FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Ondřej Pavel Ježil

Pseudofinite structures and limits

Department of Algebra

Supervisor of the master thesis: prof. RNDr. Jan Krajíček Ph.D.
Study programme: Mathematics
Study branch: Mathematics for Information Technologies

Prague 2022

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In .... date ....

Author's signature

I dedicate this thesis to my parents. I also want to thank my supervisor, professor Jan Krajíček, for both his expert knowledge and quick attendance to my numerous questions.

Title: Pseudofinite structures and limits

Author: Ondřej Ježil

Department: Department of Algebra

Supervisor: prof. RNDr. Jan Krajíček Ph.D., Department of Algebra

Abstract: For a class of graph instances of a computational problem we define a limit object, relative to some computationally restricted class of functions. The key method here is forcing with random variables where the sample set is taken as instances of some nonstandard size. We study the general theory of these limits, called in the thesis wide limits, and their connection to classical problems such as finding a large clique or with the combinatorial problems associated with the classes of total **NP** search problems **PPA** and **PPAD**. Our main results are several 0-1 laws associated with these limits and existence of a significantly large clique of the wide limit of all graph consisting of one large clique.

Keywords: pseudofinite structure, forcing, complexity, witnessing, **TFNP**, clique

Contents

Introduction

There exist several logical constructions of limits of classes of finite structures such as the ultraproduct construction and the compactness theorem. The latter was used in [Fag76] to prove the 0-1 law for structures over relational vocabularies.

In combinatorics, there are also several notions of limits of finite graphs. For example the dense graph limit defined for a sequence of graphs $\{G_k\}_{k>0}$ satisfying the condition that

$$t(F, G_n) = \frac{\hom(F, G)}{|G_n|^{|F|}}$$

converges for every fixed connected graph F which provided a framework (see [LS06]) to restate and find new proofs for results in extremal graph theory. For instance Goodman's theorem relating the number of edges to the number of triangles in a graph. There are other notions of limits of sequences of graphs and we recommend to read [NDM13] to the interested reader. Another recent use of limit objects for the results of extremal combinatorics was by Razborov in [Raz07].

These different notions of limits directly or tangentially relate to the concept of pseudofinite structures. For a first order language L we call an L-structure S pseudofinite if it satisfies the theory T_f consisting of all sentences true in all finite L-structures. Of course, the interesting case is when S is itself not finite.

In this thesis we use the concept of pseudofinite structures to define a limit of a family of finite graphs relative to some computationally restricted class of functions F. Instead of studying the density of substructures, we study these wide limits (as we shall call them) both generally and by analyzing concrete examples and tying them with the computational complexity search problems for F. The image to keep in mind is that we take a limit of a class of inputs to a specific problem and the shape of the limits reflects how some computationally restricted viewer may see a generic input.

The key method we use is arithmetical forcing with random variables, developed in [Kra11], which allows us to construct models of (weak) arithmetical theories and by restricting to a language of graphs gives us Boolean valued graphs. In these Boolean valued graphs, witnessing of existential quantifiers corresponds to the ability of F to solve search problems over the class of graphs we are considering.

After recalling important concepts in the Preliminaries chapter we define the wide limit in Chapter 1. In Chapter 2 we consider some examples and build around them a general theory. In chapters 3 and 4 we analyze more complex examples which correspond to the complexity of finding a large clique and to semantic subclasses of **TFNP** respectively.

Preliminaries

In this chapter we recall a few important notions which we will use in the next chapter to define the central construction we study. We do not review the notions formally but always provide a reference for the reader unfamiliar with these topics. Throughout this thesis we assume a basic knowledge of mathematical logic, model theory and measure theory. Important concepts for us are the nonstandard models of true arithmetic, nonstandard analysis and **NP** search problems. We discuss these in more detail in the rest of this chapter.

The ambient model $\mathcal{M}$

Let L_{all} be the language consisting of function symbols for all functions on $\mathbb{N}$, all relations and all constants. We call a model of $\mathrm{Th}_{L_{all}}(\mathbb{N})$ nonstandard if it is not isomorphic to $\mathbb{N}$. Every model of $\mathrm{Th}(\mathbb{N})$ contains an initial segment isomorphic to $\mathbb{N}$ so we can view nonstandard models as those which also contain 'infinite natural numbers', if we assume $\mathcal{M} \models \mathrm{Th}(\mathbb{N})$ contains $\mathbb{N}$ as an initial segment then the elements $\mathcal{M} \setminus \mathbb{N}$ are called nonstandard. We recommend the introduction of [Kay91] for a review of this topic. In the appendix of [Kra11] there is an explict ultraproduct construction of a model $\mathcal{M} \models \mathrm{Th}(\mathbb{N})$ which is $\aleph_1$-saturated.

This $\aleph_1$-saturated model $\mathcal{M}$ is used throughout this thesis and we call it the ambient model of arithmetic. For our applications we only need to know that the model is nonstandard and the following property holds because of the $\aleph_1$-saturation. (Note that we can encode finite sequences and sets in $\mathcal{M}$ which lets us state the property.)

Property. If $\{a_k\}_{k\geq 0}$ is a sequence with elements in $\mathbb{N}$ then there is an element $t \in \mathcal{M} \setminus \mathbb{N}$ and a sequence $\{b_k\}_{k<t} \in \mathcal{M}$ with $a_k = b_k$ for all $k \in \mathbb{N}$.

By overspill in $\mathcal{M}$ if some definable property P holds for a_k with unbounded indexes, then there is also some nonstandard $n < t$ such that b_s satisfies the property P. Moreover, if some definable property P holds for all b_k with k above some k_0 it has to also hold by induction in $\mathcal{M}$ for all nonstandard b_n. These b_n are intuitively the limit elements of the sequence $\{a_k\}_{k\geq 0}$.

Now let us introduce some notation. For $m \in \mathcal{M}$ we denote the set of numbers below m as $\langle m) := \{0, \ldots, m-1\} \in \mathcal{M}$ and $|m|$ as the bit-length of the (possibly-nonstandard) number m. These definitions can be easily made rigorous using the first order definitions of these functions.

Nonstandard analysis

The reader can refer to [Gol14] for more formal treatment of topics discussed in this section including proofs. To use the method of forcing with random variables we need to consider the concept of $\mathcal{M}$-rationals. To define them we start by simply adjoining all negative elements to the semiring $\mathcal{M}$ to obtain the integral domain $\overline{\mathcal{M}}$. $\mathcal{M}$-rationals are then simply the ordered field of fractions $\mathrm{Frac}(\overline{\mathcal{M}})$ which we denote $\mathbb{Q}^{\mathcal{M}}$.

There is a canonical injection $\mathbb{Q} \hookrightarrow \mathbb{Q}^{\mathcal{M}}$ whose image consists exactly of the 'standard fractions'. We call a $q \in \mathbb{Q}^{\mathcal{M}}$ finite if there is a standard k such that $|q| < \frac{k}{1}$, otherwise we call it infinite. We call $q \in \mathbb{Q}^{\mathcal{M}}$ infinitesimal if q^{-1} is infinite (see also Definition 1.2.2). We will use $\mathbb{Q}^{\mathcal{M}}$ in a manner similar to how hyperreal numbers are used as an alternative foundation for the concepts of real analysis. The following is an important result which we use throughout the thesis.

Theorem. There is a surjective function from the ring of finite $\mathcal{M}$-rationals to $\mathbb{R}$ which is a homomorphism of rings and the kernel is the ideal of infinitesimal numbers.

We denote the function $\mathrm{st}(-)$ and call it the **standard part of the $\mathcal{M}$-rational**.

The following result characterizes convergence of a sequence of rational numbers in the language of nonstandard analysis.

Theorem. Let $\{c_k\}_{k\geq 0}$ be a sequence of rational numbers and let $\{\tilde{c}_k\}_{t\geq k}$ be its nonstandard prolongation in $\mathcal{M}$. Then

$$\lim_{k\to\infty} c_k = r \in \mathbb{R}$$

if and only if for any nonstandard $n \leq t$ we have that $\mathrm{st}(\tilde{c}_n) = r$.

We close this section with two inequalities heavily used in the proofs throughout the thesis.

Theorem (Bernoulli's inequality)**.** Let $y \in \mathcal{M}$ and $x \in \mathbb{Q}^{\mathcal{M}}, x \geq -1$, then

$$(1+x)^y \geq 1 + yx.$$

Theorem (Exponential inequality)**.** Let $y \in \mathcal{M}$, $x \in \mathbb{Q}^{\mathcal{M}}, x \geq 0$ and $x \leq y$, then

$$\left(1 - \frac{x}{y}\right)^y \leq e^{-x}.$$

Total NP search problems and polynomial oracle time

Our goal is to tie the properties of the wide limit with some complexity theoretic statements. We will recall several notions used later on.

The class of total **NP** search problems **TFNP**, first defined in [MP91], consist of all relations on binary strings $P(x, y)$ such that: a) There is a polynomial time machine M which, given x, y, can decide whether $P(x, y)$ holds. b) There exists a constant $c > 0$ and for every x there exists at least one y satisfying $|y| \leq |x|^c$ such that $P(x, y)$ holds.

While the definition of the class **TFNP** seems natural, the inner structure looks more arbitrary and the class is generally studied through its semantic subclasses. For example it is conjectured that there is no complete problem for **TFNP** [GP18] for a suitable notion of reducibility.

Various subclasses are defined as all problems reducible to some problem corresponding usually to a combinatorial lemma, for some appropriate definition of reduction. Two main subclasses are relevant for us. The class **PPA**, polynomial parity argument, corresponds to all problems reducible to LEAF, the problem formulated as follows. An instance is given by a number k and a graph G on the vertex set $\langle 2^{|k|})$, presented by a Boolean circuit of polynomial size in $|k|$ computing its neighbourhood function, such that $\deg_G(0) = 1$ and $\forall v : \deg_G(v) \leq 2$. The task is then to find some nonzero v with $\deg_G(v) = 1$. The corresponding combinatorial principle being the handshaking lemma, which assures the problem is total.

The other class **PPAD**, directed polynomial parity argument, with the complete problem SOURCE.OR.SINK is formulated as follows. An instance is given by a number k and a directed graph G on the vertex set $\langle 2^{|k|})$, presented by a Boolean circuit of polynomial size in $|k|$ computing the neighbourhood function, such that the vertex 0 is a source and $\forall v : \deg^+_G(v), \deg^-_G(v) \leq 1$. The task is to find some nonzero vertex v which is a source or a sink. The corresponding combinatorial principle here being the directed version of the handshaking lemma.

So far, we presented what is called 'type 1' problem in [BCE$^+$95]. The ones we are interested in are 'type 2' problems which replace the input graph G with a pair (α, x) consisting of an oracle α describing the neighbourhood function $N_G(-)$ (or both $N^+_G(-)$ and $N^-_G(-)$ in the directed case) on binary string of length at most $|x|$. While the goal to solve these problems remains the same, suddenly the situation is quite different – these type 2 classes can be easily separated from $\mathbf{FP}^2$, the type 2 version of polynomial time functions, by an adversary argument. More importantly we have that **PPAD** is a strict subclass of **PPA** in the type 2 case. Intuitively one can forget the orientation to get the undirected version, but cannot consistently assign orientation to undirected edges of a large graph.

The traditional model of computation is the Turing machines and for the type 2 problems the oracle Turing machine. But to prove separations in the type 2 case, we can abstract the computation of an oracle Turing machine into a decision tree which describes the queries to an oracle. This abstraction is what to keep in mind in the following chapter when we define the class of functions F_{rud}.

1. Forcing with random variables and the wide limit

1.1 Setup

Our goal in this chapter is to provide a definition of a limit of an infinite class of finite graphs in which arbitrarily large graphs occur and the number of graphs in each cardinality tends to infinity. The following definition makes our requirements of such a class of graphs precise. From now on E denotes a binary relational symbol, so that formulas in the theory of graphs are precisely $\{E\}$-sentences.

Definition 1.1.1. Let $\{\mathcal{G}_k\}_{k>0}$ be a sequence of non-empty finite sets of (directed or undirected) finite graphs, i.e. structures of in the first order language $\{E\}$. We call it a **wide sequence** if the following holds.

- There is a strictly increasing sequence of positive whole numbers $\{g_k\}_{k>0}$ such that the underlying set of each $G \in \mathcal{G}_k$ is $\langle g_k \rangle$.
- $\lim_{k\to\infty} |\mathcal{G}_k| = \infty$

By abuse of notation we will denote the wide sequence just $\mathcal{G}_k$.

The second condition guarantees that $\mathcal{G}_n$ is an infinite set for $n > \mathbb{N}$. In fact, the condition is not strictly important to proceed with this chapter, but it more closely describes what sort of sequences we are interested in.

We will sometimes talk somewhat loosely about wide limits of a class of finite graphs $\mathcal{C}$ which abbreviates that we imagine the class to be stratified into levels $\mathcal{C}_k = \{G \in \mathcal{C}; V_G = \langle g_k \rangle\}$ determined by some canonical choice of cardinalities. If this sequence is wide we can proceed as if we started with a wide sequence in the first place. Many interesting classes of graphs form a wide sequence in this sense. For example graphs with exactly one edge, graphs with bounded degree and so on. We will get to explore many examples in depth after we define the wide limit.

The limit is defined by specifying a nonstandard prolongation $\{\tilde{\mathcal{G}}_n\}_{n<t}$ of the original wide sequence and some nonstandard $n < t$. In the next section we will consider a very rich sublanguage of L_{all} which still makes every wide sequence definable in $\mathcal{M}$ and therefore there is a unique nonstandard prolongation which we will denote just $\{\mathcal{G}_t\}_{t\in\mathcal{M}}$ because it is in fact also unbounded.

In the first order case we will define a Boolean valued graph $\lim_F \mathcal{G}_n$ through which we can investigate the first order properties of the limits. We then define its 'arithmetical expansion' $K(\mathcal{G}_n, F)$ which interprets all relational symbols from L_{all}.

We shall be also interested in second order properties of the Boolean valued graph, for example whether it contains of a large clique. For this we expand $\lim_F \mathcal{G}_n$ to $\lim_F^G \mathcal{G}_n$ and $K(\mathcal{G}_n, F)$ to $K(\mathcal{G}_n, F, G)$ with G being second order objects (relations, functions). Here the arithmetical expansion is essential, because in the (second order) language of graphs we can express the mere existence of a clique. Only with an ordering on the vertices we can actually express that such a clique is large. In this case we shall talk about second order limit.

1.2 The first order wide limit

From now on we closely follow Chapter 1 of [Kra11]. Let $\mathcal{M}$ be the $\aleph_1$-saturated model of true arithmetic discussed in the previous chapter and let $\mathcal{G}_k$ be a wide sequence of graphs and $\Omega := \mathcal{G}_n$ for $n \in \mathcal{M} \setminus \mathbb{N}$. One can check that graphs in Ω are all pseudofinite. The model $\mathcal{M}$ treats all its elements (including those which represent sets) as "finite objects" which lets us define uniform probability even on sets which are infinite from the set-theoretical perspective. Unlike in Krajíček's book, we will not define a structure in the arithmetical language L_{all} because for us the families of functions F will have their range restricted to the vertex set $\langle g_n \rangle$ of the pseudofinite member of $\mathcal{G}_n$ and functions from L_{all} could easily generate functions with range outside of this vertex set.

However nothing forbids us to interpret the relations and constants in L_{all} so we define the language L_{rel} to consist precisely of the relational and constant symbols in L_{all}

Definition 1.2.1. Let $\mathcal{A} := \{A \in \mathcal{M}; A \subseteq \Omega\}$ be the set of all subsets of Ω represented by an element in $\mathcal{M}$.

We define the **counting measure** as the uniform probability of A when we sample Ω uniformly, so we have

$$A \in \mathcal{A} \to |A| \,/\, |\Omega| \,,$$

the counting measure takes values in $\mathcal{M}$-rationals.

One can check that $\mathcal{A}$ is a Boolean algebra, but not a σ-algebra as it is not closed under all countable unions. Indeed all singleton sets are part of $\mathcal{A}$ but the set of all elements with standardly many predecessors in Ω is not in $\mathcal{A}$.

Definition 1.2.2. We call an $\mathcal{M}$-rational **infinitesimal** if it is smaller than all standard fractions $\frac{1}{k}, k \in \mathbb{N}$.

Define an ideal in $\mathcal{A}$ as $\mathcal{I} := \{A \in \mathcal{A}; |A| \,/\, |\Omega| \text{ is infinitesimal}\}$. Define the Boolean algebra $\mathcal{B} := \mathcal{A}/\mathcal{I}$. The induced measure on $\mathcal{B}$ is a real-valued measure and can be written as

$$\mu(A/\mathcal{I}) = \mathrm{st}(|A| \,/\, |\Omega|).$$

We can also check, that now μ is a measure in the ordinary sense and that $\mathcal{B}$ is an σ-algebra. In fact, the following key lemma holds.

Lemma 1.2.3. $\mathcal{B}$ is a complete Boolean algebra.

The maximal and minimal element in $\mathcal{B}$ will be denoted **1** and **0** respectively. We now define what we require of the family of functions F we already mentioned.

Definition 1.2.4. Let F be a non-empty set of $\mathcal{M}$-finite function which are elements in $\mathcal{M}$. We call it a **(random) vertex family** if it satisfies the following:

- The domain of any function $\alpha \in F$ is Ω and the range is $\langle g_n \rangle$.

Note that while every $\alpha \in F$ is represented by some element in $\mathcal{M}$, this need not be the case for the whole family F.

Now we can finally define the first order wide limit.

Definition 1.2.5. We define a $\mathcal{B}$-valued $\{E\}$-structure $\lim^F_{k\to n} \mathcal{G}_k$, with universe F and $\{E\}$-sentences being evaluated by the following inductive conditions:

- $[\![\alpha = \beta]\!] := \{\omega \in \Omega; \alpha(\omega) = \beta(\omega)\}/\mathcal{I}$.
- $[\![E(\alpha, \beta)]\!] := \{\omega \in \Omega; E_\omega(\alpha, \beta))\}/\mathcal{I}$.
- $[\![-]\!]$ commutes with $\wedge$,$\vee$,$\neg$.
- $[\![(\exists x)A(x)]\!] := \bigvee_{\alpha\in F} [\![A(\alpha)]\!]$.
- $[\![(\forall x)A(x)]\!] := \bigwedge_{\alpha\in F} [\![A(\alpha)]\!]$.

By abuse of notation we will usually denote the limit $\lim_F \mathcal{G}_n$.

We will also define the structure $K(\mathcal{G}_n, F)$ which is not important in the first order case, but makes the definition of its second order counterpart more manageable. This structure corresponds to a fragment, a substructure of a L_{rel}-reduct to be precise, of structures $K(F)$ defined in [Kra11] for some larger family F.

Definition 1.2.6. $K(\mathcal{G}_n, F)$ will denote a $\mathcal{B}$-valued $L_{rel} \cup \{E\}$-structure defined as an $L_{rel} \cup \{E\}$-expansion of $\lim_F \mathcal{G}_n$. The Boolean evaluations of atomic L_{rel}-sentences are defined as follows:

- $[\![R(\alpha_1, \ldots, \alpha_k)]\!] := \{\omega \in \Omega; R(\alpha_1, \ldots, \alpha_k)\}/\mathcal{I}$ for any k-ary $R \in L_{rel}$.

We call **the arithmetical expansion of the (first order) wide limit**.

1.3 The second order wide limit

While we can find a truth value of a sentence in the language of graphs in the limit $\lim_F \mathcal{G}_n$, we will encounter situations where this is not sufficient to analyze the wide sequence $\{\mathcal{G}_k\}_{k>0}$.

In Chapter 3 we will investigate how the existence of large cliques corresponds to the size of cliques in the limit graph. First we need some way to witness subsets of vertices – this leads us to the second order wide limit. However, in the second order case the arithmetical expansion is much more important because we cannot just measure the set-theoretical cardinality of any such clique. For specific n we could very well have $\mathrm{card}(\langle \lfloor \log n \rfloor \rangle) = \mathrm{card}(\langle \left\lfloor \frac{n}{2} \right\rfloor \rangle)$ but from the point of view of complexity theory, cliques of size $\lfloor \log n \rfloor$ and $\left\lfloor \frac{n}{2} \right\rfloor$ are dramatically different. In other words, our goal is also to have means to count the number elements of subsets or relations with values in (random variables in) $\mathcal{M}$.

When we say second order we mean two-sorted first order structures where one sort represents the usual elements (the 'first order' sort) and the other represents functions on those elements (the 'second order' sort). The 'second order' here can also represent sets and relations by $\{0,1\}$ valued functions.

Definition 1.3.1. We call a set of functions $G \subseteq \mathcal{M}$ an F-compatible **functional family** if every $\Theta \in G$ assigns to every $\omega \in \Omega$ a function $\Theta_\omega \in \mathcal{M}$ and after we define

$$\Theta(\alpha)(\omega) := \begin{cases} \Theta_\omega(\alpha(\omega)) & \alpha(\omega) \in \mathrm{dom}\,(\Theta_\omega) \\ 0 & \text{otherwise,} \end{cases}$$

we have that for every $\alpha \in F$ and $\Theta \in G$ we have $\Theta(\alpha) \in F$.

Definition 1.3.2. Let F be a vertex family and G be an F-compatible functional family. We define the structure $K(\mathcal{G}_n, F, G)$ as a two sorted $\{E\} \cup L_{rel}$-structure with sorts F and G interpreting the first order L_{rel}-sentences as in $K(\mathcal{G}_n, F)$ and treating the sort G as follows. Variables for G are treated as function symbols and can form terms with variables for F. For equality we let

$$[\![\Theta = \Xi]\!] := \{\omega \in \Omega; \Theta_\omega = \Xi_\omega\}/\mathcal{I}$$

and for the second order quantifiers we have the following inductive clauses

- $[\![(\exists X)A(X)]\!] := \bigvee_{\Theta \in G}[\![A(\Theta)]\!]$
- $[\![(\forall X)A(X)]\!] := \bigwedge_{\Theta \in G}[\![A(\Theta)]\!]$.

We define **the second order wide limit** $\lim_{F \to n}^{G} \mathcal{G}_k$ as $\{E\}$-reduct of the structure $K(\mathcal{G}_n, F, G)$, which we analogously call the **the arithmetical expansion of the (second order) wide limit**. By abuse of notation we will mostly denote the wide limit $\lim_F^G \mathcal{G}_n$.

Let us note that if we have multiple Boolean valued structures $\mathcal{S}_1, \mathcal{S}_2, \ldots$ we may add the name of the structure as a prefix to the evaluation function to get $\mathcal{S}_1[\![-]\!]$ and $\mathcal{S}_2[\![-]\!]$ to avoid ambiguity or to emphasize the structure where the evaluation takes place. This is different from the standard notation, which would be including the structure name in the superscript as $[\![\ldots]\!]^{\mathcal{S}_1}$, but it is in our case preferable for typographical reasons.

1.4 The vertex family F_{rud} and G_{rud}

Throughout this thesis we will mostly work with the vertex family F_{rud} which ties the properties of $\lim_F \mathcal{G}_n$ with decision tree complexity – but decision trees can also be seen as an abstraction of oracle machine time we mentioned in the Preliminaries chapter.

After we choose the sequence $\{\mathcal{G}_k\}_{k>0}$ and $n > \mathbb{N}$ we again put $\Omega := \mathcal{G}_n$ and define F_{rud} as follows.

Definition 1.4.1. We define a **decision tree** to be a binary tree $T \in \mathcal{M}$ with a labeling of vertices and edges ℓ. The non-leaf vertices are labeled by pairs of numbers (u, v), where $u, v \in \langle g_n \rangle$ and each edge is labeled either by 1 or 0. Each leaf vertex is then labeled by some element of $\langle g_n \rangle$.

Every sample $\omega \in \Omega$ uniquely determines a path in (T, ℓ) by interpreting the vertex labels as "is $(u, v) \in E_\omega$?" and the edge labels as true (1) and false (0) and the path then uniquely determines an output.

We define $\mathcal{T}_{rud}$ to be the set of all (T,ℓ) of depth at most $g_n^{1/t}$, for some nonstandard $t \in \mathcal{M}$, and F_{rud} to be the set of all function computed by some $(T,\ell) \in \mathcal{T}_{rud}$. For brevity we will leave out the labeling of the trees out of the notation so a tree in $\mathcal{T}_{rud}$ can be denoted just by T.

Note that if we are given a graph $G \in G_k$ on g_k vertices, we need a polynomial sized circuit with $2l$ inputs, $l := \log g_k$, to represent its edge relation. In the pseudofinite case, where $k = n$, if we are restricted to inspect/query at most $g_n^{1/t}$ edges then it corresponds to $2^{l/t}$ many queries or the subexponential oracle time.

The definition of G_{rud} is a bit more involved. The functionals in it will be computed by tuples of elements from F_{rud} in the following sense.

Definition 1.4.2. Let $\hat{\beta} = (\beta_0, \dots, \beta_{m-1}) \in \mathcal{M}$ be a m-tuple of elements in F_{rud}, for any $\alpha \in F_{rud}$ and $\omega \in \Omega$ we define

$$\hat{\beta}(\omega) = \begin{cases} \beta_{\alpha(\omega)}(\omega) & \alpha(\omega) < m \\ 0 & \text{otherwise.} \end{cases}$$

Definition 1.4.3. The family G_{rud} consists of all functionals Θ such that there is $m \in \mathcal{M}$ and some $\hat{\beta} = (\beta_0, \dots, \beta_{m-1}) \in \mathcal{M}$ that computes it.

Lemma 1.4.4. *G_{rud} is (F_{rud})-compatible.*

Proof. By induction in $\mathcal{M}$ we have that all the depth of all the trees is bounded by $g_n^{1/t}$ for some $t > \mathbb{N}$.

If we take some $\Theta \in G_{rud}$ and $\alpha \in F_{rud}$ we can compute $\Theta(\alpha)$ also by a tree in $\mathcal{T}_{rud}$ by concatenating the trees computing α and β_is. $\square$

1.5 Different choices of n

Even though we generally pose no requirements on $n > \mathbb{N}$ there are examples of wide sequences for which the limit is sensitive on the choice of the non-standard number n.

Example 1.5.1. Let

$$G_k := \begin{cases} \{(\langle k \rangle, E); |E| = 2, E(0,1)\} & k \text{ even} \\ \{(\langle k \rangle, E); |E| = 1, \neg E(0,1)\} & k \text{ odd.} \end{cases}$$

Let $n > \mathbb{N}$ then

$$\lim_{F_{rud}} \mathcal{G}_{2n+1}[\![E(0,1)]\!] = \mathbf{0}, \tag{1.1}$$

but

$$\lim_{F_{rud}} \mathcal{G}_{2n}[\![E(0,1)]\!] = \mathbf{1}. \tag{1.2}$$

Even though the concrete wide limits we will investigate in the following chapters do not depend on the specific n, it is important to note that we cannot generally remove the parameter n from the definition of the limit.

1.6 Theories of wide limits

If $\lim_F \mathcal{G}_n$ is the first order wide limit we will be interested in which exact $\{E\}$-sentences are valid in it. By **valid** we mean that their $[\![\dots]\!]$ value is $\mathbf{1}$.

Definition 1.6.1. We define $\mathrm{Th}(\lim_F \mathcal{G}_n)$ as the set of all valid $\{E\}$-sentences in $\lim_F \mathcal{G}_n$.

In the next chapter (Theorem 2.3.4) we will see that if a universal $\{E\}$-sentence φ holds for all $G \in \mathcal{G}_k$ for k big enough then $\lim_F \mathcal{G}_n[\![\varphi]\!] = \mathbf{1}$. In particular, a wide limit of undirected graphs is a Boolean valued undirected graph and a wide limit of directed graphs is a Boolean valued directed graph.

Lastly, let us recall the concept of 0-1 laws which say that a certain probability tends either to 0 or 1 and not to any intermediate value. Here, instead of probability, we can think about the Boolean values $\mathbf{0}$ and $\mathbf{1}$ and ask when does it happen that $\lim_F \mathcal{G}_n[\![\varphi]\!] \in \{\mathbf{0}, \mathbf{1}\}$ for every $\{E\}$-sentence φ. This is exactly equivalent to the situation where the theory $\mathrm{Th}(\lim_F \mathcal{G}_n)$ is complete. Later in the thesis we prove such 0-1 law for several wide limits.

2. General theory

2.1 $\mathcal{G}_k = \mathbf{EDGE}_k$

The first and simplest wide sequence we will consider of is the class of undirected graphs with exactly one edge. We put

$$\mathrm{EDGE}_k := \{(\langle k \rangle, E); |E| = 1\},$$

and one can check that this is indeed a wide sequence because $|\mathrm{EDGE}_k| = \binom{k}{2}$ which tends to infinity as k does.

Now every graph $G \in \mathrm{EDGE}_k$ has an edge. How does this reflect in the F_{rud} wide limit? The naive guess could be that there should also be an edge but another view lets us see this should not be the case. The limit encompasses the procedure of randomly sampling an exponentially large graph and then inspecting a subexponential part of its edges. In the pseudofinite case this means that we search for one edge in the nonstandard set of vertices but only search through infinitesimally small portion and so we should almost always fail.

This is exactly how we prove in the following theorem.

Theorem 2.1.1.

$$\lim_{F_{rud}} \mathrm{EDGE}_n[\![(\exists x)(\exists y)E(x,y)]\!] = \mathbf{0}.$$

Proof. Any two trees T_α and T_β computing potential witnesses α, β of the formula $E(x,y)$ on some subset of Ω can be combined then into one tree that outputs an edge on the same subset of Ω, so we can just analyze the case where the witnesses are computed by the same tree. Let $T \in \mathcal{T}_{rud}$ be a tree of depth $n^{1/t}$, for some $t > \mathbb{N}$, that outputs a pair of vertices.

Start from the root of T and always choose the path that corresponds to the nonexistence of an edge. At the end we obtain some answer, that gives us a set of at most $2 \cdot n^{1/t} + 2$ vertices the tree inspected or outputted. Now we can find at least:

$$\binom{n - 2n^{1/t} - 2}{2} = \frac{(n - 2n^{1/t} - 2)(n - 2n^{1/t} - 3)}{2} \tag{2.1}$$

different $\omega \in \Omega$ such that $T(\omega)$ is not an edge in ω.

The probability that any of those graphs is sampled is

$$\begin{aligned}
\frac{\binom{n-2n^{1/t}-2}{2}}{|\mathcal{G}_k|} &= \frac{(n - 2n^{1/t} - 2)(n - 2n^{1/t} - 3)}{n(n-1)} && (2.2)\\
&= \left(1 - \frac{2n^{1/t} + 2}{n}\right) \cdot \left(1 - \frac{2n^{1/t} + 2}{n-1}\right) && (2.3)\\
&\geq \left(1 - \frac{2n^{1/t} + 2}{n-1}\right)^2 && (2.4)\\
&\geq 1 - \frac{4n^{1/t} + 4}{n-1}. && (2.5)
\end{aligned}$$

And one can clearly see that $\mathrm{st}(1 - \frac{4n^{1/t}+4}{n-1}) = 1$. This proves that the Boolean value we are considering is $\mathbf{0}$. $\square$

2.2 Sparse $\mathcal{G}_k$

One can see that in Theorem 2.1.1 the exact shape of graphs in $\mathcal{G}_k$ does not play a crucial role. If $\mathcal{G}_k$ consisted of all graphs on k vertices containing say exactly one triangle, or any other fixed subgraph of constant size, and no other edges, we would still find that the non-existence is valid in the F_{rud} wide limit.

A more general case would be to consider a family of graphs in which there is an infinitesimally small chance that two independent uniformly random vertices have an edge between. However, this is not sufficient.

Example 2.2.1. Let $\mathcal{G}_k$ consist of those graphs on the vertex-set k which contains the edge $E(0,1)$ and then has exactly one other edge. As k increases, the fraction of edges gets smaller than any standard positive fraction. But

$$[\![(\exists x)(\exists y)E(x,y)]\!] = \mathbf{1},$$

as witnessed by x being the constant 0 and y the constant 1 both of which are computable by a $\mathcal{T}_{rud}$ tree of depth 0.

One can see that having distinguished vertices can result in the edge being found even in the sparse case. We want to distinguish from this situation by considering the sequences $\mathcal{G}_k$ satisfying the following natural definition.

Definition 2.2.2. We say that $\{\mathcal{G}_k\}_{k=0}^{\infty}$ is **isomorphism closed**, if there is k_0 such that for every $k > k_0$ if we have that $G_1 \in \mathcal{G}_k$, $V_{G_2} = \langle g_k \rangle$ and $G_1 \cong G_2$ then $G_2 \in \mathcal{G}_k$.

Theorem 2.2.3. Let an isomorphism closed $\mathcal{G}_k$ have the following property. There is a sequence $\{b_k\}_k$ and for big enough k, a uniformly sampled 2-element $\{u, v\} \subseteq g_k$ and every $G \in \mathcal{G}_k$ we have

$$\Pr[E_G(u,v)] \le b_k,$$

and some k_0 such that $\lim_{k\to\infty} k^{1/k_0} b_k = 0$. Then

$$\lim_{F_{rud}} [\![(\exists x)(\exists y)E(x,y)]\!] = \mathbf{0}.$$

Proof. Let us define the number $c_{u,v} := |\{G \in \mathcal{G}_k; E_G(u,v)\}|$, which is the number of graphs G in $\mathcal{G}_k$ satisfying $E_G(u,v)$. Of course $c_{u,u} = 0$ for every u.

Claim: Let $u \neq v, u' \neq v'$ be vertices, then $c_{u,v} = c_{u',v'}$.
proof of claim: Let $\rho := (u\,u')(v\,v')$ be a permutation with cycles $(u\,u')$ and $(v\,v')$. We can let ρ act on $\mathcal{G}_k$ by sending G to a graph $\rho(G)$ which renames the edges coordinate-wise.

The fact that $\mathcal{G}_k$ is isomorphism closed implies that ρ restricts to a bijection:

$$\rho' : \{G \in \mathcal{G}_k; E_G(u,v), \neg E_G(u',v')\} \to \{G \in \mathcal{G}_k; E_G(u',v'), \neg E_G(u,v)\}$$

which proves the claim. $\square$

Now we define a matrix with entries

$$a_{G,\{u,v\}} := \begin{cases} 1 & E_G(u,v) \\ 0 & \text{otherwise} \end{cases}$$

where the rows are indexed by one of $|\mathcal{G}_k|$ many graphs in $\mathcal{G}_k$ and the columns are indexed by the $\frac{k(k-1)}{2}$ many 2-element sets of numbers in k. We take any distinct vertices u, v and define $p := \Pr_{G \in \mathcal{G}_k}[E_G(u,v)] = \frac{c_{u,v}}{|\mathcal{G}_k|}$, by the claim the choice of u, v does not matter.

The assumption from the statement is equivalent to the equality

$$\sum_{\{u,v\}} a_{G,\{u,v\}} \leq \frac{k(k-1)}{2} b_k$$

for every G. We combine this with the claim and the definition of p to get

$$\begin{aligned}
\frac{k(k-1)}{2} |\mathcal{G}_k| \, p &= \sum_{\{u,v\}} \sum_{G \in \mathcal{G}_k} a_{G,\{u,v\}} && (2.6) \\
&= \sum_{G \in \mathcal{G}_k} \sum_{\{u,v\}} a_{G,\{u,v\}} && (2.7) \\
&\leq |\mathcal{G}_k| \frac{k(k-1)}{2} b_k && (2.8)
\end{aligned}$$

which implies

$$p \leq b_k.$$

Now let $k := n$ and let $T \in F_{tree}$ be a tree of depth $n^{1/t}$ for some $t > \mathbb{N}$, where every leaf of T is labeled by some edge. Walk down the tree T from the root by answering negatively to every edge, which gives us a set E_T of all edges T inspected or outputted and $|E_T| \leq n^{1/t} + 1$.

Now we just need to prove that the probability that T finds an edge is infinitesimally small. This is enough to prove the theorem, since the trees computing any two witnesses for x and y in the statement can be combined to construct T and if any tree T succeeds with only infinitesimally small probability, no random vertices can witness an edge on a set of non-zero measure.

We use the fact that $p \leq b_n$ to derive

$$\begin{aligned}
\Pr_{G \in \mathcal{G}_n}[T \text{ finds an edge}] &\leq \sum_{\{u,v\} \in E_T} \Pr_{G \in \mathcal{G}_n}[E_G(u,v)] && (2.9) \\
&= \sum_{\{u,v\} \in E_T} \frac{c_{u,v}}{|\mathcal{G}_n|} && (2.10) \\
&\leq \sum_{\{u,v\} \in E_T} p && (2.11) \\
&= (n^{1/t} + 1) p && (2.12) \\
&\leq (n^{1/t} + 1) b_k && (2.13) \\
&\leq n^{1/k_0} b_k && (2.14) \\
&\approx 0, && (2.15)
\end{aligned}$$

which proves the theorem. □

The assumption $\lim_{k \to \infty} k^{1/k_0} b_k = 0$ for some k_0 may seem unintuitive at first. However, it precisely captures what is "too sparse" for the trees in $\mathcal{T}_{rud}$. The following example shows that without the assumption the theorem fails.

Example 2.2.4. Let $\mathcal{G}_k$ consist of all graphs on the vertex set $\langle k\rangle$ with precisely $\lceil\frac{k(k-1)}{2\log k}\rceil$ edges.

Then we claim that

$$[\![(\exists x)(\exists y)E(x,y)]\!] = \mathbf{1}.$$

Let α and β be vertices computed by the tree of the same shape which inspects a set of any $n^{1/t}$ distinct edges for some $t > \mathbb{N}$. If it finds an edge we define α and β in any way so that they are the distinct vertices incidental with this edge. Otherwise, we let $\alpha(\omega) = \beta(\omega) = 0$.

Let T be a tree of the same shape, that computes the pair $\{\alpha, \beta\}$ then we can compute the probability where such a tree fails as the fraction of all graphs which have no edges that T inspects. Let $m = \binom{n}{2}$. We get

$$\Pr_{G\in\mathbb{G}_n}[T \text{ fails}] = \frac{\binom{m-n^{1/t}}{\lceil\frac{n(n-1)}{2\log n}\rceil}}{\binom{m}{\lceil\frac{n(n-1)}{2\log n}\rceil}} \tag{2.16}$$

$$= \frac{\frac{(m-n^{1/t})!}{\lceil\frac{n(n-1)}{2\log n}\rceil!(m-\lceil\frac{n(n-1)}{2\log n}\rceil-n^{1/t})!}}{\frac{m!}{\lceil\frac{n(n-1)}{2\log n}\rceil!(m-\lceil\frac{n(n-1)}{2\log n}\rceil)!}} \tag{2.17}$$

$$= \frac{(m-n^{1/t})!(m-\lceil\frac{n(n-1)}{2\log n}\rceil)!}{m!(m-\lceil\frac{n(n-1)}{2\log n}\rceil-n^{1/t})!} \tag{2.18}$$

$$= \prod_{i=0}^{n^{1/t}-1}\frac{m-\lceil\frac{n(n-1)}{2}\rceil-i}{m-i} \tag{2.19}$$

$$\leq \left(1-\frac{\lceil\frac{n(n-1)}{2\log n}\rceil}{\frac{n(n-1)}{2}}\right)^{n^{1/t}} \tag{2.20}$$

$$\leq \left(1-\frac{\lceil\frac{n(n-1)}{2\log n}\rceil}{\frac{n(n-1)}{2}}\right)^{n^{1/t}} \tag{2.21}$$

$$\leq \left(1-\frac{1}{\log n}\right)^{n^{1/t}} \tag{2.22}$$

And for any standard k we have

$$\left(1-\frac{1}{\log n}\right)^{n^{1/t}} \leq \left(1-\frac{1}{\log n}\right)^{k\cdot\log n} \tag{2.23}$$

$$\leq (e^{-\frac{1}{\ln 2}})^k. \tag{2.24}$$

So $\mathrm{st}(\Pr_{G\in\mathcal{G}_n}[T \text{ fails}]) = 0$ and we get

$$\mu([\![(\exists x)(\exists y)E(x,y))]\!] \geq \mu([\![E(\alpha,\beta)]\!]) \tag{2.25}$$

$$= \mathrm{st}(1-\Pr_{G\in\mathbb{G}_n}[T \text{ fails}]) \tag{2.26}$$

$$= 1. \tag{2.27}$$

2.3 Dense $\mathcal{G}_k$

Let us now consider how the density of a specific kind of substructure in the wide sequence corresponds to that substructure existing in the wide limit. The following theorem presents a sufficient condition for nonexistance to be invalid. Furthermore, this condition holds generally for all F which contain all constants.

Theorem 2.3.1. Let F contain all constants, $\mathcal{G}_k$ be a wide sequence and let $\varphi_0(\overline{x})$ be an open $\{E\}$-formula such that

$$\lim_{k\to\infty} \Pr_{\substack{G\in\mathcal{G}_k \\ \overline{a}\in\langle g_k\rangle^l}} (G \models \varphi_0(\overline{a})) \geq p.$$

Then $\mu(\llbracket(\exists x)\varphi_0(x)\rrbracket) \geq p$.

Proof. We define a matrix with components

$$C_{G,\overline{a}} = \begin{cases} 1 & G \models \varphi_0(\overline{a}) \\ 0 & \text{otherwise.} \end{cases}$$

By overspill in $\mathcal{M}$ we have that

$$\mathrm{st}\left(\frac{1}{|\mathcal{G}_n|\, g_k^l} \sum_{G\in\mathcal{G}_n} \sum_{\overline{a}\in\langle g_k\rangle^l} C_{G,\overline{\alpha}}\right) \geq p.$$

We claim that there is one $\overline{a}$ such that $\mathrm{st}(\Pr_{G\in\mathcal{G}_k}(G \models \varphi_0(\overline{a}))) \geq p$. Assume for contradiction, that for all $\overline{a}$ we have $\frac{1}{|\mathcal{G}_n|}\sum_{G\in\mathcal{G}_n} < p$. Then

$$\begin{aligned} \frac{1}{|\mathcal{G}_n|\, g_k^l} \sum_{G\in\mathcal{G}_n} \sum_{\overline{a}\in\langle g_n\rangle^l} C_{G,\overline{\alpha}} &= \frac{1}{g_k^l} \sum_{\overline{a}\in\langle g_n\rangle^l} \frac{1}{|\mathcal{G}_n|} \sum_{G\in\mathcal{G}_n} C_{G,\overline{\alpha}} && (2.28) \\ &< p, && (2.29) \end{aligned}$$

which is a contradiction after taking the standard part of each value.

Therefore there is a tuple $\overline{a}$ such that $\mu(\llbracket\varphi_0(\overline{\alpha})\rrbracket) \geq p$, let $\gamma_{\overline{a}}$ be the constant function $\omega \mapsto \overline{\alpha}$ in F and

$$\begin{aligned} \llbracket\varphi\rrbracket &= \bigvee_{\overline{\alpha}} \llbracket\varphi_0(\overline{\alpha})\rrbracket && (2.30) \\ &\geq \llbracket\varphi_0(\overline{\gamma}_{\overline{a}})\rrbracket. && (2.31) \end{aligned}$$

By taking μ of both sides we prove the theorem. $\square$

Example 2.3.2. Recall Example 2.2.4 it is not hard to notice that for a $\mathcal{G}_k$ which consists of graphs with exactly $\left\lceil \frac{k(k-1)}{2\log k} \right\rceil$ edges we have

$$\lim_{k\to\infty} \Pr_{G\in\mathcal{G}_k (u,v)\in\langle k\rangle^2} [\neg E(u,v)] = 1,$$

by the theorem it follows that we have

$$\lim_F \mathcal{G}_n \llbracket(\exists x)(\exists y)\neg E(x,y)\rrbracket = \mathbf{1}.$$

We follow with an application of the theorem on a more complex wide sequence.

Example 2.3.3. Consider

$$\mathrm{SK}_k^{1/2} := \{(\langle k\rangle, E); E \text{ has a clique of size } \lfloor k/2\rfloor, |E| = \left|E_{K_{\lfloor k/2\rfloor}}\right|\},$$

a wide sequence of all graphs with exactly one $\lfloor k/2\rfloor$ clique and no other edges. One can check that for any $\{E\}$-formula $\varphi_l(\overline{x})$ stating that $\overline{x}$ forms a clique of size l we have that

$$\lim_{k\to\infty} \Pr_{\substack{G\in\mathrm{SK}_k^{1/2}\\ \overline{a}\in\langle k\rangle^l}}[G \models \varphi(\overline{a})] \geq (1/2)^l.$$

First notice that we can compute the probability for fixed $\overline{a}$ because $\mathrm{SK}_k^{1/2}$ is isomorphism closed. So we have

$$\Pr_{G\in\mathrm{SK}_k^{1/2}}[G \models \varphi_l(\overline{a})] = \frac{\binom{k-l}{\lfloor k/2\rfloor - l}}{\binom{k}{\lfloor k/2\rfloor}} \tag{2.32}$$

$$= \prod_{i=0}^{l} \frac{k - \lfloor k/2\rfloor - i}{k-i} \tag{2.33}$$

$$= \prod_{i=0}^{l} \left(1 - \frac{\lfloor k/2\rfloor}{k-i}\right) \tag{2.34}$$

$$\geq \prod_{i=0}^{l} \left(1 - \frac{k/2}{k-i}\right) \tag{2.35}$$

$$\geq \left(1 - \frac{k/2}{k-l}\right)^l \tag{2.36}$$

$$\geq \left(1 - \frac{1}{2(1-l/k)}\right)^l, \tag{2.37}$$

and since $l \in \mathbb{N}$, so from the point of view of standard analysis a constant, we can just check that

$$\lim_{k\to\infty} 1 - l/k = 1. \tag{2.38}$$

This proves that for any F that contains all constants we have

$$\lim_F \mathrm{SK}_n^{1/2}[\![(\exists \overline{x})\varphi_l(\overline{x})]\!] > \mathbf{0}.$$

Now the following theorem describes sufficient conditions for a universal sentence to hold in the wide limit for *any* F.

Theorem 2.3.4. Let F be any vertex family, $\mathcal{G}_k$ a wide sequence and let $\varphi_0(\overline{x})$ be an open $\{E\}$-formula, such that

$$\lim_{k\to\infty} \Pr_{G\in\mathcal{G}_k}[G \models (\forall\overline{x})\varphi_0(\overline{x})] = 1.$$

Then $\lim_F \mathcal{G}_n[\![(\forall\overline{x})\varphi_0(\overline{x})]\!] = \mathbf{1}$.

Proof. We have that $\mathrm{st}(\mathrm{Pr}_{G\in\mathcal{G}_n}[G \models (\forall\overline{x})\varphi_0(\overline{x})]) = 1$ and therefore $[\![\varphi_0(\overline{\alpha})]\!] = \mathbf{1}$ for each tuple $\overline{\alpha}$ in F. Therefore

$$\begin{aligned}
[\![\varphi]\!] &= \bigwedge_{\overline{\alpha}} [\![\varphi_0(\overline{\alpha})]\!] &\quad (2.39)\\
&= \bigwedge_{\overline{\alpha}} \mathbf{1} &\quad (2.40)\\
&= \mathbf{1}. &\quad (2.41)
\end{aligned}$$

□

Example 2.3.5. Let us define

$$\begin{aligned}
\mathcal{G}_k^A &:= \{(\langle k\rangle, E); |E| = 2\}, &\quad (2.42)\\
\mathcal{G}_k^B &:= \{(\langle k\rangle, E); |E| = (k\cdot(k-1)/2) - 3\}, &\quad (2.43)\\
\mathcal{G}_k &:= \mathcal{G}_k^A \cup \mathcal{G}_k^B, &\quad (2.44)
\end{aligned}$$

and let

$$\varphi_0(a,b,c,d) := \left(\bigwedge_{\substack{x,y\in\{a,b,c,d\}\\ x,y \text{ distinct}}} (x \neq y) \right) \to \left(\bigvee_{\substack{x,y\in\{a,b,c,d\}\\ x,y \text{ distinct}}} E(x,y) \right),$$

which says that if a,b,c,d are distinct, there is an edge between one of them. The universal closure is valid on all graphs in $\mathcal{G}_k^B$ and none of the graphs in $\mathcal{G}_k^A$. Since $\binom{\binom{n}{2}}{2}/\binom{\binom{n}{2}}{n-3}$ is infinitesimal we have, that $\lim_{k\to\infty}\binom{\binom{k}{2}}{2}/\binom{\binom{k}{2}}{k-3} = 0$ we have by Theorem 2.3.4 that $\lim_F \mathcal{G}_n[\![(\forall a,b,c,d)\varphi_0(a,b,c,d)]\!] = \mathbf{1}$ for any F containing all constants.

The Theorem 2.3.4 also implies that since all $G \in \mathcal{G}_k$ satisfy that E_G is antireflexive then so does $\lim_F \mathcal{G}_n$ and if E_G are all symmetrical then so it the wide limit edge relation E. This combined with Theorem 2.2.3 proves the following.

Corollary 2.3.6 (0-1 law for too sparse F_{rud} sequence)**.** Let $\mathcal{G}_k$ be any wide sequence satisfying the statement of the Theorem 2.2.3. Then $\mathrm{Th}(\lim_{F_{rud}} \mathcal{G}_n)$ is the theory of the empty graph and therefore complete.

It is natural to ask whether we can weaken the assumption of Theorem 2.3.4 to an assumption analogous to Theorem 2.3.1. In other words, is

$$\lim_{k\to\infty} \Pr_{\substack{G\in\mathcal{G}_k\\ \overline{a}\in\langle g_k\rangle^l}} [G \models \varphi_0(\overline{a})] = 1$$

enough to imply $\lim_F \mathcal{G}_n[\![(\forall x)\varphi_0(\overline{x})]\!] = \mathbf{1}$? Unfortunately no, as we can see in the following example.

Example 2.3.7. Recall Example 2.2.4 where $\mathcal{G}_k$ consists of all graphs on $\langle k\rangle$ with exactly $\left\lceil \frac{k(k-2)}{2\log k} \right\rceil$ edges. One can easily check that

$$\lim_{k\to\infty} \Pr_{\substack{G\in\mathcal{G}_k\\ u,v\in\langle k\rangle}} [G \models \neg E(u,v)] = 1,$$

but we proved that $\lim_{F_{rud}} \mathcal{G}_n[\![(\exists x)(\exists y)E(x,y)]\!] = \mathbf{1}$ in other words we have

$$\lim_{F_{rud}} [\![(\forall x)(\forall y)\neg E(x,y)]\!] = \mathbf{0}.$$

2.4 $\mathcal{G}_k = \mathbf{ALL}_k$

Now we return to $F = F_{rud}$ and prove a theorem with a more limited use which however forces the truth value of the existential sentence in the wide limit to be $\mathbf{1}$.

Theorem 2.4.1. Let $F = F_{rud}$ and let $\varphi_0(x_0, \dots, x_{l-1})$ be an open $\{E\}$-formula. Furthermore for $0 < p \leq 1$, consider subsets $A \subseteq \langle g_k \rangle^l$ with the property that for all for all $\overline{a} \in A$ we have

$$\Pr_{G \in \mathcal{G}_k} (G \models \varphi_0(\overline{a})) \geq p$$

and

$$\{\{G \models \varphi_0(\overline{a})\} \subseteq \mathcal{G}_k; \overline{a} \in A\} \text{ are mutually independent.}$$

Moreover let A_k be the set with the largest cardinality that has this property.

If $\lim_{k \to \infty} |A_k| = \infty$, then $[\![(\exists \overline{x})\varphi_0(\overline{x})]\!] = \mathbf{1}$.

Proof. Let $\overline{x} = (x_0, \dots, x_{l-1})$. Let $T_{\overline{a}}$ be a tree of some standard depth d, that tests whether $G \models \varphi_0(\overline{a})$.

By overspill in $\mathcal{M}$ we have $n' > \mathbb{N}$ many tuples $\overline{a}_0, \dots, \overline{a}_{n'-1} \in A_n$, such that $\Pr_{G \in \mathcal{G}_k}(G \models \varphi_0(\overline{a}_i)) \geq p$, we can assume $n' < n^{1/t_0}$ for some $t_0 > \mathbb{N}$.

For $j \in [l]$ construct a tree T_j inductively as follows: Start with $T_{\overline{a}_0}$. Replace the label of every accepting leaf by $(\overline{a}_0)_j$ and remove the label of every rejecting leaf. Call this tree T_j^0. Assume we have already constructed T_j^m. Construct T_j^{i+1} by appending $T_{\overline{a}_{m+1}}$ to every undefined leaf, relabeling every satisfied leaf to $(\overline{a}_{i+1})_j$ and removing labels from every rejecting leaf. We will define T_j as $T_j^{n'}$ with undefined leaves labeled by 0. (This can be done because all instances of induction are in $\mathrm{Th}(\mathbb{N})$.) Note that $\mathrm{dp}(T_j) = d \cdot n' < n^{1/t}$ for some $t > \mathbb{N}$.

Call $\overline{\alpha}$ the tuple computed by $T_0, \dots, T_{l-1}$. We will prove that the probability of $\overline{\alpha}$ being a witness to $\varphi_0(\overline{x})$ is 1. For each $\overline{a}_i$ we have, that the probability of $G \models \varphi_0(\overline{a}_i)$ is at least p. The mutual independence of $\{G \models \varphi_0(\overline{a}_i); i \in [n']\}$ and the construction of T_j implies that T_j has a probability of $(1-p)^{n'}$ of failing, which is obviously infinitesimal. $\square$

We now use this theorem to characterize the theory of another wide limit. We denote

$$\mathrm{ALL}_k = \{G \text{ undirected graph; } V_G = \langle k \rangle\}.$$

Note that if we consider an open $\{E\}$-formula $\varphi(\overline{x})$ and convert it to DNF, we get a disjunction of conjunctions. Each such conjunction says which first order literals should be satisfied on the variables $\overline{x}$, the following theorem is proved for conjunctions of literals. By the fact that $[\![\dots]\!]$ commutes with disjunctions, we can find out a value of any existential sentence.

In the statement of the following theorem we use the notation φ^b for an $\{E\}$-formula ψ and $b \in \{0, 1\}$ to mean

$$\psi^b := \begin{cases} \psi & b = 1 \\ \neg\psi & b = 0. \end{cases}$$

Theorem 2.4.2 (Everything exists)**.** Let $\varphi(\overline{x}, \overline{y}) = \bigwedge_{i=0}^{m-1} \psi_i(\overline{x}, \overline{y}) \wedge \bigwedge_{i=0}^{l-1} \vartheta_i(\overline{y})$, where ψ_i, ϑ_i are basic formulas and ψ_i are not of the form $(y_i = y_j)^b$, $E(y_i, y_j)^b$, $x_i \neq x_i$, $E(x_i, x_i)$, $b \in \{0, 1\}$ such that

- each ψ_i are not of the form $(y_i = y_j)^b, E(y_i, y_j)^b, x_i \neq x_i, E(x_i, x_i)$ for $b \in \{0, 1\}$
- if ψ_i is of the form $(x_i = z)^b$ for z in $\overline{x}$ or $\overline{y}$ then no other ψ_j is of the form $(x_i = z)^{1-b}$ or $(z = x_i)^{1-b}$
- if ψ_i is of the form $E(x_i, z)^b$ for z in $\overline{x}$ or $\overline{y}$ then no other ψ_j is of the form $E(x_i, z)^{1-b}$ or $E(z, x_i)^{1-b}$.

If $\overline{\beta}$ is a tuple of vertices computed by F_{rud} of the same length as $\overline{y}$ then

$$\lim_{F_{rud}} \mathcal{G}_n[\![(\exists \overline{x})\varphi_0(\overline{x}, \overline{\beta})]\!] = \lim_{F_{rud}} \mathcal{G}_n[\![\bigwedge_{i=0}^{l-1} \vartheta_i(\overline{\beta})]\!],$$

specifically if the conjunction is empty then

$$\lim_{F_{rud}} \mathcal{G}_n[\![(\exists \overline{x})\varphi_0(\overline{x}, \overline{\beta})]\!] = \mathbf{1}.$$

Proof. We will construct one tree T computing the whole tuple of witnesses $\overline{\alpha}$, such a construction can be straightforwardly split into a tuple of trees with each computing the specific element.

First we concatenate all the trees used to compute $\overline{\beta}$. At each leaf we can now proceed knowing the value of $\overline{\beta}$ at the specific $\omega \in \Omega$. Now we simply construct a tree as in Theorem 2.4.1 but searching only over edges not checked previously and only to fulfill each ψ_i. Since we assume $\mathcal{G}_k = \mathrm{ALL}_k$, both of the conditions of the theorem are satisfied. So by analogous argument, we have a tree that finds a witness for all of the $\psi_i(\overline{x}, \overline{\beta})$ with probability infinitesimally close to 1.

Therefore

$$\begin{aligned} [\![(\exists \overline{x})\varphi(\overline{x}, \overline{\beta})]\!] &= [\![(\exists \overline{x}) \bigwedge_{i=0}^{m-1} \psi_i(\overline{x}, \overline{\beta})]\!] \wedge [\![\bigwedge_{i=0}^{l-1} \vartheta_i(\overline{\beta})]\!] && (2.45) \\ &= [\![\bigwedge_{i=0}^{l-1} \vartheta_i(\overline{\beta})]\!]. && (2.46) \end{aligned}$$

□

The statement of the theorem was rather technical, but we can now use it to prove the following corollaries.

Corollary 2.4.3. For each $\varphi(\overline{x})$ that is not a contradiction in the theory of graphs we have that $[\![(\exists \overline{x})\varphi(\overline{x})]\!] = \mathbf{1}$.

Proof. The conditions on ψ_i are exactly saying that the conjunction is not a contradiction. Every other formula can be rewritten as a disjunction of such conjunctions and by the theorem we can satisfy at least one of them. □

Corollary 2.4.4. For each $\varphi(\overline{x}, \overline{y})$ that is not falsifiable by $\overline{y}$ in the theory of graphs we have that $[\![(\forall \overline{y})(\exists \overline{x})\varphi(\overline{x})]\!] = \mathbf{1}$.

Proof. No $\overline{\beta}$ can falsify $(\exists \overline{x})\varphi(\overline{x}, \overline{\beta})$, this means we can invoke the theorem on one of the non-falsifiable disjuncts. □

Theorem 2.4.5. The theory
$$\mathrm{Th}(\lim_{F_{rud}} \mathrm{ALL}_n)$$
is the theory of the Rado graph and therefore complete.

Proof. In [Gai64] it was proved that the theory of the Rado graph is axiomatized by the theory of undirected graphs and the sentences $E_{i,j}$ which say that if we have a set A of i distinct vertices and a set B of j distinct vertices such that $A \cap B = \varnothing$, then there is a vertex v which has an edge with all vertices from A and with no vertices from B.

Each $E_{i,j}$ satisfies the statement of Corollary 2.4.4 and because it is also a complete theory, we have proved the theorem. □

Corollary 2.4.6. (0-1 law for ALL_k) For every $\{E\}$-sentence φ we have that
$$\lim_{F_{rud}} \mathrm{ALL}_n[\![\varphi]\!] \in \{\mathbf{0}, \mathbf{1}\}.$$

The wide sequence ALL_k is maximal in the sense that for every $\mathcal{G}_k$ with $g_k = k$ we have that $\mathcal{G}_k \subseteq \mathrm{ALL}_k$. Since we proved 0-1 law for ALL_k we can ask whether this tells us anything about the subobjects of ALL_k since they consist in some sense of all other wide sequences. From now on we no longer assume $\mathcal{G}_k$ denotes ALL_k.

Definition 2.4.7. Let $\mathcal{G}_k$ be a wide sequence. We say that $\mathcal{G}'_k$ is a **portion** of $\mathcal{G}_k$ if we have $\mathcal{G}'_k \subseteq \mathcal{G}_k$ for all k big enough which we denote $\mathcal{G}'_k \leq \mathcal{G}_k$. We say it is a **large portion** if we have
$$p := \lim_{k\to\infty} \frac{|\mathcal{G}'_k|}{|\mathcal{G}_k|} > 0,$$
which we denote $\mathcal{G}'_k \leq_l \mathcal{G}_k$. Moreover, if the limits tends to 1 we call $\mathcal{G}'_k$ a **major portion**.

By elementary probability theory we can prove the following.

Lemma 2.4.8. Let $\mathcal{G}'_k \leq_l \mathcal{G}_k$ be a wide sequence and its large portion, let $A \subseteq \mathcal{G}_n$ be an event in $\mathcal{M}$ and let $A' = \mathcal{G}'_n \cap A$. Then we have
$$\mathrm{st}\left(\Pr_{G\in\mathcal{G}_n}[A]\right) = 1 \Rightarrow \mathrm{st}\left(\Pr_{G\in\mathcal{G}'_n}[A']\right) = 1.$$

Lemma 2.4.9. Let $\mathcal{G}'_k \leq \mathcal{G}_k$ be a wide sequence and its major portion, let $A \subseteq \mathcal{G}_n$ be an event in $\mathcal{M}$ and let $A' = \mathcal{G}'_n \cap A$. Then we have
$$\mathrm{st}\left(\Pr_{G\in\mathcal{G}_n}[A]\right) = 1 \Leftrightarrow \mathrm{st}\left(\Pr_{G\in\mathcal{G}'_n}[A']\right) = 1.$$

Corollary 2.4.10. Let $\mathcal{G}_k$ be a wide sequence such that
$$\lim_{k\to\infty} \frac{|\mathcal{G}_k|}{2^{\binom{g_k}{2}}} > 0$$
then $\mathrm{Th}(\lim_{F_{rud}} \mathcal{G}_n)$ is the theory of Rado graph.

Proof. The condition on the limit assures that $\mathcal{G}_k \leq_l \mathrm{ALL}_{g_k}$. Which, by Lemma 2.4.8, is enough to replicate the proof for ALL_{g_k}. □

We will use a similar argument in Chapter 4 to instead prove something about a wide sequence by first proving it for its portion.

2.5 Isomorphism closed categorical $\mathcal{G}_k$

Take an isomorphism closed wide sequence $\mathcal{G}_k$, what are its isomorphism closed portions? It turns out it is easy to classify them because any such portion can be constructed as a union of those wide sequences which always have one isomorphism type of each $\mathcal{G}_k$ and those precisely match the isomorphism closed sequences which satisfy the following property.

Definition 2.5.1. We say that $\{\mathcal{G}_k\}_{k=0}^{\infty}$ is **categorical** if there is k_0 such that for every $k > k_0$ if we have $G_1, G_2 \in \mathcal{G}_k$ then $G_1 \cong G_2$. For a categorical wide sequence $\{G_k\}_{k=0}^{\infty}$ we denote G_k the lexicographically minimal element of $\mathcal{G}_k$.

One can see isomorphism closed $\mathcal{G}_k$ as the natural wide sequences and categorical isomorphism closed wide sequences as their building blocks. The concept of isomorphism closed categorical wide sequences can be already limiting the form of the sequence. We have the following lemma which limits sizes of $\mathcal{G}_k$ to specific values.

Lemma 2.5.2. Let $\{\mathcal{G}_k\}_{k=0}^{\infty}$ be categorical and isomorphism closed, then for large enough k

$$|\mathcal{G}_k| = \frac{g_k!}{|\mathrm{Aut}(G_k)|}.$$

Proof. Every $\rho \in S_{g_k}$ defines an isomorphism $\rho : G_k \to \rho(G_k)$, where $\rho(G_k)$ is a graph obtained from G_k by renaming every edge coordinate wise by ρ.

Claim: For any $\rho, \pi \in S_{g_k}$:

$$\rho(G_k) = \pi(G_k) \iff \exists \tau \in \mathrm{Aut}(G_k) : \rho \circ \tau = \pi.$$

Proof of claim. "$\Rightarrow$" Let $\rho(G_k) = \pi(G_k)$, therefore $\tau := \rho^{-1} \circ \pi \in \mathrm{Aut}(G_k)$ and $\rho \circ \tau = \rho \circ \rho^{-1} \circ \pi = \pi$.

"$\Leftarrow$" Let $\rho \circ \tau = \pi$. Then $\pi(G_k) = \rho(\tau(G_k)) = \rho(G_k)$. □

Notice that the τ in the statement of the claim is uniquely determined by $\rho^{-1} \circ \pi$. Therefore if we defined a quotient set $S_{g_k}/\sim$ with $\rho \sim \pi \iff \rho(G_k) = \pi(G_k)$, then $|S_{g_k}/\sim| = \frac{g_k!}{|\mathrm{Aut}(G_k)|}$.

The Lemma follows from noticing that if we start with $\{G_k\}$ and then we build $\mathcal{G}_k$ by finding isomorphic graphs on the vertex set $\langle g_k \rangle$, we can only do so by trying different permutation from S_{g_k} and these permutations find the same graph if and only if they are in the same $\sim$-class. Therefore there is a bijection between $S_{g_k}/\sim$ and $\mathcal{G}_k$. □

So far we did not encounter an isomorphism closed wide sequence $\mathcal{G}_k$ and a $\{E\}$-sentence φ for which we proved $\mathbf{0} < \lim_{F_{rud}} \mathcal{G}_n[\![\varphi]\!] < \mathbf{1}$. One can use isomorphism closed categorical wide sequences to construct such a wide sequence.

Example 2.5.3. Let

$$\mathrm{nonEDGE}_k := \{(\langle k \rangle, E); |E| = (k(k-1)/2) - 1\} \tag{2.47}$$

$$\mathcal{G}_k := \mathrm{EDGE}_k \cup \mathrm{nonEDGE}_k, \tag{2.48}$$

Then one can see that $\mu(\lim_{F_{rud}} \mathcal{G}_n[\![(\exists x)(\exists y)E(x,y)]\!]) = \frac{1}{2}$.

However, each EDGE_k and $\mathrm{nonEDGE}_k$ have complete theories. So natural follow up to this question is whether 0-1 law holds for all isomorphism closed categorical wide sequences.

Theorem 2.2.3, Corollary 2.4.10 and Lemma 2.5.1 already give some conditions on what any counterexample would have to satisfy. However, we did not find any nor did we rule out its existence.

Question 2.5.4. Is there a isomorphism closed categorical wide sequence $\mathcal{G}_k$ and an $\{E\}$-sentence φ such that

$$\mathbf{0} < \lim_{F_{rud}} \mathcal{G}_n[\![\varphi]\!] < \mathbf{1}?$$

3. Dense case

3.1 $\mathcal{G}_k = \mathbf{SK}_k^{1/2}$

Now we turn to analyze dense wide sequences in the second order case; that is the case when we shall be interested in second order properties of the wide limit. In this chapter we assume $F = F_{rud}$ and $G = G_{rud}$. Specifically we will consider the problem of finding a large clique in a graph.

Generally it is considered a computationally hard problem to find a large clique in a graph. It is an **NP**-hard problem and thus it is conjectured that it cannot be solved in polynomial time. We first turn to the following wide sequence.

Definition 3.1.1. Let

$$\mathrm{SK}_k^{1/2} = \{(\langle k\rangle, E); E \text{ consists of exactly one } k/2\text{-clique}\}.$$

Limiting inputs to $\mathrm{SK}_k^{1/2}$ makes the problem less complex, because for a vertex v to be a part of the biggest clique it is enough that it has nonzero degree. Naturally we want to see whether there is a large clique in $\lim_F^G \mathrm{SK}_n^{1/2}$, every sample has a clique of size $n/2$, but is there a set of F-vertices witnessing that?

Here to measure the size of such a clique, the second order wide limit by itself is not a sufficient object. Instead we need to turn to the arithmetical expansion $K(\mathcal{G}_n, F, G)$ and ask whether we can find an injective function from some large initial segment into a clique. This can be expressed as a second order sentence with parameters from F. It is not hard to prove the following result which also demonstrates how the sentence is formed.

Theorem 3.1.2. For every $t > \mathbb{N}$ we have

$$K(\mathrm{SK}_k^{1/2}, F, G)[\![E \text{ has a clique of size } n^{1/t}]\!] = \mathbf{1}.$$

Proof. (sketch) We need to analyze the value

$$[\![(\exists\Lambda)(\forall u)(\forall v)(((u \neq v \leq n^{1/t}) \to E(\Lambda(u), \Lambda(v))) \wedge (\Lambda : [n^{1/t}] \hookrightarrow \mathcal{M}))]\!] \quad (3.1)$$

which is equal to

$$\bigvee_\Lambda \bigwedge_u \bigwedge_v [\![(u \neq v < n^{1/t}) \to (E(\Lambda(u), \Lambda(v)) \wedge \Lambda(u) \neq \Lambda(v))]\!]. \quad (3.2)$$

So we want to find some $n^{1/t}$-tuple of trees computing some Λ, computed by some $(\gamma_0, \ldots, \gamma_{n^{1/t}-1})$, which is injective on $\langle n^{1/t}\rangle$ and its $\langle n^{1/t}\rangle$ range is a clique in E.

We define $(\gamma_0, \ldots, \gamma_{n^{1/t}-1})$ as follows. The tree T_0 computing Λ_0 inspects all the edges $(u, v) \in \langle n^{1/t}\rangle \times \langle n^{1/t}\rangle$ in some specified order and outputs the first vertex it finds with an edge and outputs 0 otherwise. The tree T_i computing γ_i extends T_{i-1} by not outputting the i-th vertex with the property but instead continues inspecting and outputs the $(i+1)$-th vertex for which it finds an edge. Also every such tree has depth at most $n^{1/t} \cdot n^{1/t} = n^{2/t} = n^{1/(t/2)}$.

One can check that the probability the tree γ_i does not find i vertices with an edge is infinitesimal and therefore it always outputs a vertex in the clique of ω. Intuitively this is because the expected ratio of the vertices in $\langle n^{1/t}\rangle$ which are also in the clique is about one half. Moreover, every γ_i outputs the i-th element element of the ordering and thus it is injective. $\square$

At first glance the lower bound $n^{1/t}$ for every nonstandard t may seem optimal given the proof method we used, but there is a way to radically improve it. The idea is to partition the set of vertices into many smaller ones and let γ_i search only in the i-th set. First we need the following lemmas.

Lemma 3.1.3. Let $S \subseteq \langle n \rangle$ such that $|S| = m > \mathbb{N}$, then

$$\operatorname{st}\left(\Pr_{G \in \mathcal{G}_n}[S \text{ contains no vertices in the clique of } G]\right) = 0.$$

Proof. There are $\binom{n-m}{\lfloor \frac{n}{2} \rfloor}$ different graphs in $\mathrm{SK}_n^{1/2}$ in which the clique does not intersect S. We then bound the probability as

$$\frac{\binom{n-m}{\lfloor \frac{n}{2} \rfloor}}{\binom{n}{\lfloor \frac{n}{2} \rfloor}} = \frac{(n-m)!(n-\lfloor \frac{n}{2} \rfloor)!}{(n)!(n-\lfloor \frac{n}{2} \rfloor - c)!} \tag{3.3}$$

$$= \prod_{i=0}^{m-1} \frac{(n-i-\lfloor \frac{n}{2} \rfloor)}{(n-i)} \tag{3.4}$$

$$= \prod_{i=0}^{m-1} \left(1 - \frac{\lfloor \frac{n}{2} \rfloor}{n-i}\right) \tag{3.5}$$

$$\leq \left(1 - \frac{\lfloor \frac{n}{2} \rfloor}{n}\right)^m \tag{3.6}$$

$$\leq \left(1 - \frac{\lfloor \frac{n}{2} \rfloor}{n}\right)^{n \cdot \frac{m}{n}} \tag{3.7}$$

$$\leq e^{-\lfloor \frac{n}{2} \rfloor \frac{m}{n}}. \tag{3.8}$$

But $\lfloor \frac{n}{2} \rfloor \frac{m}{n}$ is infinite therefore the bound is infinitesimal. □

Lemma 3.1.4. Let $a \in \mathcal{M}$, let $v_0, \ldots, v_{a-1} \in \langle n \rangle$ distinct vertices, then there exist trees T_{v_i} such that

$$\operatorname{st}\left(\Pr_{\omega \in \Omega}[\forall i : (v_i, T_{v_i}(\omega)) \in E_\omega | \forall i : v_i \text{ is in the clique}]\right) = 1.$$

Proof. The tree T_{v_i} inspects all the edges (v_i, j) where j ranges over $\langle n^{1/t} \rangle$ for some $t > \mathbb{N}$ and outputs j if $(v_i, j) \in E_\omega$. By Lemma 3.1.3 we have that only an infinitesimal number of graphs have their clique not intersect $\langle n^{1/t} \rangle$ so each T_{v_i} succeeds on all but an infinitesimally small portion of Ω. But if one T_{v_i} finds a neighbour of v_i, then all do, since nonzero degree vertices in every ω form a clique and the same $w \in \langle n^{1/t} \rangle$ is a neighbour of all v_is. □

Lemma 3.1.5. Let $S_0, \ldots, S_{a-1} \subseteq \langle n \rangle$ sets of size $a \in \mathcal{M}$ for $i \in \mathcal{M}$ then

$$\Pr_{\omega \in \Omega}\left[\bigcup_{i=0}^{a-1}[S_i \text{ contains no vertices in the clique of } \omega]\right] \leq a \cdot e^{-\lfloor \frac{n}{2} \rfloor \frac{m}{n}}.$$

Proof. Follows from the proof of Lemma 3.1.3 and union bound. □

Now we are ready to improve on Theorem 3.1.2.

Theorem 3.1.6. Let $m = 2\ln n$. Then

$$K(\mathrm{SK}_k^{1/2}, F, G)[\![E \text{ has a clique of size } \lfloor n/m \rfloor]\!] = \mathbf{1}.$$

Proof. Partition a subset of $\langle n \rangle$ to sets $S_0, \ldots, S_{\lfloor n/m \rfloor -1}$ each of size at least m. Specifically if m divides n then we partition the whole $\langle n \rangle$.

By Lemma 3.1.5 we have that with probability that we do not sample ω which have the clique intersect all S_is

$$\left\lfloor \frac{n}{m} \right\rfloor \cdot e^{-\left\lfloor \frac{n}{2} \right\rfloor \frac{m}{n}} = e^{\ln \left\lfloor \frac{n}{m} \right\rfloor - \left\lfloor \frac{n}{2} \right\rfloor \frac{m}{n}}, \tag{3.9}$$

we can bound the exponent as

$$\begin{aligned}
\ln \left\lfloor \frac{n}{m} \right\rfloor - \left\lfloor \frac{n}{2} \right\rfloor \frac{m}{n} &\leq \ln \frac{n}{m} - \frac{n}{2} \cdot \frac{m}{n} + \frac{m}{n} && (3.10)\\
&\leq \ln n - \ln m - \frac{m}{2} + \frac{m}{n} 1 && (3.11)\\
&\leq \ln n - \ln \ln n - \ln n + \frac{2 \ln n}{n} && (3.12)\\
&\leq - \ln \ln n + \frac{2 \ln n}{n} && (3.13)
\end{aligned}$$

which is negative and infinite, because $\frac{\ln x}{x} \overset{x \to \infty}{\to} 0$ as and therefore (3.9) is infinitesimal.

So with probability infinitesimally close to 1 we have in each S_i a vertex v_i which is also a part of the clique. By Lemma 3.1.4 we have that there exists a tree verifying whether a given vertex is in the clique and since $m \leq n^{1/t}$ for some t we can concatenate the trees to get a tree T_{S_i} which finds in S_i an element of the clique with probability infinitesimally close to 1.

Finally we can have a function $\Lambda \in G$ computed by $(\Lambda_0, \ldots, \Lambda_{\lfloor n/m \rfloor - 1})$ by letting Λ_i be computed by T_{S_i} we have already verified $[\![\Lambda \text{ is a clique}]\!] = \mathbf{1}$.

Because $T_{S_i}(\omega) \in S_i$ when S_i succeeds, and S_i are disjoint we have

$$[\![\Lambda : \langle \lfloor n/m \rfloor \rangle \hookrightarrow \mathcal{M}]\!] = \mathbf{1}.$$

Which proves the theorem. $\square$

One can also check that for $\mathrm{SK}_k^{1/l}$, the graphs whose edges are exactly one $\lfloor k/l \rfloor$ clique for some $l \in \mathbb{N}$, the wide limit has a clique of size $\lfloor n/m \rfloor$ for $m = l \cdot \ln(m)$ by the same technique.

Even though the size of the clique has radically increased, we still did not find a clique in E of size $n/2$. One can verify that with the method provided one cannot obtain such a clique because the probability that any of $\lfloor n/2 \rfloor$ two-element sets does not intersect the clique is too large.

Question 3.1.7. Is $\lfloor n/(2 \ln n) \rfloor$ the clique number in the Boolean valued graph $\lim_{F_{rud}}^{G_{rud}} \mathrm{SK}_n^{1/2}$, measured in $K(\mathrm{SK}_k^{1/2}, F, G)$?

3.2 $\mathcal{G}_k = \mathbf{CK}_k^{1/2}$

Now let us mention the more complex case of the wide sequence $\mathrm{CK}_k^{1/2}$

$$\mathrm{CK}_k^{1/2} = \{(\langle k \rangle, E); E \text{ contains a } \lfloor k/2 \rfloor \text{ clique.}\}$$

We are still guaranteed that every ω contains a large clique, but there is no easy way to check whether a given vertex v is contained in the large clique. To prove the following theorem we can use that $\mathrm{CK}_k^{1/2}$ is isomorphism closed to translate the case for $\mathrm{SK}_k^{1/2}$ at least for the case of cliques of standard size.

Theorem 3.2.1. Let $c \in \mathbb{N}$ then

$$K(\mathrm{CK}_k^{1/2}, F, G)[\![E \text{ has a clique of size } c]\!] > \mathbf{0}.$$

Proof. Consider the set

$$\mathcal{G}_n^+ = \{(G, U); G \in \mathcal{G}_n \text{ and } U \subseteq \langle n \rangle \text{ is a } \lfloor n/2 \rfloor \text{ clique in } G \ \}$$

and projections

$$\begin{aligned} \pi_1 &: \mathcal{G}_n^+ \to \mathcal{G}_n, (G, U) \mapsto G. \\ \pi_2 &: \mathcal{G}_n^+ \to \mathcal{P}(\langle n \rangle), (G, U) \mapsto U \end{aligned}$$

From the fact that $\mathrm{CK}_k^{1/2}$ is isomorphism closed, we have that $\left|\pi_2^{-1}[U_1]\right| = \left|\pi_2^{-1}[U_2]\right|$. We will set $\nu := \left|\pi_2^{-1}[U]\right|$ for some clique U of size $\lfloor n/2 \rfloor$ so we have $\nu \left|\mathcal{G}_n\right| = \mathcal{G}_n^+$. Let $\varphi_c(\overline{x})$ be a $\{E\}$-sentence is satisfied iff $\overline{x}$ is a c sized clique. Denote an event in the sample space $\mathcal{G}_n$

$$A_{\overline{a}} := \{G \in \mathcal{G}_n; G \models \varphi_c(\overline{a})\}.$$

and in the sample space $\mathcal{G}_n^+$

$$A_{\overline{a}}^+ := \{(G, U) \in \mathcal{G}_n^+; \overline{a} \text{ is a subclique of } U\}$$

notice that if $(G, U) \in A_{\overline{a}}^+$ then $G \in A_{\overline{a}}$ which implies (by another argument with projections) that

$$\nu \cdot |A_{\overline{a}}| \geq \left|A_{\overline{a}}^+\right|$$

so in conjunction with Example 2.3.3 we have

$$\begin{aligned} \Pr_{G \in \mathcal{G}_n}[A_{\overline{a}}] &\geq \frac{\left|A_{\overline{a}}^+\right|}{|\mathcal{G}_n^+|} \\ &= \Pr_{G \in \mathrm{SK}_k^{1/2}}[\overline{a} \text{ is in the clique}] \\ &> (1/2)^c \end{aligned}$$

and since c is constant we have $\Pr_{G \in \mathcal{G}_n}[A_{\overline{a}}] > 0$. □

Of course it remains to show that using trees can actually increase the value all the way to $\mathbf{1}$, we did not get to prove that.

Finally one can return to cliques of nonstandard size and intuitively one expects this to be hard. For a clique of size greater than $n^{1/t}$ for any $t > \mathbb{N}$ one has to check more than $n^{1/t}$ edges to even know whether a given set is a clique and the counting argument used in the previous theorem implies that just guessing the clique of nonstandard size is not enough. We therefore present the following conjecture.

Conjecture 3.2.2. Let $m \in \mathcal{M}$ such that $m \leq c \ln n$ for any $c \in \mathbb{N}$, then

$$K(\mathrm{CK}_k^{1/2}, F, G)[\![E \text{ has a clique of size } \lfloor n/m \rfloor]\!] = \mathbf{0}.$$

4. Sparse case and TFNP

4.1 $\mathcal{G}_k = *\mathbf{PATH}_k$

Now we turn our attention to a wide sequence which is made up of 'the hardest instances in LEAF'. That is, if we are given a degree 1 vertex labeled 0 and search for another degree 1 vertex, it is the hardest if there are only two degree 1 vertices and the path from 0 to the solution is as long as possible.

Definition 4.1.1. We define $*\text{PATH}_k$ (the pointed paths on k vertices) as the set of all (undirected) graphs G on the vertex set $\langle k \rangle$, where G is isomorphic to the path on k vertices and $\deg_G(0) = 1$.

One can also see that it is not fruitful to analyze the F_{rud} limit, because there are only $k-1$ edges in each $G \in *\text{PATH}_k$ and therefore, by an analogous argument to the proof of Theorem 2.2.3, we have that $\lim_{F_{rud}} *\text{PATH}_k[\![(\exists x)(\exists y)E(x,y)]\!] = \mathbf{0}$. Moreover in the type 2 version of the problem the graph is presented by an oracle which gives us the neighbour set for each vertex, so we define a new family F as follows.

Definition 4.1.2. After we fix n, we define F_{nbtree} as the set of all functions computed by some some labeled tree with the following shape:

- Each non-leaf node is labeled by some $w \in \langle n \rangle$.
- For each $\{u, v\} \subseteq \langle n \rangle$ and a node w there is an outgoing edge from w labeled $\{u, v\}$.
- Each leaf is labeled by some $m \in \langle g_n \rangle$.
- The depth of the tree is at most $g_n^{1/t}$ for some $t > \mathbb{N}$.

Computation of such a tree on an undirected graph G goes as follows. We interpret the non-leaf nodes as questions "what is the neighbour set of w?" and the edges as answers from our graph ω, and thus we follow a path in the computation tree determined by ω until we find a leaf, in which case the computation returns the label of the leaf. We will denote the set of all such trees as $\mathcal{T}_{nbtree}$.

We now shift our focus to analyzing the ability of trees from F_{nbtree} to find the non-zero degree 1 vertex in $G \in *\text{PATH}_n$. We say that a tree $T \in \mathcal{T}_{nbtree}$ fails at a graph G if $T(G)$ is not a non-zero vertex of degree one in G.

Definition 4.1.3. Let $m \leq n$ and $v \in \langle w \rangle$ and $U \subseteq \langle w \rangle$ with $|U| \leq 2$, then we define

$$\mathcal{G}_m^{v?=U} := \{G \in \mathcal{G}_m; N_G(v) = U\},$$

where N_G is the neighbour-set function of G.

Lemma 4.1.4. There are bijections for all nonstandard $m \leq n$ and distinct $u, v, w \in \langle m) \setminus \{0\}$:

$$\mathcal{G}_m^{v?=\{u,w\}} \cong \mathcal{G}_{m-2} \times [2] \tag{4.1}$$
$$\mathcal{G}_m^{v?=\{u,0\}} \cong \mathcal{G}_{m-2} \tag{4.2}$$
$$\mathcal{G}_m^{0?=\{u\}} \cong \mathcal{G}_{m-1}. \tag{4.3}$$

Proof. (sketch) For (4.1) we can just contract all of u,v,w into one vertex and relabel the rest of the graph, leaving the orientation as one remaining bit of information. This is obviously reversible and a bijection.

For (4.2) we can do the same, but the orientation is given by 0. □

Lemma 4.1.5. Let $T \in \mathcal{T}_{nbtree}$, with root labeled $v \in [m] \setminus 0$, we have for each $T_{v?=\{u,w\}}$ a tree $\tilde{T}_{v?=\{u,w\}}$ of the same depth, such that

$$\Pr_{G \in \mathcal{G}_m}(T_{v?=\{u,w\}} \text{ fails} | v? = \{u, w\}) = \Pr_{G \in \mathcal{G}_{m-2}}(\tilde{T}_{v?=\{u,w\}}). \tag{4.4}$$

For a tree T with the root labeled 0, we have a tree $\tilde{T}_{v?=\{u,w\}}$ of the same depth, such that

$$\Pr_{\mathcal{G} \in \mathcal{G}_m}(T_{v?=\{u\}} \text{ fails} | v? = \{u\}) = \Pr_{\mathcal{G}_{m-1}}(\tilde{T}_{v?=\{u\}}). \tag{4.5}$$

Proof. (sketch) To construct the tree, we just replace all vertices in labels of $T_{v?=\{u,w\}}$ by their renumbering from the bijection in (4.1).

One can then check that the trees $T_{v?=\{u,w\}}$ and $\tilde{T}_{v?=\{u,w\}}$ are isomorphic in a sense that their computation of a graph G and $\tilde{G}$ respectively, $\tilde{G}$ being the corresponding $(m-2)$-vertex graph, agree with the structure of the path and that correctness of leaves is preserved under the renumbering. Essentially they emulate the same computation but on a smaller graph. □

Lemma 4.1.6. For all nonstandard $t > \mathbb{N}, m \geq n - 2n^{1/t}$ and $k \in [n^{1/t} + 1]$ for all trees $\mathcal{T} \in \mathcal{T}_{nbtree}$ of depth k we have

$$\Pr_{G \in \mathcal{G}_m}(T \text{ fails}) \geq \prod_{i=0}^{k} \left(1 - \frac{2}{m - 2i - 2}\right).$$

Proof. We proceed by induction on k.

$k = 0$: We have that the probability of success of a straight guess is at most $\frac{1}{m-1}$. Therefore

$$\Pr_{G \in \mathcal{G}_m}(T \text{ fails}) \geq \left(1 - \frac{1}{m-1}\right) \geq \left(1 - \frac{2}{m-2}\right). \tag{4.6}$$

$(k-1) \Rightarrow k$: First we assume that the root is labeled 0. Then we have

$$\Pr_{G\in\mathcal{G}_m}[T \text{ fails}] = \sum_{u\in V\setminus\{0\}} \Pr_{G\in\mathcal{G}_{m-1}}[E(0,u)] \Pr_{G\in\mathcal{G}_{m-1}}[T_{0?=\{u\}} \text{ fails}|E(0,u)] \tag{4.7}$$

$$\geq \Pr_{G\in\mathcal{G}_{m-1}}[T_{0?=\{u\}} \text{ fails}|E(0,u)] \tag{4.8}$$

$$= \Pr_{G\in\mathcal{G}_{m-1}}[\tilde{T}_{0?=\{u\}} \text{ fails}] \tag{4.9}$$

$$\geq \prod_{i=0}^{k-1}\left(1 - \frac{2}{m-2i-2}\right) \tag{4.10}$$

$$\geq \prod_{i=0}^{k}\left(1 - \frac{2}{m-2i-2}\right). \tag{4.11}$$

Now we assume that the root is labeled $v \neq 0$. First we notice that

$$\Pr_{G\in\mathcal{G}_m}[E(v,0)] = \frac{1}{m-1} \tag{4.12}$$

$$\Pr_{G\in\mathcal{G}_m}[N(V) = 1] = \frac{1}{m-1} \tag{4.13}$$

$$\Pr_{G\in\mathcal{G}_m}[|N(V)\setminus\{0\}| = 2] = 1 - \frac{2}{m-1}, \tag{4.14}$$

the first two probabilities are obviously $\frac{1}{m-1}$ because they correspond to v being positioned on one of the ends of the non-zero segment which has length $m-1$. The event in (4.14) is the complement of the union of the first two events, which have empty intersection, giving us that stated probability.

Then we have for $p := \Pr_{G\in\mathcal{G}_m}[T \text{ fails}]$

$$p = \Pr_{G\in\mathcal{G}_m}[E(v,0)] \Pr_{G\in\mathcal{G}_m}[T \text{ fails}|E(v,0)] \tag{4.15}$$

$$+ \Pr_{G\in\mathcal{G}_m}[|N(v)\setminus\{0\}| = 2] \Pr_{G\in\mathcal{G}_m}[T \text{ fails}|\, |N(v)\setminus\{0\}| = 2] \tag{4.16}$$

$$+ (\Pr_{G\in\mathcal{G}_m}[|N(v)| = 1] \Pr_{G\in\mathcal{G}_m}[E(v,0)] \Pr_{G\in\mathcal{G}_m}[T \text{ fails}|\, |N(v)| = 1]) \tag{4.17}$$

$$\geq \Pr_{G\in\mathcal{G}_m}[|N(v)\setminus\{0\}| = 2] \Pr_{G\in\mathcal{G}_m}[T \text{ fails}|\, |N(v)\setminus\{0\}| = 2] \tag{4.18}$$

$$\geq (1 - \frac{2}{m-1}) \tag{4.19}$$

$$\sum_{\substack{u,w\in V\setminus\{0\}\\ u\neq w}} \Pr_{G\in\mathcal{G}_m}[v? = \{u,w\}] \Pr_{G\in\mathcal{G}_m}[T_{v?=\{u,w\}} \text{ fails}|v? = \{u,w\}] \tag{4.20}$$

$$\geq (1 - \frac{2}{m-1}) \Pr_{G\in\mathcal{G}_m}[T_{v?=\{u_0,w_0\}} \text{ fails}|v? = \{u_0,w_0\}] \tag{4.21}$$

$$\geq (1 - \frac{2}{m-1}) \Pr_{G\in\mathcal{G}_{m-2}}[\tilde{T}_{v?=\{u_0,w_0\}} \text{ fails}] \tag{4.22}$$

$$\geq (1 - \frac{2}{m-1}) \prod_{i=0}^{k-1}(1 - \frac{2}{m-2i-4}) \tag{4.23}$$

$$\geq (1 - \frac{2}{m-2}) \prod_{i=1}^{k}(1 - \frac{2}{m-2i-2}) \tag{4.24}$$

$$\geq \prod_{i=0}^{k}(1 - \frac{2}{m-2i-2}). \tag{4.25}$$

where in (4.21) we choose u_0, w_0 with the lowest value of

$$\Pr_{G \in \mathcal{G}_m}(T_{v?=\{u_0,w_0\}} | v? = \{u_0, w_0\}),$$

the bound follows the fact that all $\Pr_{G\in\mathcal{G}_m}(v? = \{u, w\})$ are the same for distinct non-zero u, w. In (4.22) we use the Lemma 4.1.5 and in (4.23) we use the induction hypothesis. $\square$

Corollary 4.1.7. For a tree $T \in \mathcal{T}_{nbtree}$ and $c \in \mathbb{N}$ we have that

$$\mathrm{st}\left(\Pr_{G\in\mathcal{G}_{n-c}}[T \text{ fails}]\right) = 1.$$

Proof. Since T has depth at most $n^{1/t}$ for some $t > \mathbb{N}$ we by the previous lemma that

$$\Pr_{G\in\mathcal{G}_{n-c}}[T \text{ fails}] \geq \prod_{i=0}^{n^{1/t}} \left(1 - \frac{2}{n - 2i - c - 2}\right) \tag{4.26}$$

$$\geq \left(1 - \frac{2(n^{1/t} + 1)}{n - 2n^{1/t} - c - 2}\right) \tag{4.27}$$

and the standard part of the lower bound is 1. $\square$

Finally, we can prove the following theorem.

Theorem 4.1.8.

$$\lim_{F_{nbtree}} {*}\mathrm{PATH}_n[\![(\exists v)(\exists u)(\forall w)((v \neq 0) \wedge (E(v, u)) \wedge (E(v, w) \to u = w))]\!] = \mathbf{0}$$

Proof. Expanding the value of the formula in the statement we get

$$\bigvee_\alpha \bigvee_\beta \bigwedge_\gamma [\![(\alpha \neq 0) \wedge (E(\alpha, \beta)) \wedge (E(\alpha, \gamma) \to \beta = \gamma)]\!],$$

to prove it evalues to $\mathbf{0}$ we need to find for every α, β some γ such that

$$[\![(\alpha \neq 0) \wedge (E(\alpha, \beta)) \wedge (E(\alpha, \gamma) \to \beta = \gamma)]\!] = \mathbf{0}.$$

For any α, β we define

$$\gamma(\omega) := \begin{cases} v & N(\alpha(\omega)) = \{v\} \\ w & w \in N(\alpha(\omega)) \setminus \{\beta(\omega)\}, \end{cases}$$

such a function can be computed by a tree in F_{nbtree} which we can construct by concatenation of trees computing α and β.

Let T be the tree computing α. Now we proceed by contradiction, let

$$\epsilon := \mu([\![(\alpha \neq 0) \wedge (E(\alpha, \beta)) \wedge (E(\alpha, \gamma) \to \beta = \gamma)]\!]) > 0,$$

by definition this means that

$$\epsilon = \mathrm{st}(P_n[(\alpha \neq 0) \wedge (E(\alpha, \beta)) \wedge (E(\alpha, \gamma) \to \beta = \gamma)]) > 0.$$

But by definition of γ and Corollary 4.1.7 we have

$$\begin{aligned} 0 &< \epsilon \\ &= \mathrm{st}(P_n[(\alpha \neq 0) \wedge (E(\alpha,\beta)) \wedge (E(\alpha,\gamma) \to \beta = \gamma)]) \\ &\leq \mathrm{st}(P_n[(\alpha \neq 0) \wedge (E(\alpha,\beta)) \wedge |N(\alpha)| = 1]) \\ &\leq \mathrm{st}(P_n[(\alpha \neq 0) \wedge |N(\alpha)| = 1]) \\ &= \mathrm{st}(P_n[T \text{ does not fail}]) \\ &= 0. \end{aligned}$$

A contradiction. □

Corollary 4.1.9. $\mathrm{Th}(\lim_{F_{nbtree}} *\mathrm{PATH}_n)$ is complete.

Proof. By applying Theorem 2.3.4 we have that the sentences $\neg C_k$ stating that there are not cycles of length $k \in \mathbb{N}$ are valid all in $\lim_{F_{nbtree}} *\mathrm{PATH}_n$, the sentence $D^{1,2}_{1,rest}$ stating that there is exactly one vertex of degree 1 and all other vertices have degree 2 is valid by Theorem 4.1.8.

Let $T = \{\neg C_k, k \in \mathbb{N}\} \cup \{D^{1,2}_{1,rest}\}$, and let $\mathcal{A}_1, \mathcal{A}_2 \models T$, then we can see by the handshaking lemma that $\mathcal{M}_1$ and $\mathcal{M}_2$ are both infinite. And we can see that they can be decomposed into one path starting at 0 with no end, and then more infinite paths which have the order type of $\mathbb{Z}$. The duplicator of Ehrenfeucht-Fraïssé game has a winning strategy by responding to any element on the order type $\mathbb{Z}$ with a far enough element on the path of the order $\mathbb{N}$. □

4.2 $\mathcal{G}_k = *\mathrm{PATH}^{\leq}_k$

So far we have proved that the hardest instances of LEAF do not have a solution in the F_{nbtree} limit and that they satisfy the 0-1 law. We can generalize this result to a larger class of instances.

Definition 4.2.1. We define $*\mathrm{PATH}^{\leq}_k$ (the pointed paths on k vertices of length at most k) as the set of all (undirected) graphs G on the vertex set $\langle k\rangle$, where G has a subgraph isomorphic to the path on $l \leq k$ vertices, $\deg_G(0) = 1$ and no other edges.

Immediately we have that $*\mathrm{PATH}_k$ is a portion of $*\mathrm{PATH}^{\leq}_k$, we can prove even more.

Definition 4.2.2. We define $*\mathrm{PATH}^{l}_k$ as the portion of $*\mathrm{PATH}^{\leq}_k$ where the path subgraph is of length exactly l.

Lemma 4.2.3. Let $c \in \mathbb{N}$, then

$$\lim_{k\to\infty} \frac{\left|*\mathrm{PATH}^{k-c}_k\right|}{\left|*\mathrm{PATH}^{\leq}_k\right|} = \frac{1}{ec!}$$

So $*\mathrm{PATH}^{k-c}_k$ is a large portion of $*\mathrm{PATH}^{\leq}_k$ and specially $*\mathrm{PATH}_k \leq_l *\mathrm{PATH}^{\leq}_k$.

Proof. By direct computation we have that the fraction $|{*}\mathrm{PATH}_k| \,/\, \left|{*}\mathrm{PATH}_k^{\leq}\right|$ is

$$\frac{(k-1)!/(c!)}{\sum_{i=1}^{k-1}\prod_{j=0}^{i-1}(k-j-1)} = \frac{(k-1!)}{c!\sum_{i=1}^{k-1}\frac{(k-1)!}{(k-i-1)!}} \tag{4.28}$$

$$= \frac{(k-1!)}{c!\sum_{i=1}^{k-c-1}\frac{(k-1)!}{(k-i-1)!}} \tag{4.29}$$

$$= \frac{1}{c!\sum_{i=1}^{k-1}\frac{1}{(k-i-1)!}} \tag{4.30}$$

$$= \frac{1}{c!\sum_{i=0}^{k-2}\frac{1}{i!}}, \tag{4.31}$$

and the denominator tends to $ec!$ as $k \to \infty$. □

Lemma 4.2.4. Let $T \in \mathcal{T}_{nbtree}$ be a tree, then

$$\mathrm{st}\left(\Pr_{G\in{*}\mathrm{PATH}_n^{n-c}}[T \text{ fails}]\right) = 1.$$

Proof. (sketch) By an analogous argument to how we proved Lemma 4.1.6 we get that

$$\Pr_{G\in{*}\mathrm{PATH}_n^{n-c}}[T \text{ fails}] \geq \prod_{i=0}^{n^{1/t}}\left(1 - \frac{(c+2)}{n-2i-2}\right),$$

standard part of which is also 1. The constant c appears because there is a $(1-\frac{c}{m})$ chance of finding a degree 0 in the induction step. □

Lemma 4.2.5. Let $\mathcal{G}'_k := \bigcup_{c\in\mathbb{N}} {*}\mathrm{PATH}_k^{k-c}$, then

$$\lim_{k\to\infty}\frac{|\mathcal{G}'_k|}{\left|{*}\mathrm{PATH}_k^{\leq}\right|} = 1.$$

In other words, $\mathcal{G}'_k$ is a major portion of ${*}\mathrm{PATH}_k^{\leq}$.

Proof. There is a (elementwise least) increasing sequence $\{c_k\}_{k>0}$ such that for each $k_0 > 0$ and $k \geq k_0$ we have ${*}\mathrm{PATH}_k^{k-c_{k_0}} \neq \varnothing$, moreover $\lim_{k\to\infty} c_k = \infty$.

${*}\mathrm{PATH}_k^{k-c}$ are disjoint for different choices of c, so by direct computation we have that the fraction $|\mathcal{G}'_k| \,/\, \left|{*}\mathrm{PATH}_k^{\leq}\right|$ is

$$\sum_{c=0}^{c_k}\frac{\left|{*}\mathrm{PATH}_k^{k-c}\right|}{\left|{*}\mathrm{PATH}_k^{\leq}\right|} = \sum_{c=0}^{c_k}\frac{1}{c!\sum_{i=0}^{k-2}\frac{1}{i!}} \tag{4.32}$$

$$= \left(\frac{1}{\sum_{i=0}^{k-2}\frac{1}{i!}}\right)\sum_{c=0}^{c_k}\frac{1}{c!}, \tag{4.33}$$

which tends to 1 as $k \to \infty$. □

Theorem 4.2.6.

$$\lim_{F_{nbtree}} {*}\mathrm{PATH}_n^{\leq}[\![(\exists v)(\exists u)(\forall w)((v \neq 0) \land (E(v,u)) \land (E(v,w) \to u = w))]\!] = \mathbf{0}$$

Proof. (sketch) The probability that a tree fails on any of the wide sequences $*\mathrm{PATH}_k^{k-c}$ is infinitesimal due to Lemma 4.2.4. This is true for any of their unions so for $\mathcal{G}'_k$ in particular. In Lemma 4.2.5 we proved that $\mathcal{G}'_k$ is a major portion of $*\mathrm{PATH}_k^{\leq}$, so by Lemma 2.4.9 the probability that a tree fails in $*\mathrm{PATH}^{\leq}$ is arbitrarily close to 1. ☐

So we have for a larger set of instances that their wide limit has no solution relative to F_{nbtree}. Moreover, we have the following.

Corollary 4.2.7.

$$\mathrm{Th}\left(\lim_{F_{nbtree}} *\mathrm{PATH}_n^{\leq}\right) = \mathrm{Th}\left(\lim_{F_{nbtree}} *\mathrm{PATH}_n\right)$$

Proof. Since the wide sequences $*\mathrm{PATH}_k^{k-c}$, for $c \in \mathbb{N}$, partition a major portion of $*\mathrm{PATH}_k^{\leq}$, and in each of them the probability of a tree in F_{nbtree} inspecting a degree 0 vertex is infinitesimal, we have that the wide limit of the major portion $\mathcal{G}'_k$ satisfies the theory $\mathrm{Th}(\lim_{F_{nbtree}} *\mathrm{PATH}_n)$ and therefore so does $\lim_{F_{nbtree}} *\mathrm{PATH}_n^{\leq}$. ☐

4.3 $\mathcal{G}_k = *\mathbf{DPATH}_k$

As $*\mathrm{PATH}_k$ was the wide sequence consisting of the hardest instances of LEAF the complete problem for **PPA**, we define $*\mathrm{DPATH}_k$ analogously but in the directed case so it consists of the hardest instances of SOURCE.OR.SINK the complete problem for **PPAD**.

Definition 4.3.1. We define $*\mathrm{DPATH}_k$ (the pointed directed paths on k vertices) as the set of all directed graphs G on the vertex set $\langle k\rangle$, where G is isomorphic to the path on k vertices such that $\deg_G^+(0) = 0$ and $\deg_G^-(1) = 1$.

But now, since we are working with directed graphs which have two types of neighbour sets $N_G^+(v) = \{w \in V_G; E_G(w, v)\}$ and $N_G^-(v) = \{w \in V_G; E_G(v, w)\}$, we would like to define a family F_{dtree} of those trees which can inspect either of the neighbour sets.

Definition 4.3.2. After we fix n, we define F_{dtree} as the set of all functions computed by some labeled tree with the following shape:

- Each non-leaf node is labeled by some $v \in \langle n\rangle$ and a symbol $\circ \in \{+, -\}$.
- For each $v \in \langle n\rangle$ and a node w there is an outgoing edge from w labeled $\{v\}$ and also an outgoing edge labeled $\varnothing$.
- Each leaf is labeled by some $m \in \langle g_n\rangle$.
- The depth of the tree is at most $g_n^{1/t}$ for some $t > \mathbb{N}$.

Computation of such a tree on an undirected graph G goes as follows. We interpret the non-leaf nodes as questions "what is $N_G^\circ(v)$?" and the edges as answers from our graph G, and thus we follow a path determined by G until we find a leaf, in which case the computation returns the label of the leaf.

We denote the set of such trees $\mathcal{T}_{dtree}$.

We will not cover details, but analysis of these trees in $\mathcal{T}_{dtree}$ finding the nonzero sink is more or less the same as the F_{nbtree} case for $*\mathrm{PATH}_k$, so we have the following.

Theorem 4.3.3.

$$\lim_{F_{nbtree}} *\mathrm{DPATH}_n[\![(\exists v)(\forall w)((v \neq 0) \wedge \neg E(v,w))]\!] = \mathbf{0}$$

Corollary 4.3.4. $\mathrm{Th}(\lim_{F_{nbtree}} *\mathrm{DPATH}_n)$ is complete.

In the type 2 complexity theory of $\mathbf{TFNP}^2$ we know that there is no (oracle polynomial time) reduction from LEAF to SOURCE.OR.SINK. An important question arises – is this reflected in the second order arithmetical expansion $K(*\mathrm{PATH}_n, F_{nbtree}, G_{nbtree})$? Where G_{nbtree} is defined analogously as G_{rud} but with the components in the tuple from F_{nbtree}. More specifically we ask the following.

Question 4.3.5. Consider an instance of SOURCE.OR.SINK defined by some $\Theta \in G_{nbtree}$. Let φ_Θ be the first order statement which says 'Θ has a solution'. Are all φ_Θ valid in $K(*\mathrm{PATH}_n, F_{nbtree}, G_{nbtree})$?

Concluding remarks

In this thesis we built a basic theory around wide limits of graphs, proved several general theorems and described the theories of the wide limits $\lim_F \mathrm{EDGE}_n$, $\lim_{F_{rud}} \mathrm{ALL}_n$, $\lim_{F_{nbtree}} *\mathrm{PATH}_n$, $\lim_{F_{nbtree}} *\mathrm{PATH}_n^{\leq}$ and $\lim_{F_{dtree}} *\mathrm{DPATH}_n$ and proved that they are complete (Corollary 2.3.6, Theorem 2.4.5 and corollaries 4.1.9, 4.2.7 and 4.3.4). We also proved that $\lim_{F_{rud}}^{G_{rud}} \mathrm{SK}_n^{1/2}$ contains a clique of size $\lfloor n/(2\ln n)\rfloor$ (Theorem 3.1.6) and also that nonexistence of finite cliques in not valid in the more complex wide limit $\lim_{F_{rud}}^{G_{rud}} \mathrm{CK}_n^{1/2}$ (Theorem 3.2.1).

During the development we planned to analyze wide limits the family F_{poly} of polynomial functions. In [Kra11] it was proven that forcing with F_{poly} results in quantifier elimination which implies that if an $\{E\}$-sentence holds in large enough $\mathcal{G}_k$, it has to hold in the limit. However the second order limit can still provide some information about the ability of polynomial time functions to search interesting subsets of $G \in \mathcal{G}_k$. In the end we did not get any new results about it. We want to mention that even though it would seem that the properties of F_{poly} limits directly correspond to the **P** vs. **NP** question it seems that something a bit different happens. The way the limit objects are defined, it is not enough that some polynomial time algorithm does not exist to see that we cannot witness some property in the limit, but it is important that no polynomial algorithm succeeds on any standard fraction of inputs. This more closely resembles the generic case polynomial time (see [GMMU07]).

Another natural question would be to consider structures over general languages than just the language of graphs, and allow trees to query atomic sentences. Other combinatorial structures like hypergraphs and tournaments could be considered. Furthermore, wide limits of finite universal algebras could be considered which could require a whole new theory. This all leads to the fact that generalized spectra, elementary classes of Σ_1^1-logic (see [Fag74]), with restricted vertex sets make up a wide sequence, there could be a connection to the theory of spectra of sentences.

Considering structures over arbitrary vocabularies opens the door to connecting the concept of a wide limit with the complexity of the Constraint Satisfaction Problem. Solutions to $\mathrm{CSP}(\mathcal{A})$ form a wide sequence and depending on the tractability of the class we could expect different behaviour from $\lim_F \mathrm{CSP}(\mathcal{A})_n$.

Finally, important direction in studying wide sequences would be to characterize some $\lim_F \mathcal{G}_n$ without the direct construction and therefore to prove upper and lower bounds for search problems in F.

Bibliography

[BCE+95] Paul Beame, Stephen Cook, Jeff Edmonds, Russell Impagliazzo, and Toniann Pitassi. The relative complexity of np search problems. In *Proceedings of the Twenty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '95, page 303–314, New York, NY, USA, 1995. Association for Computing Machinery.

[Fag74] Ronald Fagin. Generalized first-order spectra and polynomial-time recognizable sets. In Richard Karp, editor, *Complexity of computation*, volume 7, page 2741. SIAM-ASM Proceedings, 1974.

[Fag76] Ronald Fagin. Probabilities on finite models. *The Journal of Symbolic Logic*, 41(1):50–58, 1976.

[Gai64] Haim Gaifman. Concerning measures in first order calculi. *Israel journal of mathematics*, 2(1):1–18, 1964.

[GMMU07] Robert Gilman, Alexei G Miasnikov, Alexey D Myasnikov, and Alexander Ushakov. Report on generic case complexity. *arXiv preprint arXiv:0707.1364*, 2007.

[Gol14] Isaac Goldbring. Lecture notes on nonstandard analysis. *Ucla summer school in logic*, 2014.

[GP18] Paul W. Goldberg and Christos H. Papadimitriou. Towards a unified complexity theory of total functions. *J. Comput. System Sci.*, 94:167–192, 2018.

[Kay91] Richard Kaye. *Models of Peano arithmetic*, volume 15 of *Oxford Logic Guides*. The Clarendon Press, Oxford University Press, New York, 1991. Oxford Science Publications.

[Kra11] Jan Krajíček. *Forcing with random variables and proof complexity*, volume 382. Cambridge University Press, 2011.

[LS06] László Lovász and Balázs Szegedy. Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B*, 96(6):933–957, 2006.

[MP91] Nimrod Megiddo and Christos H Papadimitriou. On total functions, existence theorems and computational complexity. *Theoretical Computer Science*, 81(2):317–324, 1991.

[NDM13] Jaroslav Nešetřil and Patrice Ossona De Mendez. A model theory approach to structural limits. *arXiv preprint arXiv:1303.2865*, 2013.

[Raz07] Alexander A. Razborov. Flag algebras. *The Journal of Symbolic Logic*, 72(4):1239–1282, 2007.