{U}} \frac{1}{|\psi_U|^{b_U} b_U!} \cdot \sum_{\substack{y:\Omega\to[n]\\ y \text{ satisfies } P}} \Pr(A(y) \wedge B(y) \wedge T(y)) = \\ &= \prod_{U\in\mathcal{U}} \frac{1}{|\psi_U|^{b_U} b_U!} \cdot \Pr(A(z) \wedge B(z) \wedge T(z)) \cdot \sum_{\substack{y:\Omega\to[n]\\ y \text{ injective}}} 1, \end{aligned} \tag{14}$$

where $z$ is an arbitrary injective map $z : \Omega \to [n]$.

We can write

$$\Pr(A(z) \wedge B(z) \wedge T(z)) = \Pr(B(z) \wedge T(z) \,|\, A(z)) \cdot \Pr(A(z)).$$

Let $\tau_r$ be the event that $G$ is $r$-simple. One can check that $A(y) \wedge \tau_r$ implies $A(y) \wedge B(y)$. In consequence the following chain of inequalities holds

$$\Pr(T(z) \,|\, A(z)) \ge \Pr(T(z) \wedge B(z) \,|\, A(z)) \ge \Pr(T(z) \wedge \tau_r \,|\, A(z)).$$

Notice that $A(z)$ can be expressed as a purely relational open formula with free variables the elements indexed by $z$, because it only depends on the edges between vertices in the image of $z$. Using theorem 7.2 and corollary 7.2 we obtain

$$\lim_{n\to\infty} \Pr_n(T(z) \wedge \tau_r \mid A(z)) = \lim_{n\to\infty} \Pr_n(T(z) \mid A(z)) = \Gamma, \tag{15}$$

where

$$\Gamma := \prod_{\substack{U\in\mathcal{U}\\ 1\le i\le n_U}} (\lambda(k, C_{(U,i)}, r))^{b_U}.$$

Because the probability of each edge is independent, one obtains

$$\Pr_n(A(z)) = \prod_{(U,b)\in\widehat{\Omega}} \frac{\prod_{i=1}^{c} \beta_i^{|H_i(cycle(U))|}}{n^{n_U}}. \tag{16}$$

Also,

$$\sum_{\substack{y:\Omega\to[n]\\ y \text{ injective}}} 1 \simeq \prod_{(U,b)\in\widehat{\Omega}} n^{n_U}. \tag{17}$$

This way substituting eq. (15), eq. (16) and eq. (17) in eq. (14) we get

$$\lim_{n\to\infty} E\left[\prod_{U\in\mathcal{U}} \binom{X_U(n)}{b_U}\right] = \prod_{U\in\mathcal{U}} \left(\frac{\prod_{i=1}^{c} \beta_i^{|H_i(cycle(U))|} \prod_{i=1}^{n_U} \lambda_{U,i}}{|\psi_U|}\right)^{b_U} \cdot \frac{1}{b_U!}.$$

Applying the multivatiate Brun's Sieve we obtain that for any fixed natural numbers $(b_U)_{U\in\mathcal{U}}$

$$\lim_{n\to\infty} \Pr_n\left(\bigwedge_{U\in\mathcal{U}} X_U = b_U\right) = \prod_{U\in\mathcal{U}} \mathrm{Poiss}_{\xi_U}(b_i),$$

where

$$\xi_U = \frac{\prod_{i=1}^{c} \beta_i^{|H_i(cycle(U))|} \prod_{i=1}^{n_U} \lambda_{U,i}}{|\psi_U|}.$$

Notice that each $\xi_U$ lies in $\widehat{\Theta}$.

The class of $k$-agreeability of a graph depends only on the number of connected components of each $k$-morphism class. More explicitly, if $O$ is a $k$-agreeability class of radius $r$,

$$Core(G;r) \in O \iff \bigwedge_{U\in\mathcal{U}} (X_U = \langle O,U\rangle \text{ if } \langle O,U\rangle \le k),\ \text{or } (X_U \ge k{+}1 \text{ otherwise}).$$

In consequence,

$$\lim_{n\to\infty} \Pr_n(Core(G;r) \in O) = \left(\prod_{\substack{U\in\mathcal{U}\\ \langle O,U\rangle\le k}} \mathrm{Poiss}_{\xi_U}(\langle O,U\rangle)\right)\left(\prod_{\substack{U\in\mathcal{U}\\ \langle O,U\rangle\ge k+1}} \mathrm{Poiss}_{\xi_U}(\ge (k+1))\right),$$

an this last limit belongs to $\Theta$ we wanted. □

### 7.4 Proof of the Main Theorem.

We re-estate the main theorem of this section

**Theorem 7.4.** *Let $\beta = (\beta_1, \ldots, \beta_c)$, and let $\psi$ be a F.O sentence in $\mathcal{L}$. Then the function*

$$\mathfrak{F}(\beta) := \lim_{n\to\infty} \Pr(\, HG(n, p(\beta, n)) \models \psi\,)$$

*is well defined for all values of $\beta$ and it is a finite sum of expressions in $\Theta$.*

*Proof.* Let $k$ be the quantifier rank of $\psi$, and let $O_1, \ldots, O_m$ be an enumeration of all $k$-agreeability simple classes whose components are cycles of diameter at most $3^k$ with trees of radii at most $3^k$ hanging from them. Because of corollary 7.2, corollary 7.3 and theorem 7.3 respectively we have:

(1)
$$\lim_{n\to\infty} \Pr(HG(n, p(n)) \in \cup_{i=1}^m O_i) = 1,$$

(2) For any $1 \le i \le m$,
$$\lim_{n\to\infty} \Pr_n((G \models \psi) \wedge (F \models \neg\psi) \,|\, G, F \in O_i\,) = 0,$$
where $G$ and $F$ are independently chosen graphs in $HG(n, p(n, \beta))$.

(3) For any $1 \le i \le m$,
$$P_i(\beta) := \lim_{n\to\infty} \Pr((HG(n, p(n, \beta)) \in O_i)$$
is well defined for all values of $\beta$ and it is an expression in $\Theta$.

We define the events $E_1, \ldots, E_m$ as

$$E_i := (G \models \psi) \wedge (G \in O_i),$$

and the event $F$ as

$$F := (G \models \psi) \bigwedge_{i=1}^{m} (G \notin O_i).$$

Then, for any $n \in \mathbb{N}$

$$\Pr_n(G \models \psi) = \sum_{i=1}^{m} \Pr_n(E_i) + \Pr_n(F), \tag{18}$$

as the events $E_i$ together with $F$ form a partition of all the cases where $G$ satisfies $\psi$.

Fix and index $i \in \{1, \ldots, m\}$. From (2) follows that the limits

$$\lim_{n\to\infty} \Pr_n(G \models \psi \,|\, G \in O_i)$$

are either zero or one, and

$$\begin{aligned} \lim_{n\to\infty} \Pr_n(E_i) &= \lim_{n\to\infty} \Pr_n(G \in O_i) \cdot \Pr_n(G \models \phi \,|\, G \in O_i) = \\ &= \text{ either } 0 \text{ or } \lim_{n\to\infty} \Pr_n(G \in O_i). \end{aligned} \tag{19}$$

Also, as a consequence of (1) we obtain

$$\lim_{n\to\infty} \Pr_n(\bigwedge_{i=1}^{m} G \notin O_i) = 0,$$

so

$$\lim_{n\to\infty} \Pr_n(F) = \lim_{n\to\infty} \Pr_n(\bigwedge_{i=1}^{m} G \notin O_i) \cdot \Pr_n(G \models \phi \mid \bigwedge_{i=1}^{m} G \notin O_i) = 0. \tag{20}$$

Taking limits in equation 18 and using equations 19 and 20 we get

$$\lim_{n\to\infty} \Pr_n(G \models \psi) = \sum_{O_i \in \mathcal{O}} \lim_{n\to\infty} \Pr_n(G \in O_i),$$

where $\mathcal{O}$ is a (possibly empty) subset of $\{O_1, \ldots, O_m\}$. Finally, because of property (3) for each $i$ the limit $\lim_{n\to\infty} \Pr_n(G \in O_i)$ is an expression in $\Theta$. Thus $\lim_{n\to\infty} \Pr_n(G \models \psi)$ is a finite sum of expressions in $\Theta$ and the theorem follows. $\square$

## References


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

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

[3] Aleksandr Matushkin. Zero-one law for random uniform hypergraphs. *arXiv preprint arXiv:1607.07654*, 2016.

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

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

[6] Leonid Libkin. *Elements of finite model theory.* Springer Science & Business Media, 2013.

[7] Péter Csikvári. Probabilistic method, lecture notes.