Electronic Colloquium on Computational Complexity, Report No. 8 (2023)

Limits of structures and Total NP Search Problems*

Ondřej Ježil
ondrej.jezil@email.cz

Faculty of Mathematics and Physics, Charles University†

Abstract

For a class of finite graphs, we define a limit object relative to some computationally restricted class of functions. The properties of the limit object then reflect how a computationally restricted viewer "sees" a generic instance from the class. The construction uses Krajíček's forcing with random variables [7]. We prove sufficient conditions for universal and existential sentences to be valid in the limit, provide several examples, and prove that such a limit object can then be expanded to a model of weak arithmetic. We then take the limit of all finite pointed paths to obtain a model of arithmetic where the problem OntoWeakPigeon is total but Leaf (the complete problem for **PPA**) is not. This can be viewed as a logical separation of the oracle classes of total NP search problems, which in our setting implies standard nonreducibility of Leaf to OntoWeakPigeon.

1 Introduction

There exist several logical constructions of limits of classes of finite structures such as the ultraproduct and the compactness theorem. The latter was used in [2] 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

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

converges for every fixed connected graph F, where $\mathrm{hom}(F, G)$ denotes the set of all graph homomorphisms from F to G. This provided a framework (see [8]) 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 refer the interested reader to [10]. Another recent use of limit objects for the results of extremal combinatorics was by Razborov in [11].

* This work has been supported by Charles University Research Center program No.UNCE/SCI/022.

† Sokolovská 83, Prague, 186 75, The Czech Republic

ISSN 1433-8092

In this work, we define a new construction of a limit object. Given a class of finite graphs $\mathcal{G}$, whose vertex sets are initial segments of $\mathbb{N}$, we can stratify it into the sequence of sets $\{\mathcal{G}_k\}_{k=1}^{\infty}$ as follows

$$\mathcal{G}_k = \{G \in \mathcal{G}; G \text{ has } \{0, \dots, k-1\} \text{ as its vertex set}\}.$$

Our construction would yield a pseudofinite structure if $\lim_{k\to\infty}|\mathcal{G}_k| = 1$, but an ordinary application of the compactness theorem suffices for that, we therefore generally care about the case, where $\lim_{k\to\infty}|\mathcal{G}_k| = \infty$.[1] We call such a sequence of sets of graphs a *wide sequence* and the limit object its *wide limit*.

Let F be a class of functions with some computational restrictions, for example take F to be the set of functions computed by decision trees of some small depth. We define the wide limit denoted $\lim_F \mathcal{G}_n$, where n is a technical parameter to be defined later.

The wide limit $\lim_F \mathcal{G}_n$ is a Boolean-valued graph[2] — its edge relation does not only permit the truth values $\mathbf{0}$ and $\mathbf{1}$ but also many other values from some infinite complete Boolean algebra $\mathcal{B}$. This algebra is in fact also a σ-algebra with a measure μ on it, so to any statement formulated as a first order sentence φ we can assign a real number $\mu([\![\varphi]\!]) \in [0, 1]$ which measures how far is the truth value of φ (denoted $[\![\varphi]\!]$) from the value $\mathbf{0}$. The key method we use is arithmetical forcing with random variables, developed in [7], 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, validity of existential quantifiers corresponds to the ability of F to solve search problems over the class of graphs we are considering.

Our limit object can be expanded to the original model Krajíček's method would otherwise construct. We prove (Theorem 5.8) that the truth values of first order sentences concerning the object are preserved even when evaluated in the model of arithmetic relativized to the wide limit (under a mild condition on the family F).

As an application of this construction, we take the limit of all finite paths starting at the vertex 0 relative to the class of functions computed by oracle trees of subexponential depth and obtain the Boolean-valued graph $\lim_{F_{\mathrm{nb}}} {*}\mathrm{PATH}_n$ which is an infinite path with only one endpoint. This object is then expanded to a Boolean-valued model of weak second order arithmetic $K({*}\mathrm{PATH}_n, F_{\mathrm{nb}}, G_{\mathrm{nb}})$ in which every instance of OntoWeakPigeon has a solution. However, the object $\lim_{F_{\mathrm{nb}}} {*}\mathrm{PATH}_n$ in the model $K({*}\mathrm{PATH}_n, F_{\mathrm{nb}}, G_{\mathrm{nb}})$ is an instance of the **PPA**-complete problem Leaf which does not have a solution. This can be seen as a logical analogue of an oracle separation of these two classes, which is known to hold[3]. We then show the result implies a separation of those classes under stronger notion of reducibility.

[1] The case where the limit tends to some other positive number results in a structure which after collapsing to a two-valued boolean algebra becomes pseudofinite too.

[2] Generally, we can do this with any L-structures for some first order language L. The limit object is then a Boolean-valued L-structure $\lim_F \mathcal{G}_n$. In this work we restrict ourselves to the language of graphs $L = \{E\}$ to simplify the presentation.

[3] OntoWeakPigeon can be reduced to WeakPigeon which is known to be in **PPP** [6] and it is known [1] that Leaf cannot be reduced to any problem in **PPP**.

There is already an established connection between complexity of search problems and logic (namely bounded arithmetic, see [4]). The model we construct is not known nor expected to be a model of any theory which has been considered under these investigations. However, we show that open induction and open comprehension is valid in this model, and thus we show these principles along with the principle that OntoWeakPigeon is total cannot prove that the problem Leaf is total. The way the model is constructed also implies nonreducibility from Leaf to OntoWeakPigeon for subexponential time oracle machines. Moreover, one can at least in theory tweak our construction (e.g. by extending the family F_{nb}) to obtain a model of a stronger theory. This has been successfully done for several models already in [7, Chapter 10, Chapter 14, Chapter 21] using the switching lemma.

2 Preliminaries

By graphs we mean structures in a language with a single binary relation denoted E which is antireflexive and possibly symmetric if the graph in question is undirected. We will denote any particular graph by ω as it will be used in some sense as a sample of a discrete probability space. The edge relation of a particular graph ω will be denoted E_ω.

In the rest of this section we recall notions needed for Krajíček's forcing construction. Fundamental notion we use throughout the work is of nonstandard models of (true) arithmetic. Let L_{all} be the language containing the names of all relations and functions on the natural numbers and let $\mathrm{Th}_{L_{all}}(\mathbb{N})$ denote the set of true sentences in this language. By classical results of logic there exist L_{all}-structures in which all sentences from $\mathrm{Th}_{L_{all}}(\mathbb{N})$ are valid but which are not isomorphic to $\mathbb{N}$. These are called *nonstandard models* (of $\mathrm{Th}_{L_{all}}(\mathbb{N})$).

All nonstandard models of $\mathrm{Th}_{L_{all}}(\mathbb{N})$ (and even much weaker theories) contain an isomorphic copy of $\mathbb{N}$ as an initial segment. Therefore, we can assume that in fact all models we encounter satisfy $\mathbb{N} \subseteq \mathcal{M}$.

After considering a concrete nonstandard model $\mathcal{M}$ (of $\mathrm{Th}_{L_{all}}(\mathbb{N})$) we shall call the elements of $\mathcal{M} \setminus \mathbb{N}$ *nonstandard numbers*. These can be intuitively understood as "infinite natural numbers". The key feature of those elements is that all functions and relations from L_{all} are defined even on nonstandard numbers. This includes functions for coding sequences and sets by numbers, and therefore we can use notation like $a_0, \dots, a_{n-1}$ even for a nonstandard number n. The notation then means that for each $i \in \mathcal{M}$ such that $i < n$ we have an object a_i coded by a number in $\mathcal{M}$ and that this whole sequence is coded by some number $\{a_i\}_{i=0}^{n-1} \in \mathcal{M}$. For a nonstandard number $S \in \mathcal{M}$ coding a set we denote its nonstandard size (cardinality) to be $|S|$. In the case where we talk about a binary string x the notation $|x|$ denotes the bit length of x (which is nonstandard if x is).

In the next section we will fix a nonstandard model $\mathcal{M}$ which has the model theoretic property that it is $\aleph_1$-saturated. There is a self-contained construction of such model in [7, Appendix]. The only consequence of the $\aleph_1$-saturation we shall use is the following.

Property. Let $\{a_i\}_{i=0}^{\infty}$ be a sequence of standard numbers. Then there exists $t \in \mathcal{M} \setminus \mathbb{N}$

and a sequence $\{b_i\}_{i=0}^t \in \mathcal{M}$ such that for all $i \in \mathbb{N}$ it holds that $a_i = b_i$. We shall call the sequence of $\{b_i\}_{i=0}^t$ the *nonstandard prolongation of* $\{a_i\}_{i=0}^\infty$.

The language L_{all} contains symbols for all relations on $\mathbb{N}$. Since every sequence of numbers can be defined by some relation it turns out that in our case there is a unique nonstandard prolongation which matches the definition of the wide sequence (up to length which can be chosen arbitrarily high). We can therefore allow ourselves to use nonstandard numbers as indices of any sequences of objects unambiguously.

Any nonstandard model $\mathcal{M}$ can be extended to an ordered ring $\mathbb{Z}^\mathcal{M}$ by adding negative elements. This ring then can be extended to a fraction field $\mathbb{Q}^\mathcal{M}$. We shall call elements of $\mathbb{Q}^\mathcal{M}$ *$\mathcal{M}$-rationals.* The field $\mathbb{Q}^\mathcal{M}$ contains an isomorphic copy of $\mathbb{Q}$ as a substructure. We call an element in $\mathbb{Q}^\mathcal{M}$ with absolute valued greater than all $\frac{k}{1}, k \in \mathbb{N}$, *infinite* otherwise we call it *finite.* We call elements in $\mathbb{Q}^\mathcal{M}$ with absolute value smaller than all $\frac{1}{k}, k \in \mathbb{N}$ *infinitesimal.*

We will denote the set of finite $\mathcal{M}$-rationals as $\mathbb{Q}^\mathcal{M}_{fin}$ and one can check it forms an ordered ring.

Lemma (The existence of a standard part)**.** There is a function $\mathrm{st} : \mathbb{Q}^\mathcal{M}_{fin} \to \mathbb{R}$ assigning to each finite $\mathcal{M}$-rational a real number. st is a ring homomorphism and the kernel of st is exactly the ideal of infinitesimal numbers. When q is a finite $\mathcal{M}$-rational we call $\mathrm{st}(q)$ its *standard part.*

We shall use the structure $\mathbb{Q}^\mathcal{M}$ analogously to how hyperreal numbers are used in nonstandard analysis. For more details about nonstandard analysis we recommend [3] to the interested reader. The following result characterizes convergence of sequences of rational numbers using $\mathbb{Q}^\mathcal{M}$.

Theorem. Let $\{a_i\}_{i=0}^\infty$ be a sequence of rational numbers and let $r \in \mathbb{R}$. Then the following are equivalent.

- $\lim_{i\to\infty} a_i = r$
- For every $\{b_i\}_{i=0}^t$, $t \in \mathcal{M} \setminus \mathbb{N}$, which is a nonstandard prolongation of $\{a_i\}_{i=0}^\infty$, there is an nonstandard $s_0 \leq t$, such that for every nonstandard $s \leq s_0$: $\mathrm{st}(a_s) = r$.

It is important for forcing with random variables to consider discrete probability spaces of nonstandard size. We shall always use uniform distribution on the samples (although this is not necessary for the general construction). Thus, the probability of an event coded by an element $A \in \mathcal{M}$ is then just the $\mathcal{M}$-rational number $|A|/|S|$ where S is the set of samples of such a space.

We conclude this section by restating classical inequalities used in this work using the nonstandard approach.

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

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

Theorem (Exponential inequality)**.** Let $x \in \mathcal{M} \setminus \mathbb{N}$, then

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

3 Wide limits

3.1 The definition

We shall define a wide limit of every sequence of the following form.

Definition 3.1. A sequence of sets of graphs $\{\mathcal{G}_k\}_{k=1}^{\infty}$ is called *a wide sequence* if the following holds:

- Every graph $\omega \in \mathcal{G}_k$ has the vertex set $\{0, \dots, k-1\}$.
- $\lim_{k\to\infty} |\mathcal{G}_k| = \infty$.

By abuse of notation we will simply talk about a wide sequence $\mathcal{G}_k$ instead of $\{\mathcal{G}_k\}_{k=1}^{\infty}$. Since a wide limit is a Boolean-valued graph, we need to construct a Boolean algebra in which the truth evaluation of statements shall take place.

For the construction of the Boolean algebra we will closely follow [7, Chapter 1] albeit with slight changes. Let us now fix for the rest of this work an $\aleph_1$-saturated model of $\mathrm{Th}_{L_{all}}(\mathbb{N})$ which we will denote $\mathcal{M}$.

Definition 3.2. Let $n \in \mathcal{M}$. We define

$$\mathcal{A}_n = \{A \subseteq \{0, \dots, n-1\}; A \in \mathcal{M}\},$$

in words $\mathcal{A}_n$ is the set of subsets of $\{0, \dots, n-1\}$ coded by an element in $\mathcal{M}$. This is a boolean algebra and to each $A \in \mathcal{A}_n$ we assign an $\mathcal{M}$-rational $|A|/n$ which we call its *counting measure.*

Even though $\mathcal{A}_n$ is a boolean algebra with a "measure" it is not a σ-algebra. Indeed, $\mathcal{A}_n$ contains all singletons, but the countable set of those elements in $\{0, \dots, n-1\}$ with only finitely many predecessors is not definable by compactness. However, having infinite joins and meets at our disposal allows us to interpret quantifiers in the boolean valued case, so we now want to 'tweak' this Boolean algebra.

Definition 3.3. Let $\mathcal{I}$ be the ideal of $\mathcal{A}_n$ consisting of elements with infinitesimal counting measure. We define $\mathcal{B}_n = \mathcal{A}_n/\mathcal{I}$. Each element in $\mathcal{B}_n$ is of the form $A/\mathcal{I}$, where $A \in \mathcal{A}_n$, and we define $\mu(A/\mathcal{I}) = \mathrm{st}(|A|/n)$. We will denote the maximal element of $\mathcal{B}_n$ by $\mathbf{1}$ and the minimal element by $\mathbf{0}$.

One can easily check that μ is well-defined since for all $A \in \mathcal{I}$ it holds that $\mathrm{st}(|A|/n) = 0$. The measure μ is called the Loeb measure. The following then holds.

Lemma 3.4 ([7, Lemma 1.2.1])**.** $\mathcal{B}_n$ is a σ-algebra with a real valued measure μ. Moreover, $\mathcal{B}_n$ is a complete boolean algebra.

It is important to note that $\mathbf{1} \in \mathcal{B}_n$ is the only element of $\mathcal{B}_n$ with measure $\mu(\mathbf{1}) = 1$ and similarly $\mathbf{0} \in \mathcal{B}_n$ is the only element with measure $\mu(\mathbf{0}) = 0$. Also, for $B, B' \in \mathcal{B}_n$ the inequality $B \leq B'$ implies $\mu(B) \leq \mu(B')$.

We now define precisely what we mean by the family of functions F relative to which we will be taking the wide limit. This is still a part of Krajíček's construction, we just modify it to make it compatible with our setup — where we start with a wide sequence.

For every $k \in \mathbb{N}$ the set $\mathcal{G}_k$ is finite and thus can be coded by a number. Therefore, there is a nonstandard prolongation of this sequence, and we can consider the set coded by the nonstandard number $\mathcal{G}_n$, which matches the definition of the wide sequence in $\mathcal{M}$.

Definition 3.5. Let $\{\mathcal{G}_k\}_{k=1}^{\infty}$ be a wide sequence and let $n \in \mathcal{M} \setminus \mathbb{N}$. We say that F is *a family of random variables on* $\mathcal{G}_n$ if every $\alpha \in F$ is a function coded by a number in $\mathcal{M}$ with domain $\mathcal{G}_n$ and taking values in $\mathcal{M}$. We say $\alpha \in F$ is an F-vertex if for all $\omega \in \mathcal{G}_n$ it holds that $\alpha(\omega) \in \{0, \ldots, n-1\}$. The set of all F-vertices is denoted $U(F)$.

If the wide sequence $\{\mathcal{G}_k\}_{k=1}^{\infty}$ and the number $n \in \mathcal{M} \setminus \mathbb{N}$ is clear from context we just say F is a family of random variables. This is for now everything we need to recall from [7], and we can proceed to define the central object of our work.

Definition 3.6 (The wide limit)**.** Let $\{\mathcal{G}_k\}_{k=1}^{\infty}$ be a wide sequence, let $n \in \mathcal{M} \setminus \mathbb{N}$ and let F be a family of random variables on $\mathcal{G}_n$. We define *the wide limit* $\lim_{F,n}\{\mathcal{G}_k\}_{k=1}^{\infty}$ as a $\mathcal{B}_n$-valued structure in the language consisting of a single binary relation symbol $\{E\}$ as follows. The universe of the wide limit is taken as the set of all F-vertices. We now inductively define the truth values for all $\{E\}$-sentences.

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

By abuse of notation we will denote the wide limit $\lim_{F,n}\{\mathcal{G}_k\}_{k=1}^{\infty}$ by $\lim_F \mathcal{G}_n$. To stress in which boolean valued structure is the truth evaluation $[\![\,-\,]\!]$ taking place we will sometimes denote the evaluation $\mathcal{C}_1[\![\,-\,]\!]$, $\mathcal{C}_2[\![\,-\,]\!]$ for boolean valued structures $\mathcal{C}_1$ and $\mathcal{C}_2$ respectively. Furthermore, if $\mathcal{C}_1[\![\varphi]\!] = \mathbf{1}$ for some sentence φ we say φ is *valid* in $\mathcal{C}_1$.

Note that since $\mathcal{G}_n$ can be recovered from F as the domain of its elements the wide limit strictly speaking only depends on F. We keep $\mathcal{G}_n$ in the notation to cover the situation where we have a very general family of functions (e.g. the family of polynomial functions F_{PV}) which can be applied to every wide sequence. Thus, the notation $\lim_F \mathcal{G}_n$ means that F is restricted to those functions which take elements of $\mathcal{G}_n$ as an input even when F possibly contains other functions too.

The variability of the parameter n may also seem unnecessary and indeed in our applications it is the case, but generally there are examples of wide sequences where n directly affects the wide limit.

Example 3.7. Let F_{const} be the family of all constant functions with domain $\mathcal{G}_n$ and range anywhere in $\mathcal{M}$. Let

$$\mathcal{G}_k = \begin{cases} \{(\{0,\dots,k-1\},E); |E| = 2, (0,1) \in E\} & k \text{ even} \\ \{(\{0,\dots,k-1\},E); |E| = 1, (0,1) \notin E\} & k \text{ odd} \end{cases}$$

then

$$\lim_{F_{const}} \mathcal{G}_n[\![E(0,1)]\!] = \begin{cases} \mathbf{1} & n \text{ even} \\ \mathbf{0} & n \text{ odd.} \end{cases}$$

3.2 An example of a wide limit relative to shallow decision trees

Now we shall define the first nontrivial family of random variables relative to which we shall take wide limits of several sequences. The functions in the family will be computed by shallow decision trees. So the shape of the wide limit reflects what can 'superlogarithmic' trees witness in the wide sequence with probability arbitrarily close to 1.

Definition 3.8. Let $\mathcal{T}_{\mathrm{rud}}$ be a family of labeled rooted binary trees in $\mathcal{M}$ of the following form. At each vertex the tree is labeled by an element of $\{0,\dots,n-1\} \times \{0,\dots,n-1\}$ and the two outgoing edges incident to it are labeled as 0 and 1 respectively. The leaves are labeled by an element of $\mathcal{M}$. The depth of the tree is bounded by a number of a form $n^{1/t}$ (rounded to the nearest element of $\mathcal{M}$) for some $t \in \mathcal{M} \setminus \mathbb{N}$.

A *computation* of a $T \in \mathcal{T}_{rud}$ on some $\omega \in \mathcal{G}_n$ is defined as follows. Start at the root and interpret each label (i,j) of the vertex as a question whether the pair (i,j) is in the edge set E_ω and follow a path through T reading 1 as a positive answer and 0 as a negative answer. The label of the leaf visited at the end of the path is the output of T on ω, denoted $T(\omega)$.

We define F_{rud} to be the set of all functions computed by a tree $T \in \mathcal{T}_{\mathrm{rud}}$.

The simplest wide sequence we shall consider is the following sequence of sets of undirected graphs with exactly one edge.

Definition 3.9. $\mathrm{EDGE}_k = \{(\{0,\dots,k-1\},E); |E| = 1\}$

Since any $\omega \in \mathrm{EDGE}_k$ has only 1 edge in all potential $k \cdot (k-1)/2$ edges, it is not likely a shallow tree will find the edge. This is the idea behind the proof of the following theorem.

Theorem 3.10.

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

Proof. Let $\alpha, \beta \in U(F_{\mathrm{rud}})$, we proceed by proving that

$$[\![E(\alpha,\beta)]\!] = \mathbf{0}$$

which is enough to prove the theorem since even an infinite disjunction of the values $\mathbf{0}$ is $\mathbf{0}$.

Let α and β be computed by $T \in \mathcal{T}_{\text{rud}}$ and $S \in \mathcal{T}_{\text{rud}}$ respectively. Let the depth of both T and S be at most $n^{1/t}$, where $t \in \mathcal{M} \setminus \mathbb{N}$. Walk down T from the root and always prolong the path along the edge labeled 0. On this path we have a set of at most $n^{1/t}$ different pairs of vertices and a label of the leaf l_T.

We do the same for S, and we find another set of at most $n^{1/t}$ pairs of vertices and a label of the leaf l_S. l_S and l_T are then combined to one last pair (l_S, l_T). Now we just need to compute the probability that none of these $2n^{1/t}+1$ pairs of vertices are not in the edge set E_ω.

There are $\binom{n}{2}$ different graphs in EDGE_n and $\binom{n-4n^{1/t}-2}{2}$ graphs which fulfill our requirements. The probability is thus

$$\begin{aligned}\frac{\binom{n-4n^{1/t}-2}{2}}{\binom{n}{2}} &= \frac{(n-4n^{1/t}-2)(n-4n^{1/t}-3)}{n(n-1)}\\ &\geq \frac{(n-4n^{1/t}-3)^2}{n^2}\\ &\geq \left(1-\frac{4n^{1/t}+3}{n}\right)^2\\ &\geq \left(1-\frac{8n^{1/t}+6}{n}\right)\end{aligned}$$

after taking the standard part of the last line we get $\text{st}(1-\frac{8n^{1/t}+6}{n}) = 1$. Therefore, $\mu([\![E(\alpha,\beta)]\!]) = 0$ and $[\![E(\alpha,\beta)]\!] = \mathbf{0}$. □

3.3 Sufficient conditions for validity of existential and universal sentences

To analyze wide limits we need ideally to know the values of sentences which describe properties whose complexity we are interested in. Generally this can be hard, so in this section we prove sufficient conditions at least for the validity of universal and existential sentences.

We will start with the simpler condition for the validity of universal sentences. This is important also because we would like to know that a wide limit of a wide sequence of graphs (i.e. antireflexive $\{E\}$-structures) is also a graph and that a wide limit of a wide sequence of undirected graphs (directed graphs with E symmetric) is an undirected graph. All of these properties are expressible as universal sentences.

Theorem 3.11. Let $\mathcal{G}_k$ be a wide sequence and let F be any family of random variables. Let $\varphi(x_0,\ldots,x_{l-1})$ be an open $\{E\}$-formula and assume that

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

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

Proof. By induction in $\mathcal{M}$ we have that $\mathrm{st}(\Pr_{\omega\in\mathcal{G}_n}[\omega \models (\forall\overline{x})\varphi(x)]) = 1$. Therefore, we have for every tuple of F-vertices $\overline{\alpha}$ that $[\![\varphi(\overline{\alpha})]\!] = \mathbf{1}$. Now

$$\begin{aligned}[\![(\forall x)\varphi(x)]\!] &= \bigwedge_{\overline{\alpha}\in U(F)^l} [\![\varphi(\overline{\alpha})]\!] \\ &= \bigwedge_{\overline{\alpha}\in U(F)^l} \mathbf{1} \\ &= \mathbf{1}.\end{aligned}$$

□

Corollary 3.12. Let $\mathcal{G}_k$ be a wide sequence and F any family of random variables.

- If all $\omega \in \mathcal{G}_k$, $k \in \mathbb{N}$, are directed graphs ($\{E\}$-structures satisfying that E is antireflexive) then $\lim_F \mathcal{G}_n$ is a Boolean-valued $\{E\}$-structure in which the antireflexivity of E is valid (i.e. $\lim_F \mathcal{G}_n$ is a Boolean-valued graph).
- If all $\omega \in \mathcal{G}_k$, $k \in \mathbb{N}$, are undirected graphs (directed graphs where E is symmetric) then $\lim_F \mathcal{G}_n$ is an $\{E\}$-structure in which both antireflexivity and symmetry of E is valid. (i.e. $\lim_F$ is a Boolean-valued undirected graph)

Now to give a sufficient condition for the validity of an existential sentence $(\exists\overline{x})\varphi(\overline{x})$ we use the auxiliary value of *density of* $\varphi(x_0, \dots, x_{l-1})$ defined as the probability that a random graph $\omega \in \mathcal{G}_k$ and a random tuple $\overline{a} \in \{0, \dots, k-1\}^l$ satisfy $\omega \models \varphi(\overline{a})$ and show that the limiting density gives a lower bound for the measure of $[\![(\exists\overline{x})\varphi(\overline{x})]\!]$.

Theorem 3.13. Let $\mathcal{G}_k$ be a wide sequence and let F be a family of random variables which contains all constant functions. Let $\varphi(x_0, \dots, x_{l-1})$ be an open $\{E\}$-formula and let $p \in [0, 1]$. Assume that

$$\lim_{k\to\infty} \Pr_{\substack{\omega\in\mathcal{G}_k \\ \overline{a}}} [\omega \models \varphi(\overline{a})] \geq p,$$

where $\overline{a}$ is sampled uniformly over all elements of $\{0, \dots, k-1\}^l$. Then

$$\mu(\lim_F \mathcal{G}_n [\![(\exists\overline{x})\varphi(\overline{x})]\!]) \geq p.$$

In particular if $p = 1$ then $\lim_F \mathcal{G}_n [\![(\exists\overline{x})\varphi(\overline{x})]\!] = \mathbf{1}$.

Proof. Consider an array C indexed by $\omega \in \mathcal{G}_n$ and $\overline{a} \in \{0, \dots, n-1\}^l$ such that

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

By the assumption and induction in $\mathcal{M}$ we have that

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

We now claim that there exists a specific $\overline{b} \in \{0, \ldots, n-1\}^l$ such that $\mathrm{st}(\Pr_{\omega\in\mathcal{G}_n}[\omega \models \varphi(\overline{b})]) \geq p$. Assume for contradiction that the claim is false. Then

$$\begin{aligned} \frac{1}{|\mathcal{G}_n|n^l} \sum_{\omega\in\mathcal{G}_n} \sum_{\overline{a}} C_{\omega,\overline{a}} &= \frac{1}{n^l} \sum_{\overline{a}} \Pr_{\omega\in\mathcal{G}_n} [\omega \models \varphi(\overline{a})] \\ &\leq \Pr_{\omega\in\mathcal{G}_n} [\omega \models \varphi(\overline{a}_0)], \end{aligned}$$

where we pick $\overline{a}_0$ such that it maximizes $\Pr_{\omega\in\mathcal{G}_n}[\omega \models \varphi(\overline{a}_0)]$. But after taking the standard part of the inequality we obtain that

$$\mathrm{st}\left(\frac{1}{n^l|\mathcal{G}_n|} \sum_{\omega\in\mathcal{G}_n} \sum_{\overline{a}} C_{\omega,\overline{a}}\right) \leq \mathrm{st}(\Pr_{\omega\in\mathcal{G}_n} [\omega \models \varphi(\overline{a}_0)]) < p.$$

Which is a contradiction and so the claim is true. Let $\overline{\gamma}_b$ be a tuple of constant functions which is at every sample equal to $\overline{b}$. We have

$$\begin{aligned} [\![(\exists\overline{x})\varphi(\overline{x})]\!] &= \bigvee_{\overline{\alpha}\in U(F)^l} [\![\varphi(\overline{\alpha})]\!] \\ &\geq [\![\varphi(\overline{\gamma}_b)]\!] \end{aligned}$$

and by taking μ of this inequality we finally obtain that $\mu([\![(\exists\overline{x})\varphi(\overline{x})]\!]) \geq p$. $\square$

In the following example we use Theorem 3.13 to show that in the wide limit of graphs which have exactly one large clique and no other edges the nonexistence of a standard sized clique cannot be valid relative to any F with all constants.

Example 3.14. Consider the wide sequence

$$\mathrm{SK}_k^{1/2} = \{(\{0, \ldots, k-1\}, E); E \text{ has a clique of size } \lfloor k/2 \rfloor \text{ and no other edges}\}.$$

We will check that for an $\{E\}$-formula $\varphi_l(\overline{x})$ which states that $\overline{x}$ forms a clique of size l we have

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

Notice that we can compute the probability for a fixed $\overline{a}$ such that $a_i \neq a_j$ whenever $i \neq j$, since the ratio of tuples containing some vertex twice is infinitesimal. So we have

$$\begin{aligned} \Pr_{\omega\in\mathrm{SK}_k^{1/2}} [\omega \models \varphi_l(\overline{a})] &= \prod_{i=0}^{l-1} \left(1 - \frac{k - \lfloor k/2 \rfloor}{k-i}\right) \\ &\geq \left(1 - \frac{k - \lfloor k/2 \rfloor}{k-l}\right)^l \\ &\geq \left(1 - \frac{1}{2(1-l/k)} - \frac{1}{k-l}\right)^l \end{aligned}$$

and since $l \in \mathbb{N}$ we just take the limit of the inner expression. But one can see that $\lim_{k\to\infty}(1 - l/k) = 1$ and that $\lim_{k\to\infty}(1/(k-l)) = 1$.

Now by Theorem 3.13 we obtain that for *any* F that contains all constants we have

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

The following example demonstrates that Theorem 3.11 cannot be generalized to a similar hypothesis as Theorem 3.13.

Example 3.15. Let $\mathcal{G}_k$ consist of all undirected graphs on the vertex set $\{0, \dots, k-1\}$ with exactly $\lceil \frac{k(k-1)}{2\log(k)} \rceil$edges. One can see that

$$\lim_{k\to\infty} \Pr_{\substack{\omega\in\mathcal{G}_k\\ x,y}} [\omega \models \neg E(x,y)] = 1,$$

but in fact $\lim_{F_{\mathrm{rud}}} \mathcal{G}_n[\![(\forall x)(\forall y)\neg E(x,y)]\!] = \mathbf{0}$.

Let $t \in \mathcal{M} \setminus \mathbb{N}$ such that $n^{1/t}$ is not bounded above by a standard number. Let T be a tree which queries on all paths a fixed set of $n^{1/t}$ different potential edges. If we prove that any such set in $\mathcal{G}_n$ has to contain at least one edge with probability infinitesimally close to 1 then we can construct F_{rud}-vertices α and β using T such that $[\![E(\alpha,\beta)]\!] = \mathbf{1}$ by simply taking T and labeling each leaf on a path which finds an edge with one of the vertices incident to this edge.

Let S be the set of potential edges queried by T and let $m = \binom{n}{2}$. Now we have

$$\begin{aligned}
\Pr_{\omega\in\mathcal{G}_n}[S \text{ contains no edge in } \omega] &= \frac{(m - n^{1/t})!(m - \lceil \frac{m}{\log n} \rceil!)}{m!(m - \lceil \frac{m}{\log m} \rceil - n^{1/t})!} \\
&= \prod_{i=0}^{n^{1/t}-1} \frac{m - \lceil \frac{m}{\log n} \rceil - i}{m - i} \\
&\le \left(1 - \frac{\lceil \frac{m}{\log n} \rceil}{m}\right)^{n^{1/t}} \\
&\le \left(1 - \frac{1}{2\log n}\right)^{n^{1/t}}
\end{aligned}$$

standard part of which is for all $k \in \mathbb{N}$ bounded above by

$$\mathrm{st}\left(\left(1 - \frac{1}{2\log n}\right)^{k\cdot 2\log n}\right) \le e^{-k}$$

which tends to 0 as $k \to \infty$.

4 A wide limit of Leaf instances relative to oracle trees

The class of total **NP** search problems **TFNP**, first defined in [9], consists of all relations on binary strings $P(x, y)$ such that:

- (*verifiability in polynomial time*) There is a polynomial time machine M which, given x, y, can decide whether $P(x, y)$ holds.
- (*totality*) There exists a polynomial p and for every x there exists at least one y satisfying $|y| \leq p(|x|)$ such that $P(x, y)$ holds.

Two particular problems are relevant for us.

The problem Leaf is formulated as follows. An instance is given by a number k and a undirected graph ω on the vertex set $\{0, \ldots, 2^{|k|} - 1\}$, presented by a Boolean circuit of polynomial size in $|k|$ computing its neighborhood function, such that $\deg_\omega(0) = 1$ and $\forall v : \deg_\omega(v) \leq 2$. The task is then to find some nonzero v with $\deg_\omega(v) = 1$. The corresponding combinatorial principle being the handshaking lemma, which assures the problem is total.

The problem OntoWeakPigeon is formulated as follows. An instance is given by a number k and two functions $A : \{0, \ldots, 2^{|k|} - 1\} \to \{0, \ldots, 2^{|k|-1} - 1\}$ and $B : \{0, \ldots, 2^{|k|-1} - 1\} \to \{0, \ldots, 2^{|k|} - 1\}$, each presented by a Boolean circuit of polynomial size in $|k|$. The task is then to find some x such that $B(A(x)) \neq x$ or some y such that $A(B(y)) \neq y$. The corresponding combinatorial principle being the bijective weak pigeonhole principle, which assures the problem is total. The domain of A is twice as large as its range, so B and A cannot form a pair of inverse functions between their respective domains.

So far, we presented what is called 'type 1' problem in [1]. We are interested about the 'type 2' problems which replace the input function(s) with oracle(s). So in the 'type 2' Leaf problem, the input is a pair (α, x) where α is an oracle which describes the neighbor function on G with vertex set $\{0, \ldots, 2^{|x|} - 1\}$. For the 'type 2' OntoWeakPigeon problem, the input is a triple (α, β, x), where α and β are oracles describing the functions $\alpha : \{0, \ldots, 2^{|x|} - 1\} \to \{0, \ldots, 2^{|x|-1} - 1\}$ and $\beta : \{0, \ldots, 2^{|x|-1} - 1\} \to \{0, \ldots, 2^{|x|} - 1\}$.

The associated computational models for the 'type 1' problems are Turing machines and for the 'type 2' problems oracle Turing machines.

4.1 The wide limit and oracle trees

The wide sequence $*\mathrm{PATH}_k$ (pointed paths on k vertices) consists of all undirected graphs ω on the vertex set $\{0, 1, \ldots, k-1\}$ which are isomorphic to a path with $k-1$ edges connecting all vertices and $\deg_\omega(0) = 1$. Graphs in $*\mathrm{PATH}_k$ are 'the hardest instances of Leaf' so we can expect the wide limit to reflect the complexity of these instances relative to the family F we choose.

Since each $\omega \in *\mathrm{PATH}_k$ has only $k-1$ edges we can proceed similarly to the proof of Theorem 3.10 to get the following.

Lemma 4.1. $\lim_{F_{\mathrm{rud}}} *\mathrm{PATH}_n[\![(\exists x)(\exists y)E(x, y)]\!] = \mathbf{0}$

To get a result which reflects the properties of the wide sequence more faithfully we will define a new family of random variables on $*\mathrm{PATH}_n$.

Definition 4.2. We define $\mathcal{T}_{\mathrm{nb}}$ as the set of all labeled rooted trees of the following shape:

- Each non-leaf node is labeled by some $v \in \{0, \ldots, n-1\}$.
- For each $\{u, w\} \subseteq \{0, \ldots, n-1\}$ and a node v there is an outgoing edge from v labeled $\{u, w\}$ (it can be that $u = w$).
- Each leaf is labeled by some $m \in \mathcal{M}$.
- The depth of the tree is defined as the maximal number of edges in a path from the root, and we require it is at most $n^{1/t}$ (rounded to the nearest element of $\mathcal{M}$) for some $t \in \mathcal{M} \setminus \mathbb{N}$.

The *computation* of such a tree in $\mathcal{T}_{\mathrm{nb}}$ on $\omega \in *\mathrm{PATH}_n$ is defined as follows. We build a path by starting at the root and interpreting every vertex labeled by some v as a question 'what are the neighbors of the vertex v?' and we follow the output edge with the answer and continue analogously until we find a leaf. The label of the leaf is defined to be the output of the computation.

We define F_{nb} to be the set of all functions on $*\mathrm{PATH}_n$ which are computed by some $T \in \mathcal{T}_{\mathrm{nb}}$.

The trees computing the functions in F_{nb} can be thought of as a protocol describing the behavior of a machine M communicating with an oracle describing a particular $\omega \in *\mathrm{PATH}_n$. In the study of total NP search problems presented by oracles, we usually denote the size of the object by some $2^{|x|}$ where x is an additional input to the problems. If $2^{|x|} = n$ then $n^{1/t} = 2^{|x|/t}$ which for $t \in \mathcal{M} \setminus \mathbb{N}$ corresponds to protocols describing non-uniform subexponential-time computations. If we prove that no tuple of F_{nb}-vertices satisfies some open $\{E\}$-formula in $\lim_{F_{\mathrm{nb}}} *\mathrm{PATH}$ we also prove that subexponential-time oracle machines cannot solve the corresponding type 2 problem on a non-diminishing fraction of the inputs. In the rest of this section we proceed to prove that $\lim_{F_{\mathrm{nb}}} *\mathrm{PATH}_n$ has no vertex with degree 1 other than the vertex 0.

To do so, we will consider computations of trees on samples with different nonstandard lengths. For the rest of this section we put $\mathcal{G}_m = *\mathrm{PATH}_m$ for all $m \in \mathcal{M}$, but we can assume m to be smaller than n. We define $\mathcal{T}_{\mathrm{nb}}^{(m)}$ to be the subset of $\mathcal{T}_{\mathrm{nb}}$ consisting of all the trees that have the vertex labels from $\{0, \ldots, m-1\}$. For trees in $\mathcal{T}_{\mathrm{nb}}^{(m)}$ we can extend the definition of a computation to input graphs from $\mathcal{G}_m$ in a straight forward way.

Definition 4.3. We say a tree $T \in \mathcal{T}_{\mathrm{nb}}^{(m)}$ *fails* on $\omega \in \mathcal{G}_m$ if the output of T on ω has degree 2.

Definition 4.4. Let $m \in \mathcal{M}$, $v \in \{0, \ldots, m-1\}$ and $\{u, w\} \subseteq \{0, \ldots, m-1\}$ we define

$$\mathcal{G}_m^{v:\{u,w\}} = \{\omega \in \mathcal{G}_m; \omega \models E(v, u) \wedge E(v, w)\}.$$

Lemma 4.5. Let $m \in \mathcal{M}$ and let u, v and w be distinct elements of $\{1, \ldots, m-1\}$. Then there are bijections:

$$\begin{aligned} \mathcal{G}_m^{v:\{u,w\}} &\cong \mathcal{G}_{m-2} \times \{L, R\} \\ \mathcal{G}_m^{v:\{u,0\}} &\cong \mathcal{G}_{m-2} \\ \mathcal{G}_m^{v:\{u\}} &\cong \mathcal{G}_{m-1} \\ \mathcal{G}_m^{0:\{u\}} &\cong \mathcal{G}_{m-1} \end{aligned}$$

Proof. For the first case a bijection can be given as follows. Contract u, v and w to just one vertex $\min\{u, v, w\}$ and if u is closer to 0 than w pick L otherwise pick R and relabel the remaining vertices using a function 'new' which has a property that if u', v' remain and $u' < v'$ as numbers then $\mathrm{new}(u') < \mathrm{new}(w')$ and the range of new is $\{0, \ldots, m-2\}$. This can be inverted by first renaming the vertices using new^{-1} and then replacing $\min\{u, v, w\}$ by a path (u, v, w) with the orientation given either by L or R.

The second bijection is almost the same, but the orientation is clear since u is always the neighbor further from 0 since the other neighbor is 0.

The third and fourth bijections are given by just removing one end of the graph and relabeling. □

Definition 4.6. Let $m \in \mathcal{M}$ and $v \in \{0, \ldots, m-1\}$. Let u and w be elements of $\{0, \ldots, m-1\} \setminus \{v\}$ and let $T \in \mathcal{T}_{\mathrm{nb}}^{(m)}$ be a tree with the root labeled v. By $T_{v:\{u,w\}}$ we denote the induced subtree whose root is the vertex neighboring the root of T via the edge labeled $\{u, w\}$.

Lemma 4.7. Let $m \in \mathcal{M}$. Let $T \in \mathcal{T}_{\mathrm{nb}}^{(m)}$ be a tree with the root labeled $v \neq 0$. For each u and w which are distinct elements of $\{0, \ldots, m-1\} \setminus \{v\}$ there exists a tree $\tilde{T}_{v:\{u,w\}} \in \mathcal{T}_{\mathrm{nb}}^{(m-2)}$ of the same depth as $T_{v:\{u,w\}}$ such that

$$\Pr_{\omega \in \mathcal{G}_m}[T_{v:\{u,w\}} \text{ fails} \mid \omega \models E(v, u) \wedge E(v, w)] = \Pr_{\omega \in \mathcal{G}_{m-2}}[\tilde{T}_{v:\{u,w\}} \text{ fails}].$$

If T has the root labeled 0 then there exists a tree $\tilde{T}_{0:\{u\}} \in \mathcal{T}_{\mathrm{nb}}^{(m-1)}$ of the same depth as $T_{0:\{u\}}$ such that

$$\Pr_{\omega \in \mathcal{G}_m}[T_{0:\{u\}} \text{ fails} \mid \omega \models E(0, u)] = \Pr_{\omega \in \mathcal{G}_{m-1}}[\tilde{T}_{0:\{u\}} \text{ fails}].$$

Proof. In the case where the root is labeled by $v \in \{1, \ldots, m-1\}$ we can construct the tree $\tilde{T}_{v:\{u,w\}}$ by simply relabeling vertices of $T_{v:\{u,w\}}$. We use the relabeling function 'new' from the proof of Lemma 4.5. Now for every $\omega \in \mathcal{G}_m$ there is by the first bijection in Lemma 4.5 a uniquely determined $\omega' \in \mathcal{G}_{m-2}$. The computation of $\tilde{T}_{v:\{u,w\}}$ on ω' is then of the same shape as the computation of $T_{v:\{u,w\}}$ on ω assuming $\omega \models E(v, u) \wedge E(v, w)$. And $\tilde{T}_{v:\{u,w\}}(\omega')$ has the same degree in ω' as $T_{v:\{u,w\}}(\omega)$ does in ω.

The case where the root is labeled by 0 is analogous, but we instead use the relabeling from the fourth bijection in Lemma 4.5. □

Lemma 4.8. Let $T \in \mathcal{T}_{\text{nb}}^{(m)}$ of depth $d \in \mathcal{M}$ and let $d \leq m$. Then we have

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

Proof. We proceed by induction on d. The case $d = 0$ follows from

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

Now for the inductive case we assume the lemma holds for $d - 1$, and prove it for d. If the root of T is labeled 0 we proceed as follows. For a given T let u_0 be the vertex which minimizes the value $\Pr_{\omega \in \mathcal{G}_m}[T \text{ fails} \mid E(0, u_0)]$ which exists by the least number principle in $\mathcal{M}$. Then by Lemma 4.7 and the induction hypothesis

$$\begin{aligned}
\Pr_{\omega \in \mathcal{G}_m}[T \text{ fails}] &\geq \Pr_{\omega \in \mathcal{G}_m}[T \text{ fails} \mid E(0, u_0)] \\
&= \Pr_{\omega \in \mathcal{G}_{m-1}}[\tilde{T}_{0:\{u_0\}} \text{ fails}] \\
&\geq \prod_{i=0}^{d-1} \left(1 - \frac{2}{m - 2i - 3}\right) \\
&\geq \prod_{i=0}^{d} \left(1 - \frac{2}{m - 2i - 2}\right).
\end{aligned}$$

Now for the case where the root of T is labeled by nonzero v. First we note that

$$\Pr_{\omega \in \mathcal{G}_m}[v \text{ has degree } 2 \wedge \neg E(v, 0)] = 1 - \frac{2}{m-1}.$$

Now we choose distinct u_0, w_0 such that they minimize

$$\Pr_{\omega \in \mathcal{G}_m}[T_{v:\{u_0, w_0\}} \text{ fails} \mid E(v, u_0) \wedge E(v, w_0)].$$

Then by the Lemma 4.7 and the induction hypothesis we have

$$\begin{aligned}
\Pr_{\omega \in \mathcal{G}_m}[T \text{ fails}] &\geq \left(1 - \frac{2}{m-1}\right) \Pr_{\omega \in \mathcal{G}_m}[T_{v:\{u_0, w_0\}} \text{ fails} \mid E(v, u_0) \wedge E(v, w_0)] \\
&= \left(1 - \frac{2}{m-1}\right) \Pr_{\omega \in \mathcal{G}_{m-2}}[\tilde{T}_{v:\{u_0, w_0\}} \text{ fails}] \\
&\geq \left(1 - \frac{2}{m-1}\right) \prod_{i=0}^{d-1} \left(1 - \frac{2}{m - 2i - 4}\right) \\
&\geq \left(1 - \frac{2}{m-1}\right) \prod_{i=1}^{d} \left(1 - \frac{2}{m - 2i - 2}\right) \\
&\geq \prod_{i=0}^{d} \left(1 - \frac{2}{m - 2i - 2}\right).
\end{aligned}$$

□

Lemma 4.9. Let $T \in \mathcal{T}_{\mathrm{nb}}$, then $\operatorname{st}\left(\Pr_{\omega \in \mathcal{G}_n}[T \text{ fails}]\right) = 1$.

Proof. The depth of T is bounded by $n^{1/t}$ for some $t \in \mathcal{M} \setminus \mathbb{N}$. We have by Lemma 4.8 that

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

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

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

Finally, in the next theorem we prove that a formalization of 'there is a nonzero vertex of degree 1' is not valid in $\lim_{F_{\mathrm{nb}}} {*}\mathrm{PATH}_n$ and in fact its boolean value is $\mathbf{0}$.

Theorem 4.10.

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

Proof. By expanding the left-hand side of the statement we get

$$\bigvee_{\alpha \in U(F_{\mathrm{nb}})} \bigvee_{\beta \in U(F_{\mathrm{nb}})} \bigwedge_{\gamma \in U(F_{\mathrm{nb}})} [\![\alpha \neq 0 \wedge E(\alpha,\beta) \wedge (E(\alpha,\gamma) \rightarrow \gamma = \beta)]\!].$$

Therefore, it is enough if we prove that for each F_{nb}-vertices α and β there exists an F_{nb}-vertex γ such that

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

For any $\alpha, \beta \in U(F_{\mathrm{nb}})$ we can append the tree computing β to every leaf of a tree computing α. This is still a tree in $\mathcal{T}_{\mathrm{nb}}$ as its depth is at most twice the maximum of depths of the original trees. By relabeling the leaves of the resulting tree we can obtain a tree computing a function

$$\gamma(\omega) = \begin{cases} v & \text{if } \deg_\omega(\alpha(\omega)) = 1 \text{ and } v \text{ is the only neighbor of } \alpha(\omega) \\ w & \text{if } \deg_\omega(\alpha(\omega)) = 2,\ w \text{ is a neighbor of } \alpha(\omega) \text{ and } w \neq \beta(\omega). \end{cases}$$

This is obviously an F_{nb}-vertex. Let us assume for contradiction that

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

By definition this gives us

$$0 < \operatorname{st}\left(\Pr_{\omega \in \mathcal{G}_n}[\alpha(\omega) \neq 0 \wedge E_\omega(\alpha(\omega),\beta(\omega)) \wedge (E(\alpha(\omega),\gamma(\omega)) \rightarrow \gamma(\omega) = \beta(\omega))]\right)$$
$$\leq \operatorname{st}\left(\Pr_{\omega \in \mathcal{G}_n}[\alpha(\omega) \neq 0 \wedge \deg_\omega(\alpha(\omega)) = 1]\right),$$

but this is in contradiction with Lemma 4.9. $\square$

5 The expanded model with a LEAF instance without a solution and with total ONTOWEAKPIGEON

As a part of the proof of Theorem 4.10 we proved what can be reformulated as the statement that oracle instances of LEAF are not in oracle time $\mathcal{O}(2^{f(|x|)})$ with $f \in o(|x|^{1/c})$ for every $c \in \mathbb{N}$ even when we just require it to be correct on any nondiminishing ratio of inputs as $|x|$ grows. In this section we proceed to compare strength of (type 2) NP search problems not only with oracle **FP** but also with other NP search problems via relative consistency of their totality and nontotality. We will show that there is a model of weak second order arithmetic in which the problem LEAF is not total even though ONTOWEAKPIGEON is.

5.1 The structures $K(F, G)$

We will now recall the construction of second order models of weak arithmetic $K(F, G)$ defined in [7, Chapter 5]. We will take the liberty to define them as an extension of the definition of a wide limit to obtain structures $K(\mathcal{G}_n, F, G)$ [4] which under the right conditions result in a structure in some sublanguage of L_{all} with two sorts: numbers and bounded sets of numbers which contains the wide limit as an object of the second sort.

Definition 5.1. Let $L \subseteq L_{all}$. This determines a *language* L^2 which we get by adding to L second order variables $X, Y, \ldots$ whose intended interpretation are bounded sets and the equality symbol for second order variables (denoted the same as the first order one). All second order variables are treated as function symbols and can form terms with the first order terms as arguments.

We will also use the second order variables as relation symbols, and we define the atomic formula $X(x_0, \ldots, x_{k-1})$ simply to be evaluated as the formula $X(x_0, \ldots, x_{k-1}) \neq 0$.

Now we assume we fix a number n, a wide sequence $\mathcal{G}_k$ and a family of random variables on $\mathcal{G}_n$ which all together determine a wide limit $\lim_F \mathcal{G}_n$.

Definition 5.2. We define $\mathcal{M}_n \subseteq \mathcal{M}$ to be the subset of $\mathcal{M}$ consisting of all numbers bounded above by $2^{n^{1/t}}$ for some $t \in \mathcal{M} \setminus \mathbb{N}$.

Definition 5.3. We define $L_n \subseteq L_{all}$ to contain all relation symbols from L_{all} and all functions from L_{all} for which their values on any element of $\mathcal{M}_n$ is still in $\mathcal{M}_n$. We say F is L_n-closed if for every function symbol $f \in L_n$ we have that $f(\alpha_0, \ldots, \alpha_{k-1}) \in F$.

Note that $\mathcal{M}_n$ is then a substructure of the L_n-reduct of $\mathcal{M}$.

Definition 5.4. We say that G is a *family of random functions* (on $\mathcal{G}_n$) if every $\Theta \in G$ assigns to each $\omega \in \mathcal{G}_n$ a function $\Theta_\omega \in \mathcal{M}_n$.

[4] This notation is just making some parameters of the construction explicit, the models constructed can be obtained by the original method without first constructing the wide limit. Our contribution is in observing that the truth values of first order sentences concerning the wide limit is preserved between the wide limit and the structure $K(\mathcal{G}_n, F, G)$.

We say G is F-compatible if for every $\alpha \in F, \Theta \in G$ we have that the function $\Theta(\alpha)$ defined as

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

is in fact in F.

Definition 5.5. Let F be an L_n-closed family of random variables with values in $\mathcal{M}_n$. Let G be an F-compatible family of random functions. We define $K(\mathcal{G}_n, F, G)$ to be a $\mathcal{B}_n$-valued L_n^2 structure with first order sort of the universe F and second order sort of the universe G. The valuation of formulas is then given by the following inductive definition:

- $[\![\alpha = \beta]\!] = \{\omega \in \mathcal{G}_n; \alpha(\omega) = \beta(\omega)\}/\mathcal{I}$, where $\alpha, \beta \in F$
- $[\![R(\alpha_0, \ldots, \alpha_{k-1})]\!] = \{\omega \in \mathcal{G}_n; \omega \models R(\alpha_0(\omega), \ldots, \alpha_{k-1}(\omega))\}/\mathcal{I}$, where $\alpha_0, \ldots, \alpha_{k-1}$ are from F and is R a relation symbol in L_n
- $[\![\Theta = \Xi]\!] = \{\omega \in \mathcal{G}_n; \Theta_\omega = \Xi_\omega\}/\mathcal{I}$, where $\Theta, \Xi \in G$
- $[\![(\forall x)A(x)]\!] = \bigwedge_{\alpha \in F} [\![A(\alpha)]\!]$
- $[\![(\exists x)A(x)]\!] = \bigvee_{\alpha \in F} [\![A(\alpha)]\!]$
- $[\![(\forall X)A(X)]\!] = \bigwedge_{\Theta \in G} [\![A(\Theta)]\!]$
- $[\![(\exists X)A(X)]\!] = \bigvee_{\Theta \in G} [\![A(\Theta)]\!]$.

5.2 Preservation of sentences concerning the wide limit

We will now prove (under a mild condition on F) that in a structure $K(\mathcal{G}_n, F, G)$ which represents the wide limit $\lim_F \mathcal{G}_n$ by a second order object are the values of all sentences regarding the object the same as in the wide limit. This lets us construct models in which an object elementary equivalent to the wide limit might be desired.

Definition 5.6. We say that the edge relation of the wide limit $\lim_F \mathcal{G}_n$ is *represented in G by* Γ if $\Gamma \in G$ and for all $\alpha, \beta \in U(F)$ we have that

$$K(\mathcal{G}_n, F, G)[\![\Gamma(\alpha, \beta)]\!] = \lim_F \mathcal{G}_n[\![E(\alpha, \beta)]\!].$$

Definition 5.7. We say a family of random variables F has *restrictable ranges* if for every $\alpha \in F$ and $m \in \mathcal{M}_n$ there is $\tilde{\alpha}_m \in F$ such that

$$\tilde{\alpha}_m(\omega) = \begin{cases} \alpha(\omega) & \alpha(\omega) < m \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 5.8. Let φ be a $\{E\}$-sentence. Let F be L_n-closed and have restrictable ranges and let G be F-compatible. Let the edge relation of the wide limit $\lim_F \mathcal{G}_n$ be represented in G by Γ. We define $\tilde{\varphi}(\Gamma)$ to be the L_n^2-sentence obtained by replacing all the occurrences of the relation symbol E by Γ, keeping the structure of the logical connectives and replacing all quantifiers $(\forall x)(\dots)$ by $(\forall x)(x < n \to (\dots))$ and $(\exists x)(\dots)$ by $(\exists x)(x < n \wedge \dots)$.

Then we have that for all $\{E\}$-sentences that

$$\lim_F \mathcal{G}_n[\![\varphi]\!] = K(\mathcal{G}_n, F, G)[\![\tilde{\varphi}(\Gamma)]\!].$$

Proof. We will prove that for all $\{E\}$-formulas $\varphi(\overline{x})$ and all $\overline{\alpha} \in F$ we have that

$$\lim_F \mathcal{G}_n[\![\varphi(\overline{\alpha})]\!] = K(\mathcal{G}_n, F, G)[\![\tilde{\varphi}(\Gamma, \overline{\alpha})]\!].$$

We proceed by induction on the complexity of the formula. The case for atomic formulas is clear and the step for logical connectives also since $[\![\ - \ ]\!]$ commutes with them. With the induction step for negation in hand it is now enough to prove the inductive step for the universal quantifier.

We assume that the statement works for a formula of lower complexity $\varphi(y, \overline{x})$. By the restrictability of ranges in F we get that for all $\beta \in F$ there is $\tilde{\beta}_n \in U(F)$ such that

$$K(\mathcal{G}_n, F, G)[\![\tilde{\varphi}(\Gamma, \tilde{\beta}_n, \overline{\alpha})]\!] \le K(\mathcal{G}_n, F, G)[\![\beta < n \to \tilde{\varphi}(\Gamma, \beta, \overline{\alpha})]\!].$$

Now we have that for all $\overline{\alpha} \in U(F)$

$$\begin{aligned}
K(\mathcal{G}_n, F, G)[\![(\forall y)\tilde{\varphi}(\Gamma, y, \overline{\alpha})]\!] &= \bigwedge_{\alpha \in F} K(\mathcal{G}_n, F, G)[\![\beta < n \to \tilde{\varphi}(\Gamma, \beta, \overline{\alpha})]\!] \\
&= \bigwedge_{\tilde{\beta}_n \in U(F)} K(\mathcal{G}_n, F, G)[\![\tilde{\varphi}(\Gamma, \tilde{\beta}_n, \overline{\alpha})]\!] \\
&= \bigwedge_{\tilde{\beta}_n \in U(F)} \lim_F \mathcal{G}_n[\![\varphi(\tilde{\beta}_n, \overline{\alpha})]\!] \\
&= \lim_F \mathcal{G}_n[\![(\forall y)\varphi(y, \overline{\alpha})]\!].
\end{aligned}$$

□

5.3 Failure of totality of Leaf

Now we are in a situation that lets us construct a model of weak second order arithmetic that contains an instance of the problem Leaf without a solution. Consider a suitable family G_{nb} in which we can define not only the wide limit itself but instances of some other search problem. We can then ask: 'Do all these instances have a solution?' This is a way to compare the strength of the different total NP search problems by relative unprovability results. We will pick the family G_{nb} such that validity of totality of some search problem P implies the nonexistence of a suitable reduction from Leaf to P.

Definition 5.9. Let G_{nb} be the family of all random functions on $*\mathrm{PATH}_n$ such that for each $\Theta \in G_{\mathrm{nb}}$ there exists a tuple $(\gamma_0, \ldots, \gamma_{m-1}) \in \mathcal{M}$ so that $\gamma_i \in F_{\mathrm{nb}}$ and

$$\Theta(\alpha)(\omega) = \begin{cases} \gamma_{\alpha(\omega)}(\omega) & \alpha(\omega) < m \\ 0 & \text{otherwise.} \end{cases}$$

In the models $\mathcal{M}_n$ we are working with there is a pairing function $\langle i, j\rangle$ which can code pairs of numbers by a single number thus we can represent functions of any finite arity by functions from G_{nb}.

One can understand the tuples which compute the random functions from G_{nb} as tuples of protocols describing the computations of subexponential oracle machines. Such a tuple defines a function which is at every index of the tuple computed using queries to some $\omega \in *\mathrm{PATH}_n$. If we prove that every instance of a search problem P represented by such a tuple has a solution in $K(*\mathrm{PATH}_n, F_{\mathrm{nb}}, G_{\mathrm{nb}})$, and we know that Leaf in $K(*\mathrm{PATH}_n, F_{\mathrm{nb}}, G_{\mathrm{nb}})$ is not total, which implies nonexistence of a subexponential oracle machine which converts solutions of P to solutions of Leaf even on any standard fraction of instances from $*\mathrm{PATH}_n$ and thus a nonexistence of a many-one reduction from Leaf to P as defined in [1].

Lemma 5.10.

1. F_{nb} has restrictable ranges
2. F_{nb} is L_n-closed
3. G_{nb} is F_{nb}-compatible
4. G_{nb} represents the edge relation of $\lim_{F_{\mathrm{nb}}} *\mathrm{PATH}_n$.

Proof. 1, 2: Here we can proceed simply by relabeling the leaves of the trees computing the functions from F_{nb}.

3: Assume that $\Theta \in G_{\mathrm{nb}}$ is computed by a tuple $(\gamma_0, \ldots, \gamma_{m-1})$. By induction in $\mathcal{M}$ there exists $t \in \mathcal{M} \setminus \mathbb{N}$ such that $\forall i \in \{0, \ldots, m-1\}$ the depth of γ_i is at most $n^{1/t}$. Therefore, for all $\alpha \in F_{\mathrm{nb}}$ we have that $\Theta(\alpha)$ has also depth at most $n^{1/t'}$ for some $t' \in \mathcal{M} \setminus \mathbb{N}$. Thus, G_{nb} is F_{nb}-compatible.

4: Let $\gamma_{\langle i,j\rangle} \in F_{\mathrm{nb}}$ be computed by a tree in $\mathcal{T}_{\mathrm{nb}}$ which queries i and outputs 1 if the neighbor set contains j otherwise it outputs 0. Let Γ be computed by a tuple $(\gamma_{\langle i,j\rangle})_{i,j=0}^{n-1}$. Then we have

$$K(*\mathrm{PATH}_n, F_{\mathrm{nb}}, G_{\mathrm{nb}})[\![\Gamma(\alpha, \beta)]\!] = \lim_{F_{\mathrm{nb}}} *\mathrm{PATH}_n[\![E(\alpha, \beta)]\!].$$

□

Definition 5.11. The L_n^2-formula $\varphi_{\textsc{Leaf}}(X, Y, m)$ is defined as the disjunction of the following formulas

$(X(0) \neq Y(0) \vee X(0) = 0)$
$(\exists x)((x < m) \wedge (X(x) > m-1 \vee Y(x) > m-1))$
$(\exists x)((x < m) \wedge ((X(x) = x \wedge Y(x) \neq x) \vee (X(x) \neq x \wedge Y(x) = x)))$
$(\exists x)((x < m) \wedge (Y(X(x)) \neq x \wedge X(X(x)) \neq x) \vee (X(Y(x)) \neq x \wedge Y(Y(x)) \neq x)))$
$(\exists x)((0 < x < m) \wedge (X(x) = Y(x) \wedge X(x) \neq x)),$

this formula formalizes that if X and Y are functions representing the neighbor set of each $x < m$ as $\{X(x), Y(x)\} \setminus \{x\}$ and 0 has only one neighbor then there has to exist another $y < x$ which also has only one neighbor.

Theorem 5.12.

$$K(*\mathrm{PATH}_n, F, G)[\![(\exists X)(\exists Y)(\exists m)\neg\varphi_{\textsc{Leaf}}(X, Y, m)]\!] = \mathbf{1}$$

Proof. We can find $\Theta_1, \Theta_2 \in G_{\mathrm{nb}}$ such that for each $v \in \{0, \ldots, n-1\}$ we have that $\{\Theta_1(v)(\omega), \Theta_2(v)(\omega)\}$ is the neighbor set of v on $\omega \in *\mathrm{PATH}_n$. (We can just query v and split the answer between Θ_1 and Θ_2.)

By Theorem 4.10 we know that $\lim_F \mathcal{G}_n$ has one degree 1 vertex and all other vertices of degree 2 and by Lemma 5.10 we know that it can be represented by some $\Gamma \in G_{\mathrm{nb}}$. Furthermore, we can verify that

$$[\![(\Gamma(\alpha, \beta)) \equiv (\Theta_1(\alpha) = \beta \vee \Theta_2(\alpha) = \beta)]\!] = \mathbf{1},$$

thus Θ_1 and Θ_2 do not satisfy the last disjunct of $\varphi_{\textsc{Leaf}}$ otherwise it would be in contradiction with Theorem 4.10. By their construction and the definition of $*\mathrm{PATH}_k$ we have that (Θ_1, Θ_2, n) does not satisfy the other disjuncts either. ☐

5.4 Totality of OntoWeakPigeon

Definition 5.13. The L^2_n formula $\varphi_{\textsc{OntoWeakPigeon}}(X, Y, m)$ is defined as the disjunction of the following formulas

$$\begin{aligned}
&(\exists x)((x < 2m) \wedge (X(x) > m-1))\\
&(\exists y)((y < m) \wedge Y(y) > m-1))\\
&(\exists x)((x < 2m) \wedge Y(X(x)) \neq x)\\
&(\exists y)((y < m) \wedge X(Y(y)) \neq y)
\end{aligned}$$

it formalizes the bijective weak pigeonhole principle which claims that any pair of functions

$$\begin{aligned}
X :&\{0, \ldots, 2m-1\} \to \{0, \ldots, m-1\}\\
Y :&\{0, \ldots, m-1\} \to \{0, \ldots 2m-1\}
\end{aligned}$$

is not a pair of inverse bijections.

To prove that $\varphi_{\textsc{OntoWeakPigeon}}(X, Y, m)$ is valid in $K(*\mathrm{PATH}_n, F_{\mathrm{nb}}, G_{\mathrm{nb}})$ we will construct a tree which finds some x such that $Y_\omega(X_\omega(x)) \neq x$ or $X_\omega(x) > m-1$ with probability infinitesimally close to one.

Definition 5.14. Let $\Theta, \Xi \in G_{\mathrm{nb}}$, and $\zeta \in F_{\mathrm{nb}}$. We say that a tree $T \in \mathcal{T}_{\mathrm{nb}}$ *fails for* (Θ, Ξ, ζ) *on* ω if

$$\Theta_\omega(T(\omega)) < \zeta(\omega) \quad \text{and} \quad \Xi_\omega(\Theta_\omega(T(\omega))) = T(\omega).$$

In words if T does not witness the failure of Ξ being the inverse function to Θ.

Lemma 5.15. Let $\Theta, \Xi \in G_{\text{nb}}$ and $\zeta \in F_{\text{nb}}$. Then there is a tree T such that

$$\operatorname{st}\left(\Pr_{\omega \in \mathcal{G}_n}[T \text{ fails for } (\Theta, \Xi, \zeta)]\right) = 0.$$

Proof. Without loss of generality we may assume that ζ is actually constant, and its value is $r \in \mathcal{M}_n$ which we pick to be the least possible output of ζ on any sample. Furthermore, let Θ be computed by $(\theta_0, \ldots, \theta_{2r-1})$ and Ξ by $(\xi_0, \ldots, \xi_{r-1})$.

We construct T by stages and at each stage it will have some potential output. First we notice that at the beginning stage there is at least one $i \in \{0, \ldots, 2r-1\}$ such that the probability that $\theta_i < r$ or $\xi_{\theta_i} = i$ is at most $\frac{1}{2}$. The tree T_0 is thus the constant tree always outputting i.

Assume T_{d-1} have been constructed and pick any path $p \in T_{d-1}$. If p did not fail we leave it as it is otherwise we extend T_{d-1} along this path and after extending all such paths this will become the new stage T_d. The path p has a leaf with some label i. We can check whether i fails by first appending the tree θ_i to this path and then to each new leaf (labeled with a number $< r$) appending ξ_{θ_i}, let the leaves which confirm the nonfailure of i be labeled by i. Now consider a path p' extending p without determined output. We claim that there is $j \in \{0, \ldots, 2r-1\}$ such that

$$\Pr_{\mathcal{G}_n}[\theta_j < r \wedge \xi_{\theta_j} = j \mid p' \text{ is compatible with } \omega] \leq \frac{1}{2},$$

where p' being compatible with ω means that the computation along p' agrees with the edge labels which would be chosen according to ω.

To prove the claim we notice that along p' it was confirmed that already d-many distinct elements of $\{0, \ldots, 2r-1\}$ are in bijection with some d-many elements of the set $\{0, \ldots, r-1\}$. Therefore, to fail further there are only at most $(r-d)$-many other values j' in $\{0, \ldots, 2r-1\}$ for which it holds that $\xi_{\theta_{j'}} = j'$. By an analogous argument to the proof of Theorem 3.13 this is enough to show that at least for one of them the claim holds since $\frac{r-d}{2r} \leq \frac{1}{2}$. Thus, we let j to be the label of the leaf of p' which concludes the construction.

Therefore, by construction for each $d \in \mathcal{M}_n$, $d < 2r$ we have

$$\Pr_{\omega \in \mathcal{G}_n}[T_d \text{ fails for } (\Theta, \Xi, \zeta)] \leq 2^{-d}.$$

If r is in $\mathcal{M}_n \setminus \mathbb{N}$ then we put $T = T_{t'}$ for any nonstandard t' such that the depth of T is still bounded by some $n^{1/t}$, where $t \in \mathcal{M}_n \setminus \mathbb{N}$. Otherwise, we put $T = T_{2r-1}$ and since this tree can go through the whole range of Θ it can never fail. $\square$

Theorem 5.16.

$$K(*\mathrm{PATH}_n, F_{\text{nb}}, G_{\text{nb}})[\![(\forall X)(\forall Y)(\forall m)\varphi_{\textsc{OntoWeakPigeon}}(X, Y, m)]\!] = \mathbf{1}$$

Proof. By Lemma 5.15 we can construct for each (Θ, Ξ, ζ) a tree T which computes some function α which validates the third disjunct of $\varphi_{\textsc{OntoWeakPigeon}}$. $\square$

Theorem 5.17. Let $\varphi(x)$ be an L^2_n-formula with parameters from F_{nb} and G_{nb}. Then for every $m \in \mathcal{M}_n$ the open comprehension principle

$$(\exists X)(\forall y < m)(X(y) \equiv \varphi(y))$$

and the open induction principle

$$\neg\varphi(0) \vee \varphi(m) \vee (\exists x < m)(\varphi(x) \wedge \neg\varphi(x+1))$$

are both valid in $K(*\text{PATH}, F_{\text{nb}}, G_{\text{nb}})$.

Proof. This can be proven completely analogously to [7, Lemma 20.2.5]. □

Compiling the results we have about $K(*\text{PATH}, F_{\text{nb}}, G_{\text{nb}})$ we get the following.

Corollary 5.18. In the structure $K(*\text{PATH}, F_{\text{nb}}, G_{\text{nb}})$ the following are valid:

- open induction with parameters from F_{nb} and G_{nb}
- open comprehension with parameters from F_{nb} and G_{nb}
- every instance of OntoWeakPigeon has a solution
- there is an instance of Leaf which does *not* have a solution.

Concluding remarks

We have to note that the problem OntoWeakPigeon has not been considered in the context of oracle NP search problems and the proof of Theorem 5.16 cannot be adapted to prove that every instance of the stronger WeakPigeon [5] has a solution in $K(*\text{PATH}, F_{\text{nb}}, G_{\text{nb}})$ because the presence of the inverse function is essential to the construction of the witness.

A stronger problem called SourceOrSink is well established in the study of NP search problems (it is the complete problem for **PPAD**, see [1]) and can be formulated as follows: Given a directed graph ω on the vertex set $\{0, \dots, 2^{|x|} - 1\}$ with the property that any vertex v has outdegree bounded by 1 and indegree also bounded by 1 and the indegree of the zero vertex is 0 find a nonzero vertex which is either a source or a sink. In the type 2 setting the problem is given by a tuple (α, β, x), where x is a binary string and α and β functions presented by an oracle with domain $\{0, \dots, 2^{|x|} - 1\}$ computing the potential successor or predecessor of a given vertex.

It was established in [1] that Leaf is not many-one reducible to SourceOrSink and therefore this nonreducibility may be reflected in our model $K(*\text{PATH}, F_{\text{nb}}, G_{\text{nb}})$. The way SourceOrSink is presented is similar to how OntoWeakPigeon is presented, and thus a similar strategy could be potentially used to solve the following problem.

Problem. Let $\varphi_{\text{SourceOrSink}}(X, Y, m)$ be the formula formalizing that (X, Y, m) as an instance of SourceOrSink has a solution. Is it true that

$$K(*\text{PATH}, F_{\text{nb}}, G_{\text{nb}})[\![(\forall X)(\forall Y)(\forall m)\varphi_{\text{SourceOrSink}}(X, Y, m)]\!] = \mathbf{1}?$$

[5] The problem to witness that $\alpha : \{0, \dots, 2^{|x|} - 1\} \to \{0, \dots, 2^{|x|-1} - 1\}$ is not injective.

Acknowledgement

This work is based on the author's master's thesis [5] which was completed under the supervision of Jan Krajíček. The author also thanks Eitetsu Ken for comments on a draft of this paper.

References

[1] 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.

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

[3] Isaac Goldbring. Lecture notes on nonstandard analysis, 2014. Available at `https://www.math.uci.edu/~isaac/NSA%20notes.pdf` (last accessed 9th of January 2023).

[4] J. Hanika. *Search problems and Bounded Arithmetic*. PhD thesis, Charles University, Prague, 2004.

[5] Ondřej Ježil. Pseudofinite structures and limits. Master's thesis, Charles University, Prague, 2022. Also available in Electronic Colloquium on Computational Complexity.

[6] Emil Jeřábek. Integer factoring and modular square roots. *Journal of Computer and System Sciences*, 82(2):380–394, 2016.

[7] Jan Krajíček. *Forcing with random variables and proof complexity*, volume 382 of *London Mathematical Society Lecture Note Series.* Cambridge University Press, 2011.

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

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

[10] Jaroslav Nešetřil and Patrice Ossona de Mendez. A model theory approach to structural limits. *Commentationes Mathematicae Universitatis Carolinae*, 53:581–603, 11 2012.

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

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il