FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Ondřej Ježil

Pseudofinite structures

Department of Algebra

Supervisor of the master thesis: prof. RNDr. Jan Krajíček Ph.D.
Study programme: Matemathics
Study branch: Mathemathics 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

Dedication.

Title: Pseudofinite structures

Author: Ondřej Ježil

Department: Department of Algebra

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

Abstract: Abstract.

Keywords: key words

Contents

Introduction

There exist several logical constructions of limits of classes of finite structures such as the ultraproduct construction or using the compactness theorem. The latter 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 graph $\{G_k\}_{k>0}$ 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 such as 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 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 diferent 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.

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. Instead of studying the density of substructures we study these so called wide limits both generally and by analyzing concrete examples and tying them with complexity-theoretical statements.

The key method we use is arithmetical forcing with random variables which allows us to construct models of (weak) arithmetical theories and by restricting to a language of graphs gives us boolean valued graphs. An important resource for this technique is [Kra10]. In these boolean valued graphs, witnessing of existential quantifiers corresponds to complexity theoretical properties of the class of graphs we are considering.

(TODO: Comment on structure of chapters.)

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 concept for us are the nonstandard models of true arithmetic.

The ambient model $\mathcal{M}$

We call a model of $\mathrm{Th}(\mathbb{N})$ nonstandard if it is not isomorphic to $\mathbb{N}$. Every model of $\mathrm{Th}(\mathbb{N})$ of course 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 [Kra10] there is an explicting 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 just need to know that the model is nonstandard and the following property holds because of the $\aleph_1$-saturation.

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 any a_k with high enough index then there is also some nonstandard $s < t$ such that b_s satisfies the property P.

(TODO: Mention induction)

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 so-called $\mathcal{M}$-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 consist 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 is infinite. One can check that $\mathbb{Q}^{\mathcal{M}}$ fulfills the axioms of hyperreal numbers which can be used as an alternative foundation to the concepts of mathematical analysis. This implies that there is a canonical injection $\mathbb{R} \hookrightarrow \mathbb{Q}^{\mathcal{M}}$ of ordered fields. The following is an important result which we use throughout the thesis.

Theorem. Let $q \in \mathbb{Q}^{\mathcal{M}}$ finite. Then there exist unique $r \in \mathbb{R}$ and an infinitesimal $m \in \mathbb{Q}^{\mathcal{M}}$ such that $q = r + m$.

We use the notation $\mathrm{st}(q) := r$ and call r it the **standard part of** q.

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. Then

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

if and only if for any nonstandard $n \in \mathcal{M}$ we have that $\mathrm{st}(c_n) = r$.

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 $y \in \mathcal{M}$ and $x \in \mathbb{Q}^{\mathcal{M}}, x \geq 0$, 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) For every x there exists at least one y which is at most polynomially longer than x 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].

The semantic subclasses are defined as all problems reducible to some problem corresponding 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. Given a graph G on the vertex set $\langle k)$ such that $\deg_G(0) = 1$ and $\forall v : \deg_G(v) \leq 2$, find some nonzero v with $\deg_G(v) = 2$. The corresponding combinatorial principle being the handshaking lemma, which assures the problem is total. And the class **PPAD**, directed polynomial parity argument, with the complete problem SOURCE.OR.SINK formulated as follows. Given a directed graph G on the vertex set $\langle k)$ such that the vertex 0 is a source and $\forall v : \deg^+_G(v), \deg^-_G(v) \leq 1$ 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) 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, the graphs given by oracles are exponential in size and thus easily separated from $\mathbf{FP}^2$, the type 2 version of polynomial time functions. More importantly we have that $\mathbf{PPAD} \subsetneq \mathbf{PPA}$ as type 2 complexity classes. Intuitively one can forget the orientation to get the undirected version, but cannot consistently assign orientation to an undirected edges of a large graph.

The traditional model of computation are the Turing machines and for the type 2 problems oracle Turing machines. 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 is

1. Forcing with random variables and the limit

1.1 Setup

Our goal in this chapter is to provide a definition of a limit of a set an infinite set of finite graphs in which arbitrarily large graphs occur.

The following definition makes our requirements of such a class of graphs precise.

Definition 1.1.1. Let $\{\mathcal{G}_k\}_{k>0}$ be a sequence of finite sets of finite graphs. We call it a **wide sequence** if the following hold.

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

The second condition guarantees that $\mathcal{G}_n$ is an infinite set for $n > \mathbb{N}$. Many interesting classes of graphs form a wide sequence if we restrict the vertex-sets to $\langle g_k \rangle$, where $\{g_k\}_{k>0}$ can be taken as the increasing sequence of all cardinalities in such a class.

Example 1.1.2. TODO: Add some examples!

1.2 The first order limit

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 $n \in \mathcal{M} \setminus \mathbb{N}$.

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.

Definition 1.2.1. Let $\mathcal{A} := \{A \in \mathcal{M}; A \subseteq \Omega\}$ 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

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

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.

Now we define a $\mathcal{B}$-valued arithmetical model through which we define the $\mathcal{B}$-valued first order limit of $\mathcal{G}_k$ relative to a family of arithmetical functions.

Definition 1.2.4. Let $L \subseteq L_{all}$ and let F be a non-empty set of functions in $\mathcal{M}$. We call it an **L-closed family** if it satisfies the following:

- The domain of any function in F is Ω and the range is $\mathcal{M}$.
- F is closed under all L-functions and contains all L constants, where the L-functions are interpreted by composition

  $$f(\alpha_1, \dots, \alpha_k)(\omega) := f(\alpha_1(\omega), \dots, \alpha_k(\omega)),$$

  for $k \in \mathbb{N}$, $f \in L$ k-ary and $\alpha_1, \dots, \alpha_k \in F$.

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.

Definition 1.2.5. Let F be an L-closed family for some $L \subseteq L_{all}$. Then $K(F)$ will denote a $\mathcal{B}$-valued L-structure defined as follows.

The universe of $K(F)$ is F. The boolean evaluations of L-sentences are defined by the following inductive conditions:

- $[\![\alpha = \beta]\!] := \{\omega \in \Omega; \alpha(\omega) = \beta(\omega)\}/\mathcal{I}$.
- $[\![R(\alpha_1, \dots, \alpha_k)]\!] := \{\omega \in \Omega; R(\alpha_1, \dots, \alpha_k)\}/\mathcal{I}$ for any k-ary L-relation R.
- $[\![-]\!]$ commutes with $\wedge$,$\vee$,$\neg$.
- $[\![(\exists x)A(x)]\!] := \bigvee_{\alpha \in F} [\![A(\alpha)]\!]$.
- $[\![(\forall x)A(x)]\!] := \bigwedge_{\alpha \in F} [\![A(\alpha)]\!]$.

Finally, using $K(F)$ we can define the first order limit of $\mathcal{G}_k$ using the following notions.

Definition 1.2.6. We call a function $\alpha \in F$ an F-vertex if $\alpha : \Omega \to \langle g_n \rangle$.

We define a $\mathcal{B}$-valued graph $\lim^F_{k \to n} G_k$ as an $\{\Gamma\}$-structure, where Γ is a binary relation symbol, with universe $\{\alpha \in F; \alpha \text{ is an } F\text{-vertex}\}$ and Γ-sentences being evaluated by the following inductive conditions:

- $[\![\alpha = \beta]\!] := \{\omega \in \Omega; \alpha(\omega) = \beta(\omega)\}/\mathcal{I}$.

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

1.3 The second order 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 does existence of large cliques correspond to the size of cliques in the limit graph. But 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) $\mathcal{M}$.

Definition 1.3.1. Let $L \subseteq L_{all}$, we call a set of functions $G \subseteq \mathcal{M}$ an F-closed **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 $\Theta(\alpha) \in F$.

We call $\Theta \in G$ a (graph) G-**relation** if for every $\omega \in \Omega$ we have for some $k > 0$ that $\mathrm{dom}\,\Theta_\omega \supseteq \langle g_n\rangle^k$ and $\Theta_\omega : \mathrm{dom}\,\Theta_\omega \to \{0,1\}$.

Definition 1.3.2. Let $L \subseteq L_{all}$, F an L-closed family and G an F-compatible functional family. We define the L^2-structure $K(F,G)$ as a two sorted L-structure with sorts F and G interpreting L-sentences as $K(F)$ and treating the sort G as follows. First 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

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

If there is a $\Gamma \in G$ such that for every $\alpha,\beta \in F$ we have

$$\Gamma(\alpha,\beta)(\omega) := \chi_{E_\omega}(\alpha(\omega),\beta(\omega)),$$

where χ_{E_ω} is the characteristic function of E_ω, we call $K(F,G)$ the **underlying arithmetic of a second order wide limit**.

We define $\lim^G_{F,n}\{\mathcal{G}_k\}_{k>0}$ as the L^2-substructure with universe consisting of all F-vertices and all G-relations. By abuse of notation we will mostly use the notation $\lim^G_F \mathcal{G}_n$.

1.4 The L-closed family F_{rud} and G_{rud}

Throughout this thesis we will mostly work with the L-closed family F_{rud} which ties the properties of $\lim_F \mathcal{G}_n$ with decision tree complexity.

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 labelling 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 $\mathcal{M}$.

Each $\omega \in \Omega$ 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). The path then uniquely determines an output.

We define F_{rud} to be the set of all functions computed by some (T, ℓ) of depth at most $n^{1/t}$.

One can verify that F_{rud} is an L-closed family for any $L \subseteq L_{all}$.

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 $\hat{\beta} = (\beta_0, \dots, \beta_{m-1})$ 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 $n^{1/t}$ 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 F_{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

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

but

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

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.

(TODO: Theory of graphs, theory of a wide limit)

2. The limit

2.1 Basic examples

2.1.1 $\mathcal{G}_k = \mathbf{EDGE}_k$

We consider the classes of graphs

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

and we let $\mathcal{G}_k = \mathrm{EDGE}_k$ and $F = F_{tree}$.

Intuitivelly one should not be able to find the edge on a significant i.e. non-infinitesimal fraction of samples with a tree that is allowed to explore only an infinitesimal fraction of edges.

Theorem 2.1.1. We will prove that

$$[\![(\exists x)(\exists y)\Gamma(x, y)]\!] = \mathbf{0}.$$

Proof. Let $T \in F_{tree}$ be a tree of depth $n^{1/t}$, for some $t > \mathbb{N}$ that outputs a pair of numbers less than n.

Start from the root of T and always choose the path that corresponds to an edge not existing. At the end we obtain some answer, that gives us a set of at most $2 \cdot n^{1/t} + 2$ vertices. 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

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

And one can clearly see that $\mathrm{st}(1 - \frac{4n^{1/t}-4}{n}) = 1$. This proves that the boolean value we are considering is $\mathbf{0}$ since we can the two witnesses for x and y into a tree that could find an edge with depth $n^{1/t}$ for some $t > \mathbb{N}$. $\square$

2.1.2 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 verticies 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 limit graph.

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 verticies have an edge between. However, this is not sufficient.

Example 2.1.2. Let $\mathcal{G}_k$ consist of those graphs on the vertex-set k which contain the edge $E(0,1)$ and then has exactly one other edge.

As k increases, the number of edges get 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 tree of depth 0.

One can see that having distinguished verticies can ruin the sparseness implying the non-existence of edges in the limit graph. We want to distiguish from this situation by considering the sequences $\mathcal{G}_k$ satisfying the following natural definition.

Definition 2.1.3. 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 $G_1 \in \mathcal{G}_k$, $V_{G_2} = g_k$ and $G_1 \cong G_2$ then $G_2 \in \mathcal{G}_k$.

Theorem 2.1.4. 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

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

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

$$[\![(\exists x)(\exists y)\Gamma(x,y)]\!] = \mathbf{0}.$$

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. □

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

$$\frac{k(k-1)}{2} |\mathcal{G}_k| \, p = \sum_{\{u,v\}} \sum_{G \in \mathcal{G}_k} a_{G,\{u,v\}} \tag{2.6}$$

$$= \sum_{G \in \mathcal{G}_k} \sum_{\{u,v\}} a_{G,\{u,v\}} \tag{2.7}$$

$$\leq |\mathcal{G}_k| \frac{k(k-1)}{2} b_k \tag{2.8}$$

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, this gives us a set E_T of all edges T inspected or outputed and $|E_T| \leq n^{1/t} + 1$.

Now we just need to prove that the probability T find 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

$$\Pr_{G \in \mathcal{G}_n}[T \text{ finds an edge}] \leq \sum_{\{u,v\} \in E_T} \Pr_{G \in \mathcal{G}_n}[E_G(u,v)] \tag{2.9}$$

$$= \sum_{\{u,v\} \in E_T} \frac{c_{u,v}}{|\mathcal{G}_n|} \tag{2.10}$$

$$\leq \sum_{\{u,v\} \in E_T} p \tag{2.11}$$

$$= (n^{1/t} + 1)p \tag{2.12}$$

$$\leq (n^{1/t} + 1)b_k \tag{2.13}$$

$$\leq n^{1/k_0} b_k \tag{2.14}$$

$$\approx 0, \tag{2.15}$$

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 reflects what is "sparse" for the trees in T_{tree}. The following example shows that without the assumption the theorem fails.

Example 2.1.5. 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

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

Let α and β be verticies 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 they are the distinct verticies 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

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

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

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

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

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

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

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

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

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

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

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

A similar result also helps us understand the behaviour witnessing formulas over an isomorphism closed sequence.

Theorem 2.1.6. Let $\mathcal{G}_k$ isomorphism closed, let $\varphi_0(\overline{x})$ quantifier free and let

$$\lim_{F_{tree}} \mathcal{G}_n[\![(\exists\overline{x})\varphi_0^\Gamma(\overline{x})]\!] = \mathbf{1}.$$

Then in $K(\mathcal{G}_n, F_{rud}, G_{rud})$ we have

$$[\![(\exists F)(F : \langle\lfloor n/m\rfloor) \hookrightarrow \mathcal{M}) \wedge (\forall w < \lfloor n/m\rfloor)(\varphi_0^\Gamma(F(w)))]\!] = \mathbf{1}.$$

2.2 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. Let $\varphi(\overline{x})$ be an $\{E\}$-formula determining the isomorphism type of $\overline{x}$, note that for a constant tuple $\overline{\alpha} \in F_{rud}$ (computed by trees of depth 0) we have

$$\mu([\![\varphi(\overline{\alpha})]\!]) = \mathrm{st}(\Pr_{\omega\in\Omega}[\omega \models \varphi(\overline{\alpha})]).$$

Theorem 2.2.1. Let F contain all constants, $\mathcal{G}_k$ be a wide sequence and let $(\forall\overline{x})\varphi_0(\overline{x})$ be an open $\{E\}$-formula such that

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

Then $\mu([\![(\exists x)\varphi_0(x)]\!]) \geq p$.

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 $\aleph_1$-saturation we have that

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

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

$$\frac{1}{|\mathcal{G}_n|\, g_k^l} \sum_{G\in\mathcal{G}_n} \sum_{\overline{a}\in\langle g_n)^l} C_{G,\overline{\alpha}} = \frac{1}{g_k^l} \sum_{\overline{a}\in\langle g_n)^l} \frac{1}{|\mathcal{G}_n|} \sum_{G\in\mathcal{G}_n} C_{G,\overline{\alpha}} \tag{2.28}$$

$$\leq q, \tag{2.29}$$

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

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

$$[\![\varphi]\!] = \bigvee_{\overline{\alpha}} [\![\varphi_0(\overline{\alpha})]\!] \tag{2.30}$$

$$\geq [\![\varphi_0(\overline{\gamma}_{\overline{a}})]\!]. \tag{2.31}$$

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

(TODO: Add example of easy lower bound and existance of substructure.)

Theorem 2.2.2. Let F contain all constants, let $\mathcal{G}_k$ a wide sequence and let $\varphi_0(\overline{x})$ be an open $\{E\}$-sentence, such that

$$\lim_{k\to\infty} \Pr_{G\in\mathcal{G}_k}(G \models \varphi) = 1.$$

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

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

$$\begin{aligned} [\![\varphi]\!] &= \bigwedge_{\overline{\alpha}} [\![\varphi_0(\overline{\alpha})]\!] && (2.32)\\ &= \bigwedge_{\overline{\alpha}} \mathbf{1} && (2.33)\\ &= \mathbf{1}. && (2.34)\end{aligned}$$

□

Example 2.2.3. Let us define

$$\begin{aligned} \mathcal{G}_k^A &:= \{(\langle k\rangle, E); |E| = 2\}, && (2.35)\\ \mathcal{G}_k^B &:= \{(\langle k\rangle, E); |E| = (k\cdot(k-1)/2) - 3\}, && (2.36)\\ \mathcal{G}_k &:= \mathcal{G}_k^A \cup \mathcal{G}_k^B, && (2.37)\end{aligned}$$

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 then 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 2.2.2 that $\lim_F \mathcal{G}_n[\![(\forall a,b,c,d)\varphi_0^\Gamma(a,b,c,d)]\!] = \mathbf{1}$ for any F containing all constants.

It is natural to ask whether we can weaken the assumption of Theorem 2.2.2 to an assumption analogous to Theorem 2.2.1. 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 $\lim_F \mathcal{G}_n[\![(\forall x)\varphi_0^\Gamma(\overline{x})]\!] = \mathbf{1}$? Unfortunately no, as we can see in the following example.

Example 2.2.4. Recall Example 2.1.5 where $\mathcal{G}_k$ consists of all graphs on $\langle k\rangle$ with exactly $\left\lfloor \frac{k(k-2)}{2\log k} \right\rfloor$ edges. One can easily check that

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

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

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

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 2.2.5. 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 $G \models \varphi_0(\overline{a})$.

From $\aleph_1$-saturation of $\mathcal{M}$ 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 leafs 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 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 almost 0. $\square$

Example 2.2.6. Let

$$\mathcal{G}_k = \{([k], E); E \text{ at least one edge, and may have exactly } k/2 \text{ more from start}\}$$

and let $F = F_{tree}$. Then $\mu([\![(\exists x)(\exists y)\Gamma(x,y)]\!]) = \frac{1}{2}$.

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 conjuctions. Each such conjuction says which first order literals should be satisfied on the variables $\overline{x}$, the following theorem is proved for conjunctons of literals. By the fact that $[\![\dots]\!]$ commutes with disjunctions, we can find out a value of any existential sentence.

Theorem 2.2.7 (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 literals 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_{tree} of the same length as $\overline{y}$ then

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

specifically if $l = 0$ then

$$\lim_{F_{rud}} \mathcal{G}_n[\![(\exists \overline{x})\phi_0(\overline{x}, \overline{\beta})]\!] = \mathbf{1}.$$

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 tree each computing the specific element.

First we concatenate all the trees used to compute $\overline{\beta}$. At each leave we can now proceed knowing the value of $\overline{\beta}$ at the specific $\omega \in \Omega$. Now we just construct a tree as in Theorem 2.2.5 but searching only over edges not checked previously and only to fulfill each ψ_i. Luckily we have so far searched only an infinitesimal part of the edges and 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 all of the $\psi_i(\overline{x}, \overline{\beta})$ with probability infinitesimally close to 1.

Therefore

$$[\![(\exists \overline{x})\varphi^\Gamma(\overline{x}, \overline{\beta})]\!] = [\![(\exists \overline{x}) \bigwedge_{i=0}^{m-1} \psi_i(\overline{x}, \overline{\beta})]\!] \wedge [\![\bigwedge_{i=0}^{l-1} \vartheta_i(\overline{\beta})]\!] \tag{2.38}$$

$$= [\![\bigwedge_{i=0}^{l-1} \vartheta_i(\overline{\beta})]\!]. \tag{2.39}$$

□

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

Corollary 2.2.8. For each $\varphi(\overline{x})$ that is not a contradiction in the theory of graphs we have that $[\![(\exists \overline{x})\varphi^\Gamma(\overline{x})]\!] = 1$.

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 (in fact all of those which are not contradictions). □

Corollary 2.2.9. For each $\varphi(\overline{x}, \overline{y})$ that is not falsifiable by $\overline{y}$ in the theory of graphs we have that $[\![(\forall \overline{y})(\exists \overline{x})\varphi^\Gamma(\overline{x})]\!] = \mathbf{1}$.

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 conjuncts. $\square$

Theorem 2.2.10. The theory

$$\mathrm{Th}(\lim_{F_{rud}} \mathrm{ALL}_n)$$

is the theory of the Rado graph and therefore complete.

Proof. In [Gai64] it is 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 nad 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 no vertices from B.

Each $E_{i,j}$ satisfies the statement of Corollary 2.2.9 and because it is also an complete theory we have proved the theorem. $\square$

Corollary 2.2.11. (0-1 law for ALL_k) For every $\{E\}$-sentence φ we have that

$$\lim_{F_{rud}} \mathcal{G}_n[\![\varphi^\Gamma]\!] \in \{\mathbf{0}, \mathbf{1}\}.$$

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

So far the measure of every truth value we encountered was either 0 or 1. Is there a sequence $\mathcal{G}_k$ whose F_{tree}-limit and an $\{E\}$-sentence φ such that $0 < [\![\varphi^\Gamma]\!] < 1$? As in the case of edge existence for a limit of sparse graphs, it is not hard to come up with an example if we allow distinguishing elements in $\mathcal{G}_k$.

Example 2.3.1. Let

$$\mathcal{G}_k = \{(\langle k\rangle, E); E \text{ has exactly two edges, one of them being } \{0, 1\}\},$$

and let $F = F_{tree}$. Then $\mu([\![(\exists x)(\exists y)\Gamma(x, y)]\!]) = \frac{1}{2}$.

Proof. Let T_0 be a tree that always outputs 0 and T_1 be a tree that always outputs 1. We can prove that $[\![\Gamma(0, 1)]\!] \geq [\![\Gamma(\alpha, \beta)]\!]$ for any α, β. $\square$

For the case of isomorphism closed $\mathcal{G}_k$ we prove that every $\{E\}$-sentence has truth value either $\mathbf{0}$ or $\mathbf{1}$. We start with the existential case.

Example 2.3.2. NOT TRUE! Consider all graphs with one edge, or one non-edge. $\mu([\![(\exists x)(\exists y)\Gamma(x, y)]\!]) = \frac{1}{2}$.

Theorem 2.3.3. Let $\varphi_0(\overline{x})$ be an open $\{E\}$-formula, and let $\mathcal{G}_k$ be isomorphism closed and categorical then

$$[\![(\exists \overline{x})\varphi_0^\Gamma(\overline{x})]\!] \in \{\mathbf{0}, \mathbf{1}\}.$$

Proof. Let $\overline{T}$ be a tuple of trees computing $\overline{\alpha}$ such that

$$p := \mu([\![\varphi_0(\overline{\alpha})]\!]) > \mathbf{0},$$

moreover we can assume $p < 1$ otherwise we are done. We can also assume $\overline{T}$ is represented just by one tree T which outputs the whole tuple $\overline{\alpha}$.

We want to iterate T to amplify the probability of success. For every leaf l in T and put

$$A_l = \{\omega \in \Omega; T(\omega) \text{ ends up at } l \text{ and } \omega \models \varphi_0(\overline{\alpha}(\omega))\}$$

and pick some $G \in A$ and some $G' \in \Omega \setminus A$. From the categoricity we have that there exists a permutation $\rho \in S_{\langle g_n \rangle}$ such that $\rho[G] = G'$. $\square$

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. 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 computatinally hard problem to find a large clique in a graph. From complexity theoretical perspective it is an **NP**-complete problem and thus it is conjectured that is 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 a part of the biggest clique it is enough that is had 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 tuple of F-verticies witnessing that?

Here to measure the size of such a clique the second order wide limit by iteself is not a sufficient object. Instead we need to turn to the underlying arithmetic $K(F, G)$ to find an injective function from some large initial segment into a clique in the graph Γ. It is not hard to prove the following result.

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

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

Proof. (Sketch) We need to analyze the value

$$[\![(\exists \Lambda)(\forall u)(\forall v)(((u, v \le n^{1/t}) \to \Gamma(\Lambda(u), \Lambda(v))) \wedge (\Lambda : [n^{1/t}] \hookrightarrow \mathcal{M}))]\!] \tag{3.1}$$

which is equal to

$$\bigvee_{\Lambda} \bigwedge_{u} \bigwedge_{v} [\![(u \neq v < n^{1/t}) \to (\Gamma(\Lambda(u), \Lambda(v)) \wedge \Lambda(u) \neq \Lambda(v))]\!]. \tag{3.2}$$

So we want to find some $n^{1/t}$-tuple of trees computing some Λ which is injective on $\langle n^{1/t} \rangle$ and its $\langle n^{1/t} \rangle$ range is a clique in Γ.

We define $(\Lambda_0, \ldots, \Lambda_{n^{1/t}-1})$ 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. The tree T_i computing Λ_i extends the previous order to $\langle in^{1/t} \rangle \times \langle in^{1/t} \rangle$ searches it and outputs the i-th vertex with 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 Λ_i does not find i vertices with an edge is infinitesimal and therefore it always outputs a vertex in the clique of ω. Moreover, every Λ_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 Λ_i search only in the i-th set. First we need the following lemmas.

(TODO: Comment notation somewhere especially set and abs)

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

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

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

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

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 infinitesimal number of graphs have their clique not intersect $\langle n^{1/t} \rangle$ so each T_{v_i} succeeds on all but 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 $m \geq 2\ln n$ infinitesimal, then

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

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

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

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. $\square$

Even though the size of the clique has radically increased we still did not find a clique in Γ 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. Once can also check that for $\mathrm{SK}_k^{1/l}$, the graphs whose edges are exactly one $\lfloor k/l \rfloor$ clique, the wide limit has a clique of size $\lfloor n/m \rfloor$ for any $m \geq l \cdot \ln(m)$ by the same technique.

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.

One can notice that standardly large cliques should exist in Γ.

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

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

What about nonstandardly large cliques?

Theorem 3.2.2. (?) If for every sequence of elements $\{v_k \in \langle g_k \rangle\}_{k>0}$ and sequence of trees T_k such that T_n has depth at most $n^{1/t}$ for some $t > \mathbb{N}$ we have that

$$\lim_{k\to\infty} \Pr_{G \in \mathrm{CK}_k^{1/2}} [T_k(v_k) = 1 \Leftrightarrow v_k \text{ is in some } \lfloor k/2 \rfloor \text{ clique}] = 0.$$

Proof. $\square$

4. Sparse case and TFNP

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

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 $[k]$, where G is isomorphic to the path on n vertices and $\deg_G(0) = 1$.

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

Computation of such a tree on a undirected graph G goes as follows. We interpret the non-leaf nodes as questions "what is the neighbour set of v?" and the edges as answers from our graph G, and thus we follow a path determined by G until we find a vertex for which the answer is not an edge (in which case the computation returns 0) or until we find a leaf, in which case the computation returns the label of the leaf.

We now shift out focus to analysing the ability of trees from F_{nbtree} to find the non-zero degree 1 vertex in $G \in *\text{PATH}_n$. We say a tree $T \in F_{nbtree}$ 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 $v \in [w]$ and $U \subseteq [w]$ 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 $u, v, w \in [m] \setminus \{0\}$:

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

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 a 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 $T \in F_{nbtree}$, 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

$$P_m(T_{v?=\{u,w\}} \text{ fails}|v? = \{u,w\}) = P_{m-2}(\tilde{T}_{v?=\{u,w\}}). \tag{4.4}$$

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

$$P_m(T_{v?=\{u\}} \text{ fails}|v? = \{u\}) = P_{m-1}(\tilde{T}_{v?=\{u\}}). \tag{4.5}$$

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

(TODO: Elaborate) 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 $T \in F_{nbtree}$ of depth k we have

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

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

$$P(T \text{ fails}) \geq \left(1 - \frac{1}{m-1}\right) \geq \left(1 - \frac{2}{m-2}\right). \tag{4.6}$$

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

$$\begin{aligned}
P(T \text{ fails}) &= \sum_{u \in V \setminus \{0\}} P_{m-1}(0Eu) P_{m-1}(T_{0?=\{u\}} \text{ fails}|0Eu) && (4.7)\\
&\geq P_{m-1}(T_{0?=\{u\}} \text{ fails}|0Eu) && (4.8)\\
&= P_{m-1}(\tilde{T}_{0?=\{u\}} \text{ fails}) && (4.9)\\
&\geq \prod_{i=0}^{k-1} \left(1 - \frac{2}{m - 2i - 2}\right) && (4.10)\\
&\geq \prod_{i=0}^{k} \left(1 - \frac{2}{m - 2i - 2}\right). && (4.11)
\end{aligned}$$

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

$$\begin{aligned}
P_m(vE0) &= \frac{1}{m-1} && (4.12)\\
P_m(N(V) = 1) &= \frac{1}{m-1} && (4.13)\\
P_m(|N(V) \setminus \{0\}| = 2) &= 1 - \frac{2}{m-1}, && (4.14)
\end{aligned}$$

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 intersetion, giving us that stated probability.

Then we have

$$
\begin{aligned}
P_m(T \text{ fails}) &= P_m(vE0)P_m(T \text{ fails}|vE0) && (4.15)\\
&\quad + P_m(|N(v)\setminus\{0\}|=2)P_m(T \text{ fails}|\,|N(v)\setminus\{0\}|=2) && (4.16)\\
&\quad + P_m(|N(v)|=1)P_m(vE0)P_m(T \text{ fails}|\,|N(v)|=1) && (4.17)\\
&\geq P_m(|N(v)\setminus\{0\}|=2)P_m(T \text{ fails}|\,|N(v)\setminus\{0\}|=2) && (4.18)\\
&\geq (1-\frac{2}{m-1}) && (4.19)\\
&\quad \cdot \sum_{\substack{u,w\in V\setminus\{0\}\\ u\neq w}} P_m(v?=\{u,w\})P_m(T_{v?=\{u,w\}} \text{ fails}|v?=\{u,w\}) && (4.20)\\
&\geq (1-\frac{2}{m-1})P_m(T_{v?=\{u_0,w_0\}} \text{ fails}|v?=\{u_0,w_0\}) && (4.21)\\
&\geq (1-\frac{2}{m-1})P_{m-2}(\tilde{T}_{v?=\{u_0,w_0\}} \text{ fails}) && (4.22)\\
&\geq (1-\frac{2}{m-1})\prod_{i=0}^{k-1}(1-\frac{2}{m-2i-4}) && (4.23)\\
&\geq (1-\frac{2}{m-2})\prod_{i=1}^{k}(1-\frac{2}{m-2i-2}) && (4.24)\\
&\geq \prod_{i=0}^{k}(1-\frac{2}{m-2i-2}). && (4.25)
\end{aligned}
$$

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

$$P_m(T_{v?=\{u_0,w_0\}}|v?=\{u_0,w_0\}),$$

the bound follows the fact that all $P_m(v?=\{u,w\})$ 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. □

Corollary 4.1.7. For a tree $T \in F_{nbtree}$ we have that

$$P_n(T \text{ fails}) \approx 1.$$

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

$$
\begin{aligned}
P_n(T \text{ fails}) &\geq \prod_{i=0}^{n^{1/t}}\left(1-\frac{2}{n-2i-2}\right) && (4.26)\\
&\geq \left(1-\frac{2n^{1/t}}{n-2n^{1/t}-2}\right) && (4.27)\\
&\approx 1. && (4.28)
\end{aligned}
$$

□

Finally we can prove the following theorem.

Theorem 4.1.8.

$$[\![(\exists v)(\exists u)(\forall w)((v \neq 0) \wedge (\Gamma(v,u)) \wedge (\Gamma(v,w) \to u = w))]\!] = \mathbf{0}$$

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

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

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

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

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

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

by definition this means that

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

Expanding the value of the formula in the statement we get

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

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

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

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

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

by definition this means that

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

But by definition of γ and Corollary 4.1.7 we have

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

A contradiction. □

(TODO: General method to compute these, directed path, path up to n, conjecture about $K(F, G)$)

Scraps

Definition 4.1.9. 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 $G_1 \in \mathcal{G}_k$, $V_{G_2} = [g_k]$ and $G_1 \cong G_2$ then $G_2 \in \mathcal{G}_k$.

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 sequence $\{G_k\}_{k=0}^{\infty}$ we denote G_k the lexicographically minimal element of $\mathcal{G}_k$.

Lemma 4.1.10. 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 vertex v to $\rho(v)$.

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 $\pi(G_k) = \rho(\tau(G_k)) = \rho(G_k)$. $\square$

Notice that the τ in the statement of the claim is uniquely determinted 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 $[g_k]$ 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 $\mathcal{G}_k$. $\square$

Lemma 4.1.11 (Candidate for optimal search trees)**.** Let $\{\mathcal{G}_k\}_{k=0}^{\infty}$ be categorical and isomorphism closed, let $\varphi(x_0, \ldots, x_{l-1})$ be an open $\{E\}$-formula, let $\models \varphi(\overline{x}) \to \bigwedge_{i,j=0}^{l-1,l-1} x_i =^{b_{ij}} x_j$ for some $b_{ij} \in \{0, 1\}$, let $k_0 \geq 0$ and define $\{q_k\}_{k=k_0}^{\infty}$ as follows

$$q_k := \frac{g_k!}{|\mathrm{Aut}(G_k)|} \cdot \frac{|\varphi(G_k)|}{|\bigcup_{G \in \mathcal{G}} \varphi(G_k)|}.$$

Then there is $c \in \mathbb{N}$ and trees $T_0, \ldots, T_{l-1}$ of depth $n^{(r)} \cdot c$, (with $n^{(r)}$ being defined in the proof) such that for the $\overline{\alpha}$ computed by $\overline{T}$ we have $[\![\varphi(\overline{\alpha})]\!] = \mathbf{1}$.

Proof. We will use the identity from the statement to construct a search tree (iterated $T_{\overline{a}}$) which almost always finds a witness to φ.

We will analyze the problem in the finite case for big enough $k > 0$. We should only check those tuples included in $\bigcup_{G \in \mathcal{G}_k} \varphi(G)$. For example, if we are trying to find an edge then we need not check the constant tuples (a, a). Moreover, to succeed we only need to check one specific tuple in each $\varphi(G), G \in \mathcal{G}_k$.

Consider the set $S = \{(G,\overline{a}); G \in \mathcal{G}_k, G \models \varphi(\overline{a})\}$ and a projection to the second coordinate $p_2 : S \to \bigcup_{G\in\mathcal{G}_k} \varphi(G)$. Since $|S| = \frac{g_k!}{|\mathrm{Aut}(G)|} \cdot |\varphi(G_k)|$ we have that q_k is the average size of a p_2 preimage of any $\overline{a} \in \bigcup_{G\in\mathcal{G}_k} \varphi(G)$.

Claim: For all $\overline{a}, \overline{b} \in \bigcup_{G\in\mathcal{G}_k} \varphi(G)$ we have $\left|p_2^{-1}[\overline{a}]\right| = \left|p_2^{-1}[\overline{b}]\right| = q_k$.

Proof of claim. We will prove that for any $\overline{a}, \overline{b} \in \bigcup_{G\in\mathcal{G}_k} \varphi(G)$ we have $\left|p_2^{-1}[\overline{a}]\right| \leq \left|p_2^{-1}[\overline{b}]\right|$, by symmetry, they must be equal and also equal to q_k which is the average size of any singleton preimage.

Let $p_2^{-1}[\overline{a}] = \{G_0, \ldots, G_{s-1}\} \times \{\overline{a}\}$ and let $\rho = (b_0\, a_0) \ldots (b_{l-1}\, a_{l-1})$, this is a permutation from the condition on φ. Then

$$p_2^{-1}[\overline{b}] \supseteq \{\rho(G_0), \ldots, \rho(G_{s-1})\} \times \{\rho(\overline{a}) = \overline{b}\}. \quad \square$$

Now consider the multiset $M = (\bigcup_{G\in\mathcal{G}_k} \varphi(G), \mathrm{count} : \overline{a} \mapsto \left|p_2^{-1}[\overline{a}]\right|)$, we will construct the searching tree by plucking elements from this multiset in the following way.

Let $M^{(0)} := M, \mathcal{G}_k^{(0)} = \mathcal{G}_k$. For $i \geq 0$ and $M^{(i)}, \mathcal{G}_k^{(i)}$ built, take some $\overline{a} \in M^{(i)}$ with maximal $\mathrm{count}(\overline{a})$, put $\mathcal{G}_k^{(i+1)} = \mathcal{G}_k^{(i)} \setminus p_2^{-1}[\overline{a}]$ and form $M^{(i+1)}$ by removing $\overline{a}$, and for every $\overline{b} \in p_1[p_2^{-1}[\overline{a}]] \setminus \{\overline{a}\}$ setting $\mathrm{count}_{M^{(i+1)}}(\overline{b}) = \max\{0, \mathrm{count}_{M^{(i)}}(\overline{b}) - (\varphi(G_k))\}$. We also add $T_{\overline{a}}$ to the leaves of the tree we are constructing T_i and call it T_{i+1}.

For each $i \geq 0$ we have that T_i finds a witness in $G \in \mathcal{G}_k$ iff $G \notin \mathcal{G}_k^{(i)}$. So to calculate the probability of success of T_i we just need to find upper bounds on the cardinality of $\mathcal{G}_k^{(i)}$.

Define $m_i := \max\{\mathrm{count}(\overline{a}); \overline{a} \in M^{(i)}\}$. Let $k^{(0)} \geq 0$ be the greatest number such that for all $i < k^{(0)}$: $m_i = q_k$.

Define a set $M_m^{(i)} = \{\overline{a}; \mathrm{count}_{M_i}(\overline{a}) = m_i\}$. We can see, that $k^{(0)} \geq 1$ and $M^{(0)} = \bigcup_{G\in\mathcal{G}_k} \varphi(G)$. At each step $i < k^{(0)}$ we construct T_{i+1} by searching for some $\overline{a} \in M_m^{(i)}$, this results in $\left|\mathcal{G}_k^{(i+1)}\right| = |\mathcal{G}_k^i| - q_k$. We also remove one instance of every $\overline{b} \in p_1[p_2^{-1}(\overline{a})] \setminus \{\overline{a}\}$ from $M^{(i)}$ to form $M^{(i+1)}$, this results in $\left|M_m^{(i+1)}\right| \geq \left|M_m^{(i)}\right| - 1 - q_k \cdot (|\varphi(G_k) - 1|)$.

Therefore

$$k^{(0)} \geq \left\lfloor \frac{\left|M_m^{(0)}\right|}{q_k \cdot |\varphi(G_k)|} \right\rfloor \tag{4.29}$$

$$= \left\lfloor \frac{\left|\bigcup_{G\in\mathcal{G}_k} \varphi(G)\right|}{q_k \cdot |\varphi(G_k)|} \right\rfloor, \tag{4.30}$$

and $\left|\mathcal{G}_k^{(k^{(0)})}\right| = |\mathcal{G}_k| - k^{(0)} \cdot q_k = \frac{|\mathrm{Aut}(G_k)|}{g_k!} - \left\lfloor \frac{\left|\bigcup_{G\in\mathcal{G}_k}\varphi(G)\right|}{q_k \cdot |\varphi(G_k)|} \right\rfloor \cdot q_k \leq \frac{|\mathrm{Aut}(G_k)|}{g_k!} - \left\lfloor \frac{\left|\bigcup_{G\in\mathcal{G}_k}\varphi(G)\right|}{|\varphi(G_k)|} \right\rfloor$.

However the right hand side of the last inequality is rarely ≤ 0, so generally one has to continue with plucking even after $k^{(0)}$-many steps. We define $k^{(j)}$ as the greatest number such that for all $i < k^{(j)} : m_i \geq q_k - j$ and continue for $k^{(r)}$ steps, where r is the smallest number such that

$$\left|\mathcal{G}_k^{k^{(r)}}\right| = |\mathcal{G}_k| - k^{(0)} \cdot q_k - \sum_{j=1}^{r}(k^{(j)} - k^{(j-1)}) \cdot (q_k - j) \tag{4.31}$$

$$= 0. \tag{4.32}$$

However, this requires a general analysis of $k^{(j)}$ and I haven't manage to compute that.

For $k = n$ in $\mathcal{M}$ we put $n^{(r)} := k^{(r)}$. $\square$

Conclusion

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.

[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.

[Gol14] I 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.

[Kra10] Jan Krajíček. *Forcing with random variables and proof complexity*, volume 382. Cambridge University Press, 2010.

[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 Nesetril 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.