Universitat Politècnica de Catalunya

Programa de Doctorat de Matemàtica Aplicada

# First Order Logic of Random Sparse Structures

**Doctoral thesis by**
Alberto Larrauri

**Thesis advisor**
Marc Noy

Department of Mathematics, January 2023

## Abstract

This work is dedicated to the study several models of random structures from the perspective of first-order logic. We prove that the asymptotic probabilities of first-order statements converge in a general model of random structures with linear density, extending previous results by Lynch, and give an application of this result to the random SAT problem. We also inspect the set of limiting probabilities of first-order properties in sparse binomial graphs, binomial $d$-uniform hypergraphs and graphs with given degree sequences. In particular, we characterize the conditions under which this set of asymptotic probabilities is dense in the interval $[0, 1]$. Finally, we introduce the question of whether preservation theorems, namely Łoś-Tarski Theorem and Lyndon's Theorem, hold in a probabilistic sense in various models of random graphs. We obtain several positive results in different regimes of the binomial random graph and uniform graphs from addable minor-closed classes.


**Keywords.** Random graphs, first-order logic, limit laws, preservation theorems, random SAT, random structures, sparse graphs, minor-closed classes.

## Acknowledgements

This work would not have been possible without the endless support from my parents, both during my PhD studies and during all the years up until then. I thank them greatly for all the words of advice and encouragement during this period.

I would like to extend my deep gratitude to my supervisor, Marc Noy. He sparkled my interest in the study of random structures from the perspective of logic, which would later become the topic of this thesis. Discussions with him over these years have been immensely helpful and enjoyable. I also would like to thank him for his patience and understanding during these last strange years.

Next, I want to thank all the other people who have shared their knowledge about math in a broad sense with me over this period. This includes my coauthors and the people both in the Combinatorics group at Barcelona Tech, as well as in the Formal Methods group at TU Graz. Specially valuable have been the discussions with Albert Atserias, on the topics of random SAT, preservation theorems and other subjects related to logic and computation, and Gillem Perarnau, about random graphs with given degree sequences and probability in general.

Lastly, I would like to mention the people that have supported me in other ways over the course of my PhD studies. The list is long, but I am specially grateful to Roderick Bloem for for being an encouraging mentor and for his insight into the academic profession. I also want to thank my friends for bearing with me over these years, specially Hector, Santi, Malte, Filip and Bettina, and for our time together.

This work was supported by the ERC grant 648276 AUTAR from October 2019 until September 2020.

# Contents

**Introduction** **2**

**1 Preliminaries** **4**
1.1 General Notation . . . . 4
1.2 Graphs, Hypergraphs and Relational Structures . . . . 4
1.3 Logic Background . . . . 7
1.4 Probability Background . . . . 8

**2 Convergence Law for Sparse Random Structures** **12**
2.1 Inside-Out Strategies . . . . 15
2.1.1 Strategies on Trees . . . . 17
2.1.2 Strategies on Fragments . . . . 18
2.1.3 Strategies on Sparse Structures . . . . 19
2.2 The Landscape of Sparse Random Structures . . . . 20
2.3 Proof of the Convergence Law . . . . 24
2.3.1 Probabilities of Tree Classes . . . . 25
2.3.2 Probabilities of Fragment Classes . . . . 26
2.3.3 Main Result . . . . 27
2.4 Convergence Law For Derived Models . . . . 28
2.4.1 Adding Symmetries . . . . 29
2.4.2 Adding Irreflexivity Constraints . . . . 31
2.5 Application to Random SAT . . . . 31

**3 Sets of Limiting Probabilities** **36**
3.1 Binomial Graphs . . . . 37
3.1.1 Cycles and Fragments . . . . 38
3.1.2 No gap when $c \geq 1$ . . . . 42
3.1.3 Always a finite union of intervals . . . . 43
3.1.4 Transition at $c_0$ . . . . 45
3.2 Binomial d-Uniform Hypergraphs . . . . 46
3.2.1 Cycles and Fragments . . . . 47
3.2.2 No gap when $c \geq (d-2)!$ . . . . 50

3.2.3 Always a finite union of intervals . . . . . 51
3.2.4 Transition at $c_0$ . . . . . 53
3.3 Graphs with Given Degree Sequences . . . . . 54
3.3.1 Configuration Model, Cycles and Fragments . . . . . 57
3.3.2 No gap when $\nu \geq 1$ . . . . . 67
3.3.3 Always a finite union of intervals . . . . . 67
3.3.4 Transition at $\nu_0$ . . . . . 69
3.3.5 Remarks About the Convergence Law . . . . . 70

**4 Preservation Theorems for FO Logic on Random Graphs 72**
4.1 Preservation Theorems on Sparse Graphs . . . . . 74
4.1.1 Minimum Trees . . . . . 75
4.1.2 Minimum Fragments . . . . . 76
4.1.3 First Order Logic of Sparse Random Graphs . . . . . 76
4.1.4 Proof of the Main Result . . . . . 77
4.2 Preservation Theorems on Very Sparse Graphs . . . . . 77
4.3 Preservation Theorems at the Connectivity Threshold . . . . . 79
4.3.1 The Landscape at the Connectivity Threshold . . . . . 79
4.3.2 Strategies at the Connectivity Threshold . . . . . 80
4.3.3 Proof of the Main Results . . . . . 81
4.4 Preservation Theorems on Minor-Closed Classes . . . . . 84

# Introduction

Random graphs and related structures have been a central object of study in probabilistic combinatorics since the seminal work of Erdös and Renyì [20]. Besides being interesting objects in their own right, research on random graphs also finds important applications to other areas. For instance, the probabilistic method, introduced by Erdös, is a proof technique used for showing the existence of combinatorial objects with certain properties. With this approach, existence is established (non-constructively) by proving that the set of desired objects has positive probability in some probability space. One of the earliest famous results proven with this technique was the existence of graphs having both arbitrarily large girth and chromatic number at the same time. Another area of application is the study of computationally-hard problems. So called "phase transitions" have been found for some of those problems, such as $k$-SAT [25] or $k$-colorability [1]. These are phenomena where some properties of random structures change drastically when these structures experience small changes of density. Knowledge related to these phase transitions has helped to understand the limitations of current algorithms and inspired the development of new ones [11]. Lastly, random graphs and their evolution can be used to model a wide range of real-life phenomena, with applications in physics, computer science, social sciences and biology.

This thesis falls into the scope of the asymptotic study of random structures, mainly random graphs. The central problem in this context is to compute the limit probability that $\mathcal{G}_n$ satisfies $P$, given a property $P$ and a well-behaved sequence of random structures $(\mathcal{G}_n)_{n\in\mathbb{N}}$ that grow in size. An even more fundamental question is whether this limit exists at all. It is immediately clear that some regularity conditions should be imposed on $P$. For instance, when studying random graphs, one should disallow properties like "there is an even number of vertices". Hence, we want to limit ourselves to the study of *nice* graph properties, in some sense. A reasonable notion of *niceness* should, intuitively, be closed under Boolean operations: the negation of a nice property should also be nice, and the same goes for the conjunction of two nice properties. The insight of the model-theoretical approach is to group properties according to formal languages they can be defined in. This approach was initiated with the seminal result by Fagin and Glebsky (independently), which states that any first-order property of graphs (i.e., a property definable in the first-order logic of graphs) holds in the uniform graph $\mathcal{G}_n$ on $n$ vertices with limiting probability either zero or one [21, 28]. After this, the initial hope was that on simple, well-behaved random models, the probability of any first-order property would converge. This intuition was proven wrong in a surprising way by Shelah and Spencer, who showed in [61] that convergence failed in the binomial random graph for a wide range of natural decaying edge probabilities. Afterwards, work in this direction has mainly aimed to establish various convergence and non-convergence results for other languages and random models.

This thesis presents various contributions to the asymptotic study of random structures through the lens of logic. As the title indicates, our formal language of choice is first order logic, and our structures of interest are sparse (as in "opposite to dense") random structures. The text is structured as follows. In Chapter 1 we introduce some notation and preliminary notions about probability and logic that will be used throughout the rest of the thesis. Each chapter that follows corresponds to various pieces of work completed by the the author during the course of his PhD studies. In Chapter 2 we introduce a general binomial model of random sparse structures and show that the probabilities of first-order properties converge for these structures. In addition to that, we also give an application of this result to the study of the random SAT problem. In Chapter 3 we study the geometry of the set of limiting probabilities of first-order statements in several random models. These include the binomial random graph Section 3.1, the binomial $k$-uniform hypergraph Section 3.2, and random graphs with given degree sequences Section 3.3. Finally, in Chapter 4, we establish various regularity results for first-order logic, called preservation theorems, in several models of random graphs. Those include various regimes of the binomial random graph, as well as in uniform graphs from addable minor-closed classes.

# Chapter 1

# Preliminaries

## 1.1 General Notation

We adopt the convention that the set $\mathbb{N}$ of natural numbers starts at zero. Given $n \in \mathbb{N}$, we write $[n]$ for the set $\{1, 2, \ldots, n\}$, or the empty set for $n = 0$. If $m \in \mathbb{N}$ is another number, $(m)_n$ denotes its $n$-th falling factorial. That is, $m(m-1)\cdots(m-n+1)$ if $m \geq n$, or zero if $m < n$. Given a set $S$, $(S)_n$ stands for the set of tuples $(s_1, \ldots, s_n) \in S^n$ where no two elements coincide. Similarly, $\binom{S}{n}$ stands for all subsets $S' \subseteq S$ with $|S'| = n$. This way, $|S^n| = |S|^n$, $|(S)_n| = (|S|)_n$ and $\left|\binom{S}{n}\right| = \binom{|S|}{n}$. The sets $S^* = \bigcup_{n\in\mathbb{N}} S^n$ and $S_* = \bigcup_{n\in\mathbb{N}}$ contain all finite tuples with elements in $S$, with and without repetitions respectively. We commonly use boldface variables $\boldsymbol{v}$ as shorthands for tuples $v_1, \ldots, v_s$.

All asymptotics are taken in terms of the parameter $n \in \mathbb{N}$. We make use of the big-O and small-O notations with their usual meanings. Let $f(n)$ be a real function. Given $g(n)$ be another (asymptotically) positive real function, we write $f(n) = O(g(n))$ if $|f(n)| \leq Cg(n)$ holds for some constant $C > 0$ and all sufficiently large $n$, and we use $f(n) = \Omega(g(n))$ to denote $g(n) = O(|f(n)|)$. Alternatively, $f(n) = o(g(n))$ means that $\lim_{n\to\infty} |f(n)|/g(n) = 0$. Arithmetic expressions involving $O(\,)$, $\Omega(\,)$ or $o(\,)$ occurrences have the usual meaning. For example $f(n) = 1 + n + O(n^2)$ means that $f(n) - 1 - n = O(n^2)$, and so on.

## 1.2 Graphs, Hypergraphs and Relational Structures

### Graphs

Formally, a **(labeled) graph** $G$ is a pair $(V(G), E(G))$ where $V(G)$ is its **set of vertices**, and $E(G) \subseteq \binom{V(G)}{2}$ its **set of edges**. We mostly refer to labeled graphs simply as graphs. Unless stated otherwise, we deal only with finite graphs $G$, meaning that $V(G)$ is a finite set. This consideration also extends to hypergraphs and relational structures, introduced later. An object we will refer to multiple times during this thesis is the **binomial random graph** $\mathcal{G}(n, p)$, introduced by Erdös and Renyì. This is a random graph whose set of vertices is $[n]$, and each pair in $\binom{[n]}{2}$ forms an edge with probability $p$, independently.

The **excess** of $G$ is the number $\mathrm{ex}(G) = |E(G)| - |V(G)|$. Given vertices $u, v \in V(G)$, $u \sim_G v$ is a shorthand for $\{u, v\} \in E(G)$. Given two graphs $G, H$, an **isomorphism** between them is a bijection $f : V(G) \to V(H)$ such that $\{v, u\} \in E(G)$ if and only if $\{f(v), f(u)\} \in E(H)$ for all $u, v \in V(H)$. We write $G \simeq H$ when $G$ and $H$ are **isomorphic**. An **unlabeled graph** $[G]$ the isomorphism class of some graph $G$. Alternatively, $H \subseteq G$ means that $H$ is a **subgraph** of $G$ (i.e., $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$). An $H$**-copy** in $G$ is simply a subgraph $H' \subseteq G$ satisfying $H \simeq H'$. An **automorphism** of $G$ is just an isomorphism from $G$ onto itself. The number of automorphisms of $G$ is denoted by $\mathrm{aut}(G)$. Given a set $S \subseteq V(G)$, $G[S]$ denotes $G$'s **induced subgraph** on $S$. That is, $G[S]$ is the graph whose set of vertices is $S$, and whose set of edges is $E(G) \cap \binom{S}{2}$.

We consider two notions of rooted graphs, one where roots are ordered, and another where they are not. A **tuple-rooted graph** $(G, \boldsymbol{v})$ is a graph $G$ together with an ordered tuple of vertices $\boldsymbol{v} \in V(G)^*$ called **roots**. An **isomorphism** between tuple-rooted graphs $(G, \boldsymbol{v}), (H, \boldsymbol{u})$ where $\boldsymbol{v} = (v_1, \ldots, v_\ell)$, $\boldsymbol{u} = (u_1, \ldots, u_\ell)$ is an isomorphism between the underlying graphs $G, V$ that sends each $v_i$ to the corresponding root $u_i$. Similarly, a **set-rooted graph** $(G, S)$ is a graph $G$ together with a set $S \subseteq V(G)$, and an **isomorphism** between set-rooted graphs $(G, R), (H, S)$ an isomorphism between $G$ and $H$ that induces an isomorphism between $G[R]$ and $H[S]$. We usually call tuple-rooted and set-rooted graphs simply rooted graphs, and the concrete notion should be clear from context. As usual, a **unlabeled rooted** (tuple-rooted or set-rooted) **graph** is just an isomorphism class of rooted graphs.

Given a vertex $v \in V(G)$, $\deg_G(d)$ stands for its **degree**. The **distance** between two vertices $v, u$ is defined as the minimum number of edges required to reach $u$ from $v$, or infinity if $u$ and $v$ belong to different connected components, and is denoted by $d_G(v, u)$. Given sets $U, U' \subseteq V(G)$, $d_G(U, U')$ and $d_G(v, U)$ are defined in the usual way. Given $r \in \mathbb{N}$, the (closed) **disk** $D_G(v, r) \subseteq V(G)$ is the set of vertices $u$ with $d_G(u, v) \leq r$, and the **neighbourhood** $N_G(v, r)$ is the induced subgraph $G[D_G(v, r)]$, rooted at $v$. We usually drop the subscript $G$ when there is no room for ambiguity. The disks $D_G(\boldsymbol{v}, r)$ and $D_G(S, r)$ where $\boldsymbol{v}$ is a tuple of vertices and $S$ is a set of vertices are defined analogously, and the neighbourhoods $N_G(\boldsymbol{v}, r)$, $N_G(S, r)$ correspond to the induced graphs $G[D_G(\boldsymbol{v}, r)]$ and $G[D_G(S, r)]$ rooted at $\boldsymbol{v}$ and $S$, respectively.

A **rooted tree** is a pair $(T, \vartheta)$, where $T$ is a tree (i.e., a connected acyclic graph) and $\vartheta \in V(T)$ is a vertex called its **root**. The **height** of $(T, \vartheta)$ is the maximum distance from a vertex $u \in V(T)$ to the root $\vartheta$. Two rooted trees are **isomorphic**, written $(T, \vartheta) \simeq (T', \vartheta')$, if there is a root-preserving isomorphism between the corresponding trees. As before, an **unlabeled rooted tree** is an isomorphism class of rooted trees. We write $\mathbb{T}$ for the set of unlabeled rooted trees, and $\mathbb{T}(r)$ for the set of unlabeled rooted trees with height at most $r$.

During this thesis we use a notion of rooted forest where roots are ordered. Formally, a **rooted forest** $(F, \boldsymbol{\vartheta})$ is a rooted graph, where $F$ is a disjoint union of trees $T_1, \ldots, T_\ell$ together with an ordered tuple of roots $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_\ell)$, one belonging to each tree. Equivalently, to empathise the ordered nature of a rooted forest, we can also identify $(F, \boldsymbol{\vartheta})$ with the tuple of rooted trees $(T_i, \vartheta_i)_{i \in [\ell]}$.

Given $\ell \geq 3$, $\boldsymbol{C_\ell}$ stands for the $\ell$**-cycle**. A **unicycle** is a connected graph containing a single cycle. An $r$-**unicycle** is an unicycle where the cycle has at most $2r+1$ vertices, and where no vertex lies at distance greater than $r$ from the cycle. We write $\mathbb{U}$ and $\mathbb{U}(r)$ for the sets of unlabeled unicycles and unlabeled $r$-unicycles. A **fragment** is a graph whose components are all unicycles. Similarly, given $r \in \mathbb{N}$, an $r$-**fragment** is a fragment where each component is a $r$-unicycle. We write $\mathbb{F}$ and $\mathbb{F}(r)$ for the sets of unlabeled fragments and unlabeled $r$-fragments. The **fragment of a graph** $G$, denoted $\mathrm{Frag}(G)$, is the union of unicyclic components in $G$.

## Hypergraphs

Given an integer $\ell \geq 2$, an $\ell$-**uniform hypergraph** $G$ is a pair $(V(G), E(G))$, where $V(G)$ is a (finite) set of vertices and $E(G) \subseteq \binom{V(G)}{\ell}$ is a set of (hyper-)edges. Hence, 2-uniform hypergraphs are just graphs. The **excess** of $G$ is the quantity $\mathrm{ex}(G) = (\ell-1)|E(G)| - |V(G)|$. The hypergraph $G$ is called a **cycle** if it is connected and each edge share exactly two vertices with the union of all other edges, and a **unicycle** if it is connected and it contains a single cycle. Observe that, unlike the case of graphs, if $\ell > 2$ there exists a cycle with only two hyper-edges. Equivalently, $G$ is a unicycle if $\mathrm{ex}(G) = 0$, and is a cycle if, in addition to that, $\mathrm{ex}(H) \neq 0$ for any proper sub-hypergraph $H$ of $G$. All other graph notions extend to hypergraphs using their word-by-word definitions. In the context of $k$-uniform hypergraphs, $\mathbb{T}_k(r), \mathbb{U}_k(r), \mathbb{F}_k(r)$ stand for the sets of unlabeled rooted trees with height at most $r$, of unlabeled $r$-unicycles, and of unlabeled $r$-fragments, respectively.

## Relational Structures

With regards to relational structures, we deviate slightly from the commonly-used nomenclature and notation. Instead, we adopt graph-like terminology and symbols to maintain consistence with other parts of the thesis.

A **signature** $\sigma = \{E_1, \dots E_s, \mathfrak{c}_1, \dots, \mathfrak{c}_t\}$ is a (finite) collection of **relation symbols** $E_1, \dots E_s$ and **constant symbols** $\mathfrak{c}_1, \dots, \mathfrak{c}_t$, where each relation symbol $E_i$ is paired to a natural number $\mathrm{ar}(E_i)$ called its **arity**. A signature is **relational** if it only contains relation symbols. A $\sigma$-**structure** $G$ is a collection containing a **set of vertices** $V(G)$, a **set of (hyper-)edges** $E_i(G) \subseteq (V(G))^{\mathrm{ar}(E_i)}$ for each relation symbol $E_i \in \sigma$, and a vertex $\mathfrak{c}_i^G \in V(G)$ for each constant symbol $\mathfrak{c}_i \in \sigma$. The sets $E_i(V)$, and the vertices $\mathfrak{c}_i^G$ are said to **interpret** the symbols $E_i$ and $\mathfrak{c}_i$, respectively. Given a $\sigma$-structure, we write $E(G)$ for the disjoint union of all relations $E_i(G)$. Unlike the case of graphs and hypergraphs, edges in a $\sigma$-structure $G$ are ordered tuples of vertices, rather than sets of vertices. A consequence of this is that a single vertex $v$ can appear multiple times in an edge $e \in E_i(G)$. The **excess** of $G$ is defined as $\mathrm{ex}(G) = \left[\sum_{E_i} (\mathrm{ar}(E_i) - 1)|E_i(G)|\right] - V(G)$, where the sum inside the brackets ranges over all relation symbols $E_i \in \sigma$.

In the context of $\sigma$-structures, it is more comfortable to use the excess-based definitions of many graph notions. A $\sigma$-structure $G$ is called a **tree** if it is connected and $\mathrm{ex}(T) = -1$. Alternatively, if $G$ is connected but $\mathrm{ex}(G) = 0$, then $G$ is a **unicycle**. Finally, $G$ is a **cycle** if it is a unicycle and $\mathrm{ex}(H) < 0$ for all its proper substructures $H$. Observe that now it is possible to have cycles with just a single edge, which we call **loops**. All other graph notions introduced before can also be extended to $\sigma$-structures verbatim. In the setting of $\sigma$-structures, we define $\mathbb{T}_\sigma(r), \mathbb{U}_\sigma(r), \mathbb{F}_\sigma(r)$ as the sets of unlabeled rooted trees with height at most $r$, of unlabeled $r$-unicycles, and of unlabeled $r$-fragments, respectively.

Graphs and $k$-uniform hypergraphs can be interpreted as structures where the signature $\varphi$ consists of a single relation symbol $E$, which is binary for graphs and $k$-ary for $k$-uniform hypergraphs.

## 1.3 Logic Background

### First-Order Logic

We give a brief introduction to first-order (FO) logic. A more thorough exposition can be found at [43]. Fix a signature $\sigma$. We assume a countably infinite set of **variables**, which we usually denote by $x, y, z$ with subscripts and/or superscripts. A **term** is either a variable or a constant symbol $\mathfrak{c}_i \in \sigma$. The set of **first-order formulas** $\mathrm{FO}[\sigma]$ consists of all finite strings that result from applying the following rules in succession:

**FO1:** If $t_1, t_2$ are terms, then $t_1 = t_2$ is a formula.

**FO2:** If $t_1, \ldots, t_k$ are terms and $E_i \in \sigma$ is a $k$-ary relation symbol, then $E_i(t_1 \ldots t_k)$ is a formula.

**FO3:** If $\varphi, \psi \in \mathrm{FO}[\sigma]$ are formulas, so are $\neg\varphi$, $\varphi \vee \psi$ and $\varphi \wedge \psi$.

**FO4:** If $\varphi \in \mathrm{FO}[\sigma]$ is a formula and $x$ a variable, then $\exists x\varphi$ and $\forall x\varphi$ are formulas as well.

The **first-order language of graphs** $\mathrm{FO}_g$ consists of the formulas $\mathrm{FO}[\sigma_g]$, where the signature $\sigma_g$ consists of a single binary relation symbol $E$.

We will be liberal in our use of parentheses to make formulas more readable. We shorten expressions like $\forall x_1 \forall x_1 \ldots \forall x_t$ to $\forall x_1 \ldots x_t$, and similarly with $\exists x_1 \exists x_2 \ldots \exists x_t$. When discussing formulas, we use $\equiv$ as a shorthand for "is written as" (for example, in expressions like $\varphi \equiv \exists x\psi$). We use the standard shorthands $\varphi \to \psi$ and $\varphi \leftrightarrow \psi$, which stand for $\neg\varphi \vee \psi$ and $(\varphi \to \psi) \wedge (\psi \to \varphi)$. A formula described by the rule **FO1** above is called **atomic**. A formula $\varphi \in \mathrm{FO}[\sigma]$ Alternatively, $\varphi$ is in **prenex normal form** if all the quantifiers (that is, $\forall$, $\exists$) occur at the beginning of the formula.

An occurrence of some variable $x$ in a formula $\varphi \in \mathrm{FO}[\sigma]$ is called **bounded** if it lies within the scope of a quantifier. Otherwise, it is called **free**. A variable $x$ is said to be free in $\varphi$ if there is some free occurrence of $x$ in $\varphi$. We usually write $\varphi(x_1, \ldots, x_n)$ for a formula $\varphi \in \mathrm{FO}[\sigma]$ to denote that free variables in $\varphi$ are among $x_1, \ldots, x_n$. A **sentence** is a formula with no free variables.

Let $G$ be a $\sigma$-structure, and $\varphi(\boldsymbol{x}) \in \mathrm{FO}[\sigma]$ a formula, where $\boldsymbol{x} = x_1 \ldots x_l$. We define the relation $G \models \varphi(\boldsymbol{v})$, by induction on $\varphi$'s syntax. Here $\boldsymbol{v} \in V(G)^\ell$ is some tuple of elements in $G$, and $G \models \varphi(\boldsymbol{v})$ is read as "$A$ **satisfies** $\varphi(\boldsymbol{v})$". Given a term $t \equiv x_i$ for some $i \in [\ell]$, its **value** $t^G(\boldsymbol{v})$ is defined as $v_i$. Otherwise, if $t$ is some constant symbol $\mathfrak{c}_i \in \sigma$, the value $t^G(\boldsymbol{v})$ is $\mathfrak{c}_i^G$. The relation $G \models \varphi(\boldsymbol{v})$ is given by:

- If $\varphi \equiv (t_1 = t_2)$, then $G \models \varphi(\boldsymbol{v})$ holds whenever $t_1^G(\boldsymbol{v}) = t_2^G(\boldsymbol{v})$
- If $\varphi \equiv E_i(t_1, \ldots, t_{\mathrm{ar}(E_i)})$ then $G \models \varphi(\boldsymbol{v})$ holds whenever $(t_1^G(\boldsymbol{v}), \ldots, t_{\mathrm{ar}(E_i)}^G(\boldsymbol{v})) \in E_i(G)$.

- $G \models \neg\varphi(\boldsymbol{v})$ iff $G \models \varphi(\boldsymbol{v})$ does not hold.
- $G \models \varphi(\boldsymbol{v}) \wedge \psi(\boldsymbol{v})$ iff both $G \models \varphi(\boldsymbol{v})$ and $G \models \psi(\boldsymbol{v})$.
- $G \models \varphi(\boldsymbol{v}) \vee \psi(\boldsymbol{v})$ iff either $G \models \varphi(\boldsymbol{v})$ or $G \models \psi(\boldsymbol{v})$.
- If $\varphi \equiv \exists x_{\ell+1}\psi(\boldsymbol{x}, x_{\ell+1})$, then $G \models \varphi(\boldsymbol{v})$ iff $G \models \psi(\boldsymbol{v}, v_{\ell+1})$ for some $v_{\ell+1} \in V(G)$.
- If $\varphi \equiv \forall x_{\ell+1}\psi(\boldsymbol{x}, x_{\ell+1})$, then $G \models \varphi(\boldsymbol{v})$ iff $G \models \psi(\boldsymbol{v}, v_{\ell+1})$ for all $v_{\ell+1} \in V(G)$.

### Ehrenfeucht-Faïssé (EF) Games

Two $\sigma$-structures $G_0, G_1$ are said to be $k$-**equivalent** for some $k \in \mathbb{N}$, written $G_0 \equiv_k G_1$ if both satisfy the same sentences $\varphi \in \mathrm{FO}[\sigma]$ with $\mathrm{qr}(\varphi) \leq k$. A remarkable fact is that the number of $\equiv_k$-classes is finite for any $\sigma$ [43]. In principle, it seems like proving $k$-equivalence between two structures would involve determining which sentences with quantifier rank $k$ hold in each of them. However, it turns out that this relation admits a much nicer characterization in terms of pebble games.

**Definition 1.1.** Let $\boldsymbol{v}^0 = (v_1^0, \dots, v_k^0)$ and $\boldsymbol{v}^1 = (v_1^1, \dots, v_k^1)$ be tuples of vertices lying in $\sigma$-structures $G_0, G_1$. We say that $(\boldsymbol{v}^0, \boldsymbol{v}^1)$ defines a **partial isomorphism** between $G_0$ and $G_1$ if the following conditions hold:

- $v_i^0 = v_j^0$ iff $v_i^1 = v_j^1$ for all $i, j \in [k]$.
- $v_i^0 = \mathfrak{c}^{G_0}$ iff $v_i^1 = \mathfrak{c}^{G_1}$ for all $i \in [k]$ and all constants $\mathfrak{c} \in \sigma$.
- $(v_{i_1}^0, \dots, v_{i_{\mathrm{ar}(E)}}^0) \in E(G_0)$ iff $(v_{i_1}^1, \dots, v_{i_{\mathrm{ar}(E)}}^1) \in E(G_1)$ for all relations $E \in \sigma$ and all $i_0, \dots, i_{\mathrm{ar}(E)} \in [k]$.

The $k$ round **Ehrenfeucht-Faïssé** (EF) game on the $\sigma$-structures $G_0, G_1$, denoted as $\mathrm{EF}_k(G_0, G_1)$ is defined as follows. The game is held between two players, Duplicator, who tries to establish $k$-equivalence between $G_0$ and $G_1$, and Spoiler, who attempts to refute it. The number of rounds $k$ is known from the begging to both players. In the $i$-th round, Spoiler picks a vertex $v_i^j \in V(G_j)$ in either structure $j = 0, 1$, and Duplicator responds by choosing another vertex $v_i^{1-j} \in V(G_{1-j})$ in the other structure. This way, at the end of the game vertices $v_1^0, \dots, v_k^0$ have been chosen in $G_0$, and $v_1^1, \dots, v_k^1$ in $G_1$. Let $\mathfrak{c}_1, \dots \mathfrak{c}_\ell$ be all constants in $\sigma$. The game is won by Duplicator if the tuples $(\mathfrak{c}_1^{G_i}, \dots \mathfrak{c}_\ell^{G_i}, v_1^i, \dots v_k^i)$ with $i = 0, 1$ define a partial isomorphism between $G_0$ and $G_1$. Otherwise Spoiler wins.

This is a complete information game where either Spoiler or Duplicator has a winning strategy. Moreover, the following holds [43][Theorem 3.9]:

**Theorem 1.1** (Ehrenfeucht-Faïssé)**.** *Given two $\sigma$-structures $G_0$, $G_1$, Duplicator has a winning strategy in $\mathrm{EF}_k(G_0, G_1)$ for some $k \in \mathbb{N}$ if and only if $G_0 \equiv_k G_1$.*

## 1.4 Probability Background

We assume familiarity with the basics from probability theory. Given an event $A$, we write $\Pr(A)$ for its **probability**. We say that a sequence of events $A_n$ holds **with high probability** (w.h.p.) or **asymptotically almost surely** (a.a.s.) if $\Pr(A_n) = 1 - o(1)$. Random variables are assumed to be either integer or real-valued unless stated otherwise. The **mean** or **expected value** of a random variable $X$ is written $\mathrm{E}[X]$, while $\mathrm{Var}(X)$ stands for its **variance**. Given $i \in \mathbb{N}$, the $i$-th **moment** of $X$ is the quantity $E[X^i]$, whereas the $i$-th **factorial moment** of $X$ is $\mathrm{E}[(X)_i]$. Given a distribution $D$ over some measurable space, we may also use $D$ to denote a random variable with distribution $D$. Otherwise, we may explicitly write $X \sim D$ to denote that $D$ is the distribution of some random variable $X$. We write $\mathrm{Bin}(n,p)$ for the **binomial distribution** with number of trials $n \in \mathbb{N}$ and success probability $p \in [0,1]$. That is, $\Pr(\mathrm{Bin}(n,p) = i) = \binom{n}{i} p^i (1-p)^{n-i}$ for all $0 \le i \le n$. Similarly, $\mathrm{Pois}_\lambda$ denotes the **Poisson distribution** with mean $\lambda \in (0,\infty)$. I.e., the one given by $\Pr(\mathrm{Pois}_\lambda = i) = e^{-\lambda} \frac{\lambda^i}{i!}$ for all $i \in \mathbb{N}$. Given two distributions $D_1, D_2$ over measurable spaces $S_1, S_2$, the **product distribution** $D_1 \times D_2$ is the one satisfying $\Pr(D_1 \times D_2 \in A_1 \times A_2) = \Pr(D_1 \in A_1)\Pr(D_2 \in A_2)$, where $A_1 \subseteq S_1$ and $A_2 \subseteq S_2$ are measurable sets.

Below we introduce the two basic tools of the probabilistic method, named after the first and the second moment of a random variable.

**Theorem 1.2** (Markov's Inequality)**.** *Let $X$ be a non-negative random variable with finite expected value. Then $\Pr(X \ge a) \le \mathrm{E}[X]/a$ for all $a > 0$.*

**Corollary 1.1** (First Moment Method)**.** *Let $X(n)$ be sequence of non-negative random variables. If $\mathrm{E}[X(n)] = o(1)$ then $X(n) = 0$ w.h.p.*

**Theorem 1.3** (Chebyshev's Inequality)**.** *Let $X$ be a random variable with finite variance. Then $\Pr(|X - \mathrm{E}[X]| \ge a) \le \mathrm{Var}(X)/a^2$ for all $a > 0$.*

**Corollary 1.2** (Second Moment Method)**.** *Let $X(n)$ be a sequence of random variables with finite variance. Suppose that (1) $\mathrm{E}[X(n)]$ diverges to infinity and (2) $\mathrm{Var}(X(n)) = o(\mathrm{E}[X(n)]^2)$, or equivalently, $\mathrm{E}\left[X(n)^2\right] \sim \mathrm{E}[X(n)]^2$. Then a.a.s. $X(n)/\mathrm{E}[X(n)] \in [1-\epsilon, 1+\epsilon]$ for all $\epsilon > 0$. In particular, a.a.s. $X(n) > 0$.*

Given a sequence of random variables $X(n)$, and another variable $X$, we say that $X(n)$ **converges in distribution to** $X$, written $X(n) \xrightarrow{d} X$, if $\Pr(X(n) \le x) = \Pr(X \le x) + o(1)$ for all $x$ that are continuity points of $F(x) = \Pr(X = x)$. When $X$ is integer-valued, this is equivalent to $\Pr(X(n) = x) = \Pr(X = x) + o(1)$ for all integers $x$.

Whereas the results above help us estimating the size of a random variable $X(n)$, the following one is useful for determining its asymptotic distribution under some precise conditions [10, Theorem 1.23].

**Theorem 1.4** (Method of Moments)**.** *Let $k \in \mathbb{N}$. For each $n \in \mathbb{N}$, let $X_1(n), \dots, X_k(n)$ be random variables over the same measurable space. Suppose that there are real positive constants $\lambda_1, \dots, \lambda_k$ such that for all $a_1, \dots, a_k \in \mathbb{N}$*

$$\lim_{n\to\infty} \mathrm{E}\left[\prod_{i=1}^{k} \binom{X_i(n)}{a_i}\right] = \prod_{i=1}^{k} \frac{\lambda_i^{a_i}}{a_i!}.$$

*Then $(X_1(n), \dots, X_k(n)) \xrightarrow{d} \mathrm{Pois}_{\lambda_1} \times \cdots \times \mathrm{Pois}_{\lambda_k}$.*

In other situations we are able to determine the asymptotic distribution of a sequence $X(n)$ by giving another sequence of random variables $Y(n)$ with known "limit", and showing that $X(n)$ and $Y(n)$ are similar. The **total variation distance** between two variables $X, Y$ taking values on the same measurable space is defined as

$$\mathrm{d}_{TV}(X, Y) = \sup_A |\Pr(X \in A) - \Pr(Y \in A)|\,,$$

where $A$ ranges over all measurable sets. Below are some easy to derive relations between total variation distances and convergence in distribution.

**Lemma 1.1.** *Let $X(n), X'(n)$ be sequences of random variables. The following statements hold.*

*(1) If $\mathrm{d}_{TV}(X(n), X) = o(1)$ for some variable $X$, then $X(n) \xrightarrow{d} X$.*
*(2) If $X'(n) \xrightarrow{d} X$ for some variable $X$ and $\mathrm{d}_{TV}(X(n), X'(n)) = o(1)$, then $X(n) \xrightarrow{d} X$.*

A useful method to bound the total variation distance between two variables is through couplings. A **coupling** between two random variables $X, Y$ is another vector-valued random variable $(X', Y')$ where the marginal distributions of $X'$ and $Y'$ coincide with those of $X$ and $Y$.

**Lemma 1.2** (Coupling Lemma)**.** *Given two random variables $X, Y$, and a coupling $(X',\ Y')$ between them,* $\Pr(X' \neq Y') \geq \mathrm{d}_{TV}(X, Y)$.

We will use couplings mostly for determining the distribution of small neighbourhoods in various random models of graphs or relational structures. Our models of interest are *sparse*, meaning that generally small neighbourhoods "look like" random forests. This can be made precise via the notion of *local convergence* [67] (which we mention, but will not cover thoroughly in this thesis). The random forests that approximate the neighborhoods are given by so-called branching processes, which we introduce below.

**Definition 1.2.** Let $r \in \mathbb{N}$ be fixed and $\widetilde{\mathcal{X}}, \mathcal{X}$ be distributions over $\mathbb{N}$. Let $(X_i)_{i \in \mathbb{N}}$ be mutually independent random variables, where $X_i$ has distribution $\widetilde{\mathcal{X}}$ for $1 \leq i \leq r$ and distribution $\mathcal{X}$ for $i > r$. A **$r$-root branching process with root offspring distribution $\widetilde{\mathcal{X}}$ and general offspring distribution $\mathcal{X}$** is a sequence of random variables $\mathcal{BP} = (Y_i)_{i=0}^{\infty}$ satisfying the following recursive relation:

$$Y_0 = r, \qquad Y_n = \begin{cases} Y_{n-1} + X_n - 1, & \text{if } Y_{n-1} > 0 \\ 0 \text{ otherwise.} \end{cases}$$

The variables $X_i$ are referred to as **offspring variables**.

A branching process defined this way can be interpreted as the exploration of a random rooted forest $\mathcal{F}$ formed by $r$ rooted trees $\mathcal{T}_1, \dots \mathcal{T}_r$ where (1) the number of children attached to each vertex are mutually independent variables, (2) the number of children attached to the $i$-th root has distribution $\widehat{\mathcal{X}}_i$, and (3) the number of children attached to each non-root vertex has distribution $\mathcal{X}$. In the exploration, we expose $\mathcal{F}$'s vertices in breadth-first order until the whole forest has been exposed. During this procedure, an exposed vertex is called **dead** if its children have been exposed as well, and is called **active** otherwise. We begin by exposing all $\mathcal{F}$'s roots. At each step, until no active vertices are left, we pick a remaining active vertex in depth-first order and expose its children. This way, the variable $Y_n$ counts the active vertices at the end of the $n$-th round, and $X_n$ counts the newly-revealed vertices.

We make this interpretation of branching processes as random rooted forests explicit in the following definition.

**Definition 1.3.** A **Ulam-Harris forest** with $r \in \mathbb{N}$ roots is a labeled rooted forest $F = (T_i, \vartheta_i)_{i\in[r]}$ where vertices in $V(T_i)$ are words $\omega \in \mathbb{N}^*$ satisfying:

- The root $\vartheta_i$ is the singleton $i$.
- The children of a vertex $\omega \in V(T_i)$, are $\omega 1, \omega 2, \dots, \omega d_\omega$, where $d_\omega$ denotes $\omega$'s number of children.

Additionally, we consider the lexicographical order over vertices in $V(F) = \bigcup_i V(T_i)$.

The Ulam-Harris forest $F_{\mathcal{BP}}$ given by a branching process $\mathcal{BP}$ with offspring variables $X_i$ is constructed inductively by setting $d_\omega = X_i$, where $\omega$ is the $i$-th vertex of $F_{\mathcal{BP}}$. We identify branching processes $\mathcal{BP}$ with their corresponding Ulam-Harris forests $F_{\mathcal{BP}}$.

A well understood feature of branching processes is that, depending on their offspring distributions, they may yield "infinite" forests with non-zero probability. However, as stated in the next lemma, it is a well-known fact that a branching process yields a well-defined distribution over finite forests as long as its general offspring distribution has mean smaller than one.

**Lemma 1.3.** *Let $r \in \mathbb{N}$ and let $\mathcal{BP}$ be a $r$-root branching process, and let $\mathbb{T}^r$ be the set of unlabeled finite rooted forests with $r$ roots. Suppose that $\mathcal{BP}$'s general offspring distribution $X$ satisfies that $\mathrm{E}[X] < 1$. Then $\sum_{F\in\mathbb{T}^r} \Pr(\mathcal{BP} \simeq F) = 1$.*

A common issue that arises when dealing with asymptotic quantities is that of exchanging limit and sum operators. The next definition characterizes the situations where this exchange is possible.

**Definition 1.4.** Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of functions $f_n : S \to [0, \infty)$ where $S$ is a countable set. The sequence $(f_n)_{n\in\mathbb{N}}$ is **tight** if for every $\epsilon > 0$ there exists a finite $T \subset S$ satisfying $\sum_{s\notin T} f_n(s) < \epsilon$ for all $n$.

**Lemma 1.4.** *Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of functions $f_n : S \to [0, \infty)$ where $S$ is a countable set. Suppose that for each $s \in S$, $f(s) = \lim_{n\to\infty} f_n(s)$ exists and is finite. Additionally, suppose that $\sum_{s\in S} f(s)$ is finite. Then $\lim_{n\to\infty} \sum_{s\in S} f_n(s) = \sum_{s\in S} f(s)$ if and only if $(f_n)_{n\in\mathbb{N}}$ is tight.*

# Chapter 2

# Convergence Law for Sparse Random Structures

The results presented during this chapter were partially obtained during the author's master thesis and later published in [40]. The presentation of this chapter differs from [40] in that we are able to significantly simplify some proofs using different probabilistic techniques. This is further discussed at Section 2.2.

Given a formal language $\mathcal{L}$, and a sequence of random structures $(\mathcal{G}_n)_{n\in\mathbb{N}}$ growing in size, we say that $(\mathcal{G}_n)_{n\in\mathbb{N}}$ satisfies a **convergence law** with respect to $\mathcal{L}$ (or just a $\mathcal{L}$-convergence law) if $\lim_{n\to\infty} \Pr(\mathcal{G}_n$ satisfies $P)$ exists for all properties $P$ that are expressible in $\mathcal{L}$. Furthermore, if the only possible values for this limit are 0 and 1, we say that a **zero-one law** with respect to $\mathcal{L}$ holds. Convergence laws have been widely studied in the context of $\mathrm{FO}_g$ logic and the binomial random graph $\mathcal{G}(n,p)$. Here, the seminal result by Fagin [21] and Glebsky [28], independently, states that for any fixed $p \in [0,1]$, a zero-one law with respect to $\mathrm{FO}_g$ holds in $\mathcal{G}(n,p)$.

A more complex situation arises when one allows $p = p(n)$ to be a decreasing function of $n$. In this setting, a $\mathrm{FO}_g$-convergence law holds when $p(n) = o(n^{-1+\epsilon})$ for all $\epsilon > 0$, under few additional regularity conditions [49, 62, 44]. Moreover, the probabilities $p(n)$ in this range for which a zero-one law holds are known [44]. In contrast, the dense regime (i.e., $p(n) = \Omega(n^{-1+\epsilon})$ for some $\epsilon > 0$) is much harder to analyze. Here, even for seemingly tame probabilities, like $p(n) = n^{-\alpha}$ with $\alpha$ a rational number in $(0,1)$, the convergence law fails. However, if $p(n) = n^{-\alpha}$, where $\alpha \in (0,1)$ is irrational, then a zero-one law holds [61]. Apart from this, in the dense range there are fundamental obstacles to characterizing the $p(n)$ for which a zero-one law holds: when $\alpha \in (0,1)$ is a rational number, there are sequences $p(n)$ of the form $n^{-\alpha+o(1)}$ achieving a zero-one law, but those $p(n)$ are non-computable [44].

Other work on the subject studies the logical properties of other random graph models, like graphs with given degree sequences [45, 46, 30], uniform graphs from minor-closed classes [31, 38], or geometric graphs [51, 29]. Another direction is to consider different languages other than $\mathrm{FO}_g$. Most work in this regard studies fragments of second-order logic, like monadic second-order logic [35], existential monadic second-order logic [34, 42], and other quantifier classes [36, 37, 36]. For the most part, results about limit laws are stated either for uniform relational structures or for various models of random graphs. Other kinds of combinatorial objects have also been studied in a more limited fashion, including random words [47], unary functions [48], or permutations [23].

The main goal of this chapter is to establish a FO-convergence law in the sparse range of a general model random structures. This result generalizes the $\mathrm{FO}_g$-convergence law for $\mathcal{G}(n,p)$ when $p(n) \sim c/n$, obtained by Lynch in [49], as well as an unpublished result [58] that extends Lynch's result to uniform hypergraphs.

Throughout this chapter $\sigma = \{E_1, \dots, E_{|\sigma|}\}$ denotes some fixed relational signature, where $a_i = \mathrm{ar}(E_i)$ for $i = 1, \dots, s$, and all relations have arity at least two (i.e., $a_i \geq 2$ for all $i$). We refer to $\sigma$-structures simply as structures.

Let $\boldsymbol{p} = \boldsymbol{p}(n) = (p_1(n), \dots, p_{|\sigma|}(n))$ be a tuple of probabilities which depend on the parameter $n$. The binomial **random structure** $\mathcal{G}^\sigma(n, \boldsymbol{p})$ has vertex set $[n]$, and for all $1 \leq i \leq |\sigma|$, each tuple $e \in ([n])^{a_i}$ satisfies the predicate $E_i$ with probability $p_i$ independently. We are interested in dealing with **sparse** structures. More concretely, we want to study the case where the relations $E_1, \dots, E_{|\sigma|}$ grow linearly with the number of vertices. From now on, we assume that $\boldsymbol{p}$ is a tuple of decaying probabilities satisfying $p_i \sim c_i/n^{\mathrm{ar}(E_i)-1}$ for some constants $\boldsymbol{c} = (c_1, \cdots c_{|\sigma|}) \in (0,\infty)^{|\sigma|}$.

Our main result, Theorem 2.1, states that in this situation $\mathcal{G}^\sigma(n, \boldsymbol{p})$ satisfies a $\mathrm{FO}[\sigma]$-convergence law. Similarly to Lynch's result in [49], we also study the limit probability of each sentence $\varphi \in \mathrm{FO}[\sigma]$ as a function of $c_1, \dots, c_{|\sigma|}$, and give a family of well-behaved expressions to which those limits belong. We define this family below.

**Definition 2.1.** The set Expr consists of all expressions with parameters $c_1, \dots, c_{|\sigma|}$ formed by a finite application of the following rules:

1. For all $1 \leq i \leq |\sigma|$, $c_i \in$ Expr.
2. If $\lambda \in (0, \infty)$, $\omega \in$ Expr, then $\lambda\omega \in$ Expr.
3. If $\omega \in$ Expr, $n \in \mathbb{N}$, then $\Pr(\mathrm{Pois}(\omega) = n)$ and $\Pr(\mathrm{Pois}(\omega) \geq n)$ belong to Expr.
4. If $\omega_1, \omega_2 \in$ Expr, then $\omega_1\omega_2, \omega_1/\omega_2, \omega_1 + \omega_2$ all belong to Expr.

Now we are in conditions of stating the main theorem of this chapter.

**Theorem 2.1.** *Let $\boldsymbol{p} = (p_1, \dots, p_{|\sigma|})$ be a tuple of probabilities satisfying $p_i(n) \sim c_i/n^{\mathrm{ar} E_i - 1}$ for some real positive constants $\boldsymbol{c} = (c_1, \dots, c_{|\sigma|})$. Consider a* $\mathrm{FO}[\sigma]$*-sentence $\varphi$. Then the limit*

$$p_\varphi(\boldsymbol{c}) = \lim_{n\to\infty} \Pr(\mathcal{G}^\sigma(n, \boldsymbol{p}) \models \phi)$$

*exists and depends only on $\varphi$ and **c**. Moreover, for any fixed $\phi$, $p_\phi(\boldsymbol{c})$ is a finite (possibly empty) sum of expressions belonging to* Expr.

Below we sketch the proof of this result. Afterwards we give an outline of the chapter's structure.

### Sketch of the Proof

The general strategy to show a FO convergence law for some sequence of structures $(\mathcal{G}_n)_{n\in\mathbb{N}}$ is the following. Given a quantifier rank $k$, one finds classes $\mathcal{C}_1, \dots, \mathcal{C}_\ell$ satisfying (I) $G_0 \equiv_k G_1$ for all

$G_0, G_1$ in the same class $\mathcal{C}_i$, (II) $\Pr(\mathcal{G}_n \in C_i)$ converges for all $i$, and (III) w.h.p. $\mathcal{G}_n$ belongs to any of $\mathcal{C}_1, \dots, \mathcal{C}_\ell$. From those three facts follows that $\Pr(\mathcal{G} \models \varphi)$ converges for all sentences $\varphi$ with $\mathrm{qr}(\varphi) \leq k$. Indeed, by (III)

$$\Pr(\mathcal{G} \models \varphi) = \sum_{i=0}^{\ell} \Pr(\mathcal{G}_n \models \varphi \mid \mathcal{G}_n \in \mathcal{C}_i) \Pr(\mathcal{G}_n \in \mathcal{C}_i) + o(1).$$

By (I), $\Pr(\mathcal{G}_n \models \varphi \mid \mathcal{G}_n \in \mathcal{C}_i)$ is either zero or one for each $i$, and by (II) each factor $\Pr(\mathcal{G}_n \in \mathcal{C}_i)$ converges, so the statement follows.

We sketch how use this proof idea to show Theorem 2.1. We follow similar ideas to those used in [49] for establishing the $\mathrm{FO}_g$-convergence law in $\mathcal{G}(n,p)$ when $p(n) \sim c/n$. Let $\boldsymbol{p} = (p_1, \dots, p_{|\sigma|})$ be as in Theorem 2.1. First, we obtain a local description of $\mathcal{G}^\sigma(n, \boldsymbol{p})$. The $r$-**core** of the structure $\mathcal{G}^\sigma(n, \boldsymbol{p})$, denoted $\mathrm{Core}_n|_r$, is the $r$-neighbourhood of all its cycles of length at most $2r+1$. Observe that $\mathrm{Core}_n|_r$ is not necessarily a fragment: it may have complex components if there are short cycles at a small distance from each other. However, w.h.p. this is not the case. Fix $r \in \mathbb{N}$. The following hold:

(1) For each $r \in \mathbb{N}$ there is a random tree $\mathcal{T}|_r$ such that, the $r$-neighborhoods of $\ell$ uniformly chosen vertices $v_1, \dots, v_\ell \in [n]$ in $\mathcal{G}^\sigma(n, \boldsymbol{p})$ converge in distribution to $(\mathcal{T}|_r)^\ell$ (i.e., $\ell$ disjoint independent copies of $\mathcal{T}|_r$).

(2) For each $r \in \mathbb{N}$, there is a distribution $\Gamma|_r$ over fragments (recall the definition of fragment in Section 1.2) such that $\mathrm{Core}_n|_r \xrightarrow{d} \Gamma|_r$.

Fact (1) can be obtained similarly to well-known local-convergence results in $\mathcal{G}(n,p)$ [67], and (2) is a corollary of (1) and a characterization of the small-cycle distribution of $\mathcal{G}^\sigma(n, \boldsymbol{p})$. Given a quantifier rank $k \in \mathbb{N}$, we define equivalence relations $\equiv_k^{\mathrm{Ly}}$ over rooted trees and over fragments that refine logical equivalence $\equiv_k$. Roughly, using facts (1) and (2) above, one can show that w.h.p. $\mathcal{G}^\sigma(n, \boldsymbol{p})$ is $r$-**simple**, meaning that its short cycles are far away (and $\mathrm{Core}_n|_r$ is a fragment), and $(k,r)$-**rich**, meaning that, for any $\equiv_k^{\mathrm{Ly}}$-class $\mathcal{C}$ of trees, there are as many neighbourhoods "as needed" that belong to $\mathcal{C}$. Let $r = (3^k - 1)/2$. An application of EF games shows that the $\equiv_k$-class of a $r$-simple, $(k,r)$-rich structure $G$ depends only on the $\equiv_k^{\mathrm{Ly}}$-class of its $r$-core. Finally, we show that the probability that $\Pr(\mathrm{Core}_n|_r \in \mathcal{C}_i)$ converges to some expression in Expr for any $\equiv_k^{\mathrm{Ly}}$-class of fragments $\mathcal{C}_i$. Putting all of those arguments together we get that, for any sentence $\varphi \in \mathrm{FO}[\sigma]$,

$$\Pr(\mathcal{G}^\sigma(n, \boldsymbol{p}) \models \varphi) = \sum_{i=1}^{\ell} \Pr(\mathcal{G}^\sigma(n, \boldsymbol{p}) \models \varphi \mid \mathrm{Core}_n|_r \in \mathcal{C}_i) \Pr(\mathrm{Core}_n|_r \in \mathcal{C}_i) + o(1),$$

where $\mathcal{C}_1, \dots, \mathcal{C}_\ell$ is an enumeration of $\equiv_k^{\mathrm{Ly}}$-classes of fragments. From here the result follows easily using the ideas introduced above.

### Outline of the Chapter

In Section 2.1 we introduce inside-out strategies for the sparse regime of $\mathcal{G}^\sigma(n, \boldsymbol{p})$. This constitutes the model-theoretic part of Theorem 2.1's proof. In Section 2.2 we determine the asymptotic distribution of a typical $r$-neighbourhood in $\mathcal{G}^\sigma(n, \boldsymbol{p})$, and the asymptotic distribution of its $r$-core. After this, Section 2.3 is dedicated to prove our main result Theorem 2.1. Following that proof, Section 2.4 extends this convergence law to richer models of random structures where the underlying theory consists of various symmetry and anti-reflexivity axioms. Finally section 2.5 showcases an application of the chapter's results to the study of unsatisfiability certificates in random SAT instances.

## 2.1 Inside-Out Strategies

In this section we give a generalization of the so-called inside-out strategies presented in [63]. Those strategies play a central role in the proof of our FO-convergence law.

A common occurrence in random graphs is that fixed-radius neighborhoods can be divided into **abundant**, which asymptotically occur an unbounded number of times, and **rare**, which may appear in finite quantities. In the sparse regime of $\mathcal{G}^\sigma(n, \boldsymbol{p})$, abundant neighborhoods correspond to trees, while rare ones correspond to unicycles. Roughly, our first result in this section, Theorem 2.2, states that two structures with the same abundant $3^k$-neighborhoods are FO $k$-equivalent if both structures have essentially the same number of rare $3^k$-neighborhoods in each FO $k$-type. "Essentially" here means that the quantities may be different so long as both exceed $k$. Afterwards, the following subsections develop several notions of equivalence between trees, unicycles and fragments. Those, together with Theorem 2.2 are used to show some sufficient conditions for $k$-equivalence between structures where all small neighbourhoods are either trees or unicycles (Theorem 2.3).

Let $G_0, G_1$ be arbitrary structures, and let $S_0 \subseteq V(G_0), S_1 \subseteq V(G_1)$ be sets of roots. The **distance-preserving** EF **game**, $\mathrm{dEF}((G_0, S_0), (G_1, S_1); k)$ is defined the same way as $\mathrm{EF}(G_0, G_1; k)$, with the following additional winning conditions for Duplicator. Write $v^0_1, \dots v^0_{\ell+k}$, $v^1_0, \dots, v^1_{\ell+k}$ for the constant-interpretations in $G_0, G_1$ followed by the vertices chosen during the game. Then Duplicator wins if, in addition to the partial isomorphism condition described in Definition 1.1, $d(v^0_i, S_0) = d(v^1_i, S_1)$ for all $i \in [\ell+k]$, and $d(v^0_i, v^0_j) = d(v^1_i, v^1_j)$ for all $i, j \in [\ell+k]$. The expression $(G_0, S_0) \equiv^{\mathrm{dFO}}_k (G_1, S_1)$ means that Duplicator has a winning strategy in this game.

Given a structure $G$, and a set $U \subseteq V(G)$, we define the $r$-**independence number of** $U$ as the maximum size $|W|$ where $W \subseteq U$ satisfies $d(u, v) > 2r + 1$ for all $u, v \in W$.

**Theorem 2.2.** *Let $(G_0, S_0)$, $(G_1, S_1)$ be arbitrary rooted $\tau$-structures, $k \in \mathbb{N}$, $r = \frac{3^k-1}{2}$. Suppose that $S_0$ and $S_1$ contain all constant-interpretations in $G_0$ and $G_1$ respectively. Then $G_0 \equiv^{\mathrm{FO}}_k G_1$ if the following hold:*

*(1) $N(S_0, r) \equiv^{\mathrm{dFO}}_k N(S_1, r)$.*

*(2) Given $t \le r$, let $F_i(t) \subseteq V(G_i)$ be the set of vertices $v$ with $d(v, S_i) > 2t + 1$. Then for all vertices $z \in F_0(t) \cup F_1(t)$, and all $i = 0, 1$ the set*

$$K_i(z, t) = \left\{ v \in F_i(t) \mid N(z, t) \equiv^{\mathrm{dFO}}_k N(v, t) \right\}$$

*has $t$-independence number at least $k$.*

We give some intuition about the statement above. The sets of roots $U_i$ should be thought of as the "special" vertices in $G_i$ whose $r$-neighborhoods are rare. Say, for instance, the vertices forming cycles in $G(n, c/n)$. Condition (1) guarantees the rare neighborhoods in $G_0, G_1$ are equivalent. Condition (2) states that, for any vertex $x$ lying far away from the special ones, there are plenty of other vertices in $G_0, G_1$ whose neighborhoods are equivalent to that of $x$.

*Proof of Theorem 2.2.* We proceed by induction, starting from the base case, with $k = 0$, $r(k) = 0$. Let $\mathfrak{c}_1, \dots, \mathfrak{c}_\ell$ be the constant symbols in $\tau$, and let $\mathfrak{c}_j^i$ be the interpretation of $\mathfrak{c}_j$ on $G_i$ for $i = 0, 1$. We need to show that the substructures induced on $\mathfrak{c}_1^0, \dots, \mathfrak{c}_\ell^0$ and $\mathfrak{c}_1^1, \dots, \mathfrak{c}_\ell^1$ are isomorphic. This follows from condition (1), together with the fact that $S_0, S_1$ contain all the constant-interpretations.

Now, assume that the statement holds for $k-1$. We need to prove it for $k$ as well. Consider the game $\mathrm{EF}(G_0, G_1; k)$. Without loss of generality we can assume Spoiler makes their first move $x^0$ in $G_0$. Let $r' = \frac{3^{k-1}-1}{2}$. Observe that $r = 3r' + 1$. There are two cases to cover:

(I) $d(x^0, S_0) > 2r' + 1$. In this case, by (2) and $k \geq 1$, Duplicator can choose a vertex $x^1 \in V(G_1)$ satisfying both $d(x^1, S_1) > 2r' + 1$ and $N(x^0, r') \equiv_k^{\mathrm{dFO}} N(x^1, r')$.

(II) $d(x^0, S_0) \leq 2r' + 1$. Using (1) Duplicator can choose $x^1 \in N(S_1, r)$ following a winning strategy in $\mathrm{dEF}(N(S_0, r), N(S_1, r); k)$.

Let $\mathfrak{c}$ be a fresh constant symbol and let $\sigma' = \sigma \cup \{\mathfrak{c}\}$. For each $i = 0, 1$, we define $G_i'$ as the $\sigma'$-structure obtained from $G_i$ by setting $\mathfrak{c}^{G_i} = x^i$. Additionally, we define $S_i' = S_i \cup \{x^i\}$ for $i = 0, 1$. We apply the induction hypothesis to $(G_0', S_0'), (G_1', S_1')$. We prove that $G_0' \equiv_{k-1}^{\mathrm{FO}} G_1'$ by showing that $(G_0', S_0'), (G_1', S_1')$ satisfy the theorem's statement for $k-1$. Observe that this completes the proof. We need to show that both (1) and (2) hold:

**Let us begin with (1)**. That is,

$$N(S_0', r') \equiv_{k-1}^{\mathrm{dFO}} N(S_1', r'). \tag{2.1}$$

If $x^0, x^1$ were picked according to case (I) above, then $N(S_i', r')$ is the disjoint union of $N(S_i, r')$ and $N(x^i, r')$. Additionally, $N(S_0, r') \equiv_{k-1}^{\mathrm{dFO}} N(S_1, r')$ as well as $N(x^0, r') \equiv_{k-1}^{\mathrm{dFO}} N(x^1, r')$, so Equation (2.1) follows. Alternatively if $x^0, x^1$ were picked according to (II), then using that $r = 3r' + 1$ we get $N(S_i', r') \subseteq N(S_i, r)$. In this case Equation (2.1) can be deduced by observing that

- Duplicator wins the game $\mathrm{dEF}(N(S_0, r), N(S_1, r); k)$,
- $x^0, x^1$ were picked according to a winning strategy of Duplicator, and
- $S_i' = S_i \cup \{x^i\}$.

**We show property (2) now**. Let $t \leq r'$, and $F_i'(t) \subseteq V(G_i')$ be the set of vertices satisfying $d(v, S_i') > 2t + 1$, as defined in (2). Fix an arbitrary vertex $z \in F_0'(t) \cup F_1'(t)$. We need to show that for $i = 0, 1$, the $r'$-independence number of $K_i'(z, t) = \{v \in F_i'(t) \mid N(z, t) \equiv_k^{\mathrm{dFO}} N(v, t)\}$ is at least $k-1$. Observe that $F_i'(t) = F_i(t) \setminus N(x^i, t)$, so $z \in F_0(t) \cup F_1(t)$. Thus, by hypothesis the $t$-independence number of $K_i(z, t)$ is at least $k$. Observe again that $K_i'(z, t) \supseteq K_i(z, t) \setminus N(x^i, t)$. This shows that the $t$-independence number of $K_i'(z, t)$ is at least $k-1$, as we wanted.

□

### 2.1.1 Strategies on Trees

In this section we give a winning strategy for Duplicator in dEF games played on rooted trees. The goal is not to give the most general strategy, but rather to show one that is relatively simple to define but still partitions rooted trees of fixed height into a finite number of classes.

Let $(T, x)$ be a rooted tree. The **source** of an edge $e \in E(T)$ is the only vertex in $e$ satisfying $d(x, v) = d(x, e)$. Given a vertex $v \in V(T)$, we define $T(v; x)$ as the tree induced by $T$ on the vertices:

$$\{u \in V(T) \mid d(u, x) = d(u, v) + d(v, x)\},$$

and rooted at $v$. Intuitively, $T(v; x)$ corresponds to the sub-tree that "grows out" of $v$. A **branch** is a rooted tree where the root belongs to only one edge. Given an edge $e$ in a rooted tree $(T, x)$, we define $T(e; x)$ as the tree induced by $T$ on

$$\{v \in V(T) \mid d(v, x) = d(v, e) + d(e, x) + 1\} \cup \{v \in e\},$$

and rooted at the source of $e$. Observe that $T(e; x)$ is a branch.

**Definition 2.2** (Equivalence of trees)**.** Let $k \geq 0$. We define the equivalence relation $\equiv_k^{\mathrm{Ly}}$ over rooted trees of height $r$ inductively as follows. For $r = 0$, rooted trees consist simply of their roots, and we put $(T_0, x^0) \equiv_k^{\mathrm{Ly}} (T_1, x^1)$ if $x^0$ and $x^1$ represent the same constants. Now assume $\equiv_k^{\mathrm{Ly}}$ has been defined for heights up to $r - 1$. First we define the relation over branches of height $r$, and afterwards over general rooted trees. Let $(T_0, x^0), (T_1, x^1)$ be branches of height $r$, and let $e^0, e^1$ be the unique edges containing $x^0$ and $x^1$ respectively. We put $(T_0, x^0) \equiv_k^{\mathrm{Ly}} (T_1, x^1)$ if (1) $x^0$ and $x^1$ represent the same constants, (2) $e^0$ and $e^1$ are of part of the same relation, and (3) if $e^i = (v_1^i, \ldots v_\ell^i)$ for $i = 0, 1$, then for all $j \in [\ell]$ either $v_j^0 = x^0$ and $v_j^1 = x^1$ at the same time, or $T_0(v_j^0; x^0) \equiv_k^{\mathrm{Ly}} T_1(v_j^1; x^1)$. Finally, consider the case where $(T_0, x^0), (T_1, x^1)$ are arbitrary trees of height $r$. Let $\mathcal{C}$ be the set of $\equiv_k^{\mathrm{Ly}}$-classes of branches with height at most $r$. We write $(T_0, x^0) \equiv_k^{\mathrm{Ly}} (T_1, x^1)$ if (1) $x^0$ and $x^1$ represent the same constants, and (2) for all classes $C \in \mathcal{C}$ the quantity

$$\left|\left\{e \in E(T_i) \mid x^i \in e,\, T_i(e; x^i) \in C\right\}\right|$$

is the same for $i = 0$ and $i = 1$, or is at least $k$ in both cases.

**Lemma 2.1.** *Let $(T_0, x^0), (T_1, x^1)$ be rooted trees satisfying $(T_0, x^0) \equiv_k^{\mathrm{Ly}} (T_1, x^1)$. Then $(T_0, x^0) \equiv_k^{\mathrm{dFO}} (T_1, x^1)$.*

*Proof.* Note that $\equiv_k^{\mathrm{Ly}}$ imply that $T_0, T_1$ have the same height $r$. The proof is by induction on $r$. We begin with the case $r = 0$. Here $T_0, T_1$ consist simply of $x^0, x^1$, respectively. As $(T_0, x^0) \equiv_k^{\mathrm{Ly}} (T_1, x^1)$, both roots represent the same constants and the statement holds.

Now we assume the statement is true up to $r - 1$ and prove it for $r$. First we consider the case where $(T_0, x^0), (T_1, x^1)$ are branches. Let $e^0, e^1$ be the unique edges containing $x^0, x^1$ respectively, where $e^i = (v_1^i, \ldots, v_\ell^i)$. A winning strategy for Duplicator in $\mathrm{dEF}((T_0, x^0), (T_1, x^1); k)$ is as follows. Whenever Spoiler chooses $x^i$, Duplicator responds with $x^{1-i}$. Otherwise, Spoiler chooses inside $T_i' = T_i(v_j^i; x^i)$ for some $v_j^i \neq x^i$. Let $T_{1-i}' = T_{1-i}(v_j^{1-i}, x^{1-i})$. Observe that from the definition of $(T_0, x^0) \equiv_k^{\mathrm{Ly}} (T_1, x^1)$ follows that $(T_0', v_j^0) \equiv_k^{\mathrm{Ly}} (T_1', v_j^1)$. As $T_0'$, $T_1'$ have height at most $r-1$, by induction Duplicator has a winning strategy in $\mathrm{dEF}((T_0', v_j^0), (T_1', v_j^1); k)$. This way, Duplicator can play according to this strategy whenever Spoiler chooses inside $T_i'$, taking previous plays into account if necessary. Verifying this yields a winning strategy for Duplicator on the original trees $(T_0, x^0), (T_1, x^1)$ is straightforward. Finally, we prove the theorem for general trees $(T_0, x^0), (T_1, x^1)$ of height $r$ assuming it holds true for $r-1$. We describe a winning strategy for Duplicator in $\mathrm{dEF}((T_0, x^0), (T_1, x^1); k)$, as before. For this strategy, Duplicator maintains a partial matching of edges $e^0 \in E(T_0)$ containing the root $x^0$, and edges $e^1 \in E(T_1)$ containing the root $x^1$. This matching satisfies the property that $T_0(e^0; v^0) \equiv_k^{\mathrm{Ly}} T_1(e^1; v^1)$ for all pairs $e^0, e^1$. Observe that these branches have height at most $r$, so Duplicator wins the $k$-round dEF game played on them, as shown above. Whenever Spoiler picks a root $x^i$, Duplicator answers with the other $x^{1-i}$. Otherwise, Spoiler chooses a vertex inside some branch $T_i(e^i; x^i)$, where $e^i$ contains $x^i$. If $e^i$ has already been paired with another edge $e^{1-i} \in E(T_{1-i})$, then Duplicator can play according to a winning strategy between $T_0(e^0; v^0)$ and $T_1(e^1; v^1)$, taking previous plays on these branches into account. If $e^i$ has not been paired so far, then there is another unpaired edge $e^{1-i} \in E(T_{1-i})$, containing the root $x^{1-i}$ and satisfying $T_0(e^0; x^0) \equiv_k^{\mathrm{Ly}} T_1(e^1; x^1)$. This follows from the definition of $\equiv_k^{\mathrm{Ly}}$ on rooted trees and the fact that at most a new pairing is formed during each round, so at most $k-1$ pairings exist at this point. Therefore, Duplicator can add the pair $e^0, e^1$ to the matching and play following a winning strategy on $\mathrm{dEF}(T_0(e^0; x^0), T_1(e^1; x^1); k)$. Again, it is easily seen that this composition of strategies yields a winning strategy for the original game $\mathrm{dEF}((T_0, x^0), (T_1, x^1); k)$. This completes the proof of the result. □

### 2.1.2 Strategies on Fragments

Here we develop a winning strategy for dEF games played on fragments, i.e., structures formed by the disjoint union of unicycles. We begin by giving a strategy for unicycles.

Let $U$ be a unicycle whose cycle is $C \subseteq U$. Given a vertex $v \in V(C)$, we write $T(U, v)$ for the tree induced by $U$ on the set

$$\{u \in V(U) \mid d(u, C) = d(u, v)\},$$

and rooted at $v$. Intuitively, $T(U, v)$ consists of the tree that "grows out" of $v$, as before.

**Definition 2.3** (Equivalence of unicycles)**.** Let $U_0, U_1$ be two unicycles whose cycles are $C_0 \subseteq U_0$, $C_1 \subseteq U_1$. We write $U_0 \equiv_k^{\mathrm{Ly}} U_0$ if there is an isomorphism $f$ between $C_0$ and $C_1$ satisfying that $T(U_0, v^0) \equiv_k^{\mathrm{Ly}} T(U_1, f(v^0))$ for all $v \in V(C_0)$.

**Lemma 2.2.** *Let $U_0, U_1$ be two unicycles whose cycles are $C_0 \subseteq U_0$, $C_1 \subseteq U_1$, and satisfying $U_0 \equiv_k^{\mathrm{Ly}} U_1$. Then Duplicator wins the game* $\mathrm{dEF}((U_0, C_0), (U_1, C_1); k)$*.*

*Proof.* We describe a winning strategy for Duplicator. Let $f : V(C_0) \to V(C_1)$ be an isomorphism between $C_1$ and $C_1$ witnessing the fact that $U_0 \equiv_k^{\mathrm{Ly}} U_1$. Each play of Spoiler belongs uniquely to some rooted tree $T(U_i, v^i)$, where $v^i \in V(C_i)$. By definition $T(U_i, v^i) \equiv_k^{\mathrm{Ly}} T(U_{1-i}, v^{1-i})$, where $v^{1-i} = f(v^i)$, so Lemma 2.1 shows that Duplicator has a winning strategy in $\mathrm{dEF}(T(U_0, v^0), T(U_1, v^1); k)$. Hence, Duplicator can follow this strategy taking previous plays into account if necessary. It is straightforward to see that this composition of strategies indeed produces a winning strategy for the original game. □

**Definition 2.4** (Equivalence of fragments)**.** Let $H_0, H_1$ be fragments. We write $H_0 \equiv_k^{\mathrm{Ly}} H_1$ if for all $\equiv_k^{\mathrm{Ly}}$-classes of unicycles $\mathcal{C}$ the quantity

$$|\{U \text{ connected component in } H_i \mid U \in \mathcal{C}\}| .$$

is the same for $i = 0$ and $i = 1$, or is at least $k$ in both cases.

**Lemma 2.3.** *Let $H_0, H_1$ be fragments satisfying $H_0 \equiv_k^{\mathrm{Ly}} H_1$. For $i = 0, 1$, let $X_i \subseteq V(H_i)$ be the vertices which belong to cycles. Then Duplicator wins* $\mathrm{dEF}((H_0, X_0), (H_1, X_1); k)$.

*Proof.* Similarly to Lemma 2.1, Duplicator maintains a partial matching of connected components from $H_0$ and connected components from $H_1$. At most one new pair will be formed during each round, and the pairs $(U_0, U_1)$ in will be chosen in such a way that $U_0 \equiv_k^{\mathrm{Ly}} U_1$. Suppose that in some round, Spoiler picks a vertex $v^i$ in some connected component $U_i \subseteq H_i$. If $U_i$ has already been paired to another component $U_{1-i} \subseteq H_{1-i}$, then Duplicator can play according to a winning strategy in $\mathrm{dEF}((U_0, C_0), (U_1, C_1); k)$ taking previous moves into account if necessary. Here $C_i$ denotes the cycle in the unicyclic component $U_i$. If $U_i$ has not been paired yet, then by the definition of $\equiv_k^{\mathrm{Ly}}$ over fragments, there is some unpaired component $U_{1-i} \subseteq H_{i-1}$ satisfying $U_0 \equiv_k^{\mathrm{Ly}} U_1$. Hence, Duplicator can add the pair $(U_0, U_1)$ to the matching and chose his move according to a winning strategy in $\mathrm{dEF}((U_0, C_0), (U_1, C_1); k)$, as before. Again, it is easy to check that this composition of strategies yields a winning strategy for the original game. □

### 2.1.3 Strategies on Sparse Structures

**Definition 2.5.** The $r$-**core** $\mathrm{Core}(G)|_r$ of an structure $G$ is the $r$-neighbourhood of all cycles in $G$ of length at most $2r + 1$.

The final result of this section states that the $k$-type of sparse structures with some additional properties depend only on the $\equiv_k^{\mathrm{Ly}}$-class of their $r$-core, where $r = \frac{3^k - 1}{2}$. This is the main model-theoretic ingredient in the proof of our convergence law, Theorem 2.1.

**Definition 2.6.** A structure $G$ is $r$-**simple** if it contains no two cycles of length at most $2r + 1$ at a distance smaller than $2r + 1$. Equivalently, $G$ is $r$-simple if its $r$-core $\mathrm{Core}(G)|_r$ is a disjoint union of unicycles.

**Definition 2.7.** Let $G$ be a structure and $k \in \mathbb{N}$. Let $\mathfrak{T}_r$ be the set of $\equiv_k^{\mathrm{Ly}}$-class $\mathcal{C}$ of trees whose heights are at most $r$. Given $\mathcal{C} \in \mathfrak{T}_r$, $K_{\mathcal{C}}(r)$ denotes the set of vertices $v \in V(G)$ satisfying $N_G(v, r) \in \mathcal{C}$ and $d(v, \mathrm{Core}(G)|_r) > r + 1$. A structure $G$ is called $(k, r)$-**rich** if the $r$-independence number of $K_{\mathcal{C}}(r)$ is at least $k$ for all classes $\mathcal{C} \in \mathfrak{T}_r$.

**Theorem 2.3.** *Let $k \in \mathbb{N}, r = \frac{3^k - 1}{2}$. Let $G_0, G_1$ be two $r$-simple and $(k, r)$-rich structures satisfying* $\mathrm{Core}(G_0)|_r \equiv_k^{\mathrm{Ly}} \mathrm{Core}(G_1)|_r$. *Then $G_0 \equiv_k G_1$*

*Proof.* Let $S^i$ be the vertices belonging to some cycle of length at most $2r+1$ lying in $G^i$ for $i = 0, 1$. We apply Theorem 2.2 to $(G_0, S_0)$ and $(G_1, S_1)$ to prove the result. For this we need to show that both conditions of that theorem's statement hold. Condition (1) is straightforward: $N(S_0, r) \equiv_k^{\mathrm{dFO}} N(S_1, r)$ follows from the hypothesis $\mathrm{Core}(G_0)|_r \equiv_k^{\mathrm{Ly}} \mathrm{Core}(G_1)|_r$ and Lemma 2.3. Condition (2) takes more effort. Fix $t \leq r$. Let $F_i(t) \subseteq V(G_i)$ be the set of vertices $v$ with $d(v, S_i) > 2t+1$, for $i = 0, 1$, as in Theorem 2.2. Let $z \in F_0(t) \cup F_1(t)$. We need to show that the set

$$K_i(z,t) = \{v \in F_i(t) \mid N(z,t) \equiv_k^{\mathrm{dFO}} N(v,t)\}$$

has $t$-independence number at least $k$ for $i = 0, 1$. Fix $i$. Observe that $N(z,t)$ is a tree, as $z$ lies far away from $S_0, S_1$. Let $\mathcal{C}$ be the $\equiv_k^{\mathrm{Ly}}$-class of $N(z,t)$. The trees in $\mathcal{C}$ have height at most $t \leq r$, so by the $(k,r)$-richness hypothesis there are $k$ vertices $v_1, \dots, v_k$ in $G_i$ lying at distance greater than $2r+1$ from each other and from $S_i$ satisfying $N(v_j, r) \in \mathcal{C}$ for $j = 1, \dots, k$. We claim that $v_1, \dots, v_k$ all belong to $K_i(z,t)$, showing that the $r$-independence number of the set is at least $k$. Indeed, as $d(v_j, S_i) > 2r+1$, it holds $v_j \in F_i(t)$. Moreover, trees in $\mathcal{C}$ have height at most $t$, showing that $N(v_j, r) = N(v_j, t)$. Finally, by Lemma 2.1, $N(v_j, t) \equiv_k^{\mathrm{Ly}} N(z,t)$ implies $N(v_j,t) \equiv_k^{\mathrm{dFO}} N(z,t)$. Thus $v_1, \dots, v_k$ belong to $K_i(z,t)$ and condition (2) from Theorem 2.2 holds. We have shown that both hypotheses of that theorem hold, so it follows that $G_0 \equiv_k^{\mathrm{FO}} G_1$, as we wanted to prove. □

## 2.2 The Landscape of Sparse Random Structures

This section is dedicated to probabilistic results about $\mathcal{G}^\sigma(n, \boldsymbol{p})$. The main improvement over the published version of this work [40] is in the results related to the neighbourhoods of fixed vertices (mainly theorem 2.4). Whereas in [40], the asymptotic distribution of those neighbourhoods is computed via a (rather complicated) iterated application of the Method of Moments, our proof here is based on a coupling with a multi-type branching process.

**Lemma 2.4** (Small cycle distribution). *Let $H_1, \dots, H_k$ be different unlabeled cycles. For each $1 \leq i \leq k$, let $X_i = X_i(n)$ count the copies of $H_i$ in $\mathcal{G}^\sigma(n, \boldsymbol{p})$. Then $(X_1, \dots, X_k)$ converge jointly in distribution to* $\mathrm{Pois}_{\lambda_1} \times \cdots \times \mathrm{Pois}_{\lambda_k}$, *where*

$$\lambda_i = \frac{1}{\mathrm{aut}(H_i)} \prod_{1 \leq j \leq |\sigma|} c_j^{|E_j(H_i)|}.$$

*Proof.* This is a direct application of the moment's method (Theorem 1.4). Observe that $\mathrm{E}[X_i] = \lambda_i + o(1)$. Fix numbers $a_i \in \mathbb{N}$ for each $1 \leq i \leq k$. To prove the result is enough to show:

$$\mathrm{E}\left[\prod_{1 \leq i \leq k} \binom{X_i}{a_i}\right] = \prod_{1 \leq i \leq k} \frac{\lambda_i}{a_i!} + o(1).$$

Let $\hat{H}$ be the disjoint union of $a_1$ copies of $H_1$, followed by $a_2$ copies of $H_2$, ..., and $a_k$ copies of $H_k$. Let $\hat{X} = \hat{X}(n)$ count the copies of $\hat{H}$ in $\mathcal{G}^\sigma(n, \boldsymbol{p})$. This way, $\mathrm{E}\left[\hat{X}\right] = \prod_{1 \leq i \leq k} \frac{\lambda_i}{a_i!} + o(1)$. Let $\mathcal{R}$ be the set of unlabeled structures resulting from a non-disjoint union of $a_1$ copies of $H_1$, followed by $a_2$ copies of $H_2$, ..., and $a_k$ copies of $H_k$. The variable $Y = Y(n)$ counts the substructures in $\mathcal{G}^\sigma(n, \boldsymbol{p})$ isomorphic to some structure in $\mathcal{R}$. Observe that for all $H \in \mathcal{R}$, $\mathrm{ex}(H) \geq 1$, so the expected number of $H$-copies in $\mathcal{G}^\sigma(n, \boldsymbol{p})$ is $o(1)$. As there is a finite number of such structures $H$, it holds that $\mathrm{E}[Y] = o(1)$. Finally, by definition, $\mathrm{E}\left[\prod_{1\leq i\leq k} \binom{X_i}{a_i}\right] = \mathrm{E}\left[\hat{X}\right] + O(\mathrm{E}[Y])$, which equals $\prod_{1\leq i\leq k} \frac{\lambda_i}{a_i!} + o(1)$, as we wanted to prove. □

We define the set of rooted-edge types as $\mathrm{Tp}^E_\sigma = \{(i, \ell) \mid 1 \leq i \leq |\sigma|,\ 1 \leq \ell \leq \mathrm{ar}(E_i)\}$. Informally, each pair $(i, \ell) \in \mathrm{Tp}^E_\sigma$ specifies a relation symbol $E_i \in \sigma$, and a position $\ell$ within $[\mathrm{ar}(E_i)]$ for the root. We introduce a variant of multi-type branching processes [4] that approximate small neighbourhoods in $\mathcal{G}^\sigma(n, \boldsymbol{p})$

**Definition 2.8.** For each $(i, \ell) \in \mathrm{Tp}^E_\sigma$, let $X^{i,\ell}$ be a distribution over $\mathbb{N}$. Let $(X^{i,\ell}_n)$, where $n$ ranges over $\mathbb{N}$ and $(\ell, i)$ ranges over $\mathrm{Tp}^E_\sigma$, be mutually independent random variables with $X^{\ell,i}_n \sim X^{\ell,i}$. Fix $t \in \mathbb{N}$. The **$t$-root multi-type branching process corresponding to $\sigma$ with offspring distributions** $(X^{\ell,i})_{(\ell,i)\in\mathrm{Tp}^E_\sigma}$ is a sequence of random variables $(Y_n)_{n\in\mathbb{N}}$ satisfying

$$Y_0 = t, \qquad Y_n = \begin{cases} Y_{n-1} - 1 + \sum\limits_{(\ell,i)\in\mathrm{Tp}^E_\sigma} (\mathrm{ar}(E_\ell) - 1)X^{\ell,i}_n, & \text{if } Y_{n-1} > 0, \\ 0 \text{ otherwise.} \end{cases}$$

Some insight is needed to make sense of last definition. This branching process represents a random forest which is exposed in breath-first fashion, starting from the $t$ roots. At the $n$-th step, the $n$-th vertex $v$ of the forest is selected and its outgoing edges are exposed. The variable $X^{\ell,i}_n$ counts the number of type-$(\ell, i)$ edges rooted at $v$. Hence, in the $n$-th step, $\sum_{(\ell,i)}(\mathrm{ar}(E_\ell) - 1)X^{\ell,i}_n$ fresh vertices are added to the forest. The variable $Y_n$ counts the number of vertices in the forest whose outgoing edges have not been exposed yet. Thus, the branching process ends when $Y_n = 0$.

As in Definition 1.3, we identify multi-type branching processes $\mathcal{BP}$ with random forests in the way outlined above. However, we do not give these random forests explicitly - it is possible to do so via an Ulam-Harris-style construction as well -, but rather we refer to some fixed unspecified breadth-first exploration order. We define $\mathcal{BP}|_r$ as the restriction of $\mathcal{BP}$ to its first $r$ generations (i.e., the $r$-neighbourhood of the roots), and $\mathcal{BP}|^\tau$ as the restriction of $\mathcal{BP}$ to its first $\tau$ individuals (i.e., the forest induced on the first $\tau$ vertices).

**Definition 2.9.** For the remainder of this chapter the **random tree** $\mathcal{T} = \mathcal{T}(\boldsymbol{c})$ is the one-root multi-type branching process with offspring distributions $X^{\ell,i} \sim \mathrm{Pois}_{c_\ell}$ for all types $(\ell, i) \in \mathrm{Tp}^E_\sigma$. For the $t$-root multi-type process with the same offspring distributions we use $\mathcal{T}^t$. Given $(\ell, i) \in \mathrm{Tp}^E_\sigma$, the **random branch** $\mathcal{T}_{\ell,i}$ is $\mathcal{T}$ conditioned on the events $X^{\ell,i}_1 = 1$, and $X^{\ell',i'}_1 = 0$ for all $(\ell', i') \neq (\ell, i)$. The **random fragment** $\Gamma|_r = \Gamma(\boldsymbol{c})|_r$ is generated as follows: (1) given a cycle $H$ containing at most $2r + 1$ edges, $\Gamma|_r$ contains $\mathrm{Pois}_{\lambda_H}$ disjoint copies of $H$, where $\lambda_H = \frac{1}{\mathrm{aut}(H)} \prod_{1\leq i\leq|\sigma|} c_i^{|E_i(H)|}$, and (2) attach an independent copy of $\mathcal{T}|_r$ to each vertex.

The main result in this subsection is that the $r$-neighborhood of a vertex $v$ in $\mathcal{G}^\sigma(n, \boldsymbol{p})$ converges in distribution to $\mathcal{T}|_r$. This is completely analogous to the local convergence results for $\mathcal{G}(n, p)$ in the sparse regime.

**Theorem 2.4.** *Fix $t, r \in \mathbb{N}$. Let $\boldsymbol{\rho} = (\rho_1, \dots, \rho_t) \in \mathbb{N}^t$ be a $t$-tuple of different vertices. Then the tuple of neighborhoods $N(\boldsymbol{\rho}, r)$ converges in distribution to $\mathcal{T}^t|_r$.*

*Proof.* Given $\tau \in \mathbb{N}$, we define $T_n$ as the graph rooted at $\boldsymbol{\rho}$ which results from exposing all edges incident to the first $\tau$ individuals belonging to $\boldsymbol{\rho}$'s connected component, in breadth-first order. We show that

$$T_n \xrightarrow{d} \mathcal{T}^t|^\tau.$$

Observe that this implies the theorem. For each $n$, let $\mathcal{BP}_n$ be the $t$-root multi-type branching process with edge distribution $\mathrm{Bin}(n^{\mathrm{ar}(E_\ell)-1}, p_\ell(n))$ for each edge type $(\ell, i) \in \mathrm{Tp}^E_\sigma$. As $\mathrm{Bin}(n^{\mathrm{ar}(E_\ell)-1}, p_\ell(n))$ in distribution to $\mathrm{Pois}_{c_\ell}$, the branching process $\mathcal{BP}_n|^\tau$ converges in distribution to $\mathcal{T}^t|^\tau$ as well. In particular, the sequence $(\mathcal{BP}_n|^\tau)_{n\in\mathbb{N}}$ of distributions over finite rooted forests is tight because of Lemma 1.4. In order to show that $T_n$ converges in distribution to $\mathcal{T}^t|^\tau$ (when both $T_n, \mathcal{T}^t|^\tau$ are seen as random variables taking values in the set of finite unlabeled rooted graphs), it suffices to give a coupling $(T_n, \mathcal{BP}_n|^\tau)$ satisfying

$$\Pr(T_n \not\simeq \mathcal{BP}_n|^\tau) = o(1) \tag{2.2}$$

Using the Coupling Lemma (Lemma 1.2) together with Lemma 1.1 yields the result. We describe the coupling below. Roughly, the idea is that the exploration of $\mathcal{BP}_n|^\tau$ behaves like the one of $T_n$, with the difference that when exposing all edges incident to some vertex in $\mathcal{BP}_n|^\tau$, all possible edges on with vertices in $[n]$ are considered. However, when exposing $T_n$ edges containing previously-explored vertices are not taken into account. Our coupling is defined through auxiliary random variables $W^{\ell,i}_j$. For each $j \in [\tau]$, and each type $(\ell, i) \in \mathrm{Tp}^E_\sigma$, $W^{\ell,i}_j$ is a random subset of $[n]^{\mathrm{ar}(E_\ell)-1}$ where each tuple $\boldsymbol{v}$ belongs to $W^{\ell,i}_j$ with probability $p_\ell(n)$. Moreover, the variables $W^{\ell,i}_j$ are mutually independent for all $j$, $(\ell, i)$. We couple $\mathcal{BP}_n|^\tau$ to these variables by setting $X^{\ell,i}_j = |W^{\ell,i}_j|$. The variables $W^{\ell,i}_j$ are coupled with the exploration of $T_n$ as well. We say that the coupling **fails** at the $j$-th step if either of the following are satisfied.

1. (1) A vertex $v$ participating in some tuple $\boldsymbol{v} \in \bigsqcup_{(\ell,i)} \bigsqcup_{j'<j} W^{\ell,i}_{j'}$ (the symbol $\sqcup$ stands for the disjoint union of sets), participates again in some tuple $\boldsymbol{v} \in \bigsqcup_{(\ell,i)} W^{\ell,i}_j$.
2. (2) Some vertex $v$ participates in two tuples $\boldsymbol{v}, \boldsymbol{v}' \in \bigsqcup_{(\ell,i)} W^{\ell,i}_j$.
3. (3) Some vertex $v$ participates more than once in some tuple $\boldsymbol{v} \in \bigsqcup_{(\ell,i)} W^{\ell,i}_j$.

Suppose the coupling has not failed yet at step $j$. Let $v$ be the $j$-th vertex in the exploration of $T_n$. The following edges are exposed: for each type $(\ell, i) \in \mathrm{Tp}^E_\sigma$ and tuple $(u_1, \dots, u_{\mathrm{ar}(E_\ell)-1}) \in W^{\ell,i}_j$, the edge $(u_1, \dots, u_{i-1}, v, u_i, \dots u_{\mathrm{ar}(E_\ell)-1})$ is added to $E_\ell(T_n)$. Otherwise, if the coupling fails at the $j$-th step, the exploration of $T_n$ continues independently of the variables $W^{\ell,i}_j$. Observe that this is a valid coupling in the sense that both $\mathcal{BP}_n|^\tau$ and $T_n$ have the desired distribution.

Let $A = A_n$ be an event of the form $X^{\ell,i}_j = x^{\ell,i}_j$ for all $(\ell, i) \in \mathrm{Tp}^E_\sigma$, $j \in [\tau]$, and some fixed constants $x^{\ell,i}_j$. In other words, $A$ is a particular outcome of $\mathcal{BP}_n|^\tau$. Using the fact that the sequence $(\mathcal{BP}_n|^\tau)_{n\in\mathbb{N}}$ is tight, to show Equation (2.2) it is sufficient to prove

$$\Pr(T_n \not\simeq \mathcal{BP}_n|^\tau \mid A) = o(1) \tag{2.3}$$

We focus on this last identity. We condition the whole coupling on the event $A$. This way, $\widehat{W}_j^{\ell,i}$ stands for the variable $W_j^{\ell,i}$ conditioned on the event $A$. Observe that the variables $\widehat{W}_j^{\ell,i}$ are distributed as uniform subsets of $[n]^{\mathrm{ar}(E_\ell)-1}$ of size $x_j^{\ell,i}$. The probability that $T_n \not\simeq \mathcal{BP}_n$ is at most the probability that the coupling fails in the first $\tau$ rounds. We estimate the probabilities of the coupling failing in the ways (1), (2) and (3) defined above.

Let's begin with (1). At step $j$, the expected number of tuples $\boldsymbol{u} \in \widehat{W}_j^{\ell,i}$ containing some previously explored vertex tends to $jx_j^{\ell,i}(\mathrm{ar}(E_\ell)-1)/n$. Thus, by Markov's inequality, the probability of (1) is $1-O(1/n)$. Now we estimate the probability of (2). When exposing a set $\widehat{W}_j^{\ell,i}$, at most $V(T)$ where revealed previously. The probability that one of these participates in a tuple $\boldsymbol{v} \in \widehat{W}_j^{\ell,i}$ is at most $|V(T)|x_j^{\ell,i}(\mathrm{ar}(E_\ell)-1)/n + o(1)$. Also, the probability that any two tuples in $\widehat{W}_j^{\ell,i}$ share a vertex is bounded by $|V(T)|(x_j^{\ell,i})^2(\mathrm{ar}(E_\ell)-1)^2/n + o(1)$. Hence, (2) holds with probability $1-O(1/n)$ as well. Finally, we examine (3). When exposing some $\widehat{W}_j^{\ell,i}$, the probability that some vertex appears twice in a tuple $\boldsymbol{v} \in \widehat{W}_j^{\ell,i}$ is bounded by $x_j^{\ell,i}(\mathrm{ar}(E_\ell)-1)^2/n + o(1)$. This shows that (3) also holds with probability $1-O(1/n)$. The estimates for (1), (2) and (3) show that $\Pr\left(\mathcal{BP}_n|^\tau \not\simeq T_n \mid A\right) = O(1/n)$. This implies Equation (2.2) and proves the theorem. □

An easy corollary of this result is that w.h.p. bounded-radius neighbourhoods of fixed vertices are disjoint, and those neighbourhoods contain no cycles. The following lemma rephrases this slightly.

**Corollary 2.1.** *Let $\rho_1, \rho_2 \in \mathbb{N}$ be two fixed vertices, and let $r \in \mathbb{N}$. Then (1) w.h.p. the distance between $\rho_1$ and $\rho_2$ is greater than $r$ in $\mathcal{G}^\sigma(n,\boldsymbol{p})$ and (2) w.h.p. $\rho_1$ is at distance greater than $r$ from all cycles of length at most $2r+1$ in $\mathcal{G}^\sigma(n,\boldsymbol{p})$.*

As small cycles occur far away from fixed vertices in $\mathcal{G}^\sigma(n,\boldsymbol{p})$, it is reasonable that the small neighbourhoods of those vertices should be independent from the number of small cycles. Next result proves this intuition right.

**Lemma 2.5.** *Let $(\rho_1,\dots,\rho_t) \in \mathbb{N}^t$ be a set of roots, and let $F$ be a $t$-root forest of height at most $r$. Then the event $N_n(\boldsymbol{\rho}, r) \simeq F$ is asymptotically independent from the event that $\mathcal{G}^\sigma(n,\boldsymbol{p})$ has no cycles of length at most $2r+1$.*

*Proof.* Let $A_n$ be the event that $\mathcal{G}^\sigma(n,\boldsymbol{p})$ has no cycles of length at most $2r+1$. We show that $\Pr(A_n \mid N_n(\boldsymbol{\rho},r) \simeq F) = \Pr(A_n) + o(1)$. Let $F_*$ be a fixed copy of $F$ whose roots are $\boldsymbol{\rho}$ and whose vertices $V(F_*)$ lie in $\mathbb{N}$. By symmetry $\Pr(A_n \mid N_n(\boldsymbol{\rho},r) \simeq F)$ equals $\Pr(A_n \mid N_n(\boldsymbol{\rho},r) = F_*)$. Given $N_n(\boldsymbol{\rho},r) = F_*$, it is sufficient that $P_1$: the roots in $\boldsymbol{\rho}$ are at distance greater than $2r+1$ from each other and from all cycles of length at most $2r+1$, and $P_2$: the structure $G'_n$ induced by $\mathcal{G}^\sigma(n,\boldsymbol{p})$ on $[n] \setminus V(F)$ contains no cycles of length at most $2r+1$. By Corollary 2.1, $P_1$ occurs with probability $1-o(1)$. Observe that $N_n(\boldsymbol{\rho},r) \simeq F$ has probability bounded away from zero by Theorem 2.4, and $\Pr(P_1 \mid N_n(\boldsymbol{\rho},r) = F_*) = \Pr(P_1 \mid N_n(\boldsymbol{\rho},r) \simeq F)$, so $\Pr(P_1 \mid N_n(\boldsymbol{\rho},r) = F_*) = 1-o(1)$. As for $P_2$, it is easy to see that $G'_n$ is distributed like $\mathcal{G}^\sigma(n-n(F),\boldsymbol{p}')$, where $\boldsymbol{p}' = (p'_1,\dots,p'_{|\sigma|})$ satisfies $p_i \sim p'_i$. Thus $\Pr(P_2) = \Pr(A_n) + o(1)$. As $P_2$ is independent from $N_n(\boldsymbol{\rho},r) = F_*$, conditioning on that event does not change the probability. Hence, by intersection bound $\Pr(P_1 \wedge P_2 \mid N_n(\boldsymbol{\rho},r) = F_*) = \Pr(A) - o(1)$, and this quantity is smaller than $\Pr(A \mid N_n(\boldsymbol{\rho},r) = F_*)$. On the other hand, the event $P_2$ is a necessary condition for $A$, so $Pr(A \mid N_n(\boldsymbol{\rho}, r) = F_*)$, so this probability is also bounded from above by $\Pr(A) + o(1)$ as well, yielding the result. □

Having established the limiting distribution of small neighbourhoods in $\mathcal{G}^\sigma(n, \boldsymbol{p})$, we move on to studying the $r$-core $\mathrm{Core}_n|_r$ of $\mathcal{G}^\sigma(n, \boldsymbol{p})$ (recall Definition 2.5).

**Lemma 2.6.** *The $r$-core $\mathrm{Core}_n|_r$ converges in distribution to $\Gamma|_r$.*

*Proof.* Fix a $r$-fragment $G$. We prove that $\Pr(\mathrm{Core}_n|_r \simeq G) = \Pr(\Gamma|_r \simeq G) + o(1)$. We do this in two parts. First we observe that the cycle distribution in $\mathrm{Core}_n|_r$ and $\Gamma|_r$ is the same, and afterwards we show that the trees that hang from the cycles have the same distribution. Let $H$ be the union of cycles in $G$, $H_\Gamma$ be the union of cycles in $\Gamma|_r$, and $H_{\mathrm{Core}} = H_{\mathrm{Core}}(n)$ be the union of cycles in $\mathrm{Core}_n|_r$. By Lemma 2.4 and the definition of $\Gamma|_r$, we have $\Pr(H_{\mathrm{Core}} \simeq H) = \Pr(H_\Gamma \simeq H) + o(1)$. Let $H_*$ be a fixed copy of $H$ with vertex set $V(H_*) \subseteq \mathbb{N}$. By symmetry, $\Pr(\mathrm{Core}_n \simeq G \mid H_{\mathrm{Core}} \simeq H) = \Pr(\mathrm{Core}_n \simeq G \mid H_{\mathrm{Core}} = H_*)$. Let $\overline{\mathcal{G}}_n = \mathcal{G}^\sigma(n, \boldsymbol{p}) \setminus E(H_*)$. Given $H_{\mathrm{Core}} = H_*$, $\mathrm{Core}_n|_r \simeq G$ if and only if the $r$-neighbourhood of $\boldsymbol{\rho} = V(H_*)$ in $\overline{\mathcal{G}}_n$ belongs to some finite class $\mathcal{F}$ of rooted unlabeled forests. We define the following events: $P_{H_*} \equiv$ "$H_* \subset \mathcal{G}^\sigma(n, \boldsymbol{p})$", $P_{\mathrm{cycl}} \equiv$ " $\overline{\mathcal{G}}_n$ contains no cycle of length at most $2r+1$", and $P_{\mathrm{dist}} \equiv$ "vertices in $V(H_*)$ lie at distance greater than $2r+1$ from each other and from all other cycles of length at most $2r+1$ in $\overline{\mathcal{G}}_n$". The following implications hold:

$$P_{H_*} \wedge P_{\mathrm{cycl}} \wedge P_{\mathrm{dist}} \implies H_{\mathrm{Core}} = H_* \implies P_{H_*} \wedge P_{\mathrm{cycl}}.$$

When studying the neighbourhood of $H_*$, we would rather condition to $P_{H_*} \wedge P_{\mathrm{cycl}}$ than to $H_{\mathrm{Core}} = H_*$, which is a more complex event. To do this rigorously, it is required that, w.h.p., $H_{\mathrm{Core}} = H_* \iff P_{H_*} \wedge P_{\mathrm{cycl}}$. We claim this is the case. It suffices to prove $\Pr(P_{\mathrm{dist}} \mid P_{H_*} \wedge P_{\mathrm{cycl}}) = 1 - o(1)$ by the implications above. Notice that $P_{\mathrm{dist}}$ is independent from $P_{H_*}$, so $\Pr(P_{\mathrm{dist}} \mid P_{H_*} \wedge P_{\mathrm{cycl}}) = \Pr(P_{\mathrm{dist}} \mid P_{\mathrm{cycl}})$. However, $P_{\mathrm{cycl}}$ holds with probability bounded away from zero by Lemma 2.4 and $P_3$ holds a.a.s by Corollary 2.1, yielding $\Pr(P_{\mathrm{dist}} \mid P_{\mathrm{cycl}}) = 1 - o(1)$, as we wanted. This proves that $(N_{\overline{\mathcal{G}}_n}(\boldsymbol{\rho}, r) \mid H_{\mathrm{Core}} = H_*)$ converges in distribution to $(N_{\overline{\mathcal{G}}_n}(\boldsymbol{\rho}, r) \mid P_{H_*} \wedge P_{\mathrm{cycl}})$. Finally, by Theorem 2.4, we know that $(N_{\overline{\mathcal{G}}_n}(\boldsymbol{\rho}, r) \mid P_{H_*} \wedge P_{\mathrm{cycl}})$ converges in distribution to multiple copies of $\mathcal{BP}|_r$. By definition, this is the distribution of the trees growing out of $\Gamma|_r$'s cycles. Thus $\Pr(\mathrm{Core}_n \simeq G \mid H_{\mathrm{Core}} \simeq H) = \Pr(\Gamma|_r \simeq G \mid H_\Gamma \simeq H) + o(1)$ and the theorem holds. □

**Corollary 2.2.** *Let $r \in \mathbb{N}$. Then w.h.p. $\mathcal{G}^\sigma(n, \boldsymbol{p})$ is $r$-simple.*

*Proof.* The structure $\mathcal{G}^\sigma(n, \boldsymbol{p})$ fails to be $r$-simple if and only if its $r$-core $\mathrm{Core}_n|_r$ contains some complex component. However, as proven in the previous lemma, $\mathrm{Core}_n|_r$ is a fragment w.h.p. □

## 2.3 Proof of the Convergence Law

In this section we prove the FO-convergence law for $\mathcal{G}^\sigma(n, \boldsymbol{p})$ stated in Theorem 2.1. This is done by showing that (1) $k$-type of $\mathcal{G}^\sigma(n, \boldsymbol{p})$ a.a.s. depends only on the $\equiv_k^{\mathrm{Ly}}$-class of $\mathrm{Core}_n|_r$, where $r = \frac{3^k - 1}{2}$, and (2) the limit probability that $\mathrm{Core}_n|_r$ belongs to a given $\equiv_k^{\mathrm{Ly}}$-class is an expression in

Expr. Fact (1) follows from applying our result on inside-out strategies, Theorem 2.2. However, (2) requires some work. The beginning of the section focuses on this, first by studying the asymptotic probabilities of $\equiv_k^{\mathrm{Ly}}$-classes of trees, and then doing the same for fragment classes. The following is a straight-forward observation about Expr that will be useful later.

**Lemma 2.7.** *The expressions* $\omega \in \mathrm{Expr}$ *take positive values in* $(0,\infty)^{|\sigma|}$*, and are analytic. That is, there is some complex analytic extension of* $\omega$ *to some open set* $U$*, where* $(0,\infty)^{|\sigma|} \subset U \subset \mathbb{C}^{|\sigma|}$.

### 2.3.1 Probabilities of Tree Classes

Let $\mathcal{T} = \mathcal{T}(\boldsymbol{c})$ be the random tree introduced in Definition 2.8. Our goal here is to inspect the probability that $\mathcal{T}|_r$ belongs to a particular $\equiv_k^{\mathrm{Ly}}$-class of trees. Let $\mathcal{C}$ be a $\equiv_k^{\mathrm{Ly}}$-class of trees of height at most $r$. Then $\mu_{\mathcal{C},r} = \mu_{\mathcal{C},r}(\boldsymbol{c})$ denotes the probability that $\mathcal{T}|_r \in \mathcal{C}$. Analogously, $\mu_{\mathcal{C},r}^{\ell,i}$ is the probability that $\mathcal{T}^{\ell,i}|_r \in \mathcal{C}$, where $\mathcal{T}^{\ell,i}$ is the random branch, also given in Definition 2.8.

**Theorem 2.5.** *Let* $\mathcal{C}$ *be a* $\equiv_k^{\mathrm{Ly}}$*-class of trees of height at most* $r$*. Then* $\mu_{\mathcal{C},r} \in \mathrm{Expr}$.

*Proof.* Let $T = \mathcal{T}|_r$ and let $x$ be $T$'s root. The proof is by induction on $r$. For $r = 0$ there is only one $\equiv_k^{\mathrm{Ly}}$-class $\mathcal{C}$, so $T \in \mathcal{C}$ with probability 1. As $1 \in \mathrm{Expr}$, the statement holds. Now assume the theorem holds up to $r-1$. First, let $\mathcal{C}$ be a $\equiv_k^{\mathrm{Ly}}$-class of branches with height at most $r$, and let $(\ell, i)$ be the type of the initial edge of the branches in $\mathcal{C}$ (observe that this type is the same for all the branches in $\mathcal{C}$, so $(\ell, i)$ is well defined). Define $T^{\ell,i}$ as the random branch $\mathcal{T}^{\ell,i}|_r$, with root $x$. Let $e = (v_1, \ldots, v_{i-1}, x, v_i, \ldots, v_{\mathrm{ar}(E_\ell)-1})$ be the only edge incident to $x$ in $T^{\ell,i}$. The $\equiv_k^{\mathrm{Ly}}$-class of $T^{\ell,i}$ is determined by the trees that hang from the non root vertices $v_j$. More concretely, the event $T^{\ell,i} \in \mathcal{C}$ is equivalent to

$$\bigwedge_{1 \le j \le \mathrm{ar}(E_\ell)-1} T^{\ell,i}(v_j; x) \in \mathcal{C}_j,$$

for some fixed $\equiv_k^{\mathrm{Ly}}$-classes $\mathcal{C}_1, \ldots, \mathcal{C}_{\mathrm{ar}-1}$ determined by $\mathcal{C}$, that consist of trees with height at most $r-1$. The trees $T^{\ell,i}(v_j; x)$ are independent copies of $\mathcal{T}|_{r-1}$, so $\mu_{\mathcal{C},r}^{\ell,i} = \prod_j \mu_{\mathcal{C}_j,r-1}$, where $j$ ranges from 1 to $\mathrm{ar}(E_\ell) - 1$. Observe that by induction each factor $\mu_{\mathcal{C}_j,r-1}$ belongs to Expr, so the whole product does as well. Consider now the original tree $T = \mathcal{T}|_r$. Given a $\equiv_k^{\mathrm{Ly}}$-class $\mathcal{C}'$ of branches whose heights are at most $r$, the random variable $X_{\mathcal{C}'}$ counts the number of initial edges $e$ in $T$ whose branch $T(e; x)$ belongs to $\mathcal{C}'$. The branch $T(e; x)$ is distributed like $T^{\ell,i}$, where $(\ell, i)$ is the type of $e$, and is independent of all other branches. This implies $X_{\mathcal{C}'} \sim \mathrm{Pois}_{\lambda_{\mathcal{C}'}}$ independently, where $\lambda_{\mathcal{C}'} = c_\ell\, \mu_{\mathcal{C}',r}^{\ell,i}$, and $(\ell, i)$ is the initial edge type determined by $\mathcal{C}'$. We established above that $\mu_{\mathcal{C}',r}^{\ell,i}$ belongs to Expr, so that is the case for $\lambda_{\mathcal{C}'}$ as well. Now let $\mathcal{C}$ be an arbitrary $\equiv_k^{\mathrm{Ly}}$-class of trees whose height is at most $r$. The event $T \in \mathcal{C}$ is equivalent to

$$\left( \bigwedge_{\mathcal{C}' \in \mathfrak{C}^-} X_{\mathcal{C}'} = a_{\mathcal{C}'} \right) \wedge \left( \bigwedge_{\mathcal{C}' \in \mathfrak{C}^+} X_{\mathcal{C}'} > k \right),$$

for some partition $\mathfrak{C}^- \cup \mathfrak{C}^+$ of the $\equiv_k^{\mathrm{Ly}}$-classes of branches whose height is at most $r$, and some constants $a_{\mathcal{C}'} \in \mathbb{N}$. By the arguments above, all the events in this conjunction are mutually independent, and

$$\Pr(T \in \mathcal{C}) = \left( \prod_{\mathcal{C}' \in \mathfrak{C}^-} \Pr(X_{\mathcal{C}'} = a_{\mathcal{C}'}) \right) \times \left( \prod_{\mathcal{C}' \in \mathfrak{C}^+} \Pr(X_{\mathcal{C}'} > k) \right).$$

Moreover, $\Pr(X_{\mathcal{C}'} = a_{\mathcal{C}'})$ is precisely $\mathrm{Pois}_{\lambda_{\mathcal{C}'}}(a_{\mathcal{C}'})$, so it belongs to Expr by virtue of $\lambda_{\mathcal{C}'} \in \mathrm{Expr}$. Similarly, $\Pr(X_{\mathcal{C}'} > k) = 1 - \sum_{i=0}^{k} \mathrm{Pois}_{\lambda_{\mathcal{C}'}}(i)$, also belongs to Expr. Thus, $\Pr(T \in \mathcal{C}) \in \mathrm{Expr}$, concluding the proof. □

**Corollary 2.3.** *Let $k, r \in \mathbb{N}$. Then $\mathcal{G}^\sigma(n, \boldsymbol{p})$ is $(k,r)$-rich w.h.p.*

*Proof.* Given a $\equiv_k^{\mathrm{Ly}}$-class $\mathcal{C}$ of trees whose heights are at most $r$, let $K_\mathcal{C} = K_\mathcal{C}(r)$ be as in Definition 2.7. Let $A = A(n)$ be the event that the $r$-independence number of $K_\mathcal{C}$ is smaller than $k$. Fix a small $\epsilon > 0$. We show that $\Pr(A) < \epsilon + o(1)$. This proves the result. Let $a$ be the smallest number such that $\Pr(\mathrm{Bin}(a, \mu_{\mathcal{C},r}) < k) < \epsilon$. The existence of such $a$ follows from $\mathrm{E}\left[\mathrm{Bin}(a, \mu_{\mathcal{C},r})\right] = a\ \mu_{\mathcal{C},r}$, which tends to infinity with $a$, using any concentration bound for the binomial distribution. By Corollary 2.1, the vertices labeled $1, \dots, a \in \mathbb{N}$ a.a.s. lie at distance greater than $2r+1$ from each other. Let $X = X(n)$ count the vertices $v \in [a]$ satisfying $N_n(v, r) \in \mathcal{C}$. By Theorem 2.4, $X$ converges in distribution to $\mathrm{Bin}(a, \mu_{\mathcal{C},r})$. Hence

$$\Pr(A) \leq \Pr(X < k) \leq \epsilon + o(1),$$

as we wanted. This completes the proof. □

### 2.3.2 Probabilities of Fragment Classes

Here we study the probability that $\mathrm{Core}_n|_r$ belongs to a particular $\equiv_k^{\mathrm{Ly}}$-class. We use the following easy-to-derive fact about Poisson distributions.

**Lemma 2.8.** *Let $S$ be some finite set and $D$ some distribution over $S$. Let $(X_i)_{i=1}^\infty$ be a sequence of identically distributed random variables $X_i \sim D$. Let $\lambda \in (0, \infty)$, and let $Z \sim \mathrm{Pois}_\lambda$ be another variable independent from $(X_i)_{i=0}^\infty$. For each element $s$, define the variable $Y_s = \{i \mid 1 \leq i \leq Z,\, X_i = s\}$. Then $Y_s \sim \mathrm{Pois}_{\lambda\beta_s}$, where $\beta_s = \Pr(D = s)$.*

**Theorem 2.6.** *Let $\mathcal{C}$ be a $\equiv_k^{\mathrm{Ly}}$-class of $r$-fragments. Then the probability that $\mathrm{Core}_n|_r \in \mathcal{C}$ converges to an expression belonging to* Expr.

*Proof.* Instead we prove that $\Pr(\Gamma|_r \in \mathcal{C})$ belongs to Expr, where $\Gamma|_r$ is the random fragment defined at Definition 2.8. By Lemma 2.6 this is equivalent to the theorem. Let $\mathcal{D}_1, \dots, \mathcal{D}_\ell$ be the $\equiv_k^{\mathrm{Ly}}$-classes of $r$-unicycles, and let $X_i$ be count the number of components in $\Gamma|_r$ belonging to $\mathcal{D}_i$ for $1 \leq i \leq \ell$. The event that $\Gamma|_r \in \mathcal{C}$ is, by definition, equivalent to

$$\left( \bigwedge_{\mathcal{D}_i \in \mathfrak{D}^-} X_i = a_i \right) \wedge \left( \bigwedge_{\mathcal{D}_i \in \mathfrak{D}^+} X_i \geq k \right)$$

for some partition $\mathfrak{D}^- \cup \mathfrak{D}^+$ of $\mathcal{D}_1, \dots, \mathcal{D}_\ell$ and some constants $0 \le a_i < k$ for all $\mathcal{D}_i \in \mathfrak{D}^-$. In order to prove that $\Pr(\Gamma|_r \in \mathcal{C}) \in \text{Expr}$, we show that the $X_1, \dots, X_\ell$ are distributed as mutually independent Poisson variables $\text{Pois}_{\mu_1}, \dots, \text{Pois}_{\mu_\ell}$, where each $\mu_i$ belongs to Expr. This is clearly sufficient, given the definition of Expr and the conjunction of events above. For each class of unicycles $\mathcal{D}_i$ let $H_i$ be a representative and let $C_i \subseteq H_i$ be its corresponding cycle or the empty structure if $H_i$ is empty. The quantity $\text{aut}(\mathcal{D}_i)$ is defined as the number of automorphisms $\phi$ of $C_i$ that preserve the classes of the hanging trees. That is $T(U, v) \equiv_k^{\text{Ly}} T(U, \phi(v))$ for all $v \in V(C_i)$. Observe that the definition of $\text{aut}(\mathcal{D}_i)$ does not depend on the chosen representative. For each $v \in V(C_i)$, let $\mathcal{B}_v$ be the $\equiv_k^{\text{Ly}}$-class of the tree $T(U, v)$. The expression $\lambda_{\mathcal{D}_i, r}$ is given by

$$\lambda_{\mathcal{D}_i, r} = \lambda_{C_i} \frac{\text{aut}(C_i)}{\text{aut}(\mathcal{D}_i)} \prod_{v \in V(C_i)} \mu_{\mathcal{B}_v, r},$$

Note that, again, $\lambda_{\mathcal{D}_i, r}$ does not depend on the selected representative $U_i$. Observe as well that $\lambda_{\mathcal{D}_i, r}$ belongs to Expr by Theorem 2.5. Consider a copy $C_*$ of $C_i$ in the random $r$-fragment $\Gamma|_r$. The probability that $C_*$'s component, denoted $U_*$, belongs to the class $\mathcal{D}_i$ is exactly $\lambda_{\mathcal{D}_i, r}$. Indeed, the number of ways to choose the $\equiv_k^{\text{Ly}}$-class of the tree $T(v, U_*)$ for each vertex $v \in V(C_*)$ so that the component $U_*$ belongs to $\mathcal{D}_i$ is precisely $\frac{\text{aut}(C_i)}{\text{aut}(\mathcal{D}_i)}$. Also, given a choice of tree classes, the probability that all the trees $T(v, U_*)$ belong to their prescribed classes simultaneously is $\prod_{v \in V(C_i)} \mu_{\mathcal{T}_v, r}$, as each tree is an independent copy of $\mathcal{BP}|_r$. Remember that the number of $C_i$-copies in $\Gamma|_r$ follows the distribution $\text{Pois}_{\lambda_{C_i}}$, where $\lambda_{C_i} = \frac{1}{\text{aut}(C_i)} \prod_{1 \le j \le |\sigma|} c_j^{|E_j(C_i)|}$. Hence by Lemma 2.8 the variables $X_i$ have distribution $\text{Pois}_{\mu_i}$, where $\mu_i = \lambda_{C_i} \lambda_{\mathcal{D}_i, r}$, independently. As both $\lambda_{C_i}$ and $\lambda_{\mathcal{D}_i, r}$ belong to Expr, so does $\mu_i$, proving the result. □

### 2.3.3 Main Result

*Proof of Theorem 2.1.* Let $\varphi \in \text{FO}[\sigma]$ be a sentence, let $k = \text{qr}(\varphi)$, and $r = (3^k - 1)/2$. Let $\mathcal{C}_1, \dots, \mathcal{C}_\ell$ be an enumeration of all $\equiv_k^{\text{Ly}}$-classes of $r$-fragments. For each $i$, let $A_i(n)$ be the event that $\mathcal{G}^\sigma(n, \boldsymbol{p})$ is $r$-simple, $(k, r)$-rich, and $\text{Core}_n|_r \in \mathcal{C}_i$. By Corollary 2.2 and Corollary 2.3, $\mathcal{G}^\sigma(n, \boldsymbol{p})$ is $r$-simple and $(k, r)$-rich w.h.p., so $\sum_{i=1}^{\ell} \Pr(A_i(n)) = 1 - o(1)$, and

$$\Pr(\mathcal{G}^\sigma(n, \boldsymbol{p}) \models \varphi) = \sum_{i=1}^{\ell} \Pr(\mathcal{G}^\sigma(n, \boldsymbol{p}) \models \varphi \mid A_i(n)) \Pr(A_i(n)) + o(1).$$

By Theorem 2.3, $\Pr(\mathcal{G}^\sigma(n, \boldsymbol{p}) \models \varphi \mid A_i(n))$ tends to either zero or one for each $i$, depending only on the class $\mathcal{C}_i$. Let $I \subseteq [\ell]$ be the set of indices for which this expression tends to one. Then,

$$\Pr(\mathcal{G}^\sigma(n, \boldsymbol{p}) \models \varphi) = \sum_{i \in I} \Pr(A_i(n)) + o(1).$$

Finally, as $\mathcal{G}^\sigma(n, \boldsymbol{p})$ is $r$-simple and $(k, r)$-rich w.h.p., $\Pr(A_i(n)) = \Pr(\text{Core}_n|_r \in \mathcal{C}_i) + o(1)$, implying that

$$\Pr(\mathcal{G}^\sigma(n, \boldsymbol{p}) \models \varphi) = \sum_{i \in I} \Pr(Core_n|_r \in \mathcal{C}_i) + o(1).$$

However, the probabilities in the sum converge to expressions in Expr by Theorem 2.6, so the result follows. □

## 2.4 Convergence Law For Derived Models

During this section we extend the convergence given in Theorem 2.1 to slightly more complex models of random structures. In the graph setting, one may consider directed or undirected graphs, as well as graphs with or without self-edges (i.e., loops). In logical terms, those natural variants correspond to choosing whether the adjacency relation between vertices is symmetric and anti-reflexive. Each of those graph notions admits a natural random binomial model analogous to $\mathcal{G}(n,p)$: namely, take $n$ vertices and place each possible edge between them with probability $p$ independently. In a similar vein, we consider arbitrary relational structures subject to some given symmetry and anti-reflexivity axioms, and study their corresponding binomial models.

A **symmetry axiom** for the relation $E_i \in \sigma$ is a sentence of the form

$$\mathrm{Sym}(E_i,\Phi) = \bigwedge_{\varphi\in\Phi} \forall x_1,\dots,x_{\mathrm{ar}E_i}\, E_i\left(x_1,\dots,x_{\mathrm{ar}E_i}\right) \implies E_i\left(x_{\varphi(i)},\dots,x_{\varphi(\mathrm{ar}E_i)}\right),$$

where $\Phi$ is a subgroup of the symmetric group on $[\mathrm{ar}(E_i)]$. Given this axiom, $\Phi$ is called the **symmetry group** of $E_i$. An **irreflexivity** axiom for $E_i$ is a sentence of the form

$$\mathrm{Irrflx}(E_i,I) = \bigwedge_{(s,t)\in I} \forall x_1,\dots,x_{\mathrm{ar}E_i}\,(x_s = x_t) \implies \neg E_i(x_1,\dots,x_{\mathrm{ar}E_i}),$$

where $I \subseteq \binom{[\mathrm{ar}(E_i)]}{2}$ is called the set of $E_i$'s **irreflexivity constraints**.

We consider a theory $\Gamma$ consisting of symmetry axioms $\mathrm{Sym}(R_1,\Phi_1),\dots,\mathrm{Sym}(R_{|\sigma|},\ \Phi_{|\sigma|})$ together with irreflexivity axioms $\mathrm{Irrflx}(R_1,I_1),\dots,\mathrm{Irrflx}(R_{|\sigma|},I_{|\sigma|})$. Our first goal during this section is to define an appropriate random model of $\sigma$-structures satisfying this theory. Let $G$ be one of those structures. Given a tuple $\boldsymbol{v}\in V(G)^{\mathrm{ar}(E_i)}$, its orbit with respect to the action of $\Phi_i$ comprises all tuples of the form $(v_{\varphi(1)},\dots,v_{\varphi(\mathrm{ar}(E_i))})$, where $\varphi\in\Phi_i$. We write $V(G)^{\mathrm{ar}(E_i)}/\Phi_i$ for the set of all such orbits, and $[\boldsymbol{v}]$ for the orbit corresponding to a tuple $\boldsymbol{v}\in V(G)^{\mathrm{ar}(E_i)}$. Observe that a tuple $\boldsymbol{v}$ belongs to $E_i(G)$ if and only if the whole orbit $[\boldsymbol{v}]$ is part of $E_i(G)$. Thus, a natural way of randomly constructing $E_i(G)$ would be to decide whether $[\boldsymbol{v}]\subseteq E_i(G)$ for each orbit independently with some probability. However, some classes violate irreflexivity constraints and are forbidden - e.g., think of loops in loop-less graphs. We call a class $[\boldsymbol{v}]\in V(G)^{\mathrm{ar}(E_i)}/\Phi_i$ **forbidden** if it contains a tuple $(u_1,\dots,u_{\mathrm{ar}(E_i)})$ where $u_s=u_t$ for some $\{s,t\}\in A_i$. The set of forbidden classes in $V(G)^{\mathrm{ar}(E_i)}/\Phi_i$ is denoted by $\mathrm{Forb}_{V(G)}(E_i)$, or $\mathrm{Forb}_n(E_i)$ if $V(G)=[n]$.

Let $(q_1,\dots,q_{|\sigma|})$ be a tuple of probabilities. The random structure $\mathcal{G}^\sigma_\Gamma(n,\boldsymbol{q})$ is constructed by setting $[\boldsymbol{v}]\subseteq E_i(\mathcal{G}^\sigma_\Gamma(n,\boldsymbol{q}))$ with probability $q_i$ independently for all $[\boldsymbol{v}]\in[n]^{\mathrm{ar}(E_i)}/\Phi_i\setminus\mathrm{Forb}_n(E_i)$ and all $1\le i\le|\sigma|$. Observe that by definition $\mathcal{G}^\sigma_\Gamma(n,\boldsymbol{q})$ satisfies the theory $\Gamma$ and all $\sigma$-structures on $[n]$ satisfying this theory have positive probability. As before, we consider structures of linear density. That is, $\boldsymbol{p}(n)=(p_1(n),\dots,p_{|\sigma|}(n))$ where, $p_i(n)\sim c_i/n^{\mathrm{ar}(E_i)-1}$ for some $\boldsymbol{c}=(c_1,\dots,c_{|\mathrm{ar}|})$. The main result of this section is an extension of Theorem 2.1 to $\mathcal{G}^\sigma_\Gamma(n,\boldsymbol{p})$.

**Theorem 2.7.** *Consider a* FO$[\sigma]$*-sentence* $\phi$*. Then the limit*

$$p_\phi(\boldsymbol{c}) = \lim_{n\to\infty} \Pr(\mathcal{G}^\sigma_\Gamma(n,\boldsymbol{p}) \models \phi)$$

*exists for all values* $\boldsymbol{c} \in (0,\infty)^{|\sigma|}$*. Moreover, for any fixed* $\phi$*,* $p_\phi(\boldsymbol{c})$ *is a finite (possibly empty) sum of expressions belonging to* Expr*.*

Roughly, to prove this result we interpret $\mathcal{G}^\sigma_\Gamma(n,\boldsymbol{p})$ through a simpler model $\mathcal{G}^{\widehat{\sigma}}(n,\widehat{\boldsymbol{p}})$, in such a way that sentences $\phi \in \mathrm{FO}[\sigma]$ about the original model correspond to sentences $\widehat{\phi} \in \mathrm{FO}[\sigma]$ about the interpretation. This way, the events $\mathcal{G}^\sigma_\Gamma(n,\boldsymbol{p}) \models \phi$ and $\mathcal{G}^{\widehat{\sigma}}(n,\widehat{\boldsymbol{p}}) \models \widehat{\phi}$ have the same asymptotic probability.

### 2.4.1 Adding Symmetries

Given $\sigma = \{E_1,\dots,E_{|\sigma|}\}$, we define $\widehat{\sigma}$ as

$$\sigma \bigcup_{i=1}^{|\sigma|} \{E_i^{s,t} \mid 1 \le s < t \le \mathrm{ar}(E_i)\},$$

where each relation symbol $E_i$ retains its arity from $\sigma$, and $\mathrm{ar}(E_i^{s,t}) = \mathrm{ar}(E_i)$. We map each sentence $\phi \in \mathrm{FO}[\sigma]$ to another one $\widehat{\phi} \in \mathrm{FO}[\widehat{\sigma}]$ by replacing, for all $i = 1,\dots,|\sigma|$, each occurrence of $E_i(x_1,\dots,x_{\mathrm{ar}(E_i)})$, with

$$\widehat{E_i}(x_1,\dots,x_{\mathrm{ar}(E_i)}) \equiv \bigvee_{g\in\Phi_i} \tilde{E}_i(x_{g(1)},\dots,x_{g(\mathrm{ar}(E_i))}),$$

where $\tilde{E}_i(\boldsymbol{y})$ stands for the formula given by

$$\left( E_i(\boldsymbol{y}) \bigwedge_{\substack{1\le s<t\\ \le \mathrm{ar}(E_i)}} y_s \ne y_t \right) \bigvee_{\substack{1\le s<t\\ \le \mathrm{ar}(E_i)}} \left( E_i^{s,t}(\boldsymbol{y}) \wedge y_s = y_t \bigwedge_{\substack{1\le s'<t'\le \mathrm{ar}(E_i)\\ (s,t)\ne(s',t')}} y_{s'} \ne y_{t'} \right).$$

Given $1 \le s < t \le |\mathrm{ar}(E_i)|$, we define $\Phi_i^{s,t}$ as the subgroup of $\Phi_i$ consisting of all permutations $g$ which leave $s$ and $t$ fixed- i.e., $g(s) = s, g(t) = t$. For convenience, we index tuples of probabilities $\widehat{\boldsymbol{p}} \in [0,1]^{|\widehat{\sigma}|}$ following the same convention as $\widehat{\sigma}$. That is, $\widehat{p}_i$ corresponds to $E_i$, whereas $\widehat{p}_i^{s,t}$ corresponds to $E_i^{s,t}$. Given $\boldsymbol{p} = \boldsymbol{p}(n)$ we build the tuple $\widehat{\boldsymbol{p}} = \widehat{\boldsymbol{p}}(n)$ in the following way:

1. If $\mathrm{ar}(E_i) = 2$ and $\Phi_i$ is the symmetric group $S_2$, then $\widehat{p}_i(n) = \frac{p_i(n)}{1+\sqrt{1-p_i(n)}}$, and $\widehat{p}_i^{1,2}(n) = p_i(n)$.
2. Otherwise $\widehat{p}_i(n) = \frac{1}{|\Phi_i|} p_i(n)$ and $\widehat{p}_i^{s,t}(n) = \frac{1}{|\Phi_i^{s,t}|} p_i(n)$ for all $1 \le s < t \le \mathrm{ar}(E_i)$.

Observe that if $\boldsymbol{p}$ satisfies $p_i(n) \sim c_i/n^{\mathrm{ar}(E_i)-1}$ for all $i$ and some positive real constants $c_1,\dots,c_{|\sigma|}$, as is assumed throughout this chapter, then the sequence of probabilities $\widehat{\boldsymbol{p}}$ is also of the form required by Theorem 2.1 with respect to the signature $\widehat{\sigma}$. In other words, $\mathcal{G}^{\widehat{\sigma}}(n,\widehat{\boldsymbol{p}})$ satisfies a FO convergence law.

**Lemma 2.9.** *Let $\widehat{\sigma}, \widehat{\boldsymbol{p}}$ be defined as above. Let $\boldsymbol{v} \in \mathbb{N}^{\mathrm{ar}(E_i)}$ be a tuple of vertices, where at most one vertex occurs exactly twice and all the others once. Then both* $\Pr\left(\mathcal{G}^{\widehat{\sigma}}(n, \widehat{\boldsymbol{p}}) \models \widehat{E}_i(\boldsymbol{v})\right) \leq p_i(n)$ *and* $\Pr\left(\mathcal{G}^{\widehat{\sigma}}(n, \widehat{\boldsymbol{p}}) \models \widehat{E}_i(\boldsymbol{v})\right) = p_i(n) + o(1/n^{\mathrm{ar}(E_i)})$ *hold. Otherwise, if $\boldsymbol{v}$ is of another form,* $\Pr\left(\mathcal{G}^{\widehat{\sigma}}(n, \widehat{\boldsymbol{p}}) \models \widehat{E}_i(\boldsymbol{v})\right) = 0$.

*Proof.* The second part of the lemma follows directly from the definition of $\widehat{E}_i$. We prove the first part for the case where $\boldsymbol{v}$ contains no repetitions. The case where $\boldsymbol{v}$ has some vertex appearing twice can be shown analogously. By definition, $\widehat{E}_i(\boldsymbol{v})$ is equivalent to

$$\bigvee_{\substack{g \in \Phi_i \\ \boldsymbol{w}=(v_{g(1)}, \ldots, v_{g(\mathrm{ar}(E_i))})}} E_i(\boldsymbol{w}). \tag{2.4}$$

We distinguish two cases. First, if $\mathrm{ar}(E_i) = 2$ and $\Phi_i$ is the symmetric group on two elements, then $\widehat{E}_i(v_1, v_2)$ is equivalent to $E_i(v_1, v_2) \vee E_i(v_2, v_1)$, so by inclusion-exclusion

$$\Pr\left(\mathcal{G}^{\widehat{\sigma}}(n, \widehat{\boldsymbol{p}}) \models \widehat{E}_i(\boldsymbol{v})\right) = 2\widehat{p}_i - (\widehat{p}_i)^2 = p_i,$$

fulfilling the statement. Otherwise suppose that $\mathrm{ar}(E_i) > 2$. Using the union bound on Equation (2.4), we obtain

$$\Pr\left(\mathcal{G}^{\widehat{\sigma}}(n, \widehat{\boldsymbol{p}}) \models \widehat{E}_i(\boldsymbol{v})\right) \leq |\Phi_i|\widehat{p}_i = p_i.$$

Similarly, the first Bonferroni inequality yields

$$\Pr\left(\mathcal{G}^{\widehat{\sigma}}(n, \widehat{\boldsymbol{p}}) \models \widehat{E}_i(\boldsymbol{v})\right) \geq |\Phi_i|\widehat{p}_i - \frac{|\Phi_i|(|\Phi_i| - 1)}{2}(\widehat{p}_i)^2 = p_i - o(1/n^{\mathrm{ar}(E_i)}).$$

Last equality follows from the fact that $\widehat{p}_i = O(n^{1-\mathrm{ar}(E_i)})$ and $\mathrm{ar}(E_i) > 2$. This equation together with the last one show the desired result. ☐

**Theorem 2.8.** *Given $\varphi \in \mathrm{FO}[\sigma]$,*

$$\Pr\left(\mathcal{G}^{\sigma}_{\Gamma_{\mathrm{Sym}}}(n, \boldsymbol{p}) \models \varphi\right) = \Pr\left(\mathcal{G}^{\widehat{\sigma}}(n, \widehat{\boldsymbol{p}}) \models \widehat{\varphi}\right) + o(1/n).$$

*Proof.* For convenience we shorten $\mathcal{G}^{\sigma}_{\Gamma_{\mathrm{Sym}}}(n, \boldsymbol{p})$ to $\mathcal{G}_n$, and $\mathcal{G}^{\widehat{\sigma}}(n, \widehat{\boldsymbol{p}})$ to $\widehat{\mathcal{G}}_n$. In $\mathcal{G}_n$ we write $E_i[\boldsymbol{v}]$ for the sentence $\bigwedge_{\boldsymbol{u} \in [\boldsymbol{v}]} E_i(\boldsymbol{u})$, and in $\widehat{\mathcal{G}}_n$ we write $\widehat{E_i}[\boldsymbol{v}]$ for $\bigwedge_{\boldsymbol{u} \in [\boldsymbol{v}]} \widehat{E}_i(\boldsymbol{u})$ as well, where $[\boldsymbol{v}]$ stands for $\boldsymbol{v}$'s orbit in $[n]^{\mathrm{ar}(E_i)}/\Phi_i$. Note that the events $E_i[\boldsymbol{v}]$ are mutually independent in $\mathcal{G}_n$ for all the different orbits $[\boldsymbol{v}]$ and indices $1 \leq i \leq |\sigma|$, and the same holds true for the events $\widehat{E}_i[\boldsymbol{v}]$ in $\widehat{\mathcal{G}}_n$. Moreover, $\mathcal{G}_n$ can be described as the independent product of the Bernoulli variables $E_i[\boldsymbol{v}]$. By last lemma, the probability that $\widehat{\mathcal{G}}_n \models \widehat{E}_i[\boldsymbol{v}]$ is at most $p_i$, which is precisely the probability that $\mathcal{G}_n \models E_i[\boldsymbol{v}]$. This way, there is a coupling $(\mathcal{G}_n, \widehat{\mathcal{G}}_n$ where $\widehat{\mathcal{G}}_n \models \widehat{E}_i[\boldsymbol{v}]$ implies $\mathcal{G}_n \models E_i[\boldsymbol{v}]$. We claim that within this coupling $\widehat{\mathcal{G}}_n \models \widehat{E}_i[\boldsymbol{v}] \leftrightarrow \mathcal{G}_n \models E_i[\boldsymbol{v}]$ holds simultaneously for all $1 \leq i \leq |\sigma|$ and all $\boldsymbol{v}$ with probability $o(1)$. Fix $1 \leq i \leq |\sigma|$. Let $U_n \subseteq [n]^{\mathrm{ar}(E_i)}$ be the set of tuples containing less than $\mathrm{ar}(E_i) - 1$ different vertices. The size of $U_n$ is $o(n^{\mathrm{ar}(E_i)-1})$, so expected number of tuples $\boldsymbol{v} \in U_n$ satisfying $E_i[\boldsymbol{v}]$ is $o(1)$, and we do not need to take this set into account. Now, the size of

$[n]^{\mathrm{ar}(E_i)} \setminus U_n$ is $O(n^{\mathrm{ar}(E_i)})$, and by last lemma, the probability that for a tuple $\boldsymbol{v} \in [n]^{\mathrm{ar}(E_i)} \setminus U_n$, the event $E_i[\boldsymbol{v}] \wedge \neg \widehat{E}_i[\boldsymbol{v}]$ is $o(1/n^{\mathrm{ar}(E_i)})$, uniformly over the choice of $\boldsymbol{v}$. Thus, using the union bound, the probability that this occurs for any $\boldsymbol{v} \in [n]^{\mathrm{ar}(E_i)} \setminus U_n$ is $o(1)$.

We have shown that within the coupling $\Pi_n$, for all $1 \leq i \leq |\sigma|$, the sentence $\psi_i := \forall \boldsymbol{x} E_i[\boldsymbol{x}] \iff \widehat{E}_i[\boldsymbol{x}]$ holds with probability $1 - o(1)$. Observe that given a sentence $\varphi \in \mathrm{FO}[\sigma]$, $\widehat{\varphi} \iff \varphi$ follows from $\bigwedge_{1 \leq i \leq |\sigma|} \psi_i$. As $\bigwedge_{1 \leq i \leq |\sigma|} \psi_i$ holds with probability $1 - o(1)$, the result $\Pr(\widehat{G}_n \models \widehat{\varphi}) = \Pr(G_n \models \varphi) + o(1)$ follows. □

**Corollary 2.4.** *Theorem 2.7 holds if* $\Gamma$ *consists only of symmetry axioms. That is, if* $\Gamma = \Gamma_{\mathrm{Sym}}$.

*Proof.* The fact that $\mathcal{G}^{\sigma}_{\Gamma_{\mathrm{Sym}}}(n, \boldsymbol{p})$ satisfies a FO-convergence law follows from last result, using that a FO-convergence law already holds in $\mathcal{G}^{\widehat{\sigma}}(n, \widehat{p})$, because

$$\lim_{n \to \infty} \Pr(\mathcal{G}^{\sigma}_{\Gamma_{\mathrm{Sym}}}(n, \boldsymbol{p}) \models \varphi) = \lim_{n \to \infty} \Pr(\mathcal{G}^{\widehat{\sigma}}(n, \widehat{\boldsymbol{p}}) \models \widehat{\varphi}).$$

Also, $\widehat{\boldsymbol{c}}$ is a tuple of expressions in $\mathrm{Expr}_{\boldsymbol{c}}$, meaning that the limit on the right belongs to $\mathrm{Expr}_{\boldsymbol{c}}$ itself, proving the result. □

### 2.4.2 Adding Irreflexivity Constraints

In this section we complete the proof of our generalization, Theorem 2.7, by showing that it holds when $\Gamma$ includes irreflexivity axioms.

*Proof of Theorem 2.7.* Let $\mathcal{G}_n = \mathcal{G}^{\sigma}_{\Gamma}(n, \boldsymbol{p})$, $\mathcal{G}^{\mathrm{Sym}}_n = \mathcal{G}^{\sigma}_{\Gamma_{\mathrm{Sym}}}(n, \boldsymbol{p})$. Proving Theorem 2.7 amounts to showing that the asymptotic probability of $\mathcal{G}^{\mathrm{Sym}}_n \models \Gamma_{\mathrm{Irrflx}}$ is positive. If this holds, the result easily follows: By definition

$$\Pr(\mathcal{G}_n \models \varphi) = \Pr\left(\mathcal{G}^{\mathrm{Sym}}_n \models \varphi \,\middle|\, \mathcal{G}^{\mathrm{Sym}}_n \models \Gamma_{\mathrm{Irrflx}}\right) = \frac{\Pr(\mathcal{G}^{\mathrm{Sym}}_n \models \varphi \wedge \Gamma_{\mathrm{Irrflx}})}{\Pr(\mathcal{G}^{\mathrm{Sym}}_n \models \Gamma_{\mathrm{Irrflx}})}.$$

As Irrflx is a first-order property, last result tells us that both numerator and denominator in the last fraction converge to (possibly empty) sums of expressions in $\mathrm{Expr}_{\boldsymbol{c}}$. We only need to see that the denominator is bounded away from zero. For this, observe that Irrflx amounts to the fact that $\mathcal{G}^{\mathrm{Sym}}_n$ does not contain any copy of $H_1, \dots, H_\ell$, where the $H_i$ are cycles given by Irrflx. Using the method of moments, in the same fashion as Lemma 2.4, shows that the number of $H_i$-copies in $\mathcal{G}^{\mathrm{Sym}}_n$ converge jointly to a product of Poisson distributions with positive mean. This proves the result. □

## 2.5 Application to Random SAT

In this section we give an application of the general convergence law given in Theorem 2.7 to the study of satisfiability of random CNF formulas.

Given a variable $x$, both expressions $x$ and $\neg x$ are called **literals**. A **clause** is a set of literals. A clause $C$ is called **non-tautological** if $C$ does not contain two literals of the form $x$ and $\neg x$ for any variable $x$. An **assignment** over a set of variables $X$ is a map $f : X \to \{0,1\}$ . A clause $C$ is **satisfied** by an assignment $f$ if $f(x) = 1$ for some variable $x$ with $x \in C$, or $f(x) = 0$ for some variable $x$ with $\neg x \in C$. Given $\ell \in \mathbb{N}$ a **$\ell$-CNF formula** is a set of non-tautological clauses that contain exactly $\ell$ literals each. We say that a formula $F$ on the variables $x_1, \dots, x_n$ is **satisfiable** if there is an assignment $f : \{x_1, \dots, x_n\} \to \{0,1\}$ that satisfies all clauses in $F$.

The $k$-SAT problem is the problem of deciding whether a given $k$-CNF formula is satisfiable or not. For $k > 2$, this is the prototypical NP-complete problem [3][Section 2.3.5]. As such, the search for an efficient algorithm solving $k$-SAT, or a proof that such algorithm cannot exist, is one of the central problems of theoretical computer science in present times. It is known that a phase transition occurs at random $k$-CNF formulas with $n$ variables and $cn$ clauses: there are constants $c_0, c_1$ such that when $c < c_0$, typical formulas are satisfiable, and when $c > c_1$, typical formulas are unsatisfiable [14]. In fact, a well-known conjecture, proven for large $k$, is that one can take $c_0 = c_1$ [19]. The complexity of random $k$-SAT on random formulas near to this "satisfiability threshold" has been a relevant object of study for various areas in the literature, such as proof complexity [15], applied statistical physics [54], and hardness of approximation [22]. An important question in this area is whether there is a computationally "simple" (i.e., easy to recognize) property $P$ that implies unsatisfiability of $k$-CNF formulas and can be used to recognize most unsatisfiable random instances [22, 26, 5]. The most notable example of a somewhat complex property being certified by a simpler one on random structures occurs with graph connectivity. A famous result by Erdös and Rènyi states that, at the connectivity threshold $p(n) = \ln n/n + c/n$, w.h.p. the random graph $\mathcal{G}(n,p)$ is disconnected if and only if it contains some isolated vertex [20].

In [6, 5], Atserias employed a model-theoretical approach to the study of unsatisfiability certificates in random $k$-SAT. The first publication [6], considers certificates expressible in the Datalog language, and the second, [5], FO sentences. Our contribution in this section extends this second paper. In there, 3-CNF formulas are represented as relational structures. The main result of the article is that when the number of clauses in the random formula is $\Theta(n^{2+\alpha})$ for some $\alpha > 0$, where $n$ stands for the (growing) number of variables, then there is some FO sentence $\varphi$ that implies unsatisfiability and holds w.h.p. Conversely, when the number of clauses is $\Theta(n^{2-\alpha})$ for some irrational $\alpha > 0$, then no FO sentence $\varphi$ implying unsatisfiability holds w.h.p. We extend this result for the case where the number of clauses is linear in the number of variables. The main result of this section, Theorem 2.9, states that in this situation no FO property certifying unsatisfiability holds with probability bounded away from zero.

We define a binomial model of random CNF formulas along the lines of [14], but the generality in Theorem 2.7 allows for many variants.

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

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

where $|F|$ is the number of clauses in $F$. Equivalently, a random formula in $\mathcal{F}^\ell(n,p)$ is obtained by choosing each of the $2^\ell\binom{n}{\ell}$ non-tautological clauses of size $\ell$ on the variables $\{x_i\}_i$ with probability $p$ independently. We denote by $\mathcal{F}^\ell_n(\beta)$ a random sample of $\mathcal{F}^\ell(n, \beta/n^{\ell-1})$.

We consider $\ell$-CNF formulas, as defined above, as relational structures with a language $\sigma$ consisting of $\ell+1$ relation symbols $R_0,\dots,R_l$ of arity $\ell$. We do that in such a way that the expression $R_j(x_{i_1},\dots,x_{i_l})$ means that our formula contains the clause consisting of $\neg x_{i_1},\dots,\neg x_{i_j}$ and $x_{i_{j+1}},\dots x_{i_l}$. The relations $R_1,\dots,R_l$ satisfy the following axioms: (1) given $0\le j\le l$ and variables $y_1,\dots,y_l$ the fact that $R_j(y_1,\dots,y_l)$ holds is invariant under any permutation of the variables $y_1,\dots,y_j$ or $y_{j+1},\dots,y_l$, and (2) for any $0\le j\le l$ and any variables $y_1,\dots,y_l$ it holds that $R_j(y_1,\dots,y_l)$ only if all the $y_i$ are different. Call $\Gamma$ to the theory consisting of those symmetry and anti-reflexivity axioms. Then the random model $F_l(n,p)$ corresponds to the random $\sigma$-structure $\mathcal{G}^\sigma_\Gamma(n,\boldsymbol{p})$ described in Section 2.4 when $\boldsymbol{p}$ is a tuple of probabilities all equal to $p$. We obtain the following result as a particular case of Theorem 2.7.

**Theorem 2.9.** *Let $\ell>1$ be a natural number. Then for each sentence $\Phi\in \mathrm{FO}[\sigma]$ it is satisfied that the map $f_\Phi:(0,\infty)\to\mathbb{R}$ given by*

$$\beta\mapsto \lim_{n\to\infty}\Pr(\mathcal{F}^\ell_n(\beta)\models\Phi)$$

*is well defined and analytic.*

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

**Theorem 2.10.** *Let $\ell\ge 2$ be a natural number, and let $c\in(0,\infty)$ be an arbitrary real number. Let $m:\mathbb{N}\to\mathbb{N}$ be such that $m(n)\sim cn$. For each $n$ let $C_{n,1},\dots,C_{n,m(n)}$ be clauses chosen uniformly at random independently among the $2^\ell\binom{n}{\ell}$ non-tautological clauses of size $\ell$ over the variables $x_1,\dots,x_n$. For each $n$, let $\textsc{UnSat}_n$ denote the event that there is no assignment of the variables $x_1,\dots,x_n$ that satisfies all clauses $C_{n,1},\dots,C_{n,m(n)}$. Then there are two real constants $0<c_1<c_2$, such that a.a.s $\textsc{UnSat}_n$ does not hold if $c<c_1$, and a.a.s $\textsc{UnSat}_n$ holds if $c>c_2$.*

The existence of $c_1$ is proven in [14, Theorem 1]. The fact that $c_2$ exists follows from a direct application of the first moment method and is also shown for instance in [14, 24, 16]. We want to show that an analogous "phase transition" also happens in $\mathcal{F}^\ell(n,p)$ when $p\sim\beta/n^{\ell-1}$. We start by showing the following

**Corollary 2.5.** *Let $\ell\ge 2$ be a natural number. Let $c\in(0,\infty)$ be an arbitrary real number and let $m:\mathbb{N}\to\mathbb{N}$ satisfy $m(n)\sim cn$. For each $n\in\mathbb{N}$ let $\mathcal{F}^\ell(n,m(n))$ be a random formula chosen uniformly at random among all sets of $m(n)$ non-tautological clauses of size $\ell$ over the variables $x_1,\dots,x_n$. Then there are two real positive constants $0<c_1<c_2$ such that a.a.s $\mathcal{F}^\ell(n,m(n))$ is satisfiable if $c<c_1$, and a.a.s $\mathcal{F}^\ell(n,m(n))$ is unsatisfiable if $c>c_2$.*

*Proof.* For each $n\in\mathbb{N}$ let $C_{n,1},\dots,C_{n,m(n)}$ and $\textsc{UnSat}_n$ be as in the previous theorem. One can consider $\mathcal{F}^\ell(n,m(n))$ to be the result of selecting clauses $C_{n,1},\dots,C_{n,m(n)}$ uniformly at random independently among all possible clauses, given the fact that no two clauses $C_{n,i},C_{n,j}$ are equal. Hence,

$$\Pr\left(\mathcal{F}^\ell(n,m(n))\text{ is unsatisfiable }\right)=\Pr\left(\textsc{UnSat}_n \;\middle|\; \text{all the } C_{n,i} \text{ are different }\right).$$

An application of the first moment method yields that for $\ell \geq 3$ a.a.s the number of unordered pairs $\{i, j\}$ such that $C_{n,i} = C_{n,j}$ is equal to zero. In the case of $\ell = 2$, an application of the Moments Method (Theorem 1.4) proves that the number of such pairs $\{i, j\}$ converges in distribution to a Poisson variable. In either case all the $C_{n,i}$ are different with positive asymptotic probability. Thus the constants $c_1$ and $c_2$ from the previous theorem satisfy our statement. □

Let $\mathcal{F}^\ell(n, m(n))$ be as in last result. Note that because of the symmetry in the random model $\mathcal{F}^\ell(n, p(n))$ one can consider $\mathcal{F}^\ell(n, m(n))$ to be a random sample of the space $\mathcal{F}^\ell(n, p(n))$ given that the number of clauses is $m(n)$. Using this observation we can prove the following.

**Theorem 2.11.** *Let $\ell > 1$. Then there are real positive values $\beta_1 < \beta_2$ such that a.a.s $\mathcal{F}^\ell_n(\beta)$ is satisfiable for $0 < \beta < \beta_1$ and a.a.s $\mathcal{F}^\ell_n(\beta)$ is unsatisfiable and for $\beta > \beta_2$.*

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

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

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

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

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

Because of the Law of large numbers,

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

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

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

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

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

**Theorem 2.12.** *Let $\ell > 1$ be a natural number. Let $\Phi \in \mathrm{FO}[\sigma]$ be a first order sentence that implies unsatisfiability. Then for all $\beta > 0$ a.a.s $\mathcal{F}^\ell_n(\beta)$ does not satisfy $\Phi$.*

*Proof.* Let $\beta_1$ and $\beta_2$ be as in Theorem 2.11. As $\Phi$ implies unsatisfiability $\Pr(\mathcal{F}_n^\ell(\beta) \models \Phi) \leq \Pr(\mathcal{F}_n^\ell(\beta)$ is unsatisfiable ). Thus, by Theorem 2.11, we get that for all $\beta \in (0, \beta_1]$

$$\lim_{n\to\infty} \Pr(\mathcal{F}_n^\ell(\beta) \models \Phi) = 0.$$

By Theorem 2.9, last limit varies analytically with $\beta$. It vanishes in the proper interval $(0, \beta_1]$ then by the Principle of analytic continuation it has to vanish in the whole $(0, \infty)$, and the result holds. □

# Chapter 3

# Sets of Limiting Probabilities

In the previous chapter we studied the convergence of probabilities related to FO properties. In this one, we work with models $\mathcal{G}_n$ where a FO convergence law is known to hold and we study the set

$$L = \left\{ \lim_{n\to\infty} \Pr(\mathcal{G}_n \models \varphi) \;\mid\; \varphi \text{ FO sentence} \right\}.$$

When $\mathcal{G}_n$ satisfies a zero-one law, then $L = \{0, 1\}$. Otherwise, in general $L$ is a countable subset of $[0, 1]$. In this chapter we ask whether $L$ is dense in $[0, 1]$, and, more generally, we study the topological properties of its closure $\overline{L}$.

In [35] this question was considered in the context of monadic second-order (MSO) logic and uniform relational structures. Given a signature $\sigma$ containing at least one relation with arity greater than one, they show that the uniform $\sigma$-structure on $n$ vertices, $\mathcal{G}_n^\sigma$, does not satisfy a MSO convergence law. Even more, they show that any recursive real $c \in [0, 1]$ is the limit probability of some MSO property in $\mathcal{G}_n^\sigma$. More recently, this line of research was pursued in the context of the uniform graph $\mathcal{G}_n^{\mathcal{C}}$ form an arbitrary addable minor-closed class $\mathcal{C}$ in [31] and [38] independently. They show that a MSO convergence law holds in $\mathcal{G}_n^{\mathcal{C}}$ (the result in [38] applies to more general models), and also study the set $L_{\mathcal{C}}$ of limiting probabilities of MSO statements. In [38] is proven that $\overline{L_{\mathcal{C}}} \subseteq [0, c] \cup [1 - c, 1]$, where $c$ is the asymptotic probability that $\mathcal{G}_n^{\mathcal{C}}$ is not connected. In [31], they study $\overline{L_{\mathcal{C}}}$ in greater detail and find, among other results, that that $\overline{L_{\mathcal{C}}}$ always is a finite union of intervals. Furthermore, when $\mathcal{C}$ is the class of forests, then $\overline{L_{\mathcal{C}}}$ consists of exactly four intervals, and when $\mathcal{C}$ is the class of planar graphs, the set $\overline{L_{\mathcal{C}}}$ contains exactly 108 intervals. More generally, if all forbidden minors of $\mathcal{C}$ are 2-connected, $\overline{L_{\mathcal{C}}}$ is always a finite union of at least two intervals.

Throughout this chapter we study the limit probabilities of FO statements in three different random models: binomial random graphs $\mathcal{G}(n, p)$ with $p \sim c/n$ (Section 3.1), binomial $d$-uniform hypergraphs $\mathcal{G}^d(n, p)$ with $p \sim c/n^{d-1}$ and $d \geq 3$ (Section 3.2), and sparse graphs with given degree sequences (Section 3.1). Results about the binomial graph and the binomial $d$-uniform hypergraph were obtained jointly with Marc Noy and Tobias Müller, and later published in [41]. The work on graphs with given degree sequences was done in collaboration with Marc Noy and Guillem Perarnau. A relevant observation is that our results on random graphs with given degree sequences generalize our previous results about $\mathcal{G}(n,p)$ presented in [41], and use more refined techniques. However, for the sake of exposition, we present the work on $\mathcal{G}(n,p)$ with its original arguments before moving on to the generalization.

## 3.1 Binomial Graphs

In this section we consider the classical binomial random graph model $\mathcal{G}(n,p)$ from the point of view of FO logic. By Lynch's results in [49], a $\mathrm{FO}_g$ convergence law holds when $p \sim c/n$, and the limiting probability of any fixed sentence $\varphi \in \mathrm{FO}_g$ depends only on $c$. Throughout this section, let $p_\varphi(c) = \lim_{n\to\infty} \Pr(\mathcal{G}(n,c/n) \models \varphi)$, and

$$L_c = \{p_\varphi(c) \mid \varphi \text{ sentence in } \mathrm{FO}_g\}.$$

Similarly to our generalization in Chapter 2, Lynch also shows that each probability $p_\phi(c)$ is a smooth function of $c$ (more concretely, $p_\phi(c)$ is a combination of sums, products, exponentials and a set of constants). This implies in a strong way that FO logic does not capture the emergence of the giant component that occurs in $\mathcal{G}(n,c/n)$ at $c = 1$ [20] (see also [60] for a discussion including monadic second order logic).

Our main result this section is that there is a transition in the structure of $\overline{L_c}$ at a particular value of $c$. We say that $\overline{L_c}$ contains a *gap* if there is at least one subinterval $[a,b] \subseteq [0,1]$ with $a < b$ such that $\overline{L_c} \cap [a,b] = \emptyset$.

**Theorem 3.1.** *Let $\overline{L_c}$ be the closure of the of limiting probabilities of first order sentences in $G(n,c/n)$. Let $c_0 \approx 0.93$ be the unique positive solution of*

$$e^{\frac{c}{2}+\frac{c^2}{4}}\sqrt{1-c} = \frac{1}{2}. \tag{3.1}$$

*Then for every $c > 0$ the set $\overline{L_c}$ is a finite union of closed intervals. Moreover, the following holds:*

1. *$\overline{L_c} = [0,1]$ for $c \geq c_0$.*
2. *$\overline{L_c}$ has at least one gap for $0 < c < c_0$.*

This theorem follows from three intermediate results. First, $\overline{L_c} = [0,1]$ for all $c \geq 1$ (Lemma 3.8). Second, $\overline{L_c}$ is a finite union of intervals for all $c < 1$ (Lemma 3.11). Finally, $\overline{L_c}$ contains at least one gap for $c < c_0$, and equals $[0,1]$ for $c_0 \leq c < 1$ (Lemma 3.12). We give a brief outline of our techniques below.

In the regime $c \geq 1$ we show that any probability $q \in [0,1]$ can be approximated using statements of the form "There are at most $k$ cycles of length bounded by $\ell$". This follows from the asymptotic cycle distribution in $\mathcal{G}(n,c/n)$, together with the fact that Poisson distributions with large means can be approximated suitably by normal distributions. In the sub-critical regime $c < 1$, we study the set $\overline{L_c}$ through the asymptotic distribution of $\mathrm{Frag}_n$, the fragment of $\mathcal{G}(n,c/n)$. Given an unlabeled fragment $H \in \mathbb{F}$, we define $p_H = \lim_{n\to\infty} \Pr(\mathrm{Frag}_n \simeq H)$. The value of $p_H$ is computed in Lemma 3.6. We show that $\overline{L_c}$ coincides with the set of sums $\sum_{H\in\mathcal{T}} p_H$ where $\mathcal{T} \subseteq \mathbb{F}$

(Lemma 3.10). This way, we can analyze $\overline{L_c}$ as the set of partial sums of a convergent series. The main tool in this regard is the following classical result conjectured by Kakeya [33] and later proven in [55].

**Lemma 3.1** (Kakeya's Criterion)**.** *Let $\sum_{n\geq 0} p_n$ be a convergent series of non-negative real numbers. Then the following are equivalent:*

*(1)* $p_i \leq \sum_{j>i} p_j$ *for all* $i \geq 0$.

*(2)*

$$\left\{ \sum_{i\in A} p_i \colon A \subset \mathbb{N} \right\} = \left[ 0, \sum_{n\geq 0} p_n \right].$$

*Moreover, if the condition $p_i \leq \sum_{j>i} p_j$ holds for all values of $i$ large enough, then the set $\left\{\sum_{i\in A} p_i \colon A \subset \mathbb{N}\right\}$ is a finite union of intervals.*

In order to prove Lemmas 3.11 and 3.12, we consider the series of fragment probabilities $\sum_{n\geq 0} p_{H_i} = 1$, where $p_{H_0} \geq p_{H_1} \geq \dots$. Proving that $\overline{L_c} = [0,1]$ amounts to showing condition (1) above, and proving that $\overline{L_c}$ is a finite union of intervals is the same as showing that (1) holds but only for sufficiently large $i$. Those proofs involve finding good-enough lower bounds for the tails $\sum_{j>i} p_{H_j}$. In this section, as well as in the next one, which extends these results to uniform hypergraphs, we showcase the original method from [41]. This consists of using a specific family of fragments to bound $\sum_{j>i} p_{H_j}$ for sufficiently large $i > i_0$ (for example, the family of triangles with two paths attached to different vertices), and improving the bound for $1 \leq i \leq i_0$ by explicitly enumerating the first few fragments with highest probabilities. The method used in Section 3.3 for graphs with given degree sequences is slightly more refined and avoids this explicit enumeration.

Something we do not study in detail is the way the number of gaps in $\overline{L_c}$ evolves as $c$ tends to zero from above. It can be shown that the number of gaps grows to infinity as $c \to 0$, and it would be interesting to determine the growth rate. However, this is a delicate issue, since the ordering of the fragment probabilities does not depend on $c$ in straightforward manner.

### 3.1.1 Cycles and Fragments

Here we present several results related to the fragment of $\mathcal{G}(n, c/n)$, and the number of cycles in the random graph. We recall that by the fragment of a graph, we mean the union of its unicyclic components. For the most part, results here are either well-known or follow easily from existing ones. An exception is Lemma 3.6, which determines the asymptotic distribution of $\mathcal{G}(n, c/n)$'s fragment for $c < 1$.

For the rest of the section, let $X_k(n)$ denote the number of $k$-cycles in $\mathcal{G}(n, c/n)$. It is easy to show that

$$\mathrm{E}[X_k(n)] \leq \frac{c^k}{2k},$$

and

$$\lim_{n\to\infty} \mathrm{E}[X_k(n)] = \frac{c^k}{2k}.$$

In particular, the functions $k \mapsto \mathrm{E}\left[X_k(n)\right]$ form a tight sequence for $n \in \mathbb{N}$.

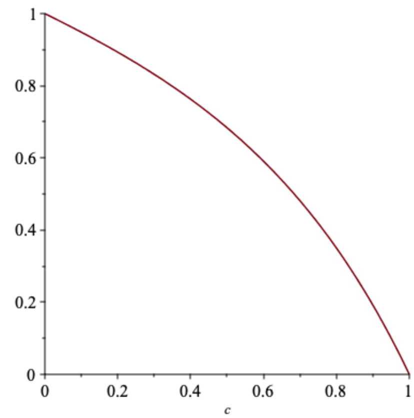



Figure 3.1: The probability that $\mathcal{G}(n, c/n)$ has no cycles as a function of $c$.

The first part of the next lemma appears already in [20] for the uniform graph $\mathcal{G}(n, M)$ on $n$ vertices and $M = M(n)$ edges, where $M \sim cn$. The second part is easily proved using the method of moments (Theorem 1.4).

**Lemma 3.2.** *For fixed $k \geq 3$, the number of $k$-cycles $X_k(n)$ in $\mathcal{G}(n, c/n)$ is distributed asymptotically as $n \to \infty$ as a Poisson law with parameter $\lambda_k = \frac{c^k}{2k}$. Moreover, for fixed $k$ the random variables $X_3(n), \dots, X_k(n)$ are asymptotically independent.*

We set

$$f(c) = \frac{1}{2} \ln \frac{1}{1-c} - \frac{c}{2} - \frac{c^2}{4}. \tag{3.2}$$

This is a function defined on $(0, 1)$ that plays an important role in our results. The function is $e^{-f(c)}$ the limiting probability that $\mathcal{G}(n, c/n)$ is acyclic; see Figure 3.1 for a plot.

**Corollary 3.1.** *When $c < 1$ the expected number of cycles in $\mathcal{G}(n, c/n)$ is $f(c)$.*

*Moreover, the limiting probability as $n \to \infty$ that $\mathcal{G}(n, c/n)$ contains no cycle is*

$$e^{-f(c)} = e^{\frac{c}{2} + \frac{c^2}{4}} \sqrt{1-c}.$$

*Proof.* As the sequence $(k \mapsto \mathrm{E}[X_k(n)])_{n \in \mathbb{N}}$ is tight,

$$\lim_{n \to \infty} \mathrm{E}\left[\sum_{k \geq 3} X_k(n)\right] = \sum_{k \geq 3} \frac{c^k}{2k} = f(c).$$

The second statement follows using this fact together with Lemma 3.2. □

The following is a well-known fact about the sub-critical regime of $\mathcal{G}(n, c/n)$. See [27, Lemma 2.10] for a proof.

**Lemma 3.3.** *Let $p(n) \sim c/n$ with $0 < c < 1$. Then a.a.s all the connected component of $\mathcal{G}(n, p)$ are either trees or unicycles.*

Let $\mathbb{F}_n$ be the set of unlabeled fragments containing $n$ edges, i.e., $\{H \in \mathbb{F} : |H| = n\}$, and let $\mathbb{F}_{\leq n} = \bigcup_{i=1}^{n} \mathbb{F}_i$. We write $\mathrm{Frag}_n$ to denote the fragment of $\mathcal{G}(n,p)$. The following result states that below the critical value $c = 1$ the expected size of $\mathrm{Frag}_n$ is asymptotically bounded.

**Lemma 3.4.** *Let $p(n) \sim c/n$ with $0 < c < 1$. Then $\lim_{n\to\infty} \mathrm{E}[|\mathrm{Frag}_n|]$ exists and is a finite quantity.*

*Proof.* This is done in [20, Theorem 5d] for the uniform model and in greater detail in [27, Lemma 2.11] for the binomial model. For future reference we sketch the main ingredients in the proof.

Let $Y_i(n)$ be the random variable equal to the number of unicyclic components in $\mathcal{G}_n$ that contain exactly $i$ edges. Then one proves that for $k$ large enough and $n \geq 0$

$$\mathrm{E}[Y_k(n)] \leq (ce^{1-c})^k e^{c/2}.$$

Furthermore, for all $k \geq 3$

$$\lim_{n\to\infty} \mathrm{E}\left[Y_k(n)\right] = C(k,k)(ce^{-c})^k,$$

where $C(k,k)$ denotes the number of labeled unicyclic graphs on $k$ vertices. In particular the sequence of maps $(k \mapsto \mathrm{E}\left[Y_k(n)\right])_{n\in\mathbb{N}}$ is tight, so the statement follows. □

Because of Lemma 3.3, when $0 < c < 1$ a.a.s. all cycles in $\mathcal{G}(n, c/n)$ are contained in unicyclic components. Since the expected number $f(c)$ of cycles in $\mathcal{G}(n, c/n)$ is asymptotically bounded we obtain the following.

**Corollary 3.2.** *Let $p(n) \sim c/n$ with $0 < c < 1$, and let $Z(n)$ be the random variable equal to the number of cycles in $\mathcal{G}(n,p)$ that belong to connected components that are not trees or unicycles. Then*

$$\lim_{n\to\infty} \mathrm{E}[Z(n)] = 0.$$

**Lemma 3.5.** *Let $p(n) \sim c/n$ with $c > 0$. Let $T$ be a finite set of unlabeled unicycles. For each $H \in T$ let $X_H(n)$ be the random variable equal to the number of connected components in $\mathcal{G}(n,p)$ isomorphic to $H$, and let $\lambda_H = \frac{(e^{-c}c)^{|H|}}{\mathrm{aut}(H)}$. Then*

$$\lim_{n\to\infty} \Pr\left(\bigwedge_{H\in T} X_H(n) = a_H\right) = \prod_{H\in T} e^{-\lambda_H} \frac{\lambda_H^{a_H}}{a_H!}.$$

*In other words, the $X_H(n)$ converge in distribution to independent Poisson variables with respective means $\lambda_H$.*

*Proof.* The proof is a slight modification of Theorem 4.8 in [10]. It follows from a straightforward application of Theorem 1.4. □

### Asymptotic distribution of the fragment for $c < 1$ and its consequences

We compute below that the asymptotic probability that the fragment $\mathrm{Frag}_n$ is isomorphic to a given union $H$ of unicycles. Recal that $\mathbb{F}$ is the class of unlabeled fragments.

**Lemma 3.6.** *Let $p(n) \sim c/n$ with $0 < c < 1$, and let $H \in \mathbb{F}$. Then*

$$\lim_{n\to\infty} \Pr\big(\mathrm{Frag}_n \simeq H\big) = e^{-f(c)} \frac{(e^{-c}c)^{|H|}}{\mathrm{aut}(H)}. \tag{3.3}$$

*Proof.* Fix such an $H$. Let $U_1, U_2, \ldots, U_i, \ldots$ be an enumeration of all unlabeled unicycles ordered by non-decreasing size. For each $i$ let $a_i$ be the number of connected components of $H$ that are copies of $U_i$, and let $W_i(n)$ be the random variable equal to the number of connected components in $\mathcal{G}_n$ that are isomorphic to $U_i$. Clearly $\mathrm{Frag}_n \simeq H$ if and only if $W_i(n) = a_i$ for all $i$. Thus,

$$\lim_{n\to\infty} \Pr\big(\mathrm{Frag}_n \simeq H\big) = \lim_{n\to\infty} \Pr\big(\bigwedge_{i=1}^{\infty} W_i(n) = a_i\big).$$

The first observation is that

$$\prod_{k=i}^{\infty} e^{-\lambda_i} \frac{\lambda_i^{a_i}}{a_i!} = e^{-f(c)} \frac{(e^{-c}c)^{|H|}}{\mathrm{aut}(H)}, \tag{3.4}$$

where $\lambda_i = \frac{(ce^{-c})^{|U_i|}}{\mathrm{aut}(U_i)}$ is the asymptotic expected number of $H_i$ copies in $\mathcal{G}(n, c/n)$ (i.e., $\lim_{n\to\infty} \mathrm{E}[W_i(n)]$). We prove this in the following. Let $X_k(n)$ count the number of unicyclic components in $\mathcal{G}(n, c/n)$ containing exactly $k$ edges. Observe that $\sum_{k\geq 1} \mathrm{E}[X_k(n)] = \sum_{i\geq 1} \mathrm{E}[W_i(n)]$ by definition of the variables $X_k(n), W_i(n)$. Also, each variable $X_k(n)$ is the sum of only finitely many variables $W_i(n)$, so

$$\sum_{k\geq 1} \lim_{n\to\infty} \mathrm{E}[X_k(n)] = \sum_{i\geq 1} \lambda_i.$$

As seen in Lemma 3.4, the sequence $(i \mapsto \mathrm{E}[X_i(n)])_{n\geq 1}$ of real maps over $\mathbb{N}$ is tight. By Lemma 1.4 we can swap limit and sum in last equation to obtain

$$\lim_{n\to\infty} \sum_{k\geq 1} \mathrm{E}[X_k(n)] = \sum_{i\geq 1} \lambda_i.$$

Observe that $\sum_{k\geq 1} X_k(n)$ counts the number of cycles in $\mathcal{G}(n, c/n)$ that belong to unicyclic components. By Corollary 3.2, the expected number of cycles in $\mathcal{G}(n, c/n)$ outside of unicyclic components is $o(1)$ when $0 \leq c < 1$. Hence, it follows that $\sum_{i\geq 1} \lambda_i = f(c)$. As a consequence,

$$\prod_{i=1}^{\infty} e^{-\lambda_i} = e^{-f(c)} = e^{-f(c)}.$$

Since $\sum_{i=1}^{\infty} |U_i| a_i = |H|$ and $\prod_{i=1}^{\infty} \mathrm{aut}(U_i)^{a_i} a_i! = \mathrm{aut}(H)$, we finally get

$$\prod_{i=1}^{\infty} \frac{\lambda_i^{a_i}}{a_i!} = \frac{(ce^{-c})^{|H|}}{\mathrm{aut}(H)},$$

and Equation (3.4) follows.

We proceed now with the second part of the proof. By last Lemma, for any $j > 1$

$$\lim_{n\to\infty} \Pr\big(\bigwedge_{i=1}^{j} W_i(n) = a_i\big) = \prod_{i=1}^{j} e^{-\lambda_i}.$$

Hence, using Equation (3.4), the theorem amounts to the fact that the following exchange of limits can be performed

$$\lim_{n\to\infty}\lim_{j\to\infty} \Pr\big(\bigwedge_{i=1}^{j} W_i(n) = a_i\big) = \lim_{j\to\infty}\lim_{n\to\infty} \Pr\big(\bigwedge_{i=1}^{j} W_i(n) = a_i\big).$$

This follows from the intersection bound, and the fact that $\Pr(\bigwedge_{i=j}^{\infty} W_i(n) = 0)$ tends to one with $j$, uniformly in $n$. To see this, observe that for each $\epsilon > 0$ there is some $j$ such that $\sum_{i\geq j} \mathrm{E}\left[W_i(n)\right] < \epsilon$ for all $n$, by the definition of tight sequence. Thus, by Markov's inequality, $\Pr(\bigwedge_{i=j}^{\infty} W_i(n) = 0) \geq 1 - \epsilon$.

□

Given $H \in \mathbb{F}$ we define $p_H = p_H(c) = \lim_{n\to\infty} \Pr\left(\mathrm{Frag}_n \simeq H\right)$. The following is a direct consequence of the fact that the expected size of $\mathrm{Frag}_n$ is bounded.

**Lemma 3.7.** *Let $p(n) \sim c/n$ with $0 < c < 1$, and let $\mathcal{T} \subset \mathbb{F}$. Then*

$$\lim_{n\to\infty} \Pr\left(\bigvee_{H\in\mathcal{T}} \mathrm{Frag}_n \simeq H\right) = \sum_{H\in\mathcal{T}} p_H.$$

*In particular,* $\sum_{H\in\mathcal{T}} p_H = 1$.

*Proof.* If $\mathcal{T}$ is finite then the statement is clearly true, since the events $\mathrm{Frag}_n \simeq H$ are disjoint for different $H$. Suppose otherwise. Let $H_1, \ldots, H_i, \ldots$ be an enumeration of $\mathcal{T}$ by non-decreasing size. Fix $\epsilon > 0$. Let $m = \lim_{n\to\infty} \mathrm{E}\big[|\mathrm{Frag}_n|\big]$, and let $M = m/\epsilon$. Then there exists $j_0$ such that $E(H_j) \geq M$ for all $j \geq j_0$. Using Markov's inequality we obtain that for any $j \geq j_0$

$$\begin{aligned}
&\lim_{n\to\infty} \left| \Pr\left(\bigvee_{H\in T} \mathrm{Frag}_n \simeq H\right) - \sum_{i=1}^{j} p_{H_i} \right| \\
&= \lim_{n\to\infty} \left| \Pr\left(\bigvee_{H\in T} \mathrm{Frag}_n \simeq H\right) - \Pr\left(\sum_{i=1}^{j} \mathrm{Frag}_n \simeq H\right) \right| \\
&\leq \lim_{n\to\infty} \Pr\left(|\mathrm{Frag}_n| > M\right) \leq \epsilon.
\end{aligned}$$

As our choice of $\epsilon$ was arbitrary this proves the statement. □

### 3.1.2 No gap when $c \geq 1$

**Lemma 3.8.** *Let $c \geq 1$. Then $\overline{L_c} = [0, 1]$.*

*Proof.* As in the previous section, let $X_k(n)$ be the number of cycles of length $k$ in $\mathcal{G}(n, c/n)$, which is asymptotically $\mathrm{Po}(c^k/(2k))$. Moreover, for fixed $k$, the random variables $X_3(n), \ldots, X_k(n)$ are asymptotically independent by Lemma 3.2. Hence for fixed $k$,

$$X_{\leq k}(n) = X_3(n) + \cdots + X_k(n) \xrightarrow{\mathrm{d}} \mathrm{Po}\left(\sum_{i=3}^{k} \frac{c^k}{2k}\right).$$

Since $c \geq 1$ the mean $\sum_{i=3}^{k} c^k/2k$ is not bounded as $k$ grows to infinity so we can pick $k$ such that this mean is as large as we like. Note that for any $k$ and $a$ the property that $X_{\leq k} \leq a$ can be expressed in FO logic. By the central limit theorem we have

$$\Pr(\mathrm{Po}(\mu) \leq \mu + x\sqrt{\mu}) \xrightarrow[\mu \to \infty]{} \Phi(x)$$

for any fixed $x \in \mathbb{R}$, where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$ is the cumulative distribution function of the standard normal law.

For $0 < p < 1$ and $\epsilon > 0$ we can find $x$ such that $\Phi(x) = p$, a value $\mu_0$ such that $\Pr(\mathrm{Po}(\mu) \leq \mu + x\sqrt{\mu}) \in (p - \epsilon, p + \epsilon)$ for all $\mu \geq \mu_0$, and then finally a $k$ such that $\sum_{i=3}^{k} \frac{c^k}{2k} \geq \mu_0$. Hence there exists a FO property $\phi$ with limiting probability within $\epsilon$ of $p$. □

### 3.1.3 Always a finite union of intervals

It is shown in [49] that whether $\mathcal{G}(n, c/n)$ satisfies $\phi$ or not depends only a.a.s. on the induced unicycles of diameter at most $3^k$, where $k = \mathrm{qr}(\phi)$ (see Theorems 4.7, 4.8 and 4.9 in [49]). This, together with the fact that for $c < 1$ a.a.s. the connected components of $\mathcal{G}_n$ are either trees or unicycles (Theorem 3.3), implies the following:

**Lemma 3.9.** *Let $p(n) \sim c/n$ with $0 < c < 1$. Let $\phi$ be a FO sentence and let $H \in \mathbb{F}$. Then*

$$\lim_{n\to\infty} \Pr\left(\mathcal{G}(n,p) \models \phi \,\middle|\, \mathrm{Frag}_n \simeq H\right) = 0 \text{ or } 1.$$

*Moreover, the value of the limit depends only on $\phi$ and $H$, and not on $c$.*

Given $0 < c < 1$ and $H \in \mathbb{F}$, we define $p_H(c) = \lim_{n\to\infty} \Pr(\mathrm{Frag}_n \simeq H)$, where $\mathrm{Frag}_n$ stands for the fragment of $\mathcal{G}(n, c/n)$, as before. We define $S_c$ as the set of partial sums of $\sum_{H\in\mathbb{F}} p_H(c)$,

$$S_c := \Big\{ \sum_{H\in\mathcal{T}} p_H(c) \colon \mathcal{T} \subseteq \mathbb{F} \Big\}.$$

As outlined at the beginning of the section, when $0 < c < 1$, we study the set $\overline{L_c}$ through Kakeya's criterion (Lemma 3.1), using the fact that $\overline{L_c}$ coincides with $S_c$. This is proven below.

**Lemma 3.10.** *Let $0 < c < 1$. Then $\overline{L_c} = S_c$.*

*Proof.* We prove the result by showing both $\overline{L_c} \subseteq S_c$ and $\overline{L_c} \supseteq S_c$.

**(I)** $\overline{L_c} \subseteq S_c$: It is a known fact [33, 55] that $S_c$ is closed and has no isolated points. Thus, $\overline{S_c} = S_c$, and it is sufficient to show $L_c \subseteq S_c$. Let $\phi \in \mathrm{FO}_g$ be a sentence. For each $H \in \mathbb{F}$

define $p_{\phi,H}(n) = \Pr(\mathcal{G}(n,\boldsymbol{d}) \models \phi \mid \mathrm{Frag}_n \simeq H)p_H(n)$. By the law of total probability it holds $p_\phi = \lim_{n\to\infty} \sum_{H\in\mathbb{F}} p_{\phi,H}(n)$. As $p_{\phi,H}(n) \leq p_H(n)$, the sequence of real maps over $\mathbb{F}$, $(H \mapsto p_{\phi,H}(n))_{n\in\mathbb{N}}$ is tight. For this reason, sum and limit can be exchanged in the r.h.s. of last equation. Moreover, by lemma 3.9, we know that $\lim_{n\to\infty} \Pr(\mathcal{G}(n,\boldsymbol{d}) \models \phi \mid \mathrm{Frag}_n \simeq H) = 0$ or 1. Let $\mathbb{F}_\phi \subseteq \mathbb{F}$ be the set of fragments for which this limit equals 1. Notice that $\lim_{n\to\infty} p_{\phi,H}(n)$ equals $p_H$ when $H \in \mathbb{F}_\phi$ and equals zero otherwise. With this, we obtain $p_\phi = \sum_{H\in\mathcal{U}} p_H$, showing that $p_\phi \in S_c$. This proves the containment $\overline{L_c} \subseteq S_c$.

**(II)** $\overline{L_c} \supseteq S_c$. Let $\mathcal{U} \subseteq \mathbb{F}$ be an arbitrary family of fragments. We give a sequence of $\mathrm{FO}_g$ sentences $\phi_{\mathcal{U},k}$ satisfying $\lim_{k\to\infty} p_{\phi_{\mathcal{U},k}} = \sum_{H\in\mathcal{U}} p_H$. For each $H \in \mathbb{F}, k \in \mathbb{N}$ let $\phi_{H,k} \in \mathrm{FO}_g$ be a sentence stating that the graph $G$ contains an isolated copy of $H$, and that no $k$-tuple of vertices outside this copy induce a cycle. Suppose that $\mathcal{U}$ is infinite. Let $(\mathcal{U}_i)_{i\in\mathbb{N}}$ be a monotonically increasing chain of finite sets $\mathcal{U}_i \subset \mathbb{F}$ satisfying $\bigcup_{i\in\mathbb{N}} \mathcal{U}_i = \mathcal{U}$. Define $\phi_{\mathcal{U},k} = \bigvee_{H\in\mathcal{U}_k} \phi_{H,k}$. The union of disjoint events $\left(\bigvee_{H\in\mathcal{U}_k} \mathrm{Frag}_n \simeq H\right)$ implies $\mathcal{G}(n,\boldsymbol{d}) \models \phi_{\mathcal{U},k}$. Let $A_k(n)$ be the event that $\mathcal{G}(n,\boldsymbol{d})$ contains some cycle of size greater than $k$ Then, $(\mathcal{G}(n,\boldsymbol{d}) \models \phi_{\mathcal{U},k}) \wedge \neg A_k(n)$ implies $\left(\bigvee_{H\in\mathcal{U}_k} \mathrm{Frag}_n \simeq H\right)$ as well. Thus,

$$|p_{\phi_{\mathcal{U},k}}(n) - \sum_{H\in\mathcal{U}_k} p_H(n)| \leq \Pr(A_k(n)).$$

By Corollary 3.4, $\lim_{k\to\infty} \Pr(A_k(n)) = 0$ uniformly for all $n$. As a consequence,

$$\lim_{k\to\infty} \left( p_{\phi_{\mathcal{U},k}} - \sum_{H\in\mathcal{U}_k} p_H \right) = 0.$$

By the definition of infinite sum this proves $\lim_{k\to\infty} p_{\phi_{\mathcal{U},k}} = \sum_{H\in\mathcal{U}} p_H$, as we wanted to show. The case were $\mathcal{U}$ is finite follows similarly, by defining $\phi_{\mathcal{U},k} = \bigvee_{H\in\mathcal{U}} \phi_{H,k}$. This proves the containment $\overline{L_c} \supseteq S_c$. □

**Lemma 3.11.** *Suppose* $0 < c < 1$. *Then* $\overline{L_c}$ *is a finite union of intervals.*

*Proof.* Let $H_1, \ldots, H_n, \ldots$ be an enumeration of $\mathbb{F}$ such that $p_{H_i}(c) \leq p_{H_j}(c)$ for all $i \leq j$. We shorten $p_{H_i}(c)$ to $p_i$. Because of Lemma 3.1 proving that $\overline{L_c}$ is a finite union of intervals amounts to showing that for all $i$ large enough

$$p_i \leq \sum_{j>i}^{\infty} p_j. \tag{3.5}$$

Let $f = f(c)$ be as defined in Equation (3.2), and let $s = ce^{-c}$, and notice that as $c < 1$ we have $s < 1$ as well. We can rewrite the $p_i$ given by Equation (3.3) as

$$p_i = e^{-f} \frac{s^{|H_i|}}{aut(H_i)}.$$

For $i \geq 1$ let $k(i)$ be the least integer such that

$$e^{-f} s^{k(i)-1} \geq p_i > e^{-f} s^{k(i)}. \tag{3.6}$$

Notice if $k \geq k(i)$ and $H_j \in \mathbb{F}_k$ then $p_j < e^{-f}s^k < p_i$ because $\mathrm{aut}(H_j) \geq 1$. For the same reason we also obtain that $|H_i| \leq k(i) - 1$. Hence to prove (3.5) it is sufficient to show that

$$p_i \leq \sum_{k \geq k(i)} \sum_{H_j \in \mathbb{F}_k} p_j. \tag{3.7}$$

Let $C_{x,y}$ denote the graph in $\mathbb{F}$ consisting of a cycle of length $x$ with a path of length $y$ attached to one of its vertices. If $y = 0$ then $\mathrm{aut}(C_{x,y}) = 2x$, and $\mathrm{aut}(C_{x,y}) = 2$ otherwise. Let $T_{x,y,z}$ be the graph consisting of a triangle with paths of length $x$, $y$, and $z$ attached to its three vertices. Note that $\mathrm{aut}(T_{x,y,z}) = 1$ if $x, y, z$ are distinct, $\mathrm{aut}(T_{x,y,z}) = 6$ if $x = y = z$, and $\mathrm{aut}(T_{x,y,z}) = 2$ otherwise. It is easy to see that $C_{3,k-3}, C_{4,k-4}, \ldots, C_{k-1,1}$ together with $T_{0,1,k-4}, T_{0,2,k-5}, \ldots, T_{0,\lfloor (k-3)/2 \rfloor, \lceil (k-3)/2 \rceil}$ form a family of different elements of $\mathbb{F}_k$. We have that for $k \geq 3$

$$\sum_{i=3}^{k-1} p_{C_{i,k-i}} = e^{-f}s^k \frac{k-3}{2}.$$

If $k$ is odd $T_{0,\lfloor (k-3)/2 \rfloor, \lceil (k-3)/2 \rceil}$ has two automorphisms, and the remaining $T_{i,k-3-i}$ with $i \geq 1$ each have only one automorphism. If $k$ is even then all of $T_{0,1,k-4}, T_{0,2,k-5}, \ldots, T_{0,\lfloor (k-3)/2 \rfloor, \lceil (k-3)/2 \rceil}$ have exactly one automorphism. This gives

$$\sum_{i=1}^{\lfloor (k-3)/2 \rfloor} p_{T_{0,i,k-3-i}} = e^{-f}s^k \frac{k-4}{2}, \qquad \text{for } k \geq 4 \ .$$

Using the last two equations it follows that for $k \geq 4$

$$\sum_{H \in \mathbb{F}_k} \frac{1}{\mathrm{aut}(H)} \geq e^{-f}s^k \frac{2k-7}{2}. \tag{3.8}$$

Hence if $i$ is such that $(2k(i) - 7)/2 > 1/s$ (that is, $k(i) > 1/s + 7/2$) then

$$\sum_{j>i} p_j \geq \sum_{H_j \in \mathbb{F}_{k(i)}} p_j \geq e^{-f}s^{k(i)} \frac{2k-7}{2} > e^{-f}s^{k(i)-1} \geq p_i.$$

Note that $k(i) > 1/s + 7/2$ whenever $|H_i| + 1 \geq 1/s + 7/2$, and this is true for sufficiently large $i$. We have seen that, for any $0 < c < 1$, it is indeed the case that $p_i < \sum_{j>i} p_j$ for all sufficiently large $i$, as was to be proved. □

### 3.1.4 Transition at $c_0$

**Lemma 3.12.** *Let $c_0$ be as defined in Equation* (3.1)*. The following hold. (1) If $0 < c < c_0$, then $\overline{L_c}$ has at least one gap, and (2) if $c_0 \leq c < 1$, then $\overline{L_c} = [0, 1]$.*

*Proof.* We begin with (1). Let $H$ be the empty fragment. Observe that $p_H(c)$ simply equals the limit probability that $\mathcal{G}(n, c/n)$ is acyclic. By the definition of $c_0$, it holds that $p_H(c) > 1/2$ for all $0 < c < c_0$. In particular $p_H(c) > \sum_{H'H} p_{H'}(c)$, so by Kakeya's criterion $\overline{L_c}$ contains at least one gap.

Now we move on to (2). Fix $c \in [c_0, 1)$. As before, $H_1, H_2, \dots$ is an enumeration of $\mathbb{F}$ satisfying $p_{H_1}(c) \leq p_{H_2}(c) \leq \dots$, and we shorten $p_{H_i}(c)$ to $p_i$. By Lemma 3.10 and Kakeya's criterion, we just need to show that $p_i \leq \sum_{j>i} p_j$ for all $i$. Observe that $s = ce^{-c}$ satisfies

$$\frac{1}{3} < s < \frac{1}{e}.$$

Given $i \geq 1$, the value $k(i)$ is defined as in Equation (3.6). Let $i$ be such that $k(i) \geq 4$. Then, using (3.8) we obtain

$$\sum_{j>i} p_j \geq \sum_{k \geq k(i)} \sum_{H_j \in \mathbb{F}_k} p_j \geq \sum_{k \geq k(i)} e^{-f} s^k \frac{2k-7}{2}.$$

And using $\sum_{k=0}^{\infty} a^k(b+ck) = \frac{b}{1-a} + \frac{ca}{(1-a)^2}$ together with $s > 1/3$ we obtain that

$$\sum_{j>i} p_j \geq e^{-f} s^{k(i)} \left( \frac{2k(i)-7}{2(1-s)} + \frac{s}{(1-s)^2} \right) \geq e^{-f} s^{k(i)} \frac{3k(i)-9}{2}.$$

In particular, since $\frac{3k-9}{2} \geq 3 > 1/s$ for all $k \geq 5$, if $p_i \leq s^4$ then $p_i < \sum_{j>i} p_j$. As a consequence, if $|H_i| \geq 4$ then $p_i < \sum_{j>i} p_j$.

The only two cases left to consider are the ones when $H_i$ is either the empty graph or the triangle. If $H_i$ is the empty graph then necessarily $i = 1$ because the empty graph is the most likely fragment. By the definition of $p_0$ critically we have $p_1 \leq 1/2$ if $c \geq c_0$, hence $p_1 \leq \sum_{j>1} p_j$. If $H_i$ is the triangle graph, then $p_i = e^{-f} s^3/6$ and

$$\sum_{j>i} p_j = \sum_{k \geq 4} \sum_{H_j \in \mathbb{F}_k} p_j \geq \sum_{k \geq 4} e^{-f} s^k \frac{2k-7}{2} \geq e^{-f} s^4 \frac{3}{2} \geq e^{-f} s^3 \frac{1}{6} = p_i,$$

as needed. Thus $p_i \leq \sum_{j>i} p_j$ for every $i$, as we needed to prove. □

## 3.2 Binomial d-Uniform Hypergraphs

In this section we extend our previous result about $\mathcal{G}(n,p)$ to random sparse hypergraphs. We consider the model $\mathcal{G}^d(n,p)$ of random $d$-uniform hypergraphs, where every $d$-edge has probability $p$ of being in $\mathcal{G}^d(n,p)$ independently. When $p = c/n^{d-1}$ the expected number of edges $p\binom{n}{d}$ is linear in $n$, justifying the qualifier 'sparse'. A phase transition where a giant component emerges also occurs in $\mathcal{G}^d(n, c/n^{d-1})$ when $c = (d-2)!$ [59]. Throughout this section we consider $d \geq 3$ as being fixed and we will refer to "$d$-uniform hypergraphs" simply as hypergraphs. The FO language of $d$-uniform hypergraphs, denoted $\mathrm{FO}_g^d$, is analogous to the FO language of graphs, but the adjacency relation is $d$-ary instead of binary, as well as anti-reflexive and completely symmetric. The following is an analog of Lynch's convergence law for random hypergraphs and can be found in [58, Proposition 6.4], or can be derived from the more general convergence law shown in Chapter 2 and [40].

**Theorem.** *Let $p(n) \sim c/n^{d-1}$. Then for each $\mathrm{FO}_g^d$ sentence $\phi$, the following limit exists:*

$$p_c(\phi) = \lim_{n\to\infty} \Pr\left(\mathcal{G}^d(n, c/n^{d-1}) \models \phi\right).$$

*Moreover, $p_c(\phi)$ is a combination of sums, products, exponentials and a set of constants $\Lambda_c$, hence it is an analytic function of $c$.*

As before we consider the set

$$L_c = \left\{ \lim_{n\to\infty} \Pr(\mathcal{G}^d(n, c/n^{d-1} \models \phi) \mid \phi \text{ FO sentence} \right\}.$$

The main result of this section reads as follows.

**Theorem 3.2.** *Let $d \geq 3$ be fixed and let $\overline{L_c}$ be the closure of the of limiting probabilities of first order sentences in $G^d(n, c/n^{d-1})$. Let $c_0$ be the unique positive solution of*

$$\exp\left(\frac{c}{2(d-2)!}\right)\sqrt{1 - \frac{c}{2(d-2)!}} = \frac{1}{2}. \tag{3.9}$$

*Then for every $c > 0$ the set $\overline{L_c}$ is a finite union of intervals. Moreover, the following holds:*

1. *$\overline{L_c} = [0, 1]$ for $c \geq c_0$.*
2. *$\overline{L_c}$ has at least one gap for $0 < c < c_0$.*

We remark that $c_0 = r(d-2)!$, where $r \approx 0.898$ is the positive solution of $\exp(r/2)\sqrt{(}1-r) = 1/2$. The difference between Equations Equation (3.1) and Equation (3.9) results from the fact that cycles have length at least 3 in the setting of graphs, whereas in the case of hypergraphs 2-cycles do exist.

The strategy for proving the main theorem completely mirrors that of the previous section. Theorem 3.2 follows from three intermediate results. First, we show that $\overline{L_c} = [0, 1]$ when $c \geq (d-2)!$ (Lemma 3.19). Then, we show that $\overline{L_c}$ consists of a finite union of intervals for all $0 < c < (d-2)!$ (Lemma 3.22). Finally, we prove that $\overline{L_c}$ contains at least one gap $0 < c < c_0$, and $\overline{L_c} = [0, 1]$ for all $c_0 \leq c < (d-2)!$ (Lemma 3.23). As in the section before, we establish that $\overline{L_c}$ equals the set of partial sums of fragment probabilities for $0 < c < (d-2)!$, and use Kakeya's criterion to study this set.

### 3.2.1 Cycles and Fragments

Recall the definitions introduced in Section 1.2 of cycle, tree, unicycle, fragment and so on in the context of $d$-uniform hypergraphs. For convenience we set $\mathcal{G}_n^d = \mathcal{G}_n^d(n, c/n^{d-1})$ when $c$ is understood from context or is not relevant.

It was proven in [59] that a phase transition in the structure of $\mathcal{G}_n^d$ occurs when $c = (d-2)!$, similar to the one for random graphs. In particular, we have the following results [59, Theorem 3.6].

**Lemma 3.13.** *Let $p(n) \sim c/n^{d-1}$ with $0 < c < (d-2)!$. Then a.a.s. all connected components of $\mathcal{G}_n^d$ are either trees or unicycles.*

The proofs of the next results are very similar to those for graphs presented in Section 3.1.1 and are omitted.

**Lemma 3.14.** *Let $p \sim c/n^{d-1}$ with $c > 0$. For each $k \geq 2$, let $X_k(n)$ be the random variable equal to the number of $k$-cycles in $\mathcal{G}_n^d$, and let $\lambda_k = \left(\frac{c}{(d-2)!}\right)^k$. Then for fixed $k \geq 2$*

*(1)* $\mathrm{E}\big[X_k(n)\big] \leq \lambda_k$,
*(2)* $\lim_{n\to\infty} \mathrm{E}\big[X_k(n)\big] = \lambda_k$,
*(3)* *$X_k(n)$ converges in distribution to a Poisson variable with mean $\lambda_k$ as $n \to \infty$.*

*Furthermore, for any fixed $k \geq 2$ the variables $X_2(n), \ldots, X_k(n)$ are asymptotically independent.*

**Corollary 3.3.** *Let $p \sim c/n^{d-1}$ with $c > 0$. Set*

$$f(c) = \sum_{k\geq 2} \left(\frac{c}{(d-2)!}\right)^k \frac{1}{2k} = \frac{1}{2} \ln \frac{1}{1 - \frac{c}{(d-2)!}} - \frac{c}{2(d-2)!}. \tag{3.10}$$

*Let $X_n$ be the random variable equal to the total number of cycles in $\mathcal{G}_n^d$. Then*

$$\lim_{n\to\infty} \mathrm{E}[X_n] = f(c),$$

*and*

$$\lim_{n\to\infty} \Pr\left(\mathcal{G}_n^d \textit{ contains no cycles}\right) = e^{-f(c)} = \exp\left(\frac{c}{2(d-2)!}\right)\sqrt{1 - \frac{c}{2(d-2)!}}.$$

**Lemma 3.15.** *Let $p \sim c/n^{d-1}$ with $0 < c < (d-2)!$. Let $Z_n$ be the random variable equal to the number of cycles in $\mathcal{G}_n^d$ that belong to connected components that are not unicycles. Then*

$$\lim_{n\to\infty} \mathrm{E}[Z_n] = 0.$$

Let $\mathcal{U}$ be the family of unlabeled $d$-hypergraphs whose connected components are unicyclic.

**Lemma 3.16.** *Let $p \sim c/n^{d-1}$ with $c > 0$. Let $\mathcal{T} \subset \mathcal{U}$ be a finite set of unicycles. For each $H \in \mathcal{T}$ let $X_{n,H}$ be the random variable that counts the connected components in $\mathcal{G}_n^d$ that are isomorphic to $H$, and set*

$$\lambda_H = \frac{\left(ce^{-c/(d-2)!}\right)^{|H|}}{aut(H)}.$$

*Then $X_{n,H}$ converges in distribution to a Poisson variable with mean $\lambda_H$ as $n \to \infty$ and the $X_{n,H}$ are asymptotically independent, that is*

$$\lim_{n\to\infty} \Pr\left(\bigwedge_{H\in\mathcal{T}} X_{n,H} = a_H\right) = \prod_{H\in T} e^{-\lambda_H} \frac{\lambda_H^{a_H}}{a_H!}.$$

Similarly to the previous section, we write $\mathrm{Frag}_n$ for the fragment of the random hypergraph $\mathcal{G}_n^d$.

**Lemma 3.17.** *Let $p \sim c/n^{d-1}$ with $0 < c < (d-2)!$. Then the limit*

$$\lim_{n\to\infty} \mathrm{E}\big[|\mathrm{Frag}_n|\big]$$

*exists and is a finite quantity.*

The same proof of Lemma 3.6 can be used to prove the following result. Remember that $\mathbb{F}^d$ stands for the class of all unlabeled fragments.

**Theorem 3.3.** *Let $p(n) \sim c/n^{d-1}$ with $0 < c < (d-2)!$. Let $H \in \mathbb{F}^d$. Then*

$$\lim_{n\to\infty} \Pr\big(\mathrm{Frag}_n \simeq H\big) = e^{-f(c)} \frac{\big(e^{-c/(d-2)!}c\big)^{|H|}}{aut(H)}.$$

### A lower bound on the number of automorphisms of unicyclic hypergraphs

Let $H$ be an hypergraph and $h \in E(H)$ an edge. We call a vertex $v$ lying in $e$ *free* if $e$ is the only edge that contains $v$. We denote by free$(h)$ the number of free vertices in $e$. Notice that

$$\mathrm{aut}(H) \geq \prod_{h\in E(H)} \mathrm{free}(h)!,$$

because free vertices inside an edge can be permuted without restriction. Given a unicycle $H$ we define the *leaves* of $H$ as the edges $e \in E(H)$ that contain only one non-free vertex.

**Lemma 3.18.** *Let $H$ be a fragment. Then,*

$$\frac{(d-2)!^{|H|}}{\mathrm{aut}(H)} \leq \frac{(d-2)^2}{(d-1)^2}.$$

*Proof.* It suffices to prove the statement for unicycles, because

$$\frac{(d-2)!^{|H|}}{\mathrm{aut}(H)} \leq \prod_i \frac{(d-2)!^{|H_i|}}{\mathrm{aut}(H_i)},$$

where the $H_i$ are the connected components of $H$.

Let $\lambda$ be the number of leaves in $H$. We show by induction that

$$\prod_{h\in E(H)} \frac{(d-2)!}{\mathrm{free}(h)!} \leq \left(\frac{d-2}{d-1}\right)^{\lambda}. \tag{3.11}$$

If $\lambda = 0$ then $H$ is a cycle and each of its edges contains exactly $d-2$ free vertices, so that

$$\prod_{h\in E(H)} \frac{(d-2)!}{\mathrm{free}(h)!} = 1,$$

and $H$ satisfies (3.11). Now let $H$ be a unicycle satisfying (3.11). Add a new edge $h'$ to $H$ to obtain another unicycle $H'$. Since $h'$ intersects $H$ in only one vertex $v$, it follows that $h'$ is a leaf of $H'$. There are two possibilities:

- $\lambda(H') = \lambda(H)$. In this case no new leaves are created with the addition of $h'$. This means that $v$ is a free vertex in one leaf $g$ of $H$ (that is, $h'$ "grows" out of $g$), and

$$\prod_{h\in E(H')} \frac{(d-2)!}{\text{free}(h)!} = \prod_{h\in E(H)} \frac{(d-2)!}{\text{free}(h)!}.$$

- $\lambda(H') = \lambda(H) + 1$. In this case $h'$ intersects an edge of $H$ that is not a leaf. The case that maximizes $\prod_{h\in E(H')} \frac{(d-2)!}{\text{free}(h)!}$ is when $h'$ grows out of a free vertex of an edge in $H$ with exactly $d-2$ free vertices. In this case

$$\prod_{h\in E(H')} \frac{(d-2)!}{\text{free}(h)!} = \frac{d-2}{d-1} \prod_{h\in E(H)} \frac{(d-2)!}{\text{free}(h)!},$$

  and $H'$ satisfies (3.11) as well.

Finally, as all unicycles can be obtained adding edges to a cycle successively, (3.11) holds for all unicycles.

To prove the original statement consider the cases $\lambda = 0$, $\lambda = 1$ and $\lambda \geq 2$.

- If $\lambda = 0$ then $H$ is a cycle of length $l \geq 2$ and $\text{aut}(H) = (d-2)!^l 2l$, yielding

$$\frac{(d-2)!^{|H|}}{\text{aut}(H)} = \frac{1}{2l} \leq \frac{(d-2)^2}{(d-1)^2},$$

  since $1/2l \leq 1/4 \leq (d-2)^2/(d-1)^2$ for all $l \geq 2, d \geq 3$.
- If $\lambda = 1$ then $H$ is a cycle with a path attached to it. In this case, $H$ has a non-trivial automorphism (a reflection of the cycle) and as a consequence $2\prod_{h\in E(H)} \text{free}(h)! \leq \text{aut}(H)$. Using this and (3.11) we get

$$\frac{(d-2)!^{|H|}}{\text{aut}(H)} \leq \frac{1}{2} \prod_{h\in E(H)} \frac{(d-2)!}{\text{free}(h)!} \leq \frac{1}{2}\left(\frac{d-2}{d-1}\right) \leq \left(\frac{d-2}{d-1}\right)^2,$$

  as we wanted.
- Finally, when $\lambda \geq 2$ the relation (3.11) suffices, since

$$\prod_{h\in E(H)} \frac{(d-2)!}{\text{free}(h)!} \leq \left(\frac{d-2}{d-1}\right)^\lambda \leq \left(\frac{d-2}{d-1}\right)^2.$$

□

### 3.2.2 No gap when $c \geq (d-2)!$

**Lemma 3.19.** *Suppose that $c \geq (d-2)!$. Then $\overline{L_c} = [0, 1]$.*

*Proof.* The arguments here mirror exactly those in Section 3.1.2. For each $k$ let $X_k(n)$ be the random variable equal to the number of $k$-cycles in $\mathcal{G}_n^d$. Then

$$X_{\leq k}(n) = X_2(n) + \cdots + X_k(n) \xrightarrow[n\to\infty]{\mathrm{d}} \mathrm{Po}\left(\sum_{i=2}^{k} \frac{(c/(d-2)!)^k}{2k}\right).$$

If $c \geq (d-2)!$ then $\sum_{i=2}^{k} \frac{(c/(d-2)!)^k}{2k}$ tends to infinity and we can use the Central Limit Theorem to approximate any $p \in (0,1)$ with FO statements of the form "$X_{\leq k}(n) \leq a$". □

### 3.2.3 Always a finite union of intervals

As with the case of graphs, the following is an implicit consequence of the FO-convergence law on $\mathcal{G}^d(n, \boldsymbol{d})$.

**Lemma 3.20.** *Let $p \sim c/n^{d-1}$ with $0 < c < (d-2)!$. Let $\phi$ be a FO sentence and let $H \in \mathbb{F}^d$. Then*

$$\lim_{n\to\infty} \Pr\left(\mathcal{G}_n^d \models \phi \,\middle|\, \mathrm{Frag}_n \simeq H\right) = 0 \text{ or } 1.$$

*Moreover, the value of the limit depends only on $\phi$ and $H$, and not on $c$.*

For each $H \in \mathbb{F}^d$ define $p_H(c) = p_H = \lim_{n\to\infty} \Pr\left(\mathrm{Frag}_n \simeq H\right)$. Consider the set

$$S_c = \left\{ \sum_{H \in \mathcal{T}} p_H(c) \,:\, \mathcal{T} \subseteq \mathbb{F}^d \right\}.$$

One can proceed exactly as in Lemma 3.10 to prove the following:

**Lemma 3.21.** *Let $0 < c < (d-2)!$. Then $\overline{L_c} = S_c$.*

Before moving on to the main lemma of this subsection, we need to introduce three families of hypergraphs having a small number of automorphisms. Those will play a similar role in our proofs to the one played by the special unicyclic graphs in Section 3.1.3.

- Let $T_{\alpha,\beta}$ denote the hypergraph consisting of a triangle (as a $d$-hypergraph) with two paths of length $\alpha$ and $\beta$ respectively attached to two of its free vertices, each one from a different edge. One can check that

$$\frac{(d-2)!^{|T_{\alpha,\beta}|}}{\mathrm{aut}(T_{\alpha,\beta})} = \frac{(d-2)!^{\alpha+\beta+3}}{\mathrm{aut}(T_{\alpha,\beta})} = \begin{cases} \left(\frac{d-2}{d-1}\right)^2 & \text{for } \alpha \neq \beta, \\ \frac{1}{2}\left(\frac{d-2}{d-1}\right)^2 & \text{otherwise.} \end{cases}$$

  Let $\mathcal{T}$ be the family of hypergraphs $\{T_{\alpha,\beta} \colon \alpha, \beta > 0\}$. Then for $k \geq 4$

$$\sum_{H \in \mathcal{T},\, |H|=k} \frac{(d-2)!^{|H|}}{\mathrm{aut}(H)} = \sum_{\alpha=1}^{\lfloor \frac{k-3}{2} \rfloor} \frac{(d-2)!^k}{\mathrm{aut}(T_{\alpha,k-3-\alpha})} = \frac{k-4}{2}\left(\frac{d-2}{d-1}\right)^2. \tag{3.12}$$

- Let $B_{\alpha,\beta}$ denote the hypergraph consisting of a two-cycle with two paths of length $\alpha$ and $\beta$ respectively attached to two of its free vertices, each one from a different edge. In this case

$$\frac{(d-2)!^{|B_{\alpha,\beta}|}}{\mathrm{aut}(B_{\alpha,\beta})} = \frac{(d-2)!^{\alpha+\beta+2}}{\mathrm{aut}(B_{\alpha,\beta})} = \begin{cases} \frac{1}{2}\left(\frac{d-2}{d-1}\right)^2 & \text{for } \alpha \neq \beta, \\ \frac{1}{4}\left(\frac{d-2}{d-1}\right)^2 & \text{otherwise.} \end{cases}$$

Let $\mathcal{B} = \{B_{\alpha,\beta} \colon \alpha, \beta > 0\}$. Then for $k \geq 3$

$$\sum_{H\in\mathcal{B},|H|=k} \frac{(d-2)!^{|H|}}{aut(H)} = \sum_{\alpha=1}^{\lfloor\frac{k-2}{2}\rfloor} \frac{(d-2)!^k}{\mathrm{aut}(B_{\alpha,k-2-\alpha})} = \frac{k-3}{4}\left(\frac{d-2}{d-1}\right)^2. \tag{3.13}$$

- We denote by $O_{\alpha,\beta}$, the hypergraph formed by attaching a path of length $\beta$ to a free vertex of a cycle of length $\alpha$. One can check that $|O_{\alpha,\beta}| = \alpha+\beta$ and that $\frac{(d-2)!^{\alpha+\beta}}{\mathrm{aut}(O_{\alpha,\beta})|} = \frac{1}{2}\left(\frac{d-2}{d-1}\right)$.
Let $\mathcal{O} = \{O_{\alpha,\beta} \colon \alpha > 1, \beta > 0\}$. Then for $k \geq 2$

$$\sum_{H\in\mathcal{O},|H|=k} \frac{(d-2)!^{|H|}}{\mathrm{aut}(H)} = \sum_{\alpha=2}^{k-1} \frac{(d-2)!^k}{\mathrm{aut}(O_{\alpha,k-\alpha})} = \frac{k-2}{2}\left(\frac{d-2}{d-1}\right). \tag{3.14}$$

Now we are in conditions of proving the main lemma of the subsection.

**Lemma 3.22.** *Suppose $0 < c < (d-2)$. Then $\overline{L_c}$ has a finite number of gaps.*

*Proof.* Let $H_1, \ldots, H_n, \ldots$ be an enumeration of $\mathbb{F}^d$ such that $p_{H_i} \leq p_{H_j}$ for all $i \leq j$. As before we shorten $p_{H_i}$ to $p_i$. Analogously to 3.1.3 we need to prove that for $i$ large enough

$$p_i \leq \sum_{j>i} p_j.$$

Let $f = f(c)$ be as defined in Equation (3.10), and let $s = \frac{c}{(d-2)!}e^{-c/(d-2)!}$. Because of Theorem 3.3 we have that

$$p_i = e^{-f} s^{|H_i|} \frac{(d-2)!^{|H_i|}}{aut(H_i)}. \tag{3.15}$$

For $i > 0$ we define $k(i)$ as the unique integer such that

$$e^{-f} s^{k(i)-1}\left(\frac{d-2}{d-1}\right)^2 \geq p_i > e^{-f} s^{k(i)}\left(\frac{d-2}{d-1}\right)^2$$

Notice that because of Lemma, (3.18), we have $|H_i| \leq k(i) - 1$.

As a consequence, if $k = k(i) \geq 4$ then

$$\begin{aligned} \sum_{j>i} p_i &\geq s^k \sum_{H\in U_k} \frac{(d-2)!^k}{\mathrm{aut}(H)} \\ &\geq s^k \frac{k-4}{2}\left(\frac{d-2}{d-1}\right)^2. \end{aligned}$$

This is obtained taking into account only the hypergraphs in $\mathcal{T}$ and using Equation (3.12). The last inequality implies that if $k(i)$ is such that $\frac{1}{s} \leq \frac{k(i)-4}{2}$ then $p_i \leq \sum_{j>i} p_j$. This clearly holds for $i$ large enough, hence $\overline{L_c}$ is a finite union of intervals, as needed to be proved. □

### 3.2.4 Transition at $c_0$

**Lemma 3.23.** *Let $c_0$ be as defined in Equation (3.9). The following hold: (1) If $0 < c < c_0$, then $\overline{L_c}$ has at least one gap, and (2) if $c_0 \leq c < (d-2)!$, then $\overline{L_c} = [0, 1]$.*

*Proof.* Statement (1) is proven exactly as in Lemma 3.12. We show (2) below.

Fix $c_0 \leq c < (d-2)!$. Let $H_1, \dots, H_n, \dots$ be an enumeration of $\mathbb{F}^d$ satisfying the same conditions as before, and let $p_i = p_{H_i}(c)$. Our goal is showing that for all $i$

$$p_i \leq \sum_{j>i} p_j. \tag{3.16}$$

If this holds, statement (2) follows from Kakeya's Criterion and Lemma 3.21.

Notice that $s = \frac{c}{(d-2)!} e^{-c/(d-2)!}$ satisfies that

$$\frac{1}{3} < s < \frac{1}{e},$$

because $0.898 \leq c/(d-2)! < 1$. The following inequalities are obtained using Equations (3.12) to (3.14) respectively, together with the formula for the sum of an arithmetic-geometric series and the fact that $1/3 < s$.

$$\sum_{H \in \mathcal{T}, |H| \geq k} p_H \geq e^{-f} s^k \frac{6k-21}{8} \left(\frac{d-1}{d-2}\right)^2 \quad \text{for } k \geq 4. \tag{3.17}$$

$$\sum_{H \in \mathcal{B}, |H| \geq k} p_H \geq e^{-f} s^k \frac{6k-15}{16} \left(\frac{d-1}{d-2}\right)^2 \quad \text{for } k \geq 3. \tag{3.18}$$

$$\sum_{H \in \mathcal{O}, |H| \geq k} p_H \geq e^{-f} s^k \frac{6k-9}{8} \left(\frac{d-1}{d-2}\right) \quad \text{for } k \geq 2. \tag{3.19}$$

Assume first that $k = k(i) \geq 5$. Then

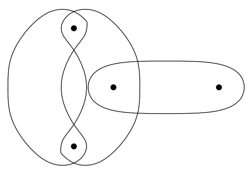
$$\begin{aligned}
\sum_{j>i} p_j &\geq e^{-f} s^k \left[\frac{18k-57}{16} \left(\frac{d-2}{d-1}\right)^2 + \frac{6k-9}{8} \left(\frac{d-2}{d-1}\right)\right] \\
&\geq e^{-f} s^k \frac{30k-75}{16} \left(\frac{d-2}{d-1}\right)^2 \geq e^{-f} s^k 3 \left(\frac{d-2}{d-1}\right)^2 \\
&\geq e^{-f} s^{k-1} \left(\frac{d-2}{d-1}\right)^2 \geq p_i,
\end{aligned}$$

as was to be proven.

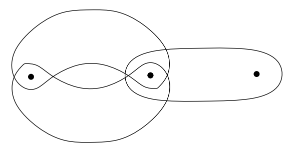
Otherwise, suppose that $k = k(i) \leq 4$. Notice that because of Lemma 3.18 necessarily $|H_i| \leq 3$. We have three cases:

- $|H_i| = 3$. In this case, the following enumeration of all (unlabeled) unicycles of size 3 gives that

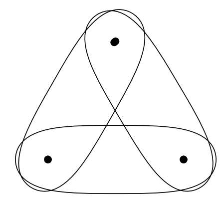
$$e^{-f} s^3 \frac{1}{2}\left(\frac{d-1}{d-2}\right) \geq p_i.$$



(a) $\frac{(d-2)!^3}{\mathrm{aut}(H)} = \frac{1}{2}\left(\frac{d-1}{d-2}\right)$.



(b) $\frac{(d-2)!^3}{\mathrm{aut}(H)} = \frac{1}{2}\left(\frac{1}{d-2}\right)$.



(c) $\frac{(d-2)!^3}{\mathrm{aut}(H)} = \frac{1}{6}$.

Proceeding as before we obtain

$$\begin{aligned}\sum_{j>i} p_j &\geq e^{-f} s^4 \left[\frac{184-57}{16}\left(\frac{d-2}{d-1}\right)^2 + \frac{64-9}{8}\left(\frac{d-2}{d-1}\right)\right] \\ &\geq e^{-f} s^4 \left[\frac{15}{16}\frac{1}{2}\left(\frac{d-2}{d-1}\right) + \frac{30}{8}\frac{1}{2}\left(\frac{d-2}{d-1}\right)\right] \\ &\geq e^{-f} s^4 \frac{3}{2}\left(\frac{d-2}{d-1}\right) \geq e^{-f} s^3 \frac{1}{2}\left(\frac{d-2}{d-1}\right) \geq p_i.\end{aligned}$$

- $|H_i| = 2$. In this case $H_i$ is the 2-cycle, and $p_i = e^{-f} s^2 \frac{1}{4}$. Using Equations (3.18) and (3.19) we obtain

$$\begin{aligned}\sum_{j>i} p_j &\geq p_{C_3} + \sum_{H \in \mathcal{B}} p_H + \sum_{H \in \mathcal{O}} p_H \\ &\geq e^{-f} s^3 \left[\frac{1}{6} + \frac{3}{16}\left(\frac{d-2}{d-1}\right)^2 + \frac{9}{8}\left(\frac{d-2}{d-1}\right)\right] \\ &\geq e^{-f} s^3 \left[\frac{4}{6}\frac{1}{4} + \frac{3}{16}\frac{1}{4} + \frac{18}{8}\frac{1}{4}\right] \geq e^{-f} s^3 3\frac{1}{4} \geq p_i.\end{aligned}$$

- $|H_i| = 0$. In this case $H_i$ is the empty graph and $p_i \geq 1/2$ by hypothesis.

□

## 3.3 Graphs with Given Degree Sequences

In this section we study the set of limit probabilities corresponding to $\mathrm{FO}_g$ sentences in the context of sparse graphs with given degree sequences. This random model is significantly more complex than the binomial random graph $\mathcal{G}(n, p)$, but we are able to prove an analogous result to Theorem 3.1. Roughly, the main theorem of this section states that the set of limit probabilities of interest is dense in $[0, 1]$ when the likelihood that the random graph is acyclic is at most $1/2$. Otherwise, this set is dense in a finite union of intervals, and has at least one gap.

Before giving a precise statement of this section's main theorem we need to properly introduce random graphs with given degree sequences. A **degree sequence on n vertices** is a sequence $\boldsymbol{d} = (d_i)_{i\in[n]}$ where $d_i$ is a non-negative integer with $d_i < n$ for all $i \in [n]$ and $\sum_{i\in[n]} d_i$ is even. We call $\boldsymbol{d}$ **feasible**, if there is some graph $G$ with $V(G) = [n]$ whose degree sequence is $\boldsymbol{d}$, meaning that $\deg v = d_v$ for all $v \in [n]$. Given a feasible degree sequence $\boldsymbol{d}$ on $n$ vertices, $\mathcal{G}(n, \boldsymbol{d})$ denotes the uniform graph with vertex set $[n]$ and whose degree sequence is $\boldsymbol{d}$.

We are interested in the asymptotic study of random graphs. For this reason, rather than working with a single degree sequence, we need to consider a family of them, containing one for each number of vertices. An **asymptotic degree sequence** (shortened to a.d.s.) is a family $\boldsymbol{d} = (\boldsymbol{d}(n))_{n\in\mathbb{N}}$ where $\boldsymbol{d}(n) = (d_i(n))_{i\in[n]}$ is a degree sequence on $n$ vertices. By convention, we set $d_i(n) = 0$ for each $i > n$. Given $n$, we define $n_k(n) = \{i \in [n] \mid d_i(n) = k\}$. We drop the argument $n$ when the dependency is clear from context in order to keep the notation light.

Our objects of study are sparse random graphs, meaning that we want to choose $\boldsymbol{d}$ so that the expected number of edges in $\mathcal{G}(n, \boldsymbol{d})$ is linear. Additionally, in order to study the $\mathcal{G}(n, \boldsymbol{d})$ from the perspective of FO logic, we need to impose some regularity conditions on $\boldsymbol{d}$. Those conditions are better stated in terms of the degree distributions corresponding to $\boldsymbol{d}$. Given $n \in \mathbb{N}$, the **degree distribution** $D_{\boldsymbol{d}}(n)$ is given by $\Pr(D_{\boldsymbol{d}}(n) = k) = n_i/n$. Equivalently, $D_{\boldsymbol{d}}(n)$ is the probability distribution of the degree of a uniform random vertex in $\mathcal{G}(n, \boldsymbol{d})$. We simply write $D(n)$ instead of $D_{\boldsymbol{d}}(n)$ when $\boldsymbol{d}$ is clear from the context.

**Definition 3.1.** An a.d.s. $\boldsymbol{d} = (\boldsymbol{d}(n))_{n\in\mathbb{N}}$ is called **well behaved** (w.b.) if the following are satisfied:

**WB1:** $\boldsymbol{d}(n)$ is feasible for all $n$.
**WB2:** $D(n) \xrightarrow{d} D$ for some random variable $D = D_{\boldsymbol{d}}$.
**WB3:** $\lim_{n\to\infty} \mathrm{E}[D(n)]$ (resp., $\lim_{n\to\infty} \mathrm{E}\left[D(n)^2\right]$), exists, is bounded and equal to $\mathrm{E}[D]$ (resp., $\mathrm{E}\left[D^2\right]$).
**WB4:** Whenever $\Pr(D = k) = 0$ for some $k$, $\Pr(D(n) = k) = 0$ for all $n$ (or equivalently, $n_k = 0$ for all $n$).

Condition **WB1** is required so that $\mathcal{G}(n, \boldsymbol{d})$ is well-defined. Conditions **WB2** and **WB3** allow us to study $\mathcal{G}(n, \boldsymbol{d})$ by looking at the limit degree distribution $D$. More concretely, **WB3** guarantees that the proportion of edges incident to vertices with very large degrees is small. In particular, $\mathrm{E}\left[D^2\right]$ being finite and equal to $\lim_{n\to\infty} \mathrm{E}\left[D^2(n)\right]$ implies that the maximum degree $\Delta(n)$ in $\boldsymbol{d}(n)$ satisfies $\Delta(n) = o(\sqrt{n^2})$. Finally, **WB4** rules out the existence of vertices with "rare degrees". Otherwise such rare-degree vertices would pose an obstacle to a $\mathrm{FO}_g$-convergence in $\mathcal{G}(n, \boldsymbol{d})$. For example, consider a situation where $\boldsymbol{d}(n)$ contains a single degree 3-vertex for odd $n$, and none for even $n$. We remark that for our purposes, condition **WB4** could be weakened replacing "for all $n$" by "for all sufficiently large $n$". However, we use the stronger version for convenience.

Before moving on, we introduce some additional notation. All of the following definitions are in terms of some well-behaved a.d.s. $\boldsymbol{d}$. We define $\lambda_i = \Pr(D = i)$, where $D$ is the limiting degree distribution corresponding to $\boldsymbol{d}$, as defined at the beginning of the section. Given $n \in \mathbb{N}$, we write $m_n$ to the twice the number of edges given by $\boldsymbol{d}(n)$. That is, $m_n = \sum_{i\in[n]} d_i(n)$. Lastly, $\rho_k(n)$ stands for the $k$-th factorial moment $\mathrm{E}\left[(D(n))_k\right]$, and for $k = 1, 2$ we put $\rho_k$ for $\mathrm{E}\left[(D)_k\right]$, which coincides with $\lim_{n\to\infty} \rho_k(n)$ by the definition of well-behavedness. The parameter $\nu = \nu_{\boldsymbol{d}}$ is defined as $\mathrm{E}\left[D(D-1)\right]/\mathrm{E}\left[D\right]$. It is well-known that $\nu$ is responsible for the apparition of a giant component in $\mathcal{G}(n, \boldsymbol{d})$. Under similar conditions to ours, [53] shows that $\mathcal{G}(n, \boldsymbol{d})$ a.a.s. contains a component of linear size if and only if $\nu > 1$ Finally, for all $n \in \mathbb{N}$, $\nu(n)$ refers to $\rho_2(n)/\rho_1(n)$. Observe that by the definition of well-behavedness $\nu(n)$ converges to $\nu$ as $n$ grows to infinity.

As with the previous section during this chapter, our starting point here is the fact that a $\mathrm{FO}_g$-convergence law holds in $\mathcal{G}(n, \boldsymbol{d})$.

**Theorem 3.4** (Lynch, [45, 46])**.** *Let $\boldsymbol{d}$ be a well-behaved a.d.s. Then a $\mathrm{FO}_g$-convergence law holds in $\mathcal{G}(n, \boldsymbol{d})$.*

If one inspects the sources [45, 46], they will find that in each of the papers different conditions are imposed on $\boldsymbol{d}$, and that those conditions are not comparable (that is, neither weaker nor stronger) to our definition of well-behavedness (Definition 3.1). In fact, due to some minor oversights the conditions from neither work [45, 46] are sufficient to deduce a convergence law in $\mathcal{G}(n, \boldsymbol{d})$ using their tools. The reason is that those works obtain results on a related multigraph model, called the configuration model, and rely on arguments lemmas that link event probabilities in the configuration model to event probabilities in $\mathcal{G}(n, \boldsymbol{d})$. However, those arguments require the second moment of $D(n)$ (the degree distribution given by $\boldsymbol{d}(n)$) to converge to a finite quantity, and this is not guaranteed in the papers. We remark this does not mean that the convergence law does not hold under the conditions of [45, 46], but simply that their particular proof methods require stronger assumptions. Further discussion on this topic can be found in Section 3.3.5. However, when $\boldsymbol{d}$ is well-behaved according to our definition $\mathrm{E}\left[D(n)^2\right]$ has a finite limit and the techniques introduced in [45, 46] yield a $\mathrm{FO}_g$-convergence law.

Throughout this section, we deal with a well-behaved a.d.s. $\boldsymbol{d}$ and we set

$$p_\varphi(\boldsymbol{d}) = \lim_{n\to\infty} \Pr(\mathcal{G}(n, \boldsymbol{d}) \models \varphi),$$

where $\varphi$ is a $\mathrm{FO}_g$-sentence. Similarly to the sections before, we study the set

$$L_{\boldsymbol{d}} = \{p_\varphi(\boldsymbol{d}) \mid \varphi \text{ sentence in } \mathrm{FO}_g\}.$$

Surprisingly, it turns out that $\nu$ is the only parameter determining whether $L_{\boldsymbol{d}}$ is dense in the whole interval $[0, 1]$. Our main result this section is the following.

**Theorem 3.5.** *Let $\boldsymbol{d}$ be a well-behaved a.d.s. and let $\nu_0$ be the only solution in $[0, 1]$ of*

$$e^{\frac{\nu}{2}+\frac{\nu^2}{4}}\sqrt{1-\nu} = 1/2. \tag{3.20}$$

*It holds that $\overline{L_{\boldsymbol{d}}}$ is a finite union of closed intervals. Moreover,*

1. *$\overline{L_{\boldsymbol{d}}} = [0, 1]$ if $\nu \geq \nu_0$.*

2. $\overline{L_{\boldsymbol{d}}}$ *has at least one gap if* $1 < \nu < \nu_0$.

The same strategy as in the previous sections works for proving this theorem, although the probabilistic arguments are more involved. Theorem 3.5 follows from three intermediate results: $\overline{L_{\boldsymbol{d}}} = [0,1]$ when $\nu \geq 1$ (Theorem 3.9), $\overline{L_{\boldsymbol{d}}}$ consists of a finite union of intervals when $0 < \nu < 1$ (Lemma 3.30), and $\overline{L_{\boldsymbol{d}}}$ contains at least one gap when $0 < \nu < \nu_0$, whereas $\overline{L_\nu} = [0,1]$ for all $\nu_0 \leq \nu < 1$ (Lemma 3.23). In the sub-critical region $0 < \nu < 1$ we also establish that $\overline{L_{\boldsymbol{d}}}$ equals the set of partial sums of fragment probabilities (Theorem 3.11), and use Kakeya's Criterion to deduce the lemmas of interest. However, this time fragment probabilities have a more complex structure that depends heavily on the a.d.s. $\boldsymbol{d}$. For example, $\boldsymbol{d}$ may not contain vertices of degree 3, limiting the possibilities for $\mathcal{G}(n,\boldsymbol{d})$'s fragment. We overcome this difficulty, roughly, by separating fragments into different blocks depending on the amount of cycles of each length they contain, and using those blocks for our bounds, instead of using individual fragments as in the previous sections.

### 3.3.1 Configuration Model, Cycles and Fragments

In this section we introduce the probabilistic results required for studying $L_{\boldsymbol{d}}$. Compared to previous sections, we do this in greater detail here, as the computations are more involved and, to the best of our knowledge, there is not a proper account in the literature for some of the results we cover.

#### Configuration Model

It turns out that $\mathcal{G}(n,\boldsymbol{d})$ is not an easy model to deal with. It is not clear how to produce a multigraph with the desired degree sequence uniformly at random. However, if we settle for multigraphs (that is, we allow double edges, and edges with both ends at the same vertex, called loops), the problem becomes easier. Given a degree sequence $\boldsymbol{d}$ on $n$ vertices, we attach $d_v$ half edges to each vertex $v \in [n]$. Afterwards, any matching of half edges produces a multigraph with the desired degree sequence. This is the so-called **configuration model** [9, 10], which we introduce in greater detail below. This model does not yield a uniform multigraph with degree sequence $\boldsymbol{d}$, but assigns the same probability to each graph (i.e., multigraph with no double edges nor loops). In other words, conditioning $\mathcal{CM}(n,\boldsymbol{d})$ to the event of being simple yields the same distribution as $\mathcal{G}(n,\boldsymbol{d})$. Hence, studying the configuration model is a good strategy for dealing with the later random graph.

Multigraphs are generalizations of graphs where multiple edges between the same vertices are allowed, as well as edges joining a vertex to itself. In addition to that, we consider multigraphs where each edge has its own identity. Formally, a **multigraph** $G$ is a triple $(V(G), E(G), r_G)$ where $V(G)$ is its **vertex set**, and $E(G)$ its **edge set**, and $r_G : E(G) \to \{\{u,v\} \mid v,u \in V(G)\}$ is a map assigning each edge $e$ to an unordered pair of vertices $\{u,v\}$, which represent its **endpoints**. We allow the possibility that $u = v$, in which case $\{u,v\}$ is a singleton rather than a pair, and $e$ is called a **loop**. Given $u,v \in V(G)$, the **multiplicity** of $\{u,v\}$ in $E(G)$ is just the size of $r_G^{-1}(\{u,v\})$. The **degree** $\deg(v)$ of a vertex $v \in V(G)$ is the number of edges $e \in E(G)$ with $r_G(e) = \{v,u\}$ for some $u \neq v$, plus twice the number of loops $e \in E(G)$ with $r_G(e) = \{v\}$. A graph $G = (V(G), E(G))$ has a natural representation as a multigraph, just by setting $r_G$ the identity map on $E(G)$. Given a multigraph $G$, a **sub-multigraph** $H$ is just a multigraph where $V(H) \subseteq V(G)$, $E(H) \subseteq E(V)$, and $r_H$ is $r_G$ restricted to $E(G)$. An **isomorphism** between two multigraphs $G, H$ is a bijection $f : V(G) \to V(H)$ where the multiplicity of $\{u, v\}$ in $G$ is the same as the multiplicity of $\{f(u), f(v)\}$ in $H$, for all $u, v \in V(G)$. Given a multigraph $G$, its number of **half-edge automorphisms** $\mathrm{aut}_{\mathrm{h.e.}}(G)$ is $\mathrm{aut}(G)2^{\ell} \prod_{u,v \in V(G)} m(u, v)!$, where $\ell$ is the number of loops in $G$, and given (non-necessarily different) vertices $u, v \in V(G)$, $m(\{u, v\})$ denotes the multiplicity of $\{u, v\}$ in $G$. Informally, $\mathrm{aut}_{\mathrm{h.e.}}(G)$ is the number of half-edge permutations in $G$ that preserve incidence to the same vertex and the matching between half-edges. As with the case of graphs, if $G, H$ are multigraphs, an $H$-copy in $G$ is just a sub-multigraph $H' \subseteq G$ that is isomorphic to $G$. Other notions related to graphs (Section 1.2) are extended to multigraphs in the intuitive way. Importantly, we define the **excess** of a multigraph $G$ to be the number $\mathrm{ex}(G) = |E(G)| - |V(G)|$. For convenience, throughout this section we attempt to treat multigraphs just as if they were normal graphs. Among other things, we omit the maps $r_G$ whenever possible.

The **random configuration** $\mathcal{CM}(n, \boldsymbol{d})$, is a uniform random matching of $[m]$ (formally, $\mathcal{CM}(n, \boldsymbol{d}) \subseteq \binom{[m]}{2})$), where $m = \sum_{i \in [n]} d_i$. We refer to the elements $e \in [m]$ as **half-edges**. We say that a half-edge $e \in [m]$ **belongs** to a vertex $v \in [n]$ if $\sum_{u<v} d_u < e \leq \sum_{u \leq v} d_u$. In other words, the first $d_0$ half-edges belong to vertex 0, the following $d_1$ belong to vertex 1, and so on. The **underlying multigraph** of $\mathcal{CM}(n, \boldsymbol{d})$ has vertex set $[n]$, its edge set consists of the pairs $\{h_1, h_2\}$ in the matching $\mathcal{CM}(n, \boldsymbol{d})$, and the endpoints of an edge $\{h_1, h_2\} \in \mathcal{CM}$ are defined as $\{v_1, v_2\}$ where the half-edges $h_1$ and $h_2$ belong to $v_1$ and $v_2$. In the following, we identify $\mathcal{CM}(n, \boldsymbol{d})$ with its underlying multigraph.

The model $\mathcal{CM}(n, \boldsymbol{d})$ assigns a probability to each multigraph $G$ whose degree sequence is $\boldsymbol{d}$ that depends only on its number of loops and its number of edges with multiplicity greater than one. In particular, when $\boldsymbol{d}$ is feasible, it holds that conditioning $\mathcal{CM}(n, \boldsymbol{d})$ on the absence of loops and multiple edges results in the same distribution as $\mathcal{G}(n, \boldsymbol{d})$ as $\mathcal{CM}(n, \boldsymbol{d})$ [10].

### Cycle Distributions

**Lemma 3.24.** *Let $H$ be a fixed multigraph, where $h = |V(H)|$, $h_i = |\{v \in V(H) \mid \deg(v) = i\}$, and $\ell = |E(H)|$. Let $X_H(n)$ be the number of $H$-copies in $\mathcal{CM}(n, \boldsymbol{d})$. Then $\Xi(H, n)(1 - O(1/n)) \leq \mathrm{E}[X_H(n)] \leq \Xi(H, n)$, where*

$$\Xi(H, n) = \frac{n^h}{\mathrm{aut}_{\mathrm{h.e.}}(H) \prod_{i=1}^{\ell} (m_n - 2i + 1)} \prod_{i \geq 0} \rho_i(n)^{h_i}.$$

*Proof.* Let $H'$ be a possible sub-configuration of $\mathcal{CM}(n, \boldsymbol{d})$ isomorphic to $H$. Then the probability that $H' \subseteq \mathcal{CM}(n, \boldsymbol{d})$ is exactly

$$\prod_{i=1}^{\ell} \frac{1}{(m_n - 2i + 1)},$$

which is obtained by dividing the number of configurations containing $H'$ by the total number of matchings of $[m_n]$. Both bounds in our statement result from estimating the number of possible sub-configurations $H'$ isomorphic to $H$. Fix a labelling $v_1, \dots, v_h$ of $V(H)$. In order to choose $H'$, we begin by picking the vertices $V(H')$, $v'_1, \dots, v'_h$ each one labeled after a vertex in $H$. In order to completely determine $H'$, we need to pick a list of $\deg(v_i)$ half-edges incident to $v'_i$ for each $1 \le i \le h$. This yields a total of $\prod_{1 \le i \le h} (a_i)_{b_i}$ choices of half-edges for $H'$, where $a_i = d_{v'_i}(n)$ and $b_i = \deg(v_i)$. Note that this is 0 unless $d_{v'_i}(n) \ge \deg(v_i)$ There are exactly $\mathrm{aut}_{\text{h.e.}}(H)$ ways of choosing vertices and half-edges that yield the same sub-configuration $H'$. Hence, the total number of possible sub-configurations of $\mathcal{CM}(n, \boldsymbol{d})$ isomorphic to $H$ is given by

$$\frac{1}{\mathrm{aut}_{\text{h.e.}}(H)} \sum_{a_1, \dots a_h \in \mathbb{N}} \sum_{\substack{\{v'_1, \dots, v'_h\} \in \binom{[n]}{h} \\ d_{v'_i}(n) = a_i, \text{ for all } 1 \le i \le h}} \prod_{1 \le i \le h} (a_i)_{\deg(v_i)}.$$

In this sum, we first pick the degrees $a_1, \dots, a_h$ of $v'_1, \dots, v'_h$ before choosing the vertices themselves. Once $a_1, \dots, a_h$ are fixed, for each $i$ there are at most $n_{a_i}$ choices for $v'_i$ (corresponding to the case where each $a_i$ is different) and at least $n_{a_i} - h + 1$ choices for this vertex (when all $a_i$ are the same). In this way,

$$\begin{aligned}
\mathrm{E}\left[X_H(n)\right] \le \; & \frac{1}{\mathrm{aut}_{\text{h.e.}}(H) \prod_{i=1}^{\ell} (m_n - 2i + 1)} \sum_{a_1, \dots, a_h \in \mathbb{N}} \left( \prod_{1 \le i \le h} (a_i)_{\deg(v_i)} \, n_{a_i} \right) \\
= \; & \frac{n^h}{\mathrm{aut}_{\text{h.e.}}(H) \prod_{i=1}^{\ell} (m_n - 2i + 1)} \prod_{1 \le i \le h} \left( \sum_{a \in \mathbb{N}} \frac{n_{a_i}}{n} (a_i)_{\deg(v_i)} \right) \\
= \; & \frac{n^h}{\mathrm{aut}_{\text{h.e.}}(H) \prod_{i=1}^{\ell} (m_n - 2i + 1)} \prod_{1 \le i \le h} \rho_{\deg(v_i)}(n) \qquad = \; \Xi(H, n),
\end{aligned}$$

as we wanted. The lower bound comes from replacing $n_{a_i}$ with $(n_{a_i} - h)$ in the first line of equations.

□

We generalize the notion of **cycle** to multi-graphs. For $\ell \ge 3$, $k$-cycles are defined as usual. For $k = 2$, a 2-cycle consists of two vertices plus two edges joining them, and for $k = 1$, a 1-cycle is just a vertex with a loop attached to it. This way, $\mathrm{aut}_{\text{h.e.}}(C_k) = 2k$, where $C_k$ is a $k$-cycle and $k \ge 1$. The following results study the number of cycles in $\mathcal{CM}(n, \boldsymbol{d})$.

**Lemma 3.25.** *Let $X_k(n)$ be the random variable counting $k$-cycles in $\mathcal{CM}(n, \boldsymbol{d})$. Then, for any finite collection $k_1, \dots, k_l$, the variables $X_{k_1}(n), \dots, X_{k_l}(n)$ converge in distribution to independent Poisson variables whose respective means are $\xi_{k_i} = \nu^{k_i}/2k_i$. In particular, the probability of $\mathcal{CM}(n, \boldsymbol{d})$ being simple is $e^{-\frac{\nu}{2} - \frac{\nu^2}{4}}$.*

*Proof.* We prove the first part of the statement. The asymptotic probability that $\mathcal{CM}(n, \boldsymbol{d})$ is simple follows easily from there. We assume $\rho_2 > 0$. Otherwise all vertices have degree 0 or 1 and the result follows trivially. We use the method of moments. Let $a_1, \dots, a_l \in \mathbb{N}$ be arbitrary. We wish to prove

$$\lim_{n \to \infty} \mathrm{E}\left[ \prod_{i=1}^{l} \binom{X_{k_i}(n)}{a_i} \right] = \prod_{i=0}^{l} \frac{\xi_{k_i}^{a_i}}{a_i!}.$$

By Theorem 1.4 this implies the result. We say that a multigraph $G$ is a **non-degenerate** union of unlabeled multigraphs $H_1,\dots,H_t$ if $G$ contains a copy $H_i'$ of $H_i$ for each $i$, $V(G) = V(H_1') \cup \dots \cup V(H_t')$ (note that this union is not necessarily disjoint), and the $H_i$ are pairwise different. Let $\mathcal{H}$ be the class of all unlabeled multigraphs that are non-degenerate unions of $a_1$ copies of $C_{k_1}$, $a_2$ copies of $C_{k_2}$, ..., and $a_l$ copies of $C_{k_l}$. Note that, in $\mathcal{H}$, only the multigraph $H_*$ corresponding to the disjoint union of cycles has zero excess, and $\mathrm{ex}(H) > 0$ for all the others. Given $H \in \mathcal{H}$, let $Y_H(n)$ be the number of $H$ copies lying in $\mathcal{CM}(n, \boldsymbol{d})$. The l.h.s. in last equation amounts to $\sum_{H\in\mathcal{H}} \frac{\mathrm{aut}_{\mathrm{h.e.}}(H)}{\mathrm{aut}_{\mathrm{h.e.}}(H_*)} \mathrm{E}\left[Y_H(n)\right]$. We show that asymptotically only $\mathrm{E}\left[Y_{H_*}(n)\right]$ contributes to the value of this sum, and this expectation has the desired value. Using the upper bound in Lemma 3.24 we get

$$\mathrm{E}\left[Y_H(n)\right] \le (1+o(1)) \frac{n^{n(H)} \prod_{i\in\mathbb{N}} (\rho_i(n))^{n_i(H)}}{(m_n)^{m(H)}\ \mathrm{aut}_{\mathrm{h.e.}}(H)} = O\left(n^{-\mathrm{ex}(H)} \prod_{i\in\mathbb{N}} \rho_i(n)^{n_i(H)}\right),$$

where $n(H) = |V(H)|, m(H) = |E(H)|$, and $n_i(H) = |\{v \in V(H) \mid \deg(v) = i\}|$.

Let $H \in \mathcal{H}$ be an arbitrary multigraph different from $H_*$. We show that $\mathrm{E}\left[Y_H(n)\right] = o(1)$. As $\boldsymbol{d}$ is well behaved, $\Delta(n) = o(n^{1/2})$, and $\rho_{i+1}(n) = o(n^{1/2}\rho_i(n))$ for all $i \ge 1$. In particular, $\rho_i(n) = o(n^{(i-2)(1/2)})$ for all $i \ge 3$, due to $\rho_1(n), \rho_2(n) = O(1)$. As $H$ contains some vertex with degree at least 3, we obtain $\prod_{i\in\mathbb{N}} \rho_i(n)^{n_i(H)} = o\left(\prod_{i=3}^{\infty} n^{n_i(H)(i-2)(1/2)}\right)$. Observe that $\sum_{i\ge 0} n_i(H)(i-2) = 2\,\mathrm{ex}(H)$, from where we get $\sum_{i\ge 3} n_i(H)(i-2) \le 2\,\mathrm{ex}(H)$ using that the minimum degree of $H$ is at least 2. This way,

$$\mathrm{E}\left[Y_H(n)\right] = o\left(n^{-\mathrm{ex}(H)} \prod_{i=3}^{\infty} n^{n_i(H)(i-2)(1/2)}\right) = o(1).$$

Now consider the case $H = H_*$. All of $H_*$'s vertices have degree 2, and $\mathrm{aut}_{\mathrm{h.e.}}(H_*) = \prod_{i=1}^{l} \frac{1}{a_i!(2k_i)^{a_i}}$. Here Lemma 3.24 yields $\mathrm{E}\left[Y_{H_*,n}\right] = \prod_{i=1}^{l} \left(\xi_{k_i}^{a_i}/a_i!\right) + o(1)$, using that $n/m_n = 1/\rho_1(n)$. Putting everything together, we obtain

$$\lim_{n\to\infty} \mathrm{E}\left[\prod_{i=1}^{l} \binom{X_{k_i}(n)}{a_i}\right] = \lim_{n\to\infty} \mathrm{E}\left[Y_{H_*}(n)\right] = \prod_{i=0}^{l} \frac{\xi_{k_i}^{a_i}}{a_i!},$$

proving our first equation and the result. □

**Lemma 3.26.** *Let $X_k(n)$ count the $k$-cycles in $\mathcal{CM}(n, \boldsymbol{d})$. Assume that $\nu < 1$. Then the sequence $(k \mapsto \mathrm{E}\left[X_k(n)\right])_{n\in\mathbb{N}}$ is tight.*

*Proof.* Clearly, adding or removing isolated vertices to $\mathcal{CM}(n, \boldsymbol{d})$ does not affect the result, so without loss of generality we may assume $\lambda_0 = 0$. Additionally, we also assume $\lambda_1 < 1$. Otherwise all vertices have degree one, $\nu = 0$, and the result follows trivially. Let $\nu < \nu' < 1$. As $\nu(n)$ tends to $\nu$, there is some value $n'$ such that $m_n > n$ and $\nu(n) < \nu'$ for all $n$ greater than $n'$. Then, from the upper bound in Lemma 3.24 follows that

$$\mathrm{E}\left[X_k(n)\right] \leq \frac{(n)_k}{2k\prod_{\ell=1}^{k}(m_n-2\ell+1)}\rho_2(n)^k \leq \frac{\nu(n)^k}{2k} \leq \frac{(\nu')^k}{2k},$$

for all $n > n'$ and all $k$. To see the second inequality, note that $m_n > n$ implies $(n-s)/(m_n - 2s+1) < n/(m_n) = 1/\rho_1(n)$. As the sum $\sum_{\ell\geq 1}\frac{(\nu')^k}{2k}$ converges, last chain of inequalities proves the result. □

**Corollary 3.4.** *Let $Z(n)$ count the cycles in $\mathcal{CM}(n,\boldsymbol{d})$. Assume that $\nu < 1$. Then,* $\mathrm{E}\left[Z(n)\right] = -1/2\,\ln(1-\nu) + o(1)$.

*Proof.* Let $X_k(n)$ count the $k$-cycles in $\mathcal{CM}(n,\boldsymbol{d})$, as in the previous lemma. In Lemma 3.25 we showed that $\mathrm{E}\left[X_k(n)\right] = \nu^k/2k + o(1)$. Hence,

$$\sum_{k=1}^{\infty}\lim_{n\to\infty}\mathrm{E}\left[X_k(n)\right] = -1/2\,\ln(1-\nu). \tag{3.21}$$

The variable $Z(n)$ corresponds to the sum $\sum_{k=1}^{\infty} X_k(n)$, so the lemma states that the limit and sum symbols can be exchanged in this equation, which is true by virtue of last lemma.

□

**Lemma 3.27.** *Assume $\nu < 1$. Let $\boldsymbol{a} = (a_\ell)_{\ell\in\mathbb{N}}$ be a sequence of natural numbers $a_\ell \in \mathbb{N}$ whose sum $\sum_{\ell\in\mathbb{N}} a_\ell$ is finite. The following hold true:*

*(1) Let $A(n)$ be the event that $\mathcal{CM}(n,\boldsymbol{d})$ contains exactly $a_\ell$ $\ell$-cycles for all $\ell\in\mathbb{N}$. Then*

$$\Pr(A(n)) = \sqrt{1-\nu}\prod_{\ell\in\mathbb{N}}\frac{(\nu^\ell/2\ell)^{a_\ell}}{a_\ell!} + o(1).$$

*In particular, the probability that $\mathcal{CM}(n,\boldsymbol{d})$ is acyclic is $\sqrt{1-\nu}+o(1)$.*

*(2) Let $B(n)$ be the event that $\mathcal{G}(n,\boldsymbol{d})$ contains exactly $a_\ell$ $\ell$-cycles for all $\ell\geq 3$. Then*

$$\Pr(B(n)) = \sqrt{1-\nu}\,e^{-\frac{\nu}{2}-\frac{\nu^2}{4}}\prod_{\ell\geq 3}\frac{(\nu^\ell/2\ell)^{a_\ell}}{a_\ell!} + o(1).$$

*In particular, the probability that $\mathcal{G}(n,\boldsymbol{d})$ is acyclic is $\sqrt{1-\nu}\,e^{-\frac{\nu}{2}-\frac{\nu^2}{4}}+o(1)$.*

*Proof.* We prove (1). Statement (2) follows from the fact that $\mathcal{G}(n,\boldsymbol{d})$ is distributed like $\mathcal{CM}(n,\boldsymbol{d})$ conditioned on the absence of 1-cycles and 2-cycles. Let $X_k(n)$ count the $k$-cycles in $\mathcal{CM}(n,\boldsymbol{d})$. This way, $A(n) = \bigwedge_{k\geq 1} X_k(n) = a_k$, and $\Pr(A(n)) = \lim_{k\to\infty} p_k(n)$, where $p_k(n) = \Pr(\bigwedge_{i=1}^{k} X_i(n) = a_i)$. Let $\epsilon > 0$ be arbitrarily small. We prove that

$$\left|\Pr(A(n)) - \sqrt{1-\nu}\prod_{\ell\in\mathbb{N}}\frac{(\nu^\ell/2\ell)^{a_\ell}}{a_\ell!}\right| < \epsilon + o(1). \tag{3.22}$$

Let $K$ be a sufficiently large number satisfying both (I)

$$\left|\left(\prod_{k=1}^{K} e^{-\nu^k/2k}\frac{(\nu^k/2k)^{a_k}}{a_k!}\right) - \left(\sqrt{1-\nu}\prod_{k\in\mathbb{N}}\frac{(\nu^k/2k)^{a_k}}{a_k!}\right)\right| < \epsilon/2,$$

and (II)

$$\Pr(\sum_{i>K} X_i(n) = 0) > 1-\epsilon/2 \text{ for all } n.$$

Property (I) can be attained because inside the absolute value, for $\nu < 1$, the parentheses on the right contain the limit (as $K$ tends to infinity) of the expression in the parentheses on the left. The existence of $k$ satisfying (II) follows from $(k \mapsto \mathrm{E}[X_k(n)])_{n\in\mathbb{N}}$ being tight, as shown in Corollary 3.4, and Markov's inequality. Indeed, there by the tightness property there is some $K$ for which $\sum_{i>K}\mathrm{E}[X_i(n)] < \epsilon/2$ uniformly in $n$, and then $\Pr(\sum_{i>K} X_i(n) = 0) > 1-\epsilon/2$ uniformly in $n$ as well.

Property (II) and the intersection bound imply that $\Pr(A(n)) > p_K(n) - \epsilon/2$ for all $n$. In addition to that, the sequence $(p_k(n))_{k\geq 1}$ is monotonically decreasing, so $\Pr(A(n)) \leq p_K(n)$ for all $n$. This way, $\lim_{n\to\infty}|\Pr(A(n)) - p_K(n)| < \epsilon/2$. However, by Lemma 3.25

$$p_K(n) = \prod_{k=1}^{K} e^{\nu^k/2k}\frac{(\nu^k/2k)^{a_k}}{a_k!} + o(1).$$

Applying property (I) here yields Equation (3.22), and completes the proof. □

As mentioned at the beginning of the section, in [53] a phase transition for $\mathcal{G}(n,\boldsymbol{d})$ was shown under very similar conditions to ours. Roughly, they show that when $\nu < 1$, w.h.p. all components in $\mathcal{CM}(n,\boldsymbol{d})$ are of sublinear size, and none of them contain more than one cycle. They show this last statement examining the exposure of a connected component in $\mathcal{CM}(n,\boldsymbol{d})$ and employing Azuma's inequality. Their arguments make use of stronger assumptions than ours on the maximum degree of $\boldsymbol{d}(n)$. However, we can follow a more cumbersome First-Moment argument to prove the same result in our setting.

**Theorem 3.6.** *Assume $\nu < 1$. Then a.a.s $\mathcal{CM}(n,\boldsymbol{d})$ has no connected component containing more than one cycle.*

*Proof.* Without loss of generality we may assume that $\lambda_0 = 0$. Also, with no loss of generality as well, we can assume $\lambda_1 < 1$. Otherwise all vertices have degree 1 and the theorem holds trivially. Note that $\lambda_0 > 1$ and $\lambda_1 < 1$ together imply $\rho_1 > 1$.

The configuration $\mathcal{CM}(n,\boldsymbol{d})$ has two cycles lying in the same component iff it has some subgraph belonging to one of the following classes:

(I) $H^{(1)}_{i,j,k}$ : A $i$-cycle and an $j$-cycle together with a path of length $k \geq 2$ joining a vertex from each cycle.

(II) $H^{(2)}_{i,j,k}$ : A $i$-cycle and an $j$-cycle sharing a path of length $k \geq 2$.

(III) $H^{(3)}_{i,j}$ : A $i$-cycle and a $j$-cycle sharing a single vertex.

We show that a.a.s. none of these sub-graphs appear in $\mathcal{CM}(n,\boldsymbol{d})$ using the first-moment method. Let us consider (I) first. Let $X^{(1)}_{i,j,k}(n)$ count the copies of $H^{(1)}_{i,j,k}$ in $\mathcal{CM}(n,\boldsymbol{d})$. The multigraph $H^{(1)}_{i,j,k}$ has $i+j+k-1$ edges and $i+j+k-2$ vertices, among which two have degree 3 and the rest have degree 2. This way, by Lemma 3.24, for sufficiently large $n$

$$\begin{aligned} \mathrm{E}\left[X^{(1)}_{i,j,k}(n)\right] &\le \frac{(n)_\ell}{\prod_{s=1}^{\ell+1}(m_n-2s+1)}\rho_2(n)^{\ell-2}\rho_3(n)^2 \\ &\le \nu(n)^{\ell-2}\frac{\rho_3(n)^2}{(m_n-2\ell-1)} \\ &\le \nu(n)^{\ell-2}\frac{\rho_3(n)^2}{2(\rho_1(n)-1)n-1}, \end{aligned} \tag{3.23}$$

where $\ell=i+j+k-2$. Second inequality follows from $(n-s)/(m_n-2s+1)\le n/m_n$, because $m_n>n$ for sufficiently large $n$, as $\rho_1>1$, and $n/m_n=1/\rho_1(n)$. Last inequality follows from $\ell\le n$ and $(m_n-2n-1)=2(\rho_1(n)-1)n-1$. Let $X^{(1)}_\ell=X^{(1)}_\ell(n)$ be the sum of all variables $X^{(1)}_{i,j,k}(n)$ with $i+j+k-2=\ell$. There are at most $\ell^2$ such choices of $i,j,k$, so $\mathrm{E}\left[X^{(1)}_\ell\right]\le \ell^2\nu(n)^{\ell-2}\frac{\rho_3(n)^2}{2(\rho_1(n)-1)n-1}$. By hypothesis $\rho_1>1$, so

$$\sum_{\ell=3}^{\infty}\mathrm{E}\left[X^{(1)}_\ell\right]=O\left(\frac{\rho_3(n)^2}{n}\sum_{\ell=3}^{\infty}\nu^\ell\ell^2\right)=O\left(\frac{\rho_3(n)^2}{n}\right).$$

Notice that, $\rho_3(n)\le\rho_2(n)\Delta(n)=o(\sqrt{n})$, so $\rho_3(n)^2/n=o(1)$. This way, using Markov's inequality we obtain that a.a.s. no class-(I) sub-graph occurs in $\mathcal{CM}(n,\boldsymbol{d})$. In an analogous way, it can be shown that the expected numbers of class-(II) and class-(III) sub-graphs in $\mathcal{CM}(n,\boldsymbol{d})$ are $O\left(\frac{\rho_3(n)^2}{n}\right)$ and $O\left(\frac{\rho_4(n)}{n}\right)$ respectively. As we have seen $\frac{\rho_3(n)^2}{n}=o(1)$, so a.a.s. $\mathcal{CM}(n,\boldsymbol{d})$ has no class-(II) sub-graph. Similarly, $\rho_4(n)\le\Delta(n)^2\rho_2(n)=o(n)$, so a.a.s. $\mathcal{CM}(n,\boldsymbol{d})$ contains no class-(III) sub-graph either. This proves the result. □

### Fragment Distribution

During this section, a **fragment** is a multigraph whose components are all unicyclic, similarly to the case of graphs. We call a fragment **simple** if it contains no loops nor double-edges, or, equivalently, if it contains no cycles of length smaller than 3. The fragment Frag$(G)$ of a multigraph $G$ is the union of its unicyclic components. Let $\mathrm{Frag}^*{}_n=\mathrm{Frag}(\mathcal{CM}(n,\boldsymbol{d}))$ and $\mathrm{Frag}_n=\mathrm{Frag}(\mathcal{G}(n,\boldsymbol{d}))$.

**Theorem 3.7.** *Suppose $\nu<1$. If $H$ is a fragment, then*

$$\lim_{n\to\infty}\Pr(\mathrm{Frag}^*{}_n\simeq H)=\frac{\sqrt{1-\nu}}{\mathrm{aut}_{\text{h.e.}}(H)}\prod_{i\ge1}\left(\frac{\lambda_i i!}{\rho_1}\right)^{h_i},$$

*where $h_i=|\{v\in V(H)\mid \deg(v)=i\}|$.*

*Proof.* Let $h=|V(H)|$ and $h_i=|\{v\in V(H)\mid\deg(v)=i\}|$. Let $V(H)=\{v_1,\dots,v_h\}$. For each $v_i\in V(H)$ fix some ordering of the half-edges incident to $v_i$. Define $\mathcal{H}(n)$ as the set of possible isolated $H$-copies in $\mathcal{CM}(n,\boldsymbol{d})$. In order to pick a copy $H'\in\mathcal{H}(n)$, we first select the vertices

$v'_1, \ldots, v'_h$. As we want the copy to be isolated, we require $d_{v'_i}(n) = \deg(v_i)$ for all $1 \le i \le h$. In order to completely determine $H'$ we give an ordering of the half-edges incident to each vertex $v'_i$. Afterwards, half-edges should be matched according to the half-edge orderings defined for $H$. Observe that there are exactly $\mathrm{aut}_{\mathrm{h.e.}}(H)$ ways of picking vertices and half-edge orderings that result in the same sub-configuration $H'$. Hence,

$$|\mathcal{H}(n)| = \frac{\prod_{i\ge 0}(n_i)_{h_i}(i!)^{h_i}}{\mathrm{aut}_{\mathrm{h.e.}}(H)}.$$

Given $H' \in \mathcal{H}(n)$, let $A(H', n)$ be the event that $H' \subseteq \mathcal{CM}(n, \boldsymbol{d})$ and $\mathcal{CM}(n, \boldsymbol{d}) \setminus V(H')$ is acyclic. Observe that the events $A(H', n)$ are disjoint. Let $P(n)$ be the event that no component in $\mathcal{CM}(n, \boldsymbol{d})$ contains more than one cycle. Then the event $(\mathrm{Frag^*}_n \simeq H \mid P(n))$ coincides with the union of the events $A(H', n)$. Thus, by Theorem 3.6

$$\Pr(\mathrm{Frag^*}_n \simeq H) = \sum_{H' \in \mathcal{H}(n)} \Pr(A(H', n)) + o(1).$$

By symmetry, the probability $A(H', n)$ is the same for all $H' \in \mathcal{H}(n)$. Fix an $H$-copy $H'_n \in \mathcal{H}(n)$ for each $n$. Using the expression for $|\mathcal{H}(n)|$, we obtain

$$\Pr(\mathrm{Frag^*}_n \simeq H) = \frac{\prod_{i\ge 0}(n_i)_{h_i}(i!)^{h_i}}{\mathrm{aut}_{\mathrm{h.e.}}(H)} \Pr(A(H'_n, n)) + o(1).$$

Let us examine now the probability of $A(H'_n, n)$. Let $\widehat{G}_n$ be the random multigraph $\mathcal{CM}(n, \boldsymbol{d}) \setminus V(H'_n)$. By definition

$$\Pr(A(H'_n, n)) = \frac{1}{\prod_{i=1}^{h}(2m_n - 2i + 1)} \Pr\left(\widehat{G}_n \text{ is acyclic} \,\middle|\, H'_n \subseteq \mathcal{CM}(n, \boldsymbol{d})\right).$$

For each $n$, let $\widehat{\boldsymbol{d}}(n-h)$ be a degree sequence obtained by removing the vertices $V(H'_n)$ from $[n]$ and relabeling the remaining vertices as $[n-h]$. Note that $(\widehat{G}_n \mid H'_n \subseteq \mathcal{CM}(n, \boldsymbol{d})) \sim \mathcal{CM}(n-h, \widehat{\boldsymbol{d}})$. Clearly, the asymptotic degree sequence $\widehat{\boldsymbol{d}}$ is well-behaved. Additionally, it is easy to see that the first and second moments of the related degree distribution have the same limits as those of $\boldsymbol{d}$ (that is, $\rho_1$ and $\rho_2$ as $h = O(1)$. By Lemma 3.27,

$$\Pr\left(\widehat{G}_n \text{ is acyclic} \,\middle|\, H'_n \subseteq \mathcal{CM}(n, \boldsymbol{d})\right) = \sqrt{1-\nu} + o(1).$$

Putting everything together, we obtain

$$\begin{aligned}\Pr(\mathrm{Frag^*}_n \simeq H) &= \frac{\sqrt{1-\nu}\prod_{i\ge 0}(n_i)_{h_i}(i!)^{h_i}}{\mathrm{aut}_{\mathrm{h.e.}}(H)\prod_{i=1}^{h}(2m_n - 2i + 1)} + o(1) \\ &= \frac{\sqrt{1-\nu}}{\mathrm{aut}_{\mathrm{h.e.}}(H)} \prod_{i\ge 1}\left(\frac{\lambda_i i!}{\rho_1}\right)^{h_i} + o(1),\end{aligned}$$

as in the theorem's statement. This completes the proof. □

The following Corollary states that the fragment of $\mathcal{G}(n, \boldsymbol{d})$ is asymptotically distributed like the simple fragment of $\mathcal{CM}(n, \boldsymbol{d})$.

**Corollary 3.5.** *Assume that $\nu < 1$. Let $G$ be a simple fragment. Then*

$$\lim_{n\to\infty} \Pr(\mathrm{Frag}_n \simeq G) = \frac{\sqrt{1-\nu}\, e^{-\nu/2-\nu^2/4}}{\mathrm{aut}(G)} \prod_{i\geq 1} \left(\frac{\lambda_i i!}{\rho_1}\right)^{g_i},$$

*where $g_i = |\{v \in V(G) \mid \deg(v) = i\}|$.*

*Proof.* Let $A(n)$ be the event that $\mathcal{CM}(n, \boldsymbol{d})$ is simple (i.e., it contains no loops nor multiple edges). By definition, $\Pr(\mathrm{Frag}_n \simeq G) = \Pr({\mathrm{Frag}^*}_n \simeq G \mid A(n))$. When $\mathcal{CM}(n, \boldsymbol{d})$ has no complex components, the event $({\mathrm{Frag}^*}_n \simeq G) \wedge A(n)$ is equivalent to ${\mathrm{Frag}^*}_n \simeq G$. This way, using Theorem 3.6 we obtain

$$\Pr(\mathrm{Frag}_n \simeq G) = \Pr({\mathrm{Frag}^*}_n \simeq G) \Pr(A(n)) + o(1).$$

By Corollary 3.4 $\Pr(A(n)) = e^{-\nu/2-\nu^2/4} + o(1)$. This, together with the previous theorem and the fact that $\mathrm{aut}(G) = \mathrm{aut}_{\mathrm{h.e.}}(G)$ when $G$ is simple, proves the result. □

From now on let $p^*_H(n) = \Pr({\mathrm{Frag}^*}_n \simeq H)$, $p_G(n) = \Pr(\mathrm{Frag}_n \simeq G)$, $p^*_H = \lim_{n\to\infty} p^*_H(n)$, and $p_G = \lim_{n\to\infty} p_G(n)$, for all unlabeled fragments $H$, and unlabeled simple fragments $G$. Our next goal is to show that the numbers $p^*_H$ define a distribution over unlabeled fragments, meaning that $\sum_H p^*_H = 1$. For this, we define a random fragment $\Gamma$ satisfying $\Pr(\Gamma \simeq H) = p^*_H$.

**Definition 3.2** (Random Fragment)**.** Suppose $\nu < 1$. Define the distributions $\widehat{D}$ and $\widehat{D}'$ over $\mathbb{N}$ by $\Pr(\widehat{D} = i-1) = \frac{\lambda_i i}{\rho_1}$ and $\Pr(\widehat{D}' = i-2) = \frac{\lambda_i i(i-1)}{\rho_2}$, respectively. Let $(\mathcal{BP}_r)_{r\in\mathbb{N}}$ be mutually independent branching processes, where $\mathcal{BP}_r$ has $r$ roots, its root offspring distribution is $\widehat{D}'$ and its general offspring distribution is $\widehat{D}$. Let $(Z_i)_{i\geq 1}$ be mutually independent Poisson variables, which are mutually independent with $(\mathcal{BP}_r)_{r\in\mathbb{N}}$ as well, and where $\mathrm{E}[Z_i] = \frac{\nu^i}{2i}$. Let $R = \sum_i iZ_i$. The **random fragment** $\Gamma = \Gamma(\boldsymbol{d})$ associated to the a.d.s. $\boldsymbol{d}$, is obtained as follows: First we generate the random rooted forest $\mathcal{BP}_R$. Observe that the offspring distribution $\widehat{D}$ has mean $\nu < 1$, so $\mathcal{BP}_R$ yields a distribution over finite rooted forests by Lemma 1.3. Afterwards, we form various cycles using the roots $1, \ldots, R$ as follows: For every $k$, we form $Z_k$ $k$-cycles, one with each successive $k$-tuple of roots with labels in the interval $\left(\sum_{i=1}^{k-1} iZ_i,\ \sum_{i=1}^{k} iZ_i\right]$.

Showing that this random fragment $\Gamma$ satisfies the desired identities $\Pr(\Gamma \simeq H) = p^*_H$ is not completely straight-forward. In order to compute $\Pr(\Gamma \simeq H)$, we need to count how many concrete outcomes of $\Gamma$ yield a fragment isomorphic to $H$. The following definition has that purpose.

**Definition 3.3.** A **Ulam-Harris (UH) fragment** $G$ is an ordered labeled fragment, where the vertices in $V(G)$ are words $\omega \in \mathbb{N}^*$, equipped with the usual lexicographical order, and satisfying:

– The vertices belonging to cycles are labeled by singleton words (i.e., numbers) $1, \ldots, r$.
– Vertices in the same cycle are labeled by consecutive numbers, according to the cyclic order.

– Vertices belonging to $i$-cycles have correspond to smaller numbers than vertices belonging to $j$-cycles, for any $i < j$.
– The rooted forest resulting from marking vertices lying in cycles as roots and removing all edges forming cycles, is a UH Definition 1.3 forest.

This way, cycles in a Ulam-Harris fragment are ordered, from shortest to largest, vertices inside cycles are ordered in a cyclic manner, and vertices in attached trees are ordered in a breadth-first fashion.

Observe that each outcome of $\Gamma$ has a natural representation as a UH fragment: roots are ordered in the appropriate way, and the random forest hanging from the roots is an U.H. forest as in Definition 1.3. Moreover, isomorphic U.H. fragments have the same probability of being the outcome of $\Gamma$. Using this fact we are able to prove the following lemma.

**Lemma 3.28.** *Assume $\nu < 1$. Let $\Gamma$ be the random fragment given in Definition 3.2, and let $H$ be an arbitrary unlabeled fragment. Then* $\Pr(\Gamma \simeq H) = p^*_H$.

*Proof.* Fix $H$, and let $G$ be a UH fragment isomorphic to $H$. Let $h = n(G), h_i = n_i(G)$. We define $c_i$ as number of $i$-cycles in $G$, $h'$ as the number of vertices belonging to cycles in $G$, and $h'_i$ as the number of degree-$i$ vertices belonging to these cycles. The graph $\Gamma$ is completely determined by the random variables $(Z_i)_{i\in\mathbb{N}}$ and the random labeled forest $\mathcal{BP}_R$. Letting $R = \sqrt{1-\nu}$, we obtain

$$\begin{aligned}\Pr(\Gamma = G) &= \left(R\prod_{i\in\mathbb{N}}\frac{(\nu^i/2i)^{c_i}}{c_i!}\right)\left(\prod_{i\in\mathbb{N}}\Pr(\widehat{D}' = i-2)^{h'_i}\right)\left(\prod_{i\in\mathbb{N}}\Pr(\widehat{D} = i-1)^{h_i-h'_i}\right)\\ &= \left(R\frac{\nu^{h'}}{\prod_{i\in\mathbb{N}}(2i)^{c_i}c_i!}\right)\left(\frac{\prod_{i\in\mathbb{N}}(\lambda_i i(i-1))^{h'_i}}{\rho_2^{h'}}\right)\left(\frac{\prod_{i\in\mathbb{N}}(\lambda_i i)^{h_i-h'_i}}{\rho_1^{h-h'}}\right)\\ &= R\left(\frac{\prod_{i\in\mathbb{N}}(\lambda_i i(i-1))^{h'_i}(\lambda_i i)^{h_i-h'_i}}{\rho_1^h\prod_{i\in\mathbb{N}}(2i)^{c_i}c_i!}\right).\end{aligned} \tag{3.24}$$

Now we count all the different U.H. fragments $G'$ isomorphic to $H$. There are three operations that can be performed on $G$ to produce an isomorphic U.H. fragment:

1 Inside a tree attached to a cycle, we can arbitrarily change the order of any vertex's children.
2 Arbitrarily change the order of cycles of the same length.
3 Inside a cycle, we can change the ordering of the vertices for another cyclic ordering.

Operations (1) and (2) can be performed in $\prod_{i\geq 1}(i-2)!^{h'_i}(i-1)!^{h_i-h'_i}$, and $\prod_{i\geq 1}c_i!$ different ways, respectively. The number of ways to apply (3) depends on the length of the cycle. For cycles of length $k \geq 3$, there are $2k$ cyclic orderings of their vertices. However, for $k = 1, 2$, vertices inside a $k$-cycle admit only $k$ cyclic orderings. Thus operation (3) can be performed in $\frac{1}{2^{c_1+c_2}}\prod_{i\geq 1}(2i)^{c_i}$ ways. Among all these ways of permuting vertices in $G$, $\mathrm{aut}(H)$ ones yield $G$ itself. Thus, the number of different U.H. fragments isomorphic to $H$ is

$$\frac{\left(\prod_{i\geq 1}(i-2)!^{h'_i}(i-1)!^{h_i-h'_i}\right)\left(\prod_{i\geq 1}2i^{c_i}c_i!\right)}{\mathrm{aut}(H)2^{c_1+c_2}}.$$

Observe that $\mathrm{aut}_{\text{h.e.}}(H) = \mathrm{aut}(H)2^{c_1+c_2}$. This way, the amount above multiplied with last expression in Equation (3.24) yields $\Pr(\Gamma \simeq H) = p^*_H$, as we wanted. □

**Theorem 3.8** (Combinatorial Interpretation of The Fragment Distribution)**.** *Suppose $\nu < 1$. Let $\mathcal{H}$ be the class of unlabeled fragments. Then $\sum_{H\in\mathcal{H}} p^*_H = 1$.*

*Proof.* By last lemma it is sufficient to show that the model $\Gamma(\boldsymbol{d})$ defined in Definition 3.2 induces a distribution over unlabeled fragments. The total number of cycles in $\Gamma$, $R$, and the number of $i$-cycles, $Z_i$, have well-defined distributions over $\mathbb{Z}$. In addition to that, the general offspring distribution of $\mathcal{BP}_r$ is $\widehat{D}$, whose mean is $\nu < 1$, so by Lemma 1.3, this branching process induces a well-defined distribution over finite rooted forests. This completes the proof. □

**Corollary 3.6.** *Assume $\nu < 1$. Let $\mathbb{F}$ be the class of unlabeled fragments. Then the sequences $(H \mapsto p^*_H(n))_{n\in\mathbb{N}}$ and $(H \mapsto p_H(n))_{n\in\mathbb{N}}$ of real maps over $\mathbb{F}$ are tight. In particular, for all functions $\omega(n)$ tending to infinity, $\Pr(|\mathrm{Frag^*}_n| \geq \omega(n)) = o(1)$ and $\Pr(|\mathrm{Frag}_n| \geq \omega(n)) = o(1)$.*

*Proof.* The last part of the statement follows from the definition tight sequence. The fact that $(H \mapsto p^*_H(n))_{n\in\mathbb{N}}$ is tight follows from last theorem. To see that $(H \mapsto p_H(n))_{n\in\mathbb{N}}$ is tight as well, note that by definition $p_H(n) \leq \Pr(\mathrm{Frag^*}_n \simeq H)/\Pr(\mathcal{CM}(n,\boldsymbol{d})\text{ is simple})$, and $1/\Pr(\mathcal{CM}(n,\boldsymbol{d})\text{ is simple}) \leq e^{\nu/2+\nu^4/4} + o(1)$. □

### 3.3.2 No gap when $\nu \geq 1$

The following can be proven exactly as Lemma 3.8, using the results from the previous section.

**Theorem 3.9.** *Assume $\nu \geq 1$. Then $\overline{L_{\boldsymbol{d}}} = [0,1]$.*

### 3.3.3 Always a finite union of intervals

Here we study $\overline{L_{\boldsymbol{d}}}$ in the case $\nu < 1$. Once more, as in the previous sections, it holds that the $\equiv_k$-class of $\mathcal{G}(n,\boldsymbol{d})$ is determined w.h.p. by its fragment $\mathrm{Frag}_n$ when $\nu < 1$. This is (implicitly) stated in [46][Lemma 3.12]. However, the results in [45, 46] contain slight inaccuracies. These are discussed in Section 3.3.5.

**Theorem 3.10.** *Suppose that $\nu < 1$. Let $H \in \mathbb{F}$ be some fragment, and $\varphi \in \mathrm{FO}_g$ be a sentence. Then*

$$\lim_{n\to\infty} \Pr(\mathcal{G}(n,\boldsymbol{d}) \models \varphi \mid \mathrm{Frag}_n \simeq H) = 0 \text{ or } 1.$$

Recall that when $\nu < 1$ and $H \in \mathbb{F}$ is some fragment, $p_H(n)$ stands for the probability $\Pr(\mathrm{Frag}_n \simeq H)$, and $p_H = \lim_{n\to\infty} p_H(n)$. Similarly to the previous sections, it turns out that $L_{\boldsymbol{d}}$ equals the set of partial sums of fragments probabilities. This allows us to study $L_{\boldsymbol{d}}$ using Kakeya's criterion. The following can be proven exactly as in Lemma 3.10, using last theorem and the fact that the fragment distribution is tight (Corollary 3.6).

**Theorem 3.11.** *Assume $\nu < 1$. Then*

$$\overline{L_{\boldsymbol{d}}} = \left\{ \sum_{H\in\mathcal{U}} p_H \;\middle|\; \mathcal{U} \subseteq \mathbb{F} \right\}.$$

As in previous sections, now the desired results about $\overline{L_{\boldsymbol{d}}}$ follow from analysing the set of fragment probabilities and using Kakeya's Criterion. However, a difficulty here is that fragment probabilities depend on many features of $\boldsymbol{d}$ other than the parameter $\nu$. In order to circumvent this issue, we use the following lemma.

**Lemma 3.29.** *Suppose that $\nu < 1$. Define $Q = Q(\nu) = \sqrt{1-\nu}\, e^{-\nu/2-\nu^2/4}$. Let $\boldsymbol{a} = (a_n)_{n\geq 3}$ be a sequence of natural numbers $a_n \in \mathbb{N}$ whose sum $\sum_{n\geq 3} a_n$ is finite. Let $\mathbb{F}_{\boldsymbol{a}}$ be the set of fragments*

$$\{H \in \mathbb{F} \mid H \text{ contains exactly } a_i \text{ } i\text{-cycles for all } i \geq 3\}.$$

*Then*

$$\sum_{H\in\mathbb{F}_{\boldsymbol{a}}} p_H = Q \prod_{i\geq 3} \frac{(\nu^i/2i)^{a_i}}{a_i!}. \tag{3.25}$$

*In particular, $p_H$ is maximized when $H$ is the empty fragment.*

*Proof.* Let $A(n)$ be the event that $\mathcal{G}(n, \boldsymbol{d})$ contains exactly $a_i$ $i$-cycles for each $i \geq 3$. By Lemma 3.27 it holds that

$$\Pr(A(n)) = Q \prod_{i\geq 3} \frac{(\nu^i/2i)^{a_i}}{a_i!} + o(1).$$

For each $H \in \mathbb{F}$, let

$$q_H(n) = \Pr(A(n) \mid \mathrm{Frag}_n \simeq H) \Pr(\mathrm{Frag}_n \simeq H).$$

By the law of total probability $\Pr(A_n) = \sum_{H\in\mathbb{F}} q_H(n)$. Moreover, observe that $q_H(n) \leq p_H(n)$ for all $H$, so the sequence of maps $(H \mapsto q_H(n))_{n\in\mathbb{N}}$ is tight. This way

$$\lim_{n\to\infty} \Pr(A(n)) = \sum_{H\in\mathbb{F}} \lim_{n\to\infty} q_H(n).$$

By Theorem 3.6, we know that w.h.p. all cycles in $\mathcal{G}(n, \boldsymbol{d})$ lie in $\mathrm{Frag}_n$. This implies that $q_H(n) = p_H + o(1)$ if $H \in \mathbb{F}_{\boldsymbol{a}}$ and $q_H(n) = o(1)$ otherwise. Using this in last equality shows Equation (3.25).

We show now the last part of the lemma, which states that $p_H$ is maximized for the empty fragment $H$. Let $H \in \mathbb{F}$ be not empty, and let $\boldsymbol{a} = (a_i)_{i\geq 3}$ be the sequence where $a_i$ is the number of $i$-cycles in $H$ for each $i \geq 3$. By Equation (3.25),

$$p_H \leq Q \prod_{i\geq 3} \frac{(\nu^i/2i)^{a_i}}{a_i!}.$$

However, as $\nu < 1$, the expression on the right is at most $Q$, which is the probability of the empty fragment. This completes the proof. □

For the remainder of this subsection, we number the fragments in $\mathbb{F}$ as $H_1, H_2, \ldots$ in such a way that $p_{H_i} \geq p_{H_j}$ for all $i < j$. For convenience we define $p_i = p_{H_i}$. For each $i > 1$, let $k = k(i)$ be the number satisfying

$$Q\frac{\nu^k}{2k} \geq p_i > Q\frac{\nu^{k+1}}{2(k+1)},$$

where $Q = \sqrt{1-\nu}\, e^{-\nu/2-\nu^2/4}$ as in last lemma. We impose the condition $i > 1$, because for $i = 1$, by Lemma 3.29, $H_1$ corresponds to the empty fragment and $p_1 = Q$, so $k(1)$ would not be well-defined. Observe that Lemma 3.29 also implies $k(i) \geq 3$ for all $i > 1$. Finally, the probabilities $p_i$ are non-increasing and have limit zero, so $k(i)$ is non-decreasing and tends to infinity with $i$.

**Lemma 3.30.** *Assume $\nu < 1$. Then $\overline{L_{\boldsymbol{d}}}$ is a finite union of intervals in* $[0, 1]$.

*Proof.* Let $i_0$ be an index for which $k_0 = k(i_0)$ satisfies $\sum_{j=3}^{k_0-2} 1/j \geq 4/\nu$. The harmonic series $\sum 1/j$ is divergent, so this value $i_0$ exists. We prove that $p_i \leq \sum_{j>i} p_j$ for $i \geq i_0$. By Kakeya's Criterion, this implies the result. Let $i > i_0$, $k = k(i)$. For each $3 \leq \ell \leq \lfloor \frac{k+1}{2} \rfloor$, let $\mathbb{F}_\ell$ be the set of unlabeled fragments containing a $\ell$-cycle, a $(k-\ell+1)$-cycle, and no other cycle. By Lemma 3.29, it holds that $\sum_{H\in\mathbb{F}_\ell} p_H$ equals $Q\frac{\nu^{k+1}}{4j(k-\ell+1)}$ for $\ell \neq (k+1)/2$, and half this value for $\ell = (k+1)/2$. In all cases, this quantity is less than $Q\frac{\nu^{k+1}}{2(k+1)}$, so $p_H < p_i$ for all $H \in \mathbb{F}_\ell$. This shows that the elements of $\mathbb{F}_\ell$ contribute to the tail $\sum_{j>i} p_j$. In other words, $\mathbb{F}_\ell \subset \{H_j \mid i > j\}$. The same expression holds true when substituting $\mathbb{F}_\ell$ for the disjoint union of all the sets $\mathbb{F}_\ell$. This way,

$$\sum_{j>i} p_j \geq \sum_{\ell=3}^{\lfloor \frac{k+1}{2} \rfloor} \sum_{H\in\mathbb{F}_\ell} p_H = \frac{Q\,\nu^{k+1}}{8} \sum_{\ell=3}^{k-2} \frac{1}{\ell(k-\ell+1)} \geq \frac{Q\,\nu^{k+1}}{8k} \sum_{\ell=3}^{k-2} \frac{1}{\ell} \geq \frac{Q\,\nu^k}{2k}.$$

Last inequality holds because of our choice of $k = k(i)$. By hypothesis, $p_i \leq Q\nu^k/2k$, so $p_i \leq \sum_{j>i} p_j$, as we wanted to show. This proves the result. ☐

### 3.3.4 Transition at $\nu_0$

**Lemma 3.31.** *Let $\nu_0$ be as defined in Equation* (3.20)*. The following hold. (1) If $0 < \nu < \nu_0$, then $\overline{L_{\boldsymbol{d}}}$ has at least one gap, and (2) if $\nu_0 \leq \nu < 1$, then $\overline{L_{\boldsymbol{d}}} = [0, 1]$.*

*Proof.* As in the previous subsection, let $H_1, H_2, \dots$ be an enumeration of the class of fragments $\mathbb{F}$ satisfying $p_{H_1} \geq p_{H_2} \geq \dots$, and let $p_i = p_{H_i}$ for all $i$. By Kakeya's theorem, $\overline{L_{\boldsymbol{d}}} = [0, 1]$ if and only if

$$p_i \leq \sum_{j>i} p_j, \tag{3.26}$$

being true for all $i$. We begin by showing (1). Recall that $\nu_0$ is defined as the only solution to $Q(\nu_0) = 1/2$ lying in $[0, 1]$. As $Q(\nu)$ is monotonically decreasing in $[0, 1]$ and $0 < \nu < \nu_0$, it holds $Q > 1/2$. By Lemma 3.29, $H_1$ corresponds to the empty fragment, so $p_1 = Q > 1/2 > \sum_{j>1} p_j$. using that $\sum_{j\geq 1} p_j = 1$. Hence, eq. (3.26) does not hold for $i = 1$ and $\overline{L_{\boldsymbol{d}}}$ contains at least one gap. Now we proceed to show (2). In this case, $\nu_0 \leq \nu < 1$, and eq. (3.26) holds for $i = 1$, because $Q \leq 1/2$. We show that eq. (3.26) holds for $i > 1$ as well. Fix $i > 1$. For all $\ell \geq 3$, we define $\mathbb{F}_\ell$ as the set of unlabeled fragments containing an $\ell$-cycle and no other cycles. By Lemma 3.29, $\sum_{H\in\mathbb{F}_\ell} p_H = Q\nu^\ell/(2\ell)$. Similarly to last theorem, with this we obtain

$$\sum_{j>i} p_j \geq \sum_{\ell>k} \sum_{H\in\mathbb{F}_\ell} p_H = \sum_{\ell>k} Q\frac{\nu^\ell}{2\ell} \geq \frac{Q\nu^k}{2k} \sum_{\ell\geq 1} \left(\frac{\nu(k+1)}{k}\right)^\ell,$$

where $k = k(i)$. Last inequality above follows from the quotient $\left(\frac{Q\nu^{\ell+1}}{2(\ell+1)}\right) \Big/ \left(\frac{Q\nu^{\ell}}{2(\ell)}\right)$ being at most $\frac{\nu k}{k+1}$ for all $\ell > k$. In addition to that, we have $k(i) \geq 3$ for all $i > 1$, so $k/(k+1) \geq 3/4$. Joining this with the inequality above we obtain $\sum_{j>i} p_j \geq \frac{Q\nu^k}{2k} \sum_{\ell>1} (3\nu/4)^{\ell}$. Note that $\nu_0 \geq 3/4$, so last sum is at least $\frac{9}{16}\frac{1}{1-9/16} = 9/7$. Putting this together with last inequality we obtain

$$\sum_{j>i} p_j \geq \frac{Q\nu^k}{2k}\frac{9}{7} > \frac{Q\nu^k}{2k}.$$

But by definition of $k(i)$, it holds $p_i \leq \frac{Q\nu^k}{2k}$, so we have proven eq. (3.26) for our choice of $i$. This proves the result. □

### 3.3.5 Remarks About the Convergence Law

In this section we discuss the convergence law studied by Lynch in [46, 45]. The main result there states that for any sentence $\phi \in \mathrm{FO}_g$, the limit of $\Pr(\mathcal{G}(n, \boldsymbol{d}) \models \phi)$ exists. We note that in those works conditions different than ours are imposed on the asymptotic degree sequence $\boldsymbol{d}$. More precisely, there is no condition stating that the second moment $\mathrm{E}\left[D_n^2\right]$ converges to a finite quantity. Instead, this is replaced by the existence of some cutoff function $\omega(n)$ satisfying $\Delta(n) \leq \omega(n)$. In [45], $\omega(n) = n^{\alpha}$, where $\alpha < 1/4$, while in [46], $\omega(n)$ was sub-polynomial (that is, $\omega(n) = o(n^{\alpha})$ for all $\alpha > 0$). Observe that neither cutoff is enough to guarantee that $\mathrm{E}\left[D_n^2\right]$ converges to a finite quantity (in fact, no diverging cutoff function suffices). We shall see that these cutoff requirements are not enough to justify the proofs presented in [46, 45]. However, under our conditions the techniques shown in those papers can be applied and a convergence law holds.

The approach followed in [46, 45] consists of proving a FO convergence law for the configuration model $\mathcal{CM}(n, \boldsymbol{d})$ and transferring this result to $\mathcal{G}(n, \boldsymbol{d})$ afterwards. Configurations $F$ are seen as relational structures whose elements $e \in F$ are the half-edges, equipped with two relations: a matching $M$, relating half-edges that are joined, and an equivalence relation $\equiv$ that relates half-edges belonging to the same vertex. Hence, the first-order language for configurations is $\mathrm{FO}[\sigma_{\mathcal{CM}}]$, where $\sigma_{\mathcal{CM}} = \{M, \equiv\}$, and sentences $\phi \in \mathrm{FO}[\sigma_{\mathcal{CM}}]$ are interpreted in the obvious way. It is easy to see that for any graph sentence $\phi \in \mathrm{FO}_g$, there is another one $\phi_F \in \mathrm{FO}[\sigma_{\mathcal{CM}}]$ such that any simple configuration (i.e., configuration with no loops of multiple edges) $F$ satisfies $F \models \phi_F$ iff $G \models \phi$, where $G$ is the underlying graph of $F$.

Suppose that a $\mathrm{FO}[\sigma_{\mathcal{CM}}]$-convergence law holds for $\mathcal{CM}(n, \boldsymbol{d})$. In [46, 45] is argued that a $\mathrm{FO}_g$-convergence law for $\mathcal{G}(n, \boldsymbol{d})$ follows. The property of being simple is expressible via a sentence $\psi \in \mathrm{FO}[\sigma_{\mathcal{CM}}]$. Let $\phi \in \mathrm{FO}_g$ be an arbitrary sentence and let $\phi_F \in \mathrm{FO}[\sigma_{\mathcal{CM}}]$ be a sentence for configurations chosen as described in the previous paragraph. Then:

$$\Pr(\mathcal{G}(n, \boldsymbol{d}) \models \phi) = \frac{\Pr(\mathcal{CM}(n, \boldsymbol{d}) \models \psi \wedge \phi_F)}{\Pr(\mathcal{CM}(n, \boldsymbol{d}) \models \psi)}.$$

This alone yields a convergence law for $\mathcal{G}(n, \boldsymbol{d})$ as long as $\Pr(\mathcal{CM}(n, \boldsymbol{d}) \models \psi)$ has a positive limit. In [46, 45] is stated that this is the case. However, this is not true when $\mathrm{E}\left[D_n^2\right]$ diverges.

**Lemma 3.32.** *Let $\boldsymbol{d} = \boldsymbol{d}(n)$ be a smooth asymptotic degree sequence with $\mathrm{E}\left[D_n^2\right]$ diverging to* *infinity. Then $\mathcal{CM}(n, \boldsymbol{d})$ a.a.s. contains a loop.*

*Sketch of the proof.* Let $X(n)$ count the loops in $\mathcal{CM}(n, \boldsymbol{d})$. Then

$$\mathrm{E}\left[X(n)\right] = \frac{1}{2}\sum_{v\in[n]} \frac{d_v(n)(d_v(n)-1)}{2m_n - 1} = \frac{1}{2}\frac{\rho_2(n)}{\rho_1(n) - 1/n}.$$

Using that $\rho_2(n)$ diverges and $\boldsymbol{d}$ is smooth we get that $\mathrm{E}\left[X(n)\right]$ tends to infinity. The result follows now after showing $\mathrm{Var}\left(X(n)\right) = o(\mathrm{E}\left[X(n)^2\right])$ and applying the second moment method. □

Nevertheless, the arguments in [46, 45] work when $\boldsymbol{d}$ is well-behaved as we assumed throughout this section. The proof follows the analogous one for $G(n, c/n)$ given in [49]. We sketch the arguments here. Given $r, n$, let $\mathrm{Frag}^*{}_n|_r$ be the $r$-neighbourhood of all cycles of length at most $2r+1$ lying in $\mathcal{CM}(n, \boldsymbol{d})$. For each $r$, there are distributions $\Gamma|_r$ and $\mathcal{T}|_r$, over unlabeled $r$-fragments and unlabeled rooted trees, respectively, satisfying:

(1) $\mathrm{Frag}^*{}_n|_r$ converges in distribution to $\Gamma|_r$.
(2) For any fixed $\ell$, the $r$-neighborhood of $\ell$ uniformly-chosen vertices in $\mathcal{CM}(n, \boldsymbol{d})$ converges in distribution to $(\mathcal{T}|_r)^\ell$.

From those two facts we can derive that, for any $r$, $\mathcal{CM}(n, \boldsymbol{d})$ is $r$-simple w.h.p. in the sense of Definition 2.6. We can also conclude that, for all $k, r$, $\mathcal{CM}(n, \boldsymbol{d})$ is $(k, r)$-rich in the sense of Definition 2.7, but considering only $\equiv_k^{\mathrm{Ly}}$-classes of trees $\mathcal{C}$ that have non-zero probability according to $\mathcal{T}|_r$. This is enough to show that the $\equiv_k^{\mathrm{Ly}}$-class of $\mathrm{Frag}^*{}_n|_r$ determines the $\equiv_k$-class of $\mathcal{CM}(n, \boldsymbol{d})$ using the strategy given in Theorem 2.3, and to prove a convergence law along the lines of Theorem 2.1. We remark that we are being slightly informal here: in order to properly reproduce these arguments we need to define the notions of tree, cycle, fragment, $\equiv_k^{\mathrm{Ly}}$, and so on over configurations. However, the intuitive definitions fortunately work.

# Chapter 4

# Preservation Theorems for FO Logic on Random Graphs

Preservation theorems are classical result from first-order logic stating the equivalence of certain *semantic classes* and *syntactic classes* of sentences. A remarkable aspect of those results is that the semantic classes they talk about are not decidable ( i.e., there is no algorithmic procedure determining whether a general sentence belongs to the class or not), while the syntactic classes are.

A sentence $\varphi \in \mathrm{FO}[\sigma]$ is called **monotone** in a relation symbol $R \in \sigma$, if $G \models \varphi$ implies $G' \models \varphi$ for any $\sigma$-structures (finite or infinite) $G, G'$ where $G'$ is obtained from $G$ by adding tuples to $R(G)$. We write $G \leq_R G'$ to represent this situation. When $\varphi \in \mathrm{FO}_g$ and $R$ is the adjacency relation, we simply say $\varphi$ is positive. A $\varphi \in \mathrm{FO}[\sigma]$ is **preserved under extensions** if $G \models \varphi$ implies $G' \models \varphi$ for any $\sigma$-structures (finite or infinite) $G, G'$ where $G'$ contains some induced $G$-copy. This situation is represented as $G \sqsubset G'$. An **homomorphism** between $\sigma$-structures $G, G'$ is a map $f : V(G) \to V(G')$ that maps $\mathfrak{c}^G$ to $\mathfrak{c}^{G'}$ for each constant symbol $\mathfrak{c} \in \sigma$ and maps tuples from $R(G)$ to tuples from $R(G')$ for all relations $R \in \sigma$. We write $G \to G'$ to denote that there is some homomorphism from $G$ to $G'$. A sentence $\varphi \in \mathrm{FO}[\sigma]$ is **preserved under homomorphisms** if $G \models \varphi$ implies $G' \models \varphi$, for all (finite or infinite) $\sigma$-structures $G, G'$ with $G \to G'$.

The properties of FO sentences introduced above are *semantic* in the sense that they are defined by imposing some constraints on their classes of models. Perhaps unsurprisingly, those properties are undecidable. For example, given a sentence $\varphi \in \mathrm{FO}[\sigma]$, we define the sentence $\psi \in \mathrm{FO}[\sigma']$ as $\varphi \wedge \forall x \neg R(x)$, where $\sigma' = \sigma \cup \{R\}$ and $R$ is a fresh unary relation symbol. It is easy to see that $\psi$ is monotone in $R$ if and only if $\varphi$ does not have any non-empty model. However, satisfiability of FO-sentences is undecidable, as shown by Church [13] and Turing [66] independently. Similar tricks can be used to show the undecidability of closedness under extensions and closedness under homomorphisms.

A sentence $\varphi \in \sigma[\sigma]$ is **positive** in a relation $R \in \sigma$ if $R$ does not appear in $\psi$ for all subformulas $\neg\psi$ of $\varphi$. If $\varphi$ is positive in all relations $R \in \sigma$, then $\varphi$ is simply called positive. The sentence $\varphi$ is said to be **existential** if it is in prenex normal form, and all its quantifiers are existential. Clearly, it is easy to recognize whenever a formula is positive or existential. The following three results are the so-called **preservation theorems** in FO logic:

**Theorem 4.1** (Lyndon's Theorem [50]). *Let $\varphi \in \mathrm{FO}[\sigma]$ be a sentence which is monotone in a relation symbol $R \in \sigma$. Then $\varphi$ is equivalent to another sentence $\psi \in \mathrm{FO}[\sigma]$ which is positive in $R$.*

**Theorem 4.2** (Łoś-Tarski Theorem [32]). *Let $\varphi \in \mathrm{FO}[\sigma]$ be a sentence preserved under extensions. Then $\varphi$ is equivalent to an existential sentence $\psi \in \mathrm{FO}[\sigma]$.*

**Theorem 4.3** (Homomorphism Preservation Theorem [57]). *Let $\varphi \in \mathrm{FO}[\sigma]$ be a sentence preserved under homomorphisms. Then $\varphi$ is equivalent to an existential positive sentence $\psi \in \mathrm{FO}[\sigma]$.*

It is well-known that many features of FO logic do not survive the restriction to finite models. The most salient example is the failure of the Compactness Theorem (an important tool in model theory) on finite structures [43]. Among the three preservation theorems introduced above, only the Homomorphism Preservation Theorem (shortened to HMT) is still true when restricted to the finite case. We give the precise statements below. Given a class $\mathcal{C}$ of $\sigma$-structures, a sentence $\varphi \in \mathrm{FO}[\sigma]$ is said to be **positive in** $R \in \sigma$ **on** $\mathcal{C}$ if $G \models \varphi$ implies $G' \models \varphi$ whenever $G \leq_R G'$ for $G, G' \in \mathcal{C}$. The notions of sentence **preserved under homomorphisms on** $\mathcal{C}$ and **preserved under extensions on** $\mathcal{C}$ are defined similarly. Two sentences $\varphi, \psi \in \mathrm{FO}[\sigma]$ are said to be **equivalent on** $\mathcal{C}$ if $\varphi \leftrightarrow \psi$ holds in all structures $G \in \mathcal{C}$. We say that either of the preservation theorems above (Theorems 4.1 to 4.3) holds in a class $\mathcal{C}$ when they remain true after restricting to $\mathcal{C}$ the notions of monotonicity, preservation under homomorphisms, preservation under extensions, and equivalence between sentences.

**Theorem 4.4** (Lyndon's Theorem fails on finite structures [2, 64]). *There exists a signature $\sigma$ and a sentence $\varphi \in \mathrm{FO}[\sigma]$ which is monotone in $R \in \sigma$ on finite structures and is not finitely (i.e., on finite structures) equivalent to any sentence $\psi \in \mathrm{FO}[\sigma]$ positive in $R$.*

**Theorem 4.5** (Łoś-Tarski Theorem fails on finite structures [65]). *There exists a signature $\sigma$ and a sentence $\varphi \in \mathrm{FO}[\sigma]$ which is preserved under extensions on finite structures and is not finitely equivalent to any existential sentence $\psi \in \mathrm{FO}[\sigma]$.*

**Theorem 4.6** (Finite HPT [57]). *The Homomorphism Preservation Theorem holds on finite structures.*

Even more, counterexamples for Lyndon's and Łoś-Tarski Theorems over finite graphs have been found recently in [39] and [12] respectively. However, it turns out that those counterexamples quite involved, and the underlying reasons why those preservation theorems fail are not straightforward. There is a line of research which tries to establish precise conditions under which preservation theorems hold in a particular class $\mathcal{C}$ [7, 8, 18, 17]. Most of this work focuses on the HMT [8, 18], except for [8], which deals with Łoś-Tarski Theorem.

Given that the obstacles to Lyndon's and Łoś-Tarski theorems on finite structures seem to be somewhat artificial, a natural question is whether those results are still "mostly true" in some sense. The purpose of this chapter is to study those two preservation theorems in the context of random graphs. Given a random graph model $(\mathcal{G}_n)_{n\in\mathbb{N}}$, we say that **Lyndon's (Łoś-Tarski) Theorem holds a.a.s. in** $\mathcal{G}_n$ if for any sentence $\varphi \in \mathrm{FO}_g$ which is monotone (preserved under extensions) on finite graphs there is another positive sentence (existential sentence) $\psi \in \mathrm{FO}_g$ such that a.a.s. $\mathcal{G}_n \models \varphi \leftrightarrow \psi$. Observe that if the asymptotic probability of $\varphi$ is either zero or one for all monotone sentences, or those preserved under extensions, then the respective preservation theorem holds a.a.s. trivially.

During this chapter we give various positive results in this regard, particularly for $\mathcal{G}_n = \mathcal{G}(n,p)$ in the regimes $p \sim c/n$ Section 4.1, $p \sim cn^{-1-1/\ell}$ Section 4.2, and $p = \Theta(\log n/n)$ Section 4.3, the main cases where a $\mathrm{FO}_g$-convergence law holds, but a zero-one law does not. Additionally, we also study the case where $\mathcal{G}_n$ is chosen uniformly from an addable minor-closed class Section 4.4.

## 4.1 Preservation Theorems on Sparse Graphs

The main theorem of this section is the following.

**Theorem 4.7.** *Let $\varphi \in \mathrm{FO}_g$ be a sentence which is either monotone or preserved under extensions on finite structures. Then there is an existential positive sentence $\psi \in \mathrm{FO}_g$ such that*

$$\lim_{n\to\infty} \Pr(\mathcal{G}(n,p) \models \varphi \leftrightarrow \psi) = 1,$$

*for all probabilities $p = p(n)$ satisfying $p \sim c/n$ for some $c > 0$. In particular, both Lyndon's and Łoś-Tarski Theorems hold a.a.s. in $\mathcal{G}(n,p)$.*

We sketch the proof below. Some definitions are needed for that.

**Definition 4.1.** Similarly to Definition 2.5, the $r$-**core** $\mathrm{Core}(G)|_r$ of a graph $G$ as the $r$-neighbourhood of all its cycles of length at most $2r+1$.

**Definition 4.2.** Given two graphs $G, H$, we write $G \leq_k^{\mathrm{mo}} H$ if $H$ satisfies all sentences $\varphi \in \mathrm{FO}_g$ that are monotone on finite graphs with $\mathrm{qr}(\varphi) \leq k$ and hold in $G$. Similarly, we write $G \leq_k^{\mathrm{ext}} H$ if the same holds for sentences $\varphi$ closed under extensions on finite structures, instead of monotone.

The main ideas for showing Theorem 4.7 are as follows. We focus on the case where $\varphi$ is monotone on finite graphs. The situation where $\varphi$ is preserved under extensions is handled in a similar way. Let $k = \mathrm{qr}(\varphi)$ Similarly as in Chapter 2, Lynch showed that from the perspective of $\mathrm{FO}_g$ logic of quantifier rank $k$ w.h.p. the only distinguishing factor between random trials $G_0, G_1 \sim \mathcal{G}(n, c/n)$ is their $r$-core, where $r = (3^k - 1)/2$. We show that whenever $\mathrm{Core}(G_0)|_r \subseteq \mathrm{Core}(G_1)|_r$ then w.h.p. $G_0 \leq_k^{\mathrm{mo}} G_1$. This is done by giving explicit graphs $G_0', G_1'$ such that w.h.p. $G_i' \equiv_k G_i$ for $i = 0, 1$ and $G_0' \leq_E G_1'$ (i.e., $G_1'$ can be obtained from $G_0'$ by adding edges). This shows that $\varphi$ is a.a.s. equivalent to $\mathrm{Core}_n|_r \in \mathcal{F}$, where $\mathrm{Core}_n|_r$ stands for the $r$-core of $\mathcal{G}(n, c/n)$ and $\mathcal{F}$ is some upwards closed (with respect to subgraph inclusion) family of graphs. The last piece of the proof is showing that $\mathcal{F}$ has a finite number of minimal elements $H_1, \dots, H_\ell$. In this situation, $\varphi$ would be equivalent w.h.p. to the property "There exists a $H_i$-copy for some $1 \leq i \leq \ell$" which clearly can be expressed via an existential positive sentence. In order to prove that $\mathcal{F}$ has a finite number of minimal elements, we use the fact that the $\equiv_k$-type of $\mathcal{G}(n, c/n)$ is given by the $\equiv_k^{\mathrm{Ly}}$-class of $\mathrm{Core}_n|_r$ ($\equiv_k^{\mathrm{Ly}}$ is a relation given in [49], similarly to Section 2.1), and show that each of those $\equiv_k^{\mathrm{Ly}}$-classes has a unique minimal element.

Before moving on to the proof of Theorem 4.7 we briefly introduce some facts about $\mathcal{G}(n, c/n)$ given in [49]. The ideas mirror those of Chapter 2: W.h.p. $\mathcal{G}(n, c/n)$ is $r$-simple and $(k, r)$-rich (in a sense analogous to Definition 2.6 and Definition 2.7), and the $\equiv_k$-class of $\mathcal{G}(n, c/n)$ depends only on the class of its $r$-fragment (for $r = (3^k - 1)/2$) according to some equivalence relation $\equiv_k^{\mathrm{Ly}}$ analogous to those introduced in Section 2.1. As an auxiliary result, we also show that each $\equiv_k^{\mathrm{Ly}}$-class has a unique minimal element.

### 4.1.1 Minimum Trees

**Definition 4.3** (Equivalence of trees, graph case)**.** Given $k \in \mathbb{N}$, we define the equivalence relation $(T_0, x^0) \equiv_k^{\mathrm{Ly}} (T_1, x^1)$ between rooted trees of the same height by induction, as follows. If both $T_0, T_1$ have height 0, they consist simply of their roots and $(T_0, x^0) \equiv_k^{\mathrm{Ly}} (T_1, x^1)$ holds. Now assume $T_0, T_1$ have height $r > 0$, and $\equiv_k^{\mathrm{Ly}}$ has been defined for smaller values of $r$. Then $(T_0, x_0) \equiv_k^{\mathrm{Ly}} (T_1, x^1)$ if for all $\equiv_k^{\mathrm{Ly}}$-class $\mathcal{C}$ of trees with height at most $r - 1$, the quantity

$$|\{v \in V(T_i) \mid \{x^i, v\} \in E(T_i),\, T_i(v; x^i) \in \mathcal{C}\}|$$

is the same for $i = 0, 1$, or is at least $k$ in both cases. Here $T_i(v; x^i)$ denotes the rooted sub-tree that "hangs" from $v$ in $(T_i, x^i)$, as in the word by word definition from Section 2.1.1.

The following is just an easier version of Lemma 2.1. The same proof works.

**Lemma 4.1.** *Let $(T_0, x^0), (T_1, x^1)$ be rooted trees satisfying $(T_0, x^0) \equiv_k^{\mathrm{Ly}} (T_1, x^1)$. Then $(T_0, x^0) \equiv_k^{\mathrm{dFO}} (T_1, x^1)$.*

**Definition 4.4** (Minimum trees)**.** Given a $\equiv_k^{\mathrm{Ly}}$ class $\mathcal{C}$ of rooted trees, we define the representative $\min(\mathcal{C})$ by induction on the height of $\mathcal{C}$'s elements. If $\mathcal{C}$ is formed by height zero trees, then $\min(\mathcal{C})$ is simply the isolated root $x^{\mathcal{C}}$. Otherwise, suppose that $\mathcal{C}$'s height is $r > 0$. Then, by definition of $\equiv_k^{\mathrm{Ly}}$, there is a partition $\mathfrak{C}^-, \mathfrak{C}^+$ of all tree $\equiv_k^{\mathrm{Ly}}$-classes of height at most $r - 1$, and numbers $a_{\mathcal{C}_i} \in \mathbb{N}$ for each $\mathcal{C}_i \in \mathfrak{C}^-$, such that a tree $(T, x)$ belongs to $\mathcal{C}$ if and only if the quantity

$$|\{v \mid x \sim v,\, T(v; x) \in \mathcal{C}_i\}|$$

equals $a_{\mathcal{C}_i}$ for all $\mathcal{C}_i \in \mathfrak{C}^-$, and is at least $k$ for all $\mathcal{C}_i \in \mathfrak{C}^+$. In this situation, we construct $\min(\mathcal{C})$ by attaching $a_{\mathcal{C}_i}$ copies of $\min(\mathcal{C}_i)$ to the root for each $\mathcal{C}_i \in \mathfrak{C}^-$, and $k$ copies of $\min(\mathcal{C}_i)$ for each $\mathcal{C}_i \in \mathfrak{C}^+$.

**Lemma 4.2.** *Suppose that $\mathcal{C}$ is the $\equiv_k^{\mathrm{Ly}}$-class of $(T, x)$. Then $(T, x)$ is an extension of $\min(\mathcal{C})$.*

*Proof.* Let $(T_{\mathcal{C}_i}, x^{\mathcal{C}_i}) = \min(\mathcal{C})$. We proceed by induction on $T$'s height. When $T$'s height is zero the result trivially holds. Otherwise, suppose $T$'s height is $r > 0$ and the statement is true for smaller values. Let $v_1, \ldots, v_\ell$ be all $v^{\mathcal{C}}$'s children in $(T_{\mathcal{C}}, x^{\mathcal{C}})$, and let $\mathcal{C}_i$ be the $\equiv_k^{\mathrm{Ly}}$-class of

$T^{\mathcal{C}}(v_i)$, for all $i = 1, \ldots, \ell$. By definition, of $(T_{\mathcal{C}}, x^{\mathcal{C}})$ it actually holds that $T^{\mathcal{C}}(v_i)$ is isomorphic to $(T_{\mathcal{C}_i}, x^{\mathcal{C}_i})$, for all $i$. By definition of $(T_{\mathcal{C}}, x^{\mathcal{C}})$, the number of children $v \sim x$ in $(T, x)$ satisfying $T(v; x) \in \mathcal{C}_i$ for any fixed $\equiv_k^{\mathrm{Ly}}$-class is not greater than the number of children $v \sim x^{\mathcal{C}}$ in $(T_{\mathcal{C}}, x^{\mathcal{C}})$ with the same property. Thus, we can find different vertices $u_1, \ldots, u_\ell$, each satisfying $u_i \sim x$ in $(T, x)$, and $T(u_i; x) \equiv_k^{\mathrm{Ly}} T_{\mathcal{C}}(v_i; x^{\mathcal{C}})$. By the induction hypothesis, there are embeddings $f_i$ from $T_{\mathcal{C}}(v_i; x^{\mathcal{C}})$ into $T(u_i; x)$ for all $1 \leq i \leq \ell$. An embedding $f$ from $(T_{\mathcal{C}}, x^{\mathcal{C}})$ into $(T, x)$ can be defined by setting $f(v) = x$ if $v = x^{\mathcal{C}}$, and $f(v) = f_i(v)$ if $v$ belongs to the tree $T_{\mathcal{C}}(v_i)$. This proves the statement. □

### 4.1.2 Minimum Fragments

We define the relation $\equiv_k^{\mathrm{Ly}}$ between unicycles and fragments exactly in the same way as in Chapter 2.

**Definition 4.5** (Minimum unicycles and fragments)**.** Let $\mathcal{C}$ be a $\equiv_k^{\mathrm{Ly}}$-class of unicycles, and $U$ be a representative of $\mathcal{C}$, whose cycle is $C$. The **minimum unicycle** $\min(\mathcal{C})$ is formed by taking a $C$-copy and attaching to each vertex $v \in V(C)$ a copy of $T_{\mathcal{C}_v}$, where $\mathcal{C}_v$ is the $\equiv_k^{\mathrm{Ly}}$-class of $T(U, v)$ (i.e., the tree that "hangs" out of $v$ in the unicycle $U$, as in Section 2.1.2). Similarly, let $\mathcal{C}$ be a $\equiv_k^{\mathrm{Ly}}$-class of fragments. The **minimum fragment** $\min(\mathcal{C})$ is formed by taking $n_{\mathcal{C}}(\mathcal{C}_i)$ copies of the unicycle $\min(\mathcal{C}_i)$ for each $\equiv_k^{\mathrm{Ly}}$-class of unicycles $\mathcal{C}_i$, where $n_{\mathcal{C}}(\mathcal{C}_i)$ stands for the number of components in the class $\mathcal{C}_i$ are contained in an arbitrary representative of $\mathcal{C}$, or just $k$ if this number exceeds $k$.

The following are easy consequences of Lemma 4.2 together with the definition of $\equiv_k^{\mathrm{Ly}}$ over unicycles and fragments.

**Lemma 4.3.** *Let $U$ be a unicycle and let $\mathcal{C}$ be its $\equiv_k^{\mathrm{Ly}}$-class. Then $\min(\mathcal{C})$ can be embedded into $U$.*

**Lemma 4.4.** *Let $H$ be a fragment and let $\mathcal{C}$ be its $\equiv_k^{\mathrm{Ly}}$-class. Then $\min(\mathcal{C})$ can be embedded into $H$.*

### 4.1.3 First Order Logic of Sparse Random Graphs

The we define the notions of $r$-**simple** and $(k, r)$-**rich** graph word by word as in Definition 2.6 and Definition 2.7, for relational structures. The following results are proven in [49], and can be shown as in Chapter 2.

**Lemma 4.5.** *Given $r, k \in \mathbb{N}$ and $p(n) \sim c/n$, w.h.p. $\mathcal{G}(n, p)$ is $r$-simple and $(k, r)$-rich.*

**Lemma 4.6.** *Let $k \in \mathbb{N}, r = (3^k - 1)/2$. Let $G_0, G_1$ be $r$-simple $(k, r)$-rich graphs with $\mathrm{Core}(G_0)|_r \equiv_k^{\mathrm{Ly}} \mathrm{Core}(G_1)|_r$. Then $G_0 \equiv_k G_1$.*

### 4.1.4 Proof of the Main Result

*Proof of Theorem 4.7.* Let $k = \mathrm{qr}(\varphi)$. Given a $r$-fragment $H$ (i.e., a fragment where all cycles have length at most $2r+1$ and all the attached trees have height at most $r$), define $G[H]$ as the disjoint union of $H$ and $k$ copies of $\min(\mathcal{C})$, for each $\equiv_k^{\mathrm{Ly}}$-class $\mathcal{C}$ of trees whose height is at most $r = (3^k - 1)/2$. Observe that by definition, $\mathrm{Core}(G[H])|_r = H$, and $G[H]$ is both $r$-simple and $(k, r)$-rich.

Given two $\equiv_k^{\mathrm{Ly}}$-classes $\mathcal{C}_1, \mathcal{C}_2$ of fragments, we write $\mathcal{C}_1 \leq \mathcal{C}_2$ if $\min(\mathcal{C}_2)$ is an extension of $\min(\mathcal{C}_2)$. Observe that $\leq$ establishes a partial order over $\equiv_k^{\mathrm{Ly}}$-classes of fragments. Let $\mathcal{C}_1 \leq \mathcal{C}_2$ be two $\equiv_k^{\mathrm{Ly}}$-classes of $r$-fragments. We claim that $G[\min(\mathcal{C}_1)] \models \varphi$ implies $G[\min(\mathcal{C}_2)] \models \varphi$. We distinguish between the case where $\varphi$ is preserved under extensions and the one where $\varphi$ is monotone. Suppose $\varphi$ is preserved under extensions on finite structures. Clearly, $G[\min(\mathcal{C}_1)], G[\min(\mathcal{C}_2)]$ are both finite graphs and $\mathcal{C}_1 \leq \mathcal{C}_2$ implies $G[\min(\mathcal{C}_2)]$ extends $G[\min(\mathcal{C}_1)]$, so the claim holds. Alternatively, suppose now that $\varphi$ is monotone on finite structures. Let $s = |V(\min(\mathcal{C}_2))| - |V(\min(\mathcal{C}_1))|$. Let $G'$ be the graph resulting from adding $s$ new isolated vertices to $G[\min(\mathcal{C}_1)]$. It is easily seen that $G' \equiv_k G[\min(\mathcal{C}_2)]$: both graphs are $r$-simple, $(k, r)$-rich, and share the same fragment. Observe that the $G[\min(\mathcal{C}_2)]$ can be obtained from $G'$ by edge addition: using $\min(\mathcal{C}_1)$ and the newly-added $s$ isolated vertices one can reconstruct $\min(\mathcal{C}_2)$, and the remainder of $G'$ is already isomorphic to $G[\min(\mathcal{C}_2)] \setminus \min(\mathcal{C}_2)$. Let $\Omega$ be the set of $\equiv_k^{\mathrm{Ly}}$-classes of $r$-fragments $H$ satisfying $G[H] \models \varphi$ (observe that by Lemma 4.6, if $H_1 \equiv_k^{\mathrm{Ly}} H_2$ then $G[H_1] \equiv_k G[H]$).

The previous claim yields that $\Omega$ is downwards closed with respect to $\leq$. Let $\mathcal{C}_1, \ldots, \mathcal{C}_\ell$ be the minimal elements in $\Omega$ (recall the number of $\equiv_k^{\mathrm{Ly}}$-classes of $r$-fragments is finite, so $\Omega$ is a finite set as well). Let $\psi \in \mathrm{FO}_g$ be an existential positive sentence stating that "$G$ contains a copy of $\min(\mathcal{C}_i)$ for some $1 \leq i \leq \ell$". Given an $r$-fragment $H$, it is easy to see that its $\equiv_k^{\mathrm{Ly}}$-class lies in $\Omega$ if and only if $H$ extends $\min(\mathcal{C}_i)$ for some $1 \leq i \leq \ell$. Thus, $G[H] \models \phi \leftrightarrow \psi$ for all $r$-fragments $H$. Let $\mathrm{Core}_n|_r$ stand for the $r$-core of $\mathcal{G}(n, p)$. By definition of $r$-core, $\mathcal{G}(n, p)$ contains a copy of $\min(\mathcal{C}_i)$ if and only if $\mathrm{Core}_n|_r$ contains a copy of $\min(\mathcal{C}_i)$. This means that $\mathcal{G} \models \psi$ is equivalent to $G[\mathrm{Core}_n|_r] \models \psi$. Additionally, by Lemma 4.5 and Lemma 4.6, $\mathcal{G}(n, p)$ w.h.p. $\mathcal{G}(n, p) \equiv_k G[\mathrm{Core}_n|_r]$ holds. Finally $\mathrm{qr}(\varphi) = k$, so

$$\lim_{n \to \infty} \Pr(\mathcal{G}(n, p) \models \phi \leftrightarrow \psi) = \lim_{n \to \infty} \Pr(G[\mathrm{Core}_n|_r] \models \phi \leftrightarrow \psi).$$

However, the expression in last limit equals one, as stated above, showing the result. □

## 4.2 Preservation Theorems on Very Sparse Graphs

In this section we briefly discuss the regime $p \sim cn^{-\frac{t+1}{t}}$ for some fixed integer $t > 0$. A FO-convergence law can be established here using the same techniques as for $p \sim c/n$ [49]. In fact, computations are simpler in this regime as w.h.p. all components are trees of size at most $t + 1$. We outline the arguments below. In order to do this we temporarily redefine the notions of *simple* and *rich* graphs to fit the new edge density.

**Definition 4.6.** A graph $G$ is **simple** if all its components are trees of size at most $t + 1$. We say that $G$ is $k$-**rich** if, given any tree $T$ of size at most $t$, there are at least $k$ components in $G$ isomorphic to $T$.

Given a graph $H$, we denote by $\overline{H}^k$ the graph resulting from removing all components in $H$ with more than $k$ copies until every component appears at most $k$ times. The following is an straight-forward application of EF games.

**Theorem 4.8.** *Let $G_0, G_1$ be two graphs which are simple and $k$-rich. Additionally, let $H_i$ be the union of all components with size $t+1$ in $G_i$ for $i = 0, 1$. Suppose that $\overline{H_0}^k \simeq \overline{H_1}^k$. Then $G_0 \equiv_k G_1$.*

**Lemma 4.7.** *A.a.s. $\mathcal{G}(n,p)$ is simple. Additionally, for any fixed $k$ a.a.s. $\mathcal{G}(n,p)$ is $k$-rich.*

Last two results together imply that the $k$-type of $\mathcal{G}(n,p)$ is a.a.s. determined by its components of size $t+1$ in this regime. Compare this with the regime $p \sim c/n$, where the $k$-type of $\mathcal{G}(n,p)$ was entirely dependent on its $(3^k - 1)/2$-core.

The main result of this section is as follows.

**Theorem 4.9.** *Let $\varphi \in \mathrm{FO}_g$ be a sentence which is either monotone or preserved under extensions on finite structures. Then there is an existential positive sentence $\psi \in \mathrm{FO}_g$ such that*

$$\lim_{n\to\infty} \Pr(\mathcal{G}(n,p) \models \varphi \leftrightarrow \psi) = 1,$$

*for all probabilities $p = p(n)$ satisfying $p \sim cn^{\frac{-t-1}{t}}$ for some $c > 0$. In particular, both Lyndon's and Łoś-Tarski Theorems hold a.a.s. in $\mathcal{G}(n,p)$.*

An observation is that the Łoś-Tarski Theorem on $\mathcal{G}(n,p)$ in this range can be established using the results from [8]. Indeed, among other results, they show that Łoś-Tarski Theorem holds on the class of finite forests, and $\mathcal{G}(n,p)$ w.h.p. is a forest when $p(n) \sim cn^{\frac{-t-1}{t}}$. Nevertheless, we give an independent proof.

*Proof.* Proof of Theorem 4.9 The proof is analogous to the one for Theorem 4.7. We describe the changes without going into detail. Let $k = \mathrm{qr}(\varphi)$. Given a graph $H$, we define $G[H]$ as the disjoint union of $H$ with $k$ copies of each tree with size at most $t$. It can be seen that if $H_1$ extends $H_0$, then $G[H_0] \models \varphi$ implies $G[H_1] \models \varphi$. Let $\mathcal{F}$ be the set of minimal forests (w.r.t. the extension relation) satisfying $G[F_i] \models \varphi$, and whose trees have all size $t+1$. It is easy to check that $\mathcal{F}$ is actually finite: If $F$ has more than $k$ copies of some component, then $F$ extends $\overline{F}^k$ without equality and $F \equiv_k \overline{F}^k$ by Theorem 4.8. Thus no element in $\mathcal{F}$ contains more than $k$ copies of a component, and so $\mathcal{F}$ is finite, as there is only a finite number of components to choose from (the trees of size $t+1$). Let $\psi \in \mathrm{FO}_g$ be a sentence stating "$G$ contains a copy of some forest in $\mathcal{F}$". It is routine to check that

$$\lim_{n\to\infty} \Pr(\mathcal{G}(n,p) \models \varphi \leftrightarrow \psi) = 1,$$

showing the result. □

## 4.3 Preservation Theorems at the Connectivity Threshold

In this section we establish preservation theorems near to the connectivity threshold of $\mathcal{G}(n,p)$, that is, when $p(n) = \Theta(\ln n/n)$. Here we need to be more precise in our description of $p$. We consider $p(n)$ of the form $p_{s,t,c} + o(1/n)$, where $p_{s,t,c} = \frac{\ln n + t \ln\ln n + c}{sn}$, for some integers $1 \leq s \leq t+1$ and real $c \in \mathbb{R}$. It was shown in [44] that probabilities of this form are the only thresholds of FO properties in the range $p = \Theta(\ln n/n)$. The precise statement is that a FO zero-one law holds in $\mathcal{G}(n,p)$ whenever for all $1 \leq s \leq t+1$, the expression $\ln n + t \ln\ln n - snp(n)$ tends to either $\infty$ or $-\infty$. Moreover, it was proven in [62] that a FO convergence law holds for all probabilities $p = p_{s,t,c} + o(1/n)$. We fix $s, t, c$ and $p$ of this form for the rest of the section.

In this regime, the situation regarding preservation theorems turns out to be more involved than in the previous ones. Our main results are the following

**Theorem 4.10.** *Let $p(n) = p_{s,t,c} + o(1/n)$, and let $\varphi \in \mathrm{FO}_g$ be a sentence preserved under extensions on finite graphs. Then $\lim_{n\to\infty} \Pr(\mathcal{G}(n,p) \models \varphi)$ is either zero or one.*

**Theorem 4.11.** *Let $p(n) = p_{s,t,c} + o(1/n)$, and let $\varphi \in \mathrm{FO}_g$ be a monotone sentence on finite graphs. Suppose $s > 1$. Then, $\lim_{n\to\infty} \Pr(\mathcal{G}(n,p) \models \varphi)$ is either zero or one.*

**Theorem 4.12.** *Let $p(n) = p_{s,t,c} + o(1/n)$, and let $\varphi \in \mathrm{FO}_g$ be a monotone sentence on finite graphs. Suppose $s = 1$. Then, there is a positive sentence $\psi \in \mathrm{FO}_g$ for which $\lim_{n\to\infty} \Pr(\mathcal{G}(n,p) \models \varphi \leftrightarrow \psi) = 1$.*

### 4.3.1 The Landscape at the Connectivity Threshold

The landscape of $\mathcal{G}(n,p)$ in this range is as follows. Asymptotically, cycles of constant size occur an unbounded number of times, and those small cycles lie far away from each other. The degree of most vertices grows with $n$, and vertices of bounded degree are rare. This is reflected in the fact that bounded-degree vertices can only be found far away from cycles, and are grouped in "clusters" of size at most $s$ (we clarify the terms below). In the following we give a more precise description of $\mathcal{G}(n,p)$ before moving on to the main results.

Given a graph $G$ and $k, r > 0$, $(k,r)$-**cluster** (of low-degree vertices) is a maximal set $U \subseteq V(G)$ such that $N(U,r)$ is connected, and where all vertices $v \in U$ satisfy $\deg(v) < k$. For our purposes, the relevant features of a cluster $U$ are the degrees of all vertices $v \in U$, as well as the paths connecting $U$ in $N(U,r)$. We capture this information through the following notion. A **low-degree tree** $(T,U)$ is tree $T$, together with a set of roots $U \subseteq V(T)$, where each leaf $v \in V(T)$ is either adjacent to a root $u \in U$, or is a root itself. The **width** of $(T,U)$ is the minimum radius $r$ for which all vertices $v \in V(T)$ satisfy $d(U,v) \leq r$. We define $\mathrm{LDT}_s(k,r)$ as the set of unlabeled low-degree trees $(T,U)$ with $|U| = s$, where all roots have degree smaller than $k$, and whose width is at most $r$, and $\mathrm{LDT}_s^t(k,r) \subset \mathrm{LDT}_s(k,r)$ as the subset containing those $(T,U)$ with $|E(T)| = t$.

Observe that $\mathrm{LDT}_s(k,r)$ contains only a finite number of low-degree trees $(T,U)$. Indeed, it is easy to see that $T$ is the union of some paths $P_1, \dots, P_\ell$ with both ends in $U$ and length at most $2r+1$, and some additional edges incident to $U$. The paths $P_1, \dots, P_\ell$ must form a tree, so $\ell = |U| - 1 = s - 1$, and the total size of $V(T)$ is at most $(2r+k+1)s$.

Given a graph $G$, and $k, r > 0$, a low-degree tree $(T', U')$ is the **pattern** of a $(k,r)$-cluster $U$ if $N(U,r)$ is a tree, and $(T', U') \simeq (T, U)$, where $T$ is the smallest connected subgraph of $N(U,r)$ containing $N(U,1)$ (i.e., containing $U$ and all edges incident to $U$). Observe that, by definition, $(T,U)$ belongs to $\mathrm{LDT}_s(k,r)$, where $s = |U|$.

**Theorem 4.13** ([62]). *Let $k \in \mathbb{N}$, $r = (3^k - 1)/2$, The following properties hold in $\mathcal{G}(n,p)$ w.h.p.:*

*(1) For all $\ell \leq 2r+1$, there are at least $k$ copies of each $\ell$-cycle.*
*(2) No two cycles of length at most $2r+1$ lie at distance smaller than $2r+1$ from each other.*
*(3) For each $\ell \leq (2r+1)s$, all vertices lying in $\ell$-cycles have degree at least $k+2$. Moreover the $2r+1$-neighbourhoods of all $\ell$-cycles contain no vertex of degree smaller than $k+1$.*
*(4) There are at least $k$ vertices $v_1, \ldots, v_k$ at distance greater than $2r+1$ from each other that satisfy $\deg(u) \geq k+2$ for all $u$ with $d(v_i, u) < r$.*
*(5) For all $s' < s$, all $t' > t$ and all $(T,U)$ belonging to either $\mathrm{LDT}_{s'}(r, k+1)$ or $\mathrm{LDT}_s^{t'}(r, k+1)$, the number of $(k+1, r)$-clusters with pattern $(T,U)$ is at least $k$.*
*(6) There are no $(k+1, r)$-clusters of size greater than $s$.*
*(7) For all $t' < t$ and all $(T,U) \in \mathrm{LDT}_s^{t'}(k+1, r)$, there are no $(k+1, r)$-clusters with pattern $(T,U)$.*

### 4.3.2 Strategies at the Connectivity Threshold

In this subsection we show that w.h.p. the $\equiv_k$-class of $\mathcal{G}(n,p)$ is given by its set of small $(k+1,r)$-clusters. The following two auxiliary results can be easily shown using EF games.

A rooted tree $(T, x)$ is called **perfectly height-balanced** if all its leaves are at the same distance from $x$. The **minimum internal degree** of $(T, x)$ is the minimum number of children of any of its non-leaf vertices.

**Lemma 4.8.** *Fix $k, r > 0$. Let $(T_0, x_0)$, $(T_1, x_1)$ be two perfectly height-balanced rooted trees of height $r$ whose minimum internal degrees exceed $k$. Then $(T_0, x_0) \equiv_k^{\mathrm{dFO}} (T_1, x_1)$.*

**Lemma 4.9.** *Let $G$ be a graph and $k > 0$. For each $v \in V(G)$ let $(H_v^0, x_v)$, $(H_v^1, x_v)$ be two rooted graphs satisfying $(H_v^0, x) \equiv_k^{\mathrm{dFO}} (H_v^1, x)$. Then $G_0 \equiv_k^{\mathrm{dFO}} G_1$ holds, where for $i = 0,1$ $G_i$ stands for the result of attaching for each $v \in V(G)$ the graph $(H_v^i, x_v)$ to $G$ by identifying $x_v$ with $v$.*

**Lemma 4.10.** *Let $G_0, G_1$ be graphs, $k > 0, r = (3^k - 1)/2$. The following facts hold:*

- *Let $U_0 \subseteq V(G_0)$, $U_1 \subseteq V(G_1)$ be $(k+1, r)$-clusters with the same pattern $(T,U)$. Then $N(U_0, r) \equiv_k^{\mathrm{dFO}} N(U_1, r)$.*
- *Suppose both Suppose both $G_0, G_1$ satisfy properties (2), (3) from Theorem 4.13. Let $C_0 \subseteq G_0, C_1 \subseteq G_1$ be cycles of the same length $\ell \leq 2r+1$. Then $N(C_0, r) \equiv_k^{\mathrm{dFO}} N(C_1, r)$.*

*Proof.* We begin with the first item. Let $(T,U)$ be the pattern of both $U_0, U_1$. By the definition of cluster, for both $i = 0,1$, $N(U_i, r)$ is formed by taking a copy of $T$ and attaching to each vertex $v \in V(T)$ a perfectly height-balanced tree $T_v$ with height $h = r - d(v, U)$ and minimum internal degree at least $k+1$. Thus, by Lemma 4.8 and Lemma 4.9, $N(U_0, r) \equiv_k^{\mathrm{dFO}} N(U_1, r)$. The second item is shown similarly: Properties (2) and (3) imply the neighbourhood $N(C_i, r)$ consists of $C_i$

plus a perfectly height-balanced tree $T_v$ attached to each vertex, whose height is $r$ and minimum internal degree is at least $k+1$. By Lemma 4.8 and Lemma 4.9 again, $N(C_0, r) \equiv N(C_1, r)$. □

**Theorem 4.14.** *Fix $k > 0, r = (3^k - 1)/2$. Let $G_0, G_1$ be graphs satisfying properties (1)-(8) from Theorem 4.13. Suppose that for all low-degree trees $(T, U) \in \mathrm{LDT}_s^t(k+1, r)$ the number of $(k+1, r)$-clusters in $G_i$ with pattern $(T, U)$ is the same for $i = 0, 1$ or is at least $k$ in both cases. Then $G_0 \equiv_k G_1$.*

*Proof.* For $i = 0, 1$, let $S_i \subseteq V(G_i)$ be the set of vertices that either lie on a cycle of length at most $2r+1$, or whose degree is at most $k$. We apply Theorem 2.3 to $(G_0, S_0), (G_1, S_1)$. For this we need to show that both conditions of the Theorem are satisfied. First, we see that $N(S_0, r) \equiv_k^{\mathrm{dFO}} N(S_1, r)$ holds. This follows by using the previous lemma and the fact that $G_0, G_i$ have the same number (up to $k$) of $\ell$-cycles for each $\ell \leq 2r+1$ and of $(k+1, r)$-clusters whose pattern is $(T, U)$, for each low-degree tree $(T, U)$. To check the second condition of Theorem 2.3, observe that if $v$ is a vertex in $G_i$ with $d(S_i, v) > 2\ell + 1$, for some $\ell \leq r$, then by properties (2) and (3) from Theorem 4.13 $N(v, \ell)$ is a perfectly-height balanced tree of height $\ell$ and minimum internal degree at least $k+1$. Thus, by Lemma 4.8 the $\equiv_k^{\mathrm{dFO}}$-class of $N(v, \ell)$ is the same for all $v$ with $d(S_i, v) > 2\ell + 1$. In addition to that, property (4) implies that there are at least $k$ vertices with that property in $G_i$ lying at distance greater than $2r+1$ from each other. This shows the second condition of Theorem 2.3 Finally, as $(G_0, S_0), (G_1, S_1)$ fulfill both conditions in Theorem 2.3, $G_0 \equiv_k G_1$, as we wanted to prove. □

### 4.3.3 Proof of the Main Results

Given a graph $G$, the function $\chi_{k,r}^G : \mathrm{LDT}_s^t(k, r) \to [k]$ maps each low-degree tree $(T, U)$ to the number of $(k, r)$-clusters in $G$ with pattern $(T, U)$, or to $k$ if this number is larger. Theorem 4.14, together with Theorem 4.13, show that a.a.s. the $\equiv_k$-type of $\mathcal{G}(n, p)$ is determined by the function $\chi_{k+1,r}^{\mathcal{G}(n,p)}$.

In order to apply a similar proof technique to the one used for Theorem 4.7, our next task is to build finite graphs that imitate the typical behaviour of $\mathcal{G}(n, p)$ from the perspective of FO logic of fixed quantifier depth. This is achieved in the following two definitions.

**Definition 4.7.** A graph $G$ is $(k, r)$-homogeneous if its minimum degree is at least $k$, it contains no $\ell$-cycles for each $\ell \leq 2r+1$ (i.e., its girth is grater than $2r+1$), and its $r$-independence number is at least $k$.

A simple way to convince oneself that $(k, r)$-homogeneous graphs exist for all $(k, r)$ is by noting that, asymptotically, a random $k$-regular graph has girth greater than $2r+1$ with probability bounded away from zero.

**Lemma 4.11.** *Fix $k > 0, r = (3^k - 1)/2$. Let $H_0, H_1$ be two $(k, r)$-homogeneous graphs. Then $H_0 \equiv_k H_1$.*

For the rest of the section $\mathbb{H}_{k,r}$ denotes a fixed $(k, r)$-homogeneous graph, and $x_{k,r} \in V(\mathbb{H}_{k,r})$ is an arbitrary fixed vertex. Given a graph $G$, a set $S \subseteq V(G)$ and a rooted graph $(H, x)$ we write

$G \odot_S^{k+1} (H,x)$ for the result of performing the following operation on each vertex $u \in S$: create $k$ disjoint copies of $(H,x)$ and join them to $u$ through $k$ edges incident to their respective roots. For convenience we shorten $G \odot_{V(G)}^{k+1} (H,x)$ to $G \odot^{k+1} (H,x)$.

**Definition 4.8.** Let $k > 0, r = (3^k - 1)/2$, and $(\mathbb{H}, x) = (\mathbb{H}_{k+1,r}, x_{k+1,r})$. Consider a map $\chi : \mathrm{LDT}_s^t(k+1,r) \to [k]$. The graph $G[\chi]$ is the disjoint union of the following:

- $k$ copies of $\mathbb{H}$.
- For each $\ell \leq 2r+1$, $k$ copies of $C_\ell \odot^{k+1} (\mathbb{H}, x)$, where $C_\ell$ stands for the $\ell$-cycle.
- For all $s' < s$, all $t' > t$, and all $(T,U)$ belonging to either $\mathrm{LDT}_{s'}(k+1,r)$ or $\mathrm{LDT}_s^{t'}(k+1,r)$, $k$ copies of $T \odot_S^{k+1} (\mathbb{H}, x)$, where $S = V(T) \setminus U$.
- for each $(T,U) \in \mathrm{LDT}_s^t(k+1,r)$, $\chi(T,U)$ copies of $T \odot_S^{k+1} (\mathbb{H}, x)$, where $S = V(T) \setminus U$.

We call the subgraph of $G[\chi]$ formed by the components in the first three items, the **almost sure part** of $G[\chi]$. The remaining part of the $G[\chi]$, formed by the components described in the last item, is called its **identifying part**.

It is routine to check that the graph $G[\chi]$ satisfies properties (1)-(7) from Theorem 4.13, independently of the map $\chi$. This, Theorem 4.13 itself, and Theorem 4.14 imply the following corollary.

**Corollary 4.1.** *A.a.s.,* $\mathcal{G}(n,p) \equiv_k G[\chi_{k+1,r}^{\mathcal{G}(n,p)}]$.

**Lemma 4.12.** *Fix* $k > 0, r = (3^k - 1)$. *Let* $\chi_0, \chi_1 : \mathrm{LDT}_s^t(k+1,r)$ *be two functions satisfying* $\chi_0 \leq \chi_1$ *(i.e.,* $\chi_0(T,U) \leq \chi_1(T,U)$ *pointwise), and let* $\varphi \in \mathrm{FO}_g$ *be a sentence with* $\mathrm{qr}(\varphi) = k$ *that is either monotone on finite graphs, or preserved under extensions on finite graphs. Then* $G[\chi_1] \models \varphi$ *implies* $G[\chi_0] \models \varphi$.

*Proof.* Let $(\mathbb{H}, x) = (\mathbb{H}_{k+1,r}, x_{k+1})$, as in the definition of $G[\chi]$. The main idea is that the "excess" components of the form $T \odot_S^{k+1} (\mathbb{H}, x)$ in $G[\chi_1]$ can be made to look like $\mathbb{H}$ by joining some extra copies of $\mathbb{H}$ to $T$. Indeed, given any tree $T$, the construction $\mathbb{T} = T \odot^{k+1} (\mathbb{H}, x)$ yields a $(k+1,r)$-homogeneous graph. Thus, by Lemma 4.11, $\mathbb{T} \equiv_k \mathbb{H}$. One can obtain $\mathbb{T}$ from $T \odot_S^{k+1} (\mathbb{H}, x)$ by joining the roots of $k+1$ fresh copies of $\mathbb{H}$ to $v$, for all vertices $v \in V(T) \setminus S$. With this in mind, we proceed to the proof.

Let us begin with the case where $\varphi$ is preserved under extensions on finite graphs. We define the graph $M$ as the union of $(\chi_1(T,U) - \chi_0(T,U))$ disjoint copies of $T \odot^{k+1} (\mathbb{H}, x)$ to the graph $G[\chi_0]$, for each $(T,U) \in \mathrm{LDT}_s^t(k+1,r)$. Each component of $M$ is $(k+1,r)$-homogeneous and $G[\chi_0]$ already had $k$ components with this property (the ones isomorphic to $\mathbb{H}$), so $G[\chi_0] \equiv_k G'[\chi_0] \cup M$. We claim that $G[\chi_0] \cup M$ extends $G[\chi_1]$. To see this, note the following: (i) The almost sure parts of $G[\chi_0]$ and $G[\chi_1]$ are isomorphic. (ii) For each $(T,U) \in \mathrm{LDT}_s^t(k+1,r)$ the identifying part of $G[\chi_1]$ contains $\chi_1(T,U)$ components of the form $T \odot_{V(T)\setminus U}^{k+1} (\mathbb{H}, x)$. Out of these, $\chi_0(T,U)$ components can be embedded into isomorphic components lying in the identifying part of $G[\chi_0]$. The remaining $\chi_1(T,U) - \chi_0(T,U)$ components can be embedded into components of the form $T \odot^{k+1} (\mathbb{H}, x)$ in $M$. Hence, if $G[\chi_1]$ satisfies $\varphi$, so does $G[\chi_0] \cup M$ by virtue of the second graph being an extension of the first. The fact that $G[\chi_0] \cup M \equiv_k G[\chi_0]$ proves the statement from here.

Now suppose that $\varphi$ is monotone on finite graphs. Let

$$\ell = (k+1)s \sum_{(T,U)\in \mathrm{LDT}^t_s(k+1,r)} \left(\chi_1(T,U) - \chi_0(T,U)\right).$$

Define $N$ as the union of $\ell$ disjoint copies of $\mathbb{H}$. Clearly, $G[\chi_1] \equiv_k G[\chi_1] \cup N$. We claim that $G[\chi_1] \cup N \leq G[\chi_0] \cup M$. Indeed, one can form $G[\chi_0] \cup M$ by adding edges to $G[\chi_1] \cup N$ in the following way: for each $(T,U) \in \mathrm{LDT}^t_s(k+1,r)$, take $\chi_1(T,U) - \chi_0(T,U)$ copies of $T \odot^{k+1}_{V(T)\setminus U} (\mathbb{H}, x)$ lying in $G[\chi_0]$ and connect each of them to $(k+1)|U|$ copies of $\mathbb{H}$ in $N$ to form copies of $T \odot^{k+1} (\mathbb{H}, x)$. Thus, if $\varphi$ holds in $G[\chi_1] \cup N$, it must also hold in $G[\chi_0] \cup M$. This, together with $G[\chi_1] \cup N \equiv_k G[\chi_1]$ and $G[\chi_0] \cup M \equiv_k G[\chi_0]$, proves the result. □

*Proof of Theorem 4.10.* Last lemma showed that if $\varphi \in \mathrm{FO}_g$ is a sentence closed under extensions with $\mathrm{qr}(\varphi) = k > 0$, and $\chi_0 \leq \chi_1$ are two functions from $\mathrm{LDT}^t_s(k+1,r)$ into $[k]$, then $G[\chi_1] \models \varphi$ implies $G[\chi_0] \models \varphi$. In addition to that, it is easily seen that the other implication also holds. Indeed, by definition $G[\chi_1]$ extends $G[\chi_0]$, so $G[\chi_0] \models \varphi$ implies $G[\chi_1] \models \varphi$ too. In particular, this shows that $G[\chi_1]$ satisfies $\varphi$ if and only if $G[\overline{\chi}]$ does so as well, where $\overline{\chi}$ is the identically zero function. Hence, $\varphi$ either holds in all graphs of the form $G[\chi]$ or in none of them. By corollary 4.1, this implies the limit $\lim_{n\to\infty} \Pr(\mathcal{G}(n,p) \models \varphi)$ is either zero or one, as we wanted to prove. □

*Proof of Theorem 4.11.* We proceed similarly to last proof. Let $\varphi \in \mathrm{FO}_g$ be monotone on finite graphs with $\mathrm{qr}(\varphi) = k > 0$, and let $\chi_0 \leq \chi_1$ be two functions from $\mathrm{LDT}^t_s(k+1,r)$ into $[k]$. We claim that $\varphi$ holds in $G[\chi_0]$ if and only if it does in $G[\chi_1]$. The 'if' direction is given by Lemma 4.12. We prove the 'only if' direction now. Let $\ell = |V(G[\chi_1])| - |V(G[\chi_0])|$, and let $G_0'$ be the result of adding $\ell$ isolated vertices to $G[\chi_0]$. As $s > 0$, $G[\chi_0]$ already contains at least $k$ isolated vertices, so $G_0' \equiv_k G[\chi_0]$. As shown in the previous proof $G[\chi_1]$ extends $G[\chi_0]$, implying that $G_0' \leq G[\chi_1]$. Thus $G[\chi_0] \models \varphi$ forces $G[\chi_1] \models \varphi$, as desired. Hence, as in the last proof, we get that $\varphi$ either holds in all graphs $G[\chi]$ or in none of them, so $\lim_{n\to\infty} \Pr(\mathcal{G}(n,p) \models \varphi)$ equals zero or one, by Corollary 4.1. □

*Proof of Theorem 4.12.* Let $\varphi \in \mathrm{FO}_g$ be monotone on finite graphs with $\mathrm{qr}(\varphi) = k > 0$. We assume $k > t$. Otherwise, as $s = 1$, $\mathrm{LDT}^t_s(k+1,r)$ is empty and $\lim_{n\to\infty} \Pr(\mathcal{G}(n,p) \models \varphi)$ is either zero or one by Corollary 4.1, so the result holds. In the case where $s = 1$, $k > t$ the only low-degree $(T,U) \in \mathrm{LDT}^t_s(k+1,r)$ is the one consisting of a root attached to exactly $t$ non-roots. In this situation, a map $\chi : \mathrm{LDT}^t_1(k+1,r) \to [k]$ simply counts (up to $k$) the number of degree-$t$ vertices in $G[\chi]$. Thus we can regard $\chi$ simply as a number in $[k]$. By Lemma 4.12, $G[\chi] \models \varphi$ if and only if $\chi < \ell$ for some $\ell$. Also, observe that $G[\chi]$ has no vertices of degree lesser than $t$. This indicates that, on graphs of the form $G[\chi]$, $\varphi$ is equivalent to the statement $P$ = "There are at most $\ell$ vertices of degree not larger than $t$". By Theorem 4.13, $\mathcal{G}(n,p)$ a.a.s. contains no vertex of degree lesser than $t$, so a.a.s. $P$ holds in $\mathcal{G}(n,p)$ if and only if it does so on $G[\chi^{\mathcal{G}(n,p)}_{k+1,r}]$. This, together with Corollary 4.1, yields that $\varphi$ is a.a.s. equivalent to $P$ on $\mathcal{G}(n,p)$. In order to prove the theorem we just need to show that $P$ can be expressed with a positive sentence. This is accomplished below.

$$\psi = \forall x_1 \dots x_\ell \; \exists y_1 \dots y_{t+1} \left( \bigvee_{1 \le i < j \le \ell} x_i = x_j \right) \vee \left( \bigvee_{i \in [\ell]} \bigwedge_{j \in [t+1]} x_i \sim y_j \right).$$

□

## 4.4 Preservation Theorems on Minor-Closed Classes

A graph $H$ is a **minor** of another graph $G$ if $H$ can be obtained from $G$ by deleting vertices, deleting edges, and identifying adjacent vertices. A class of finite graphs $\mathcal{C}$ is called **minor-closed** if whenever a graph $G$ belongs to $\mathcal{C}$ then all its minors do so as well. A celebrated result by Robertson and Seymour states that any minor-closed class $\mathcal{C}$ can be characterized by a finite list of **excluded minors** $H_1, \dots, H_\ell$, meaning that $G \in \mathcal{C}$ if and only if $H_i$ is not a minor of $G$ for all $1 \le i \le \ell$ [56]. A class of graphs $\mathcal{C}$ is **addable** if (1) $G \in \mathcal{C}$ whenever all of $G$'s components belong to $\mathcal{C}$ as well, and (2) $H \in \mathcal{C}$ whenever $H$ results from adding an edge between two components of some graph $G \in \mathcal{C}$.

We fix an addable minor-closed class $\mathcal{C}$ of graphs for the rest of the section. The **random graph** $\mathcal{G}_n^{\mathcal{C}}$ is chosen uniformly at random among those in $\mathcal{C}$ containing exactly $n$ vertices. Observe that if $\mathcal{C}$ is non-empty, it always contains the graph consisting of $n$ isolated vertices, so $\mathcal{G}_n^{\mathcal{C}}$ is well-defined. The random model $\mathcal{G}_n^{\mathcal{C}}$ was extensively studied in [52, 38]. Logical limit laws in this model were considered in [31], both for monadic second-order logic and first-order logic. Among other results, they show that a $\mathrm{FO}_g$-convergence law holds in $\mathcal{G}_n^{\mathcal{C}}$. Our main theorem this section is that the a.a.s. version of Łoś-Tarski Theorem holds in $\mathcal{G}_n^{\mathcal{C}}$ in the form of a zero-one law.

**Theorem 4.15.** *Let $\varphi \in \mathrm{FO}_g$ be a sentence which is preserved under extensions on finite structures. Then $\Pr(\mathcal{G}_n^{\mathcal{C}} \models \varphi)$ converges to either zero or one.*

In contrast to this result is the fact, proven in [8], that Łoś-Tarski Theorem does not hold on minor-closed classes in general. More concretely, it does not hold on the class of planar graphs which is both addable and minor-closed [68].

We define $\mathrm{Big}_n$ as the largest component inside $\mathcal{G}_n^{\mathcal{C}}$, and $\mathrm{Small}_n$ as the union of all other components. Theorem 4.15 can be proven exploiting a single fact about $\mathrm{Big}_n$'s structure, stated below. Given connected graphs $G, H$, and a vertex $v \in V(G)$, we say that $H$ contains a **pendant copy** of $(G, v)$ if $H$ contains a $G$-copy $G'$ that is connected to the rest of $H$ through a single edge incident to $v' \in V(G')$, the vertex corresponding to $v$.

**Lemma 4.13** ([52])**.** *Let $k \in \mathbb{N}$. Given a connected graph $H \in \mathcal{C}$ and a vertex $v \in V(H)$, w.h.p.* $\mathrm{Big}_n$ *contains at least $k$ pendant copies of $(H, v)$.*

Given rooted graphs $(G, v), (H, u)$, we write $(G, v) \equiv_k (H, u)$ if Duplicator wins the variant of the game $\mathrm{EF}_{k+1}(G, H)$ where the first moves on $G$ and $H$ are forced to be $v$ and $u$. Equivalently, we can interpret rooted graphs $(G, v)$ as relational structures with an adjacency predicate and a constant symbol, representing the root. Under this interpretation the new definition of $(G, v) \equiv_k (H, u)$ coincides with the usual notion of logical equivalence. It follows that the number of $\equiv_k$-classes of rooted graphs is finite. Next results show that the $\equiv_k$-class of $\mathrm{Big}_n$ is determined w.h.p.

**Lemma 4.14** ([31, Theorem 3.1] )**.** *Fix $k \in \mathbb{N}$. Let $(H_1, v_1), \dots, (H_\ell, v_\ell)$ be representatives of all $\equiv_k$-classes of rooted graphs $(H, v)$, where $H$ ranges over all connected graphs in $\mathcal{C}$. Let $(\mathbb{H}_k, \vartheta)$ be the rooted graph where $\mathbb{H}_k \in \mathcal{C}$ is the graph formed by attaching a pendant copy of each $(H_i, v_i)$ to a central vertex $\vartheta$. Let $G \in \mathcal{C}$ be a connected graph containing a pendant copy of $(\mathbb{H}_k, \vartheta)$. Then $G \equiv_k \mathbb{H}_k$.*

Putting together last two lemmas yields the following.

**Corollary 4.2.** *Fix $k \in \mathbb{N}$. Let $\mathbb{H}_k$ be as in last lemma. Then w.h.p.* $\mathrm{Big}_n \equiv_k \mathbb{H}_k$.

Now we are in conditions of proving this section's main result.

*Proof of Theorem 4.15.* Let $k = \mathrm{qr}(\varphi)$, and let $(\mathbb{H}_k, \vartheta)$ be as in Lemma 4.14. We show that, if $\mathbb{H}_k \models \varphi$, w.h.p. $\mathcal{G}_n^{\mathcal{C}} \models \varphi$, and $\mathcal{G}_n^{\mathcal{C}} \not\models \varphi$ w.h.p. otherwise. We begin with the 'if' direction. By last corollary, w.h.p. $\mathcal{G}_n^{\mathcal{C}} \equiv_k \mathrm{Small}_n \cup \mathbb{H}_k$. The graph $\mathrm{Small}_n \cup \mathbb{H}_k$ is clearly and extension of $\mathbb{H}_k$, so $\mathbb{H}_k \models \varphi$ implies that w.h.p. $\mathcal{G}_n^{\mathcal{C}} \models \varphi$. We prove the converse now. Define the random graph $\mathbb{G}_n$ as the result of joining each component of $\mathrm{Small}_n$ to a fresh single vertex through an arbitrary edge, and joining this fresh vertex to $\vartheta$, the central vertex of $\mathbb{H}_k$. Observe that $\mathbb{G}_n$ contains a pendant copy of $(\mathbb{H}_k, \vartheta)$, so by Lemma 4.14, $\mathbb{G}_n \equiv_k \mathbb{H}_k$. Additionally, $\mathbb{G}_n$ is clearly an extension of $\mathrm{Small}_n \cup \mathbb{H}_k$. However, as stated before, w.h.p. $\mathcal{G}_n^{\mathcal{C}} \equiv_k \mathrm{Small}_n \cup \mathbb{H}_k$. Hence, if $\mathbb{H}_k \not\models \varphi$, then w.h.p. $\mathcal{G}_n^{\mathcal{C}} \not\models \varphi$, as we wanted. This completes the proof. $\square$

# Bibliography

[1] Dimitris Achlioptas and Ehud Friedgut. “A Sharp Threshold for K-Colorability”. In: *Random Structures and Algorithms* 14.1 (Jan. 1999).

[2] Miklos Ajtai and Yuri Gurevich. “Monotone versus Positive”. In: *Journal of the ACM* 34.4 (Oct. 1987).

[3] Sanjeev Arora and Boaz Barak. *Computational Complexity: A Modern Approach*. First. Cambridge University Press, Apr. 2009.

[4] Krishna B. Athreya and Peter E. Ney. “Multi-Type Branching Processes”. In: *Branching Processes*. Ed. by Krishna B. Athreya and Peter E. Ney. Die Grundlehren Der Mathematischen Wissenschaften. Berlin, Heidelberg: Springer, 1972.

[5] A. Atserias. “Definability on a Random 3-CNF Formula”. In: *20th Annual IEEE Symposium on Logic in Computer Science (LICS' 05)*. June 2005.

[6] Albert Atserias. “On Sufficient Conditions for Unsatisfiability of Random Formulas”. In: *Journal of the ACM* 51.2 (Mar. 2004).

[7] Albert Atserias and Anuj Dawar. “On Preservation under Homomorphisms and Unions of Conjunctive Queries”. In: ().

[8] Albert Atserias, Anuj Dawar, and Martin Grohe. “Preservation Under Extensions on Well-Behaved Finite Structures”. In: *Automata, Languages and Programming*. Ed. by Luís Caires, Giuseppe F. Italiano, Luís Monteiro, Catuscia Palamidessi, and Moti Yung. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2005.

[9] Edward A Bender and E.Rodney Canfield. “The Asymptotic Number of Labeled Graphs with given Degree Sequences”. In: *Journal of Combinatorial Theory, Series A* 24.3 (May 1978).

[10] Béla Bollobás. *Random Graphs*. Second. Cambridge Studies in Advanced Mathematics. Cambridge: Cambridge University Press, 2001.

[11] A. Braunstein, M. Mezard, and R. Zecchina. “Survey Propagation: An Algorithm for Satisfiability”. In: *Random Structures and Algorithms* 27 (2005).

[12] Yijia Chen and Jörg Flum. "Forbidden Induced Subgraphs and the Łoś–Tarski Theorem". In: *2021 36th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*. June 2021.

[13] Alonzo Church. "A Note on the Entscheidungsproblem". In: *The Journal of Symbolic Logic* 1.1 (Mar. 1936).

[14] V. Chvatal and B. Reed. "Mick Gets Some (the Odds Are on His Side) (Satisfiability)". In: *Proceedings., 33rd Annual Symposium on Foundations of Computer Science*. IEEE Computer Society, Oct. 1992.

[15] Vašek Chvátal and Endre Szemerédi. "Many Hard Examples for Resolution". In: *Journal of the ACM* 35.4 (Oct. 1988).

[16] Vašek Chvátal and Endre Szemerédi. "Many Hard Examples for Resolution". In: *Journal of the ACM* 35.4 (Oct. 1988).

[17] Anuj Dawar. "Finite Model Theory on Tame Classes of Structures". In: *Mathematical Foundations of Computer Science 2007*. Ed. by Luděk Kučera and Antonín Kučera. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2007.

[18] Anuj Dawar. "Homomorphism Preservation on Quasi-Wide Classes". In: *Journal of Computer and System Sciences*. Workshop on Logic, Language, Information and Computation 76.5 (Aug. 2010).

[19] Jian Ding, Allan Sly, and Nike Sun. "Proof of the Satisfiability Conjecture for Large k". In: *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*. STOC '15. New York, NY, USA: Association for Computing Machinery, June 2015.

[20] P. Erdős and A. Renyi. *On the Evolution of Random Graphs*. 1960.

[21] Ronald Fagin. "Probabilities on Finite Models". In: *The Journal of Symbolic Logic* 41.1 (1976).

[22] Uriel Feige. "Relations between Average Case Complexity and Approximation Complexity". In: *Proceedings of the Thiry-Fourth Annual ACM Symposium on Theory of Computing*. STOC '02. New York, NY, USA: Association for Computing Machinery, May 2002.

[23] John Foy and Alan R. Woods. "Probabilities of Sentences about Two Linear Orderings". In: *Feasible Mathematics: A Mathematical Sciences Institute Workshop, Ithaca, New York, June 1989*. Ed. by Samuel R. Buss and Philip J. Scott. Progress in Computer Science and Applied Logic. Boston, MA: Birkhäuser, 1990.

[24] John Franco and Marvin Paull. “Probabilistic Analysis of the Davis Putnam Procedure for Solving the Satisfiability Problem”. In: *Discrete Applied Mathematics* 5.1 (Jan. 1983).

[25] Ehud Friedgut and appendix by Jean Bourgain. “Sharp Thresholds of Graph Properties, and the $k$-Sat Problem”. In: *Journal of the American Mathematical Society* 12.4 (May 1999).

[26] Joel Friedman, Andreas Goerdt, and Michael Krivelevich. “Recognizing More Unsatisfiable Random K-SAT Instances Efficiently”. In: *SIAM Journal on Computing* 35.2 (Jan. 2005).

[27] Alan Frieze and Michal Karonski. *Introduction to Random Graphs*. Cambridge: Cambridge University Press, 2016.

[28] Yu. V. Glebskii, D. I. Kogan, M. I. Liogon’kii, and V. A. Talanov. “Range and Degree of Realizability of Formulas in the Restricted Predicate Calculus”. In: *Cybernetics* 5.2 (Mar. 1969).

[29] Simi Haber, Tal Hershko, and Tobias Müller. “Disproving the Limit Law for Random Geometric Graphs”. In: *Preprint* ().

[30] Simi Haber and Michael Krivelevich. “The Logic of Random Regular Graphs”. In: *Journal of Combinatorics* 1.4 (2010).

[31] Peter Heinig, Tobias Müller, Marc Noy, and Anusch Taraz. “Logical Limit Laws for Minor-Closed Classes of Graphs”. In: *Journal of Combinatorial Theory, Series B* 130 (May 2018).

[32] Wilfrid Hodges. *A Shorter Model Theory*. 1st edition. Cambridge ; New York: Cambridge University Press, Apr. 1997.

[33] Soichi Kakeya. “On the Partial Sums of an Infinite Series”. In: *Tohoku Sci. Rep* 3.4 (1914).

[34] Matt Kaufmann. “A Counterexample to the 0-1 Law for Existential Monadic Second-Order Logic”. In: *CLI Internal Note 32* (1987).

[35] Matt Kaufmann and Saharon Shelah. “On Random Models of Finite Power and Monadic Logic”. In: *Discrete Mathematics* 54.3 (May 1985).

[36] P. Kolaitis and M. Vardi. “The Decision Problem for the Probabilities of Higher-Order Properties”. In: *Proceedings of the Nineteenth Annual ACM Conference on Theory of Computing - STOC ’87*. New York, New York, United States: ACM Press, 1987.

[37] Phokion G. Kolaitis and Moshe Y. Vardi. “0–1 Laws and Decision Problems for Fragments of Second-Order Logic”. In: *Information and Computation*. Special Issue: Selections from 1988 IEEE Symposium on Logic in Computer Science 87.1 (July 1990), pp. 302–338.

[38] Eryk Kopczyński and Anuj Dawar. “Logical Properties of Random Graphs from Small Addable Classes”. In: *Logical Methods in Computer Science* Volume 15, Issue 3 (July 2019).

[39] Denis Kuperberg. “Positive First-order Logic on Words and Graphs”. In: *arXiv:2201.11619* (Feb. 2022).

[40] Alberto Larrauri. “Probabilities of First-Order Sentences on Sparse Random Relational Structures: An Application to Definability on Random CNF Formulas”. In: *Journal of Logic and Computation* 31.2 (Mar. 2021).

[41] Alberto Larrauri, Tobias Müller, and Marc Noy. “Limiting Probabilities of First Order Properties of Random Sparse Graphs and Hypergraphs”. In: *Random Structures & Algorithms* 60.3 (May 2022).

[42] Jean-Marie Le Bars. “The 0–1 Law Fails for Monadic Existential Second-Order Logic on Undirected Graphs”. In: *Information Processing Letters* 77.1 (Jan. 2001).

[43] Leonid Libkin. *Elements of Finite Model Theory*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004.

[44] Tomasz Luczak and Joel Spencer. “When Does the Zero-One Law Hold?” In: *Journal of the American Mathematical Society* (1991).

[45] J.F. Lynch. “Convergence Law for Random Graphs with Specified Degree Sequence”. In: *18th Annual IEEE Symposium of Logic in Computer Science, 2003. Proceedings.* June 2003.

[46] James F Lynch. “Convergence Law for Random Graphs With Specified Degree Sequence”. In: *ACM Transactions on Computational Logic* 6.4 (2005).

[47] James F Lynch. “Convergence Laws for Random Words”. In: *Australasian Journal of Combinatorics* (1993).

[48] James F. Lynch. “Probabilities of First-Order Sentences about Unary Functions”. In: *Transactions of the American Mathematical Society* 287.2 (1985).

[49] James F. Lynch. “Probabilities of Sentences about Very Sparse Random Graphs”. In: *Random Structures & Algorithms* 3.1 (1992).

[50] Roger C. Lyndon. “Properties Preserved under Homomorphism.” In: *Pacific Journal of Mathematics* 9.1 (Jan. 1959).

[51] Gregory L. McColm. "First Order Zero–One Laws for Random Graphs on the Circle". In: *Random Structures & Algorithms* 14.3 (1999).

[52] Colin McDiarmid. "Random Graphs from a Minor-Closed Class". In: *Combinatorics, Probability and Computing* 18.4 (July 2009).

[53] Michael Molloy and Bruce Reed. "A Critical Point for Random Graphs with a given Degree Sequence". In: *Random Structures & Algorithms* 6.2-3 (Mar. 1995).

[54] Rémi Monasson and Riccardo Zecchina. "Statistical Mechanics of the Random K-Satisfiability Model". In: *Physical Review E* 56.2 (Aug. 1997).

[55] J. Nymann and R. Sáenz. "On a Paper of Guthrie and Nymann on Subsums of Infinite Series". In: *Colloquium Mathematicae* 83.1 (2000).

[56] Neil Robertson and P. D. Seymour. "Graph Minors. XX. Wagner's Conjecture". In: *Journal of Combinatorial Theory, Series B*. Special Issue Dedicated to Professor W.T. Tutte 92.2 (Nov. 2004).

[57] Benjamin Rossman. "Homomorphism Preservation Theorems". In: *Journal of the ACM* 55.3 (Aug. 2008).

[58] Nicolau C. Saldanha and Márcio Telles. *Spaces of Completions of Elementary Theories and Convergence Laws for Random Hypergraphs*. Feb. 2016.

[59] Jeanette Schmidt-Pruzan and Eli Shamir. "Component Structure in the Evolution of Random Hypergraphs". In: *Combinatorica* 5.1 (Mar. 1985).

[60] Saharon Shelah and Joel Spencer. "Can You Feel the Double Jump?" In: *Random Structures & Algorithms* 5.1 (1994).

[61] Saharon Shelah and Joel Spencer. "Zero-One Laws for Sparse Random Graphs". In: *Journal of the American Mathematical Society* (1988).

[62] Joel Spencer and Lubos Thoma. *On the Limit Values of Probabilities for the First Order Properties of Graphs*. Tech. rep. In DIMACS Ser. Discrete Math. Theoret. Comput. Sci, 1997.

[63] Joel H Spencer. *The Strange Logic of Random Graphs*. 2001.

[64] A.P. Stolboushkin. "Finitely Monotone Properties". In: *Proceedings of Tenth Annual IEEE Symposium on Logic in Computer Science*. June 1995.

[65] W. W. Tait. "A Counterexample to a Conjecture of Scott and Suppes". In: *The Journal of Symbolic Logic* 24.1 (Mar. 1959).

[66] Alan Turing. "On Computable Numbers, with an Application to the Entscheidungsproblem". In: *Journal of Math* 58 (1936).

[67] Remco van der Hofstad. *Random Graphs and Complex Networks, Vol 2*. Cambridge: Cambridge University Press, 2017.

[68] K. Wagner. “Über eine Eigenschaft der ebenen Komplexe”. In: *Mathematische Annalen* 114.1 (Dec. 1937).