Limit probability of first order sentences on random graphs

András Németh

Faculty of Science
Eötvös Loránd University

Advisors:
Dr. Katalin Friedl
Dr. Gábor Tardos

2008.

Acknowladgements

I would like to thank Gábor Tardos for inviting me to learn and work at Simon Fraser University in the beautiful city of Vancouver, for drawing my attention and introducing me to this subject and for all his guidance and ideas during this work.

I also would like to thank Katalin Friedl for undertaking to be my official MSc Thesis supervisor at Eötvös Loránd University and for her hard work thoroughly reading and helping to improve and finalize these pages. Without her valuable corrections, clarification ideas and general comments this thesis would be much harder to comprehend.

Contents

Chapter 1

Introduction

In this thesis we are going to investigate first order sentences on random graphs. More precisely we are interested in how the truth value of a first order sentence can change as a function of a density parameter.

$G(n, \alpha)$ denotes the random graph with n vertices in which every possible edge is present with probability $n^{-\alpha}$ independently of each other. Shelah and Spencer proved in [4] that if α is irrational then for any first order sentence φ the limit $\lim_{n\to\infty} P(G(n, \alpha) \models \varphi)$ exists and is either one or zero. If the above limit is 1, then we say φ *almost surely holds* in $G(n, \alpha)$.

Here we are interested in how the limit of the probability of a fixed first order sentence changes as we change the value of α. For this purpose, let us define the following function on non-negative irrationals:

$$f_\varphi(x) = \lim_{n\to\infty} P(G(n, x) \models \varphi) \in \{0, 1\}$$

The main question is that for which $f : \mathbb{R}^+ \setminus \mathbb{Q} \to \{0, 1\}$ functions can we found a formula φ such that $f = f_\varphi$. After some definitions we state four necessary conditions, that is four properties that must hold for f_φ for any formula φ. Maybe surprisingly, the behavior of f_φ is closely related to a rational approximation sequence.

Definition 1.0.1. For $\alpha \geq \beta \geq 0$, where β is rational let us define:

$$\tau(\alpha, \beta) = \max\{\frac{p}{q} \in \mathbb{Q} \mid \frac{p}{q} \leq \alpha, \frac{p-1}{q} \leq \beta\}$$

We define the *strong approximation sequence* of $\alpha > 0$ as $\tau_0(\alpha)$, $\tau_1(\alpha)$, ... where $\tau_0(\alpha) = 0$ and $\tau_i(\alpha) = \tau(\alpha, \tau_{i-1}(\alpha))$.

Notice that this definition is correct, that is the set $S = \{\frac{p}{q} \in \mathbb{Q} \mid \frac{p}{q} \leq \alpha, \frac{p-1}{q} \leq \beta\}$ is not empty and has a maximal element. Indeed, $\beta \in S$ and for any $\frac{r}{s} > \beta$ and for all the $\frac{u}{v}$ elements of S larger then $\frac{r}{s}$ we have $v < \frac{1}{\frac{r}{s} - \beta}$, thus there are only finitely many such elements.

As established in [7] but also apparent from the results in Chapter 3, the strong approximation sequence tends to α (from below), and it reaches it in finitely many steps if α is rational. This motivates the notion of *length*. We say a rational α is of length k if it is reached in k steps in the above sequence, that is $\alpha = \tau_k(\alpha)$ but $\alpha > \tau_i(\alpha)$ for

$0 \leq i < k$. We denote by $LEN(k)$ the set of rationals of length at most k, that is the rationals α for which $\alpha = \tau_k(\alpha)$. We define another closely related sequence by changing $\frac{p}{q} \leq \alpha$ to $\frac{p}{q} < \alpha$:

Definition 1.0.2. For $\alpha > \beta \geq 0$, where β is rational let us define:

$$\nu(\alpha, \beta) = \max\{\frac{p}{q} \in \mathbb{Q} \mid \frac{p}{q} < \alpha, \frac{p-1}{q} \leq \beta\}$$

Let $\nu_0(\alpha) = 0$ and $\nu_i(\alpha) = \nu(\alpha, \nu_{i-1}(\alpha))$.

As long as $\tau_i(\alpha) < \alpha$ we have $\nu_i(\alpha) = \tau_i(\alpha)$, but ν_i never reaches α, only it tends to it. It was also proved in [7] that the set $LEN(k)$ is well-ordered under $>$, and the greatest element of $LEN(k)$ below $\alpha > 0$ is $\nu_k(\alpha)$.

Now we are ready to state the three known rules which govern where the value of f_φ can change:

Very Sparse Condition f_φ is constant on each interval $(1 + \frac{1}{i+1}, 1 + \frac{1}{i})$ and on $(2, \infty)$.

Very Dense Condition There exists a $k_0 = k_0(\varphi) > 0$ such that f_φ is constant on $(0, \frac{1}{k_0})$.

Locally Constant Condition There exists an $l = l(\varphi)$ such that f_φ is constant on $(\nu_l(\alpha), \alpha)$ for any $0 < \alpha$.

Clearly, for the Locally Constant Condition to hold it is enough to require that f_φ is constant on the irrationals of $(\nu_k(\alpha), \alpha)$ for $\alpha \in LEN(k)$, as for other values of α we obtain subintervals. (Assuming $k > 0$, as for $k = 0$, the condition simply says f_φ is constant on $\alpha > 0$.)

The first two conditions were already proved in [4], the third was established in [7].

A function f defined on positive irrationals satisfying the Locally Constant Condition can be extended to positive rationals as:

$$f^-(x) = \lim_{\epsilon \to +0} f(x - \epsilon)$$

This is well defined as the Locally Constant Condition guarantees a positive length constant segment left of any positive number.

This enables us to talk about the computational complexity of such functions by defining the language $L_f = \{0^p 1^q \mid p, q > 0, f^-(\frac{p}{q}) = 1\}$. Now we are ready to state the last necessary condition on the functions that come up as f_φ (see [7]):

Complexity Condition L_{f_φ} is in PH.

Here PH denotes the union of all complexity classes in the polynomial hierarchy. We will give the definition of the polynomial hierarchy in Section 2.3.

In [7] Spencer and Tardos conjectured that the four conditions above are also sufficient, that is if a function f satisfies all four of the above conditions, then there is a first order formula φ such that $f = f_\varphi$. The subject of this thesis is the proof for an important part of this conjecture. We will show that it is true for the dense region of the α parameter, more precisely for any f satisfying all four conditions there is a first order formula φ such that $f|_{[0,1/2]} = f_\varphi|_{[0,1/2]}$.

The structure of this thesis is as follows. In Chapter 2 we give all the necessary notions and earlier results that we will use in the rest.

In Chapter 3 we take a detour to elementary number theory. We will examine another approximation sequence and prove that it is equivalent to the strong approximation sequence defined above. We need to do this because later we will work out a way to construct and first order characterize graphs whose size will correspond to the elements of this new sequence. If the new sequence would not tend to α fast enough relatively to the strong approximation sequence then this construction would not be usable to prove that the above four conditions are also sufficient. We also give here some facts about the speed of the approximation by these sequences. The results in this chapter are my own work.

In Chapter 4 we give the above promised graph construction and the way to first order characterize subgraphs in α-graphs isomorphic to the constructed graphs. The main concept of this chapter is the so called *hybrid* construction. The idea of this construction and the main steps for the first order characterization were already invented by Gábor Tardos. My contribution was to simplify in one way and generalize in other ways these ideas to exactly fit the number theoretical results of Chapter 3.

The main result in Chapter 4 is that for any d we can create formulae independently of α which characterize the occurrences of various subgraphs in an α-graph among which subgraphs there will be a specific one with v vertices and e edges where $\frac{v}{e} = \tau_d(\alpha)$. We know that there is a PH algorithm that computes $f^-(\frac{p}{q})$ if p and q are given unary. This can be easily modified to a PH algorithm which computes $f^-(\frac{v}{e})$ if given a graph with v vertices an e edges. Thus all we need is to create a first order formula which somehow simulates the execution of this PH algorithm on the above mentioned subgraph. By Fagin's Theorem (see Section 2.3) we know that there is such *second order* formula. So we need to work out a toolkit to be able to simulate second order formulae on a specific subgraph with first order formulae on the whole α-graph. Exactly this is what happens in Chapter 5. The foundation-stone of this is an idea by Gábor Tardos on how to represent multivariate functions on small vertex sets. My contribution in this chapter is to prove that this idea indeed works, to define the notion of set dresses which allows for representing relations on larger sets and to build the framework around these to first order simulate second order formulae.

Finally in Chapter 6 we put all these tools together to prove the main result promised above.

Chapter 2

Preliminaries

2.1 Graph extensions

We follow [7] with the notations concerning graph extensions. We call a pair (H, G) a graph extension where H is a finite subgraph of the (possibly infinite) graph G. We call H the base of the extension. As a special case we allow H to be empty (no vertices, thus no edges), and this way we can look at any graph as a graph extension with empty base. If we write (S, G) where S is a set of vertices of G, we mean the extension (H, G) where H is the graph with vertex set S and no edges. A graph extension is *trivial* if $H = G$ and *finite* if G is also finite. An *intermediate graph* of an extension (H, G) is a finite subgraph H' of G such that $H \leq H'$ (so (H', G) is an extension and (H, H') is a finite extension). We also call an intermediate graph of (H, G) a finite extension of H in G. A function $i : V(G) \to V(G')$ is an isomorphism from (H, G) to (H', G') if it is an isomorphism from G to G' and $i|_{V(H)}$ is an isomorphism from H to H'.

The size of a finite extension (H, G) is the pair (v, e), where $v = |V(G) - V(H)|$ and $e = |E(G) - E(H)|$. A simple but fundamental fact about extensions in random graphs is the following:

Lemma 2.1.1. *Let (H, H') be a finite extension of size (v, e). For every isomorphism i from H to a subgraph of $G(n, \alpha)$ let $\mathcal{X}_{i,H'}$ be the set of all isomorphisms j from H' to a subgraph of $G(n, \alpha)$ such that j is an extension of i (as an isomorphism). Then $E(|\mathcal{X}_{i,H'}|) = \frac{(n-|H|)!}{(n-|H'|)!} n^{-\alpha e} \sim n^{v-\alpha e}$.*

Proof. The images of the vertices in $V(H') - V(H)$ can be chosen $\frac{(n-|H|)!}{(n-|H'|)!}$ ways. A choice is good if for all edges present in H' but not present in H the image of its endpoints are connected in $G(n, \alpha)$. For each edge, the probability of this is $n^{-\alpha}$ independently from all other edges, so the probability of a given function to be a good isomorphism is $n^{-\alpha e}$. □

This lemma motivates the following definitions. Let us fix an irrational α. We call an extension of size (v, e) *dense*, if $\alpha e \geq v$ and *sparse* if $\alpha e \leq v$. Notice that (as α is irrational) an extension cannot be sparse and dense at the same time, except for a trivial extension, which we do consider both sparse and dense. If we say that a graph H is sparse/dense, we mean that $(\emptyset, H)$ is sparse/dense, similarly when we talk about the size of H, we mean the size of $(\emptyset, H)$.

Applying the above lemma to the extension $(\emptyset, H)$ it is clear that if the graph H is dense, then almost surely in $G = G(n, \alpha)$ there is no subgraph isomorphic to H. On the other hand the converse is not true: only because H is sparse, we cannot almost surely found in G a subgraph isomorphic to H. One obvious reason can be that H has a dense subgraph. It will turn out in Theorem 2.2.2 that this is the only possible reason.

As motivated above, we will call a finite extension (H, H') *safe*, if for all intermediate graph H'' the extension (H, H'') is sparse. We will call a finite extension *rigid*, if for all intermediate graph H'' the extension (H'', H') is dense.

The following lemma summarizes some poperties of graph extensions:

Lemma 2.1.2.

1. *Let H be an intermediate graph of the dense extension (H_1, H_2). If one of (H_1, H) and (H, H_2) is sparse, then the other is dense.*

2. *Let H be an intermediate graph of the sparse extension (H_1, H_2). If one of (H_1, H) and (H, H_2) is dense, then the other is sparse.*

3. *If for all intermediate graph $H \neq H_2$ of the dense extension (H_1, H_2) we have that (H_1, H) is sparse, then (H_1, H_2) is rigid.*

4. *If (H_1, H_2) is not safe, then there is an intermediate graph $H \neq H_1$ for which (H_1, H) is rigid.*

5. *If (H_1, H_2) is a rigid extension then for any finite graph H we have $(H \cup H_1, H \cup H_2)$ is rigid.*

6. *If (H_1, H_2) is a safe extension then for any finite graph H we have $(H \cap H_1, H \cap H_2)$ is safe.*

7. *If (H_1, H_2) and (H_2, H_3) are rigid then so is (H_1, H_3).*

8. *If (H_1, H_2) and (H_2, H_3) are safe then so is (H_1, H_3).*

Proof. 1. and 2.: The size (v, e) of (H_1, H_2) is the sum of the sizes (v_1, e_1) of (H_1, H) and (v_2, e_2) of (H, H_2). Suppose $H_1 \neq H_2$. If both $v_1 - \alpha e_1 \geq 0$ and $v_2 - \alpha e_2 \geq 0$ then we also have $v - \alpha e \geq 0$, but $=$ is not possible so $v - \alpha e > 0$, which contradicts (H_1, H_2) being dense. If $H_1 = H_2$ then for the unique intermediate graph $H = H_1 = H_2$ both (H_1, H) and (H, H_2) are dense. Changing $\geq$ to $\leq$, $>$ to $<$ and "dense" to "sparse" in the above argument we get 2.

3.: By 1 for all $H \neq H_2$ intermediate graph (H, H_2) is dense, and (H_2, H_2) is dense always, thus (H_1, H_2) is rigid indeed.

4.: Take a H intermediate graph for which (H_1, H) is not sparse, but which is minimal with that property, that is for all $H' \neq H$ intermediate graphs of (H_1, H) we have that (H_1, H') is sparse. As (H_1, H_1) is sparse, but not all intermediate graphs of (H_1, H_2) are sparse there is such H. Using 3. we have that (H_1, H) is rigid.

5.: Let H' be an intermediate graph of $(H \cup H_1, H \cup H_2)$. Notice that the size of $(H', H \cup H_2)$ is the same as that of $(H' \cap H_2, H_2)$. As the second extension is dense as $H' \cap H_2$ is an intermediate graph of the rigid extension (H_1, H_2) the former extension is also dense, thus $(H \cup H_1, H \cup H_2)$ is rigid indeed.

6.: Let H' be an intermediate graph of $(H \cap H_1, H \cap H_2)$. Notice that the size of $(H \cap H_1, H')$ is the same as that of $(H_1, H' \cup H_1)$. As the second extension is sparse as $H' \cup H_1$ is an intermediate graph of the safe extension (H_1, H_2) the former extension is also safe, thus $(H \cup H_1, H \cup H_2)$ is safe indeed.

7.: Let H be an intermediate graph of (H_1, H_3). The extension $(H_2 \cup H, H_3)$ is dense as $H_2 \cup H$ is an intermediate graph of the rigid extension (H_2, H_3). Using that (H_1, H_2) is rigid by item 5. $(H, H_2 \cup H)$ is also rigid, thus dense. As the size of (H, H_3) is the sum of the sizes of the dense extensions $(H_2 \cup H, H_3)$ and $(H, H_2 \cup H)$ we have that (H, H_3) is dense as needed.

8.: Let H be an intermediate graph of (H_1, H_3). The extension $(H_1, H_2 \cap H)$ is sparse as $H_2 \cap H$ is an intermediate graph of the safe extension (H_1, H_2). Using that (H_2, H_3) is safe by item 6. $(H_2 \cap H, H)$ is also safe, thus sparse. As the size of (H_1, H) is the sum of the sizes of the sparse extensions $(H_1, H_2 \cap H)$ and $(H_2 \cap H, H)$ we have that (H_1, H) is sparse as needed. □

2.2 The almost sure theory of random graphs

For a fixed α the set of sentences which hold almost surely that is the set of formulae $\{\varphi \mid \lim_{n\to\infty} P(G(n, \alpha) \models \varphi) = 1\}$ obviously forms a consistent and complete theory. This is called the almost sure theory of $G(n, \alpha)$. It is clear that the sentence "the graph has more than k vertices" is in the theory for every k, so no finite graph will actually satisfy all the almost sure sentences. So for us it will be easier to deal with infinite graphs which models the whole theory. We shall call such a graph an *α-graph*. As there are no finite models by Gödel's completeness theory there must exist infinite models, so there exist α-graphs for every α.

We are not going to use this, but remark here that by the Löwenheim-Skolem theorem, there is also a countable model for the almost sure theory. In fact, proven in [5], if $\alpha > 1$, there is exactly one such countable model, but if $\alpha < 1$, there are continuum non-isomorphic countable models.

To give an axiomatization of the almost sure theory we need one more notion. We will call an extension of a subgraph in a bigger graph *generic* if it does not have small rigid extensions, except maybe for those of the base graph. More precisely:

Definition 2.2.1. Let H be a subgraph of a graph G. We say the finite extension H' of H in G is *k-generic* if for any rigid extension H'' of H' in G of size (v, e) with $v \leq k$ there is no edge in $E(H'') - E(H')$ having an endpoint in $V(H') - V(H)$.

As proved in [6], the following axiomatization of the almost sure theory of $G(n, \alpha)$ can be given:

Theorem 2.2.2. *The two axiom schemes below give an axiomatization of the almost sure theory of $G(n, \alpha)$:*

A_H *(sparsity axiom; H is a finite, but non-empty dense graph) G does not contain a subgraph isomorphic to H.*

$B^k_{H_0,H_1}$ *(safe extension axiom; (H_0, H_1) is a finite safe extension, $k \in \mathbb{Z}^+$) Every isomorphism from H_0 to a subgraph H_0' of G can be extended to an isomorphism from H_1 to a subgraph H_1' of G, such that H_1' is a generic extension of H_0' in G.*

The first axiom scheme simply states the above observation about the lack of dense subgraphs. The second axiom scheme not only tells that safe extensions are always present, but also guarantees the existence of generic safe extensions.

2.3 Logic, complexity theory and Fagin's theorem

We will denote signatures as $\iota = (J_1/r_1, ..., J_l/r_l)$ where J_1, ..., J_l are the predicate symbols of the signature and r_i is the arity of J_i. We will not have functions in our structures and formulae. Let us fix a signature ι. Let $\mathfrak{A}$ be an ι-model with universe A. We will denote with $J_i^{\mathfrak{A}}$ the interpretation of the J_i predicate in $\mathfrak{A}$, thus $J_i^{\mathfrak{A}} \subseteq A^{r_i}$. If φ is a closed ι-formula we use the standard notation $\mathfrak{A} \models \varphi$ to say that $\mathfrak{A}$ models φ. If φ is not closed and $\sigma : V \to A$ is a variable assignment such that all free variables of φ is contained in V then we will use the notation $\mathfrak{A}[\sigma] \models \varphi$ to say that $\mathfrak{A}$ models φ with the σ variable assignment. For any first order variables x_1, x_2, ..., x_n and a_1, a_2, ..., $a_n \in A$ we will use the notation $\{x_1 \mapsto a_1, x_2 \mapsto a_2, ..., x_n \mapsto a_n\}$ to refer to the assignment $\sigma : \{x_1, x_2, ..., x_n\} \to A$ where $\sigma(x_i) = a_i$. If $\sigma_1 : V_1 \to A$ and $\sigma_2 : V_2 \to A$ are two variable assignment where V_1 and V_2 are disjoint variable sets then we will denote with $\sigma_1 \cup \sigma_2$ the union of the two assignment, that is the $\sigma : V_1 \cup V_2 \to A$ where $\sigma(x) = \sigma_1(x)$ if $x \in V_1$ and $\sigma(x) = \sigma_2(x)$ otherwise.

We will define formulae as:

$$\varphi(x, y, z, w) = \text{“some first order formula”}$$

All free variables of the formula on the right hand side must be listed on the left in the parentheses, although we can also list unused variables (for example $\varphi(x, y, z, w) = E(x, y) \wedge (y = z)$ is valid if E is a binary predicate symbol of the signature). The length of the variable list in the parentheses is called the arity of the formula, for example φ above is a 4-ary formula. The order of the variables in the list is important as for t_1, ..., t_n arbitrary terms we will use the notation $\varphi(t_1, ..., t_n)$ to refer to the formula where we substitute all occurrences of the first variable with t_1, the second with t_2, etc. For example, if φ was defined as above then $\varphi(x', y', z', w')$ refers to the formula $E(x', y') \wedge (y' = z')$.

If it is clear from the context that we are talking about the model $\mathfrak{A}$ then for $a_1, ..., a_n \in A$ we will use $\varphi(a_1, ..., a_n)$ instead of $\mathfrak{A}[\{x_1 \mapsto a_1, ..., x_n \mapsto a_n\}] \models \varphi(x_1, x_2, ..., x_n)$. If φ was defined as an n-ary formula, than $\varphi(a_1, ..., a_k, \_, ..., \_)$ denotes the following $(n-k)$ ary relation on A:

$$\begin{aligned}\varphi(a_1, ..., a_k, \_, ..., \_) = \{&(b_1, ..., b_{n-k}) \in A^{n-k} \mid \\ &\mathfrak{A}[\{x_1 \mapsto a_1, ..., x_k \mapsto a_k, x_{k+1} \mapsto b_1, ..., x_n \mapsto b_{n-k}\}] \models \varphi(x_1, x_2, ..., x_n)\}\end{aligned}$$

We will work with parameterized formulae, when we do not want to explicitly list all the free variables of the formula. The below formalism helps to keep the notation simpler in that case. For a set P of variables we can say φ is an *n-ary P-formula* and define it as:

$$\varphi(x_1, x_2, ..., x_n) = \text{“some first order formula”}$$

if the set of free variables in the formula on the right hand side is contained in $P \cup \{x_1, x_2, ..., x_n\}$. We also call a 0-ary P-formula a *closed* P-formula. In case of P-formulae the notation $\varphi(t_1, ..., t_n)$ denotes the formula that we get from φ by substituting x_1 with t_1, ... x_n with t_n and we leave the variables in P untouched.

For an n-ary P-formula φ and variable assignment $\sigma : P \to A$ and a_1, ..., $a_n \in A$ we will use the notation $G[\sigma] \models \varphi(a_1, ..., a_n)$ meaning $G[\sigma \cup \{x_1 \to a_1, ..., x_n \to a_n\}] \models \varphi(x_1, ..., x_n)$. Instead of $G[\sigma] \models \varphi(a_1, ..., a_n)$ we can also say $\varphi(a_1, ..., a_n)$ holds in $G[\sigma]$.

When we speak about complexity classes we will fix the alphabet $\{0, 1\}$, so a language is a subset of $\{0, 1\}^*$. We fix an encoding of pairs of words, that is an injection:

$$\langle \_, \_ \rangle : \{0, 1\}^* \times \{0, 1\}^* \to \{0, 1\}^*$$

We define the polynomial hierarchy following [9]. For a language $L \subseteq \{0, 1\}^*$ and a polynomial p we define the following two languages:

$$\forall^p L = \{x \in \{0, 1\}^* \mid \forall y \in \{0, 1\}^{p(|x|)}(\langle x, y \rangle \in L)\}$$
$$\exists^p L = \{x \in \{0, 1\}^* \mid \exists y \in \{0, 1\}^{p(|x|)}(\langle x, y \rangle \in L)\}$$

Given a class $\mathcal{C}$ of languages, we define the following two language classes:

$$\forall^P \mathcal{C} = \{\forall^p L \mid L \in \mathcal{C} \text{ and } p \text{ is a polynomial}\}$$
$$\exists^P \mathcal{C} = \{\exists^p L \mid L \in \mathcal{C} \text{ and } p \text{ is a polynomial}\}$$

Now we can recursively define the polynomial hierarchy (P denotes the class of the polynomial time decision problems):

$$\Sigma_0^P = \Pi_0^P = P$$
$$\Sigma_{i+1}^P = \exists^P \Pi_i^P$$
$$\Pi_{i+1}^P = \forall^P \Sigma_i^P$$

Observe that $\Sigma_1^P = NP$ and $\Pi_1^P = coNP$. It is easy to see that $\Sigma_i^P \cup \Pi_i^P \subseteq \Sigma_{i+1}^P \cap \Pi_{i+1}^P$. We set $PH = \bigcup_{i=0}^{\infty} \Sigma_i^P = \bigcup_{i=0}^{\infty} \Pi_i^P$.

To state Fagin's Theorem we need to fix an encoding of structures of a given signature $\iota = (J_1/r_1, ..., J_k/r_k)$. If we are given and ordering $<_A$ of the universe A then we can encode any relation $R \subseteq A^d$ with a bit string of length $|A|^d$ by setting the jth bit 1 iff the jth tuple of A^d in the lexicographical order induced by $<_A$ is in R. Let us denote this encoding as $EncR_{<_A}(R)$. Using this we can give the encoding of ι-structures ($\cdot$ means concatenation):

$$Enc^{\iota}_{<_A}(\mathfrak{A}) = 1^{|A|}0 \cdot EncR_{<_A}(J_1^{\mathfrak{A}}) \cdot EncR_{<_A}(J_2^{\mathfrak{A}}) \cdot ... \cdot EncR_{<_A}(J_l^{\mathfrak{A}})$$

Let us denote with L^{ι} the language of encodings of finite ι-structures, that is:

$$L^{\iota} = \{Enc^{\iota}_{<_A}(\mathfrak{A}) \mid \mathfrak{A} \text{ is a finite structure of signature } \iota,\ A \text{ is the universe of } \mathfrak{A},\ <_A \text{ is any ordering on } A\}$$

Definition 2.3.1. We say that a language $L \subseteq L^\iota$ is order invariant if for any structure $\mathfrak{A}$ and any two ordering $<_A$ and $<'_A$ on its universe we have $Enc^\iota_{<_A}(\mathfrak{A}) \in L$ iff $Enc^\iota_{<'_A}(\mathfrak{A}) \in L$.

For an arbitrary second order ι-formula ψ let us define:

$$L^\iota_\psi = \{Enc^\iota_{<_A}(\mathfrak{A}) \mid \mathfrak{A} \text{ is a finite structure of signature } \iota,\ A \text{ is the universe of } \mathfrak{A},$$
$$<_A \text{ is any ordering on } A \text{ and } \mathfrak{A} \models \psi\}$$

Notice that L^ι_ψ is order invariant.

Theorem 2.3.2 (Fagin, 1974, [3])**.** *For any signature ι and any second order ι-sentence ψ we have that $L^\iota_\psi \in PH$. On the other hand for any $L \subseteq L^\iota$ order invariant language if $L \in PH$ then there is a second order formula ψ such that $L = L^\iota_\psi$.*

Remark 2.3.3. Fagin originally proved in the [3] that existential second order sentences correspond to decision problems in NP in the above claimed way. The theorem as stated above is a trivial generalization of his work.

Chapter 3

Rational approximations

In this chapter, we are going to consider a way to approximate non-negative real numbers, referred to as *weak approximation sequence*. Our first goal is to prove that this sequence is equivalent to the strong approximation sequence (Definition 1.0.1). The reason we are interested in this sequence is its strong relation to the hybrid construction given in the next chapter. Just to give a rough idea now we can say a bit imprecisely that if we are able to first order characterize extensions of size (v, e) and (v', e') with $\frac{v}{e} \leq \alpha$ and $\frac{v'}{e'} \leq \alpha$ then we will be able to characterize extensions of size $(kv + lv' + 1, ke + le')$ if $\frac{kv+lv'+1}{ke+le'} < \alpha$ but $\frac{(k-1)v+lv'+1}{(k-1)e+le'} > \alpha$ and $\frac{kv+(l-1)v'+1}{ke+(l-1)e'} > \alpha$. This motivates the Definition 3.1.4 below. The above sketched connection will be precisely stated in Lemma 4.6.1.

In the first section we establish the promised equivalence between the new sequence and the strong approximation sequence. In the remaining two sections we will prove some results about the approximation speed of these sequences. The results of Section 3.3 will be used in Chapter 5.

In this chapter almost always whenever we refer to a rational $\frac{p}{q}$ the numbers p and q are going be relatively prime, and in many points it is actually important to choose the reduced form. Thus whenever we write $\frac{x}{y}$ we implicitly claim that x and y are coprime. When we choose integers to represent a rational, this notation means that we chose them to be relatively prime, or if x and y have not been just chosen, then for some reason we know that they have to be coprime. Although in some rare cases we will need to write fractions without this assumption. In these cases we will either use x/y or the special $*\frac{x}{y}*$ notation.

3.1 The weak approximation sequence

Before defining the weak approximation sequence we need some preparation.

Definition 3.1.1. A rational $\frac{p}{q}$ is *reached* from the rationals $\frac{r_1}{s_1}$, $\frac{r_2}{s_2}$, ...,$\frac{r_t}{s_t}$ with non-negative integer coefficients k_1, ..., k_t if $p = k_1r_1 + k_2r_2 + ... + k_tr_t + 1$ and $q = k_1s_1 + k_2s_2 + ... + k_ts_t$. The number $\frac{p}{q}$ is *reachable* from $\frac{r_1}{s_1}, \frac{r_2}{s_2}, ..., \frac{r_t}{s_t}$ if there are non-negative integer coefficients as above.

Remark 3.1.2. It is important that we not only required that:

$$\frac{p}{q} = *\frac{k_1r_1 + k_2r_2 + ... + k_tr_t + 1}{k_1s_1 + k_2s_2 + ... + k_ts_t}*$$

So, for example, $\frac{3}{4}$ is not reached from $\frac{1}{2}$ and $\frac{2}{3}$ with coefficients 1 and 2, although $*\frac{1\times 1+2\times 2+1}{1\times 2+2\times 3}* = *\frac{6}{8}* = \frac{3}{4}$. (We can reach $\frac{3}{4}$ from $\frac{1}{2}$ alone, though.)

Definition 3.1.3. A rational $\frac{p}{q} \le \alpha$ is *α-reached* from the rationals $\frac{r_1}{s_1}$, $\frac{r_2}{s_2}$, ..., $\frac{r_t}{s_t}$ with non-negative integer coefficients k_1, ..., k_t if it is reached and for all $1 \le i \le t$ where $k_i > 0$ we have $k_1r_1 + k_2r_2 + ... + k_tr_t + 1 - r_i > \alpha(k_1s_1 + k_2s_2 + ... + k_ts_t - s_i)$. The number $\frac{p}{q}$ is *α-reachable* from $\frac{r_1}{s_1}$, $\frac{r_2}{s_2}$, ..., $\frac{r_t}{s_t}$ if there are non-negative integer coefficients as above.

For example $\frac{9}{13}$ is α-reachable from $\frac{1}{2}$ and $\frac{2}{3}$ if $\frac{9}{13} \le \alpha < \frac{7}{10}$. Now we are ready to define our new approximation sequence, which is actually a sequence of sets:

Definition 3.1.4. Let S be any set of rationals.

$$H(\alpha, S) := \{\frac{p}{q} \in \mathbb{Q} \mid \frac{p}{q} \text{ can be } \alpha\text{-reached from a rational } \frac{r}{s} \in S \text{ or from two rationals } \frac{r}{s}, \frac{r'}{s'} \in S\}$$

We define the *weak approximation set sequence* of α as $H_0(\alpha) = \{\frac{0}{1}\}$, and $H_i(\alpha) = H(\alpha, H_{i-1}(\alpha)) \cup H_{i-1}(\alpha)$.

It is easy to see that $max(H_i(\alpha)) \le \tau_i(\alpha)$. We can prove it by induction observing that if $\frac{p}{q}$ is reachable from a set of rationals then $*\frac{p-1}{q}*$ is less or equal to the largest rational in the set. The rest of this section is devoted to prove the following theorem:

Theorem 3.1.5. *For any $0 \le \alpha < 1$ and $i \in \mathbb{N}$ it holds that $\tau_i(\alpha) \in H_i(\alpha)$.*

To prove the theorem, we will need several tools. First of all, we will often use the following function:

Definition 3.1.6. For rationals $\frac{p}{q}$ and $\frac{r}{s}$ let $l(\frac{p}{q}, \frac{r}{s}) = rq - ps = qs(\frac{r}{s} - \frac{p}{q})$.

Notice that this is well defined by our assumption that all rationals are in reduced form. Observe, that $l(\frac{p}{q}, \frac{r}{s}) < 0$ iff $\frac{r}{s} < \frac{p}{q}$ and it also holds if we change "<" to "=" or ">". Also, trivially $l(\frac{p}{q}, \frac{r}{s}) = -l(\frac{r}{s}, \frac{p}{q})$.

The following is a simple property of the l function.

Lemma 3.1.7. *For any non-negative rationals $\frac{p}{q}, \frac{r_1}{s_1}$, ..., $\frac{r_t}{s_t}$ and integer c if $gcd(k_1r_1 + k_2r_2 + ... + k_tr_t + c, k_1s_1 + k_2s_2 + ... + k_ts_t) = 1$ we have:*

$$l\left(\frac{k_1r_1 + ... + k_tr_t + c}{k_1s_1 + ... + k_ts_t}, \frac{p}{q}\right) = k_1 l\left(\frac{r_1}{s_1}, \frac{p}{q}\right) + ... + k_t l\left(\frac{r_t}{s_t}, \frac{p}{q}\right) - cq$$

Proof. Elementary computation. □

One of the central observations in the proof is the following lemma:

Lemma 3.1.8. *Let $\frac{r_1}{s_1} \le \frac{r_2}{s_2} \le \alpha$ such that $l(\frac{r_1}{s_1}, \frac{r_2}{s_2}) = 1$ and $\frac{a}{b} \le \alpha$ be such that $\frac{r_1}{s_1} \le *\frac{a-1}{b}* \le \frac{r_2}{s_2}$. If $\alpha b - a < \alpha s_1 - r_1$ and $\alpha b - a < \alpha s_2 - r_2$ then $\frac{a}{b} \in H(\{\frac{r_1}{s_1}, \frac{r_2}{s_2}\}, \alpha)$.*

Proof. Let us solve the following system of linear equations:

$$k_1 r_1 + k_2 r_2 = a - 1$$
$$k_1 s_1 + k_2 s_2 = b$$

Using Cramer's rule, we get:

$$k_1 = \frac{\begin{vmatrix} a-1 & r_2 \\ b & s_2 \end{vmatrix}}{\begin{vmatrix} r_1 & r_2 \\ s_1 & s_2 \end{vmatrix}} = \frac{(a-1)s_2 - r_2 b}{l\left(\frac{r_2}{s_2}, \frac{r_1}{s_1}\right)}$$

$$k_2 = \frac{\begin{vmatrix} r_1 & a-1 \\ s_1 & b \end{vmatrix}}{\begin{vmatrix} r_1 & r_2 \\ s_1 & s_2 \end{vmatrix}} = \frac{r_1 b - (a-1)s_1}{l\left(\frac{r_2}{s_2}, \frac{r_1}{s_1}\right)}$$

By $\frac{r_1}{s_1} \leq *\frac{a-1}{b}* \leq \frac{r_2}{s_2}$ we have that both numerators are non-positive integers. The denominators are -1, so k_1 and k_2 are non-negative integers. This already proves that $\frac{a}{b}$ is reachable from $\frac{r_1}{s_1}$ and $\frac{r_2}{s_2}$. To establish α-reachability, we need that for $i = 1, 2$ if $k_i > 0$ then $k_1 r_1 + k_2 r_2 + 1 - r_i > \alpha(k_1 s_1 + k_2 s_2 - s_i)$. But this is true, as:

$$(k_1 r_1 + k_2 r_2 + 1 - r_i) - \alpha(k_1 s_1 + k_2 s_2 - s_i) =$$

$$\alpha s_i - r_i - (\alpha b - a) > 0$$

□

To prove the main theorem, we will need to yet another sequence.

Lemma 3.1.9. *For a positive rational $\frac{p}{q}$ with $q > 1$ there uniquely exist two non-negative rationals $\frac{p^-}{q^-}$ and $\frac{p^+}{q^+}$ such that $q^- < q$, $q^+ < q$, $l(\frac{q^-}{p^-}, \frac{p}{q}) = 1$ and $l(\frac{p}{q}, \frac{q^+}{p^+}) = 1$. We will call $\frac{p^-}{q^-}$ the one down of $\frac{p}{q}$ and denote it by $OD(\frac{p}{q})$. Similarly, $\frac{p^+}{q^+}$ called the one up of $\frac{p}{q}$ and denoted by $OU(\frac{p}{q})$. It also holds that $p = p^+ + p^-$ and $q = q^+ + q^-$. (Of course, the value of p^- and p^+ above also depends on q, not only on p, and the same way q^- and q^+ depends on p, so it is not that we defined a "+" and a "-" operation.)*

Proof. We need to solve the Diophantine equation $pq^- - p^-q = 1$. Observe that in this case $gcd(p^-, q^-) = 1$ holds automatically. The equation yields $pq^- \equiv 1 \ (mod\ q)$ which has a unique solution in the range $[0, q)$ as $gcd(p, q) = 1$. As $q > 1$, $q^- = 0$ is not a solution to the congruence. So we found the unique q^-, and then $p^- = (pq^- - 1)/q$. We can do a similar calculation for p^+, q^+ In that case we solve $pq^+ \equiv -1 \ (mod\ q)$, whose only solution in $[0, q)$ is $q - q^-$. Then $p^+ = (pq^+ + 1)/q$, and indeed $p^+ + p^- = (pq^+ + 1)/q + (pq^- - 1)/q = pq/q = p$. □

If we iteratively apply OD starting from a rational $0 < \frac{p}{q} < 1$ we obtain a decreasing sequence of rationals with smaller and smaller denominators. Finally, we have to get a rational with denominator 1, which by being smaller then $\frac{p}{q}$, must be 0. We will call the finite sequence $\frac{p}{q}$, $OD(\frac{p}{q})$, $OD(OD(\frac{p}{q}))$, ..., 0 the *one down sequence* of $\frac{p}{q}$.

Let us state two important facts about the one ups of the elements of a one down sequence:

Lemma 3.1.10. *Let $\frac{p_0}{q_0} = \frac{p}{q}$, $\frac{p_1}{q_1}$, ..., $\frac{p_n}{q_n} = 0$ be the one down sequence of $\frac{p}{q}$. Let $\frac{p_i'}{q_i'} = OU(\frac{p_i}{q_i})$. Then for any $0 \le i < j \le n$ we have:*

1) $\frac{p_i'}{q_i'} \le \frac{p_j'}{q_j'}$

2) $q_i' \ge q_j'$

Proof. Let $\frac{r}{s} = OD(\frac{u}{v})$ for any rational $\frac{u}{v}$. First we claim that $l(\frac{r}{s}, OU(\frac{u}{v})) = 1$. Let $\frac{u'}{v'} = OU(\frac{u}{v})$. By Lemma 3.1.9 we know that $u' = u - r$ and $v' = v - s$. This is enough to prove our claim as:

$$l(\frac{r}{s}, \frac{u'}{v'}) = (u-r)s - r(v-s) = ur - rs = l\left(\frac{r}{s}, \frac{u}{v}\right) = 1$$

So what we know about $OU(\frac{r}{s})$? Either it is $OU(\frac{u}{v})$, or it is larger, as it is easy to see that among all rationals x having $l(\frac{r}{s}, x) = 1$ the largest is $OU(\frac{r}{s})$. Also, by definition, among these rationals $OU(\frac{r}{s})$ has the smallest denominator. If we apply these observations to $\frac{u}{v} = \frac{p_i}{q_i}$ and $\frac{r}{s} = \frac{p_{i+1}}{q_{i+1}}$ we get the two claims of the lemma. □

We will need a simple statement about the function l:

Lemma 3.1.11. *There are no rationals $\frac{p_1}{q_1} < \frac{p_2}{q_2} < \frac{p_3}{q_3} < \frac{p_4}{q_4}$ such that $l(\frac{p_1}{q_1}, \frac{p_3}{q_3}) = 1$ and $l(\frac{p_2}{q_2}, \frac{p_4}{q_4}) = 1$.*

Proof. Assume the contrary. We have:

$$\frac{1}{q_1 q_2} \le \frac{p_2}{q_2} - \frac{p_1}{q_1} < \frac{p_3}{q_3} - \frac{p_1}{q_1} = \frac{l(\frac{p_1}{q_1}, \frac{p_3}{q_3})}{q_1 q_3} = \frac{1}{q_1 q_3}$$

This implies $q_3 < q_2$. On the other hand:

$$\frac{1}{q_3 q_4} \le \frac{p_4}{q_4} - \frac{p_3}{q_3} < \frac{p_4}{q_4} - \frac{p_2}{q_2} = \frac{l(\frac{p_2}{q_2}, \frac{p_4}{q_4})}{q_2 q_4} = \frac{1}{q_2 q_4}$$

This means $q_2 < q_3$, a contradiction. □

It is an easy observation that the one ups of the elements of the strong approximation sequence are above α.

Lemma 3.1.12. *For any $\alpha \in [0, 1)$ and $i \in \mathbb{Z}^+$ we have $OU(\tau_i(\alpha)) > \alpha$. Also, if $\frac{p}{q}$ is an element of the one down sequence of $\tau_i(\alpha)$, then $OU(\frac{p}{q}) > \alpha$.*

Proof. Let $\frac{r}{s} = \tau_i(\alpha)$ and $\frac{r^+}{s^+} = OU(\frac{r}{s})$. Notice that $0 < \tau_i(\alpha) < 1$, so $s > 1$, thus the one up indeed exists. Observe that $*\frac{r^+-1}{s^+}* \le *\frac{r-1}{s}* \le \tau_{i-1}(\alpha)$. The second inequality holds by definition of τ and the first is also true as:

$$(r-1)s^+ - (r^+ - 1)s = rs^+ - r^+ s + s - s^+ = s - 1 - s^+ \ge 0$$

This means that if $\frac{r^+}{s^+} \le \alpha$ then $\frac{r^+}{s^+}$ is in the set whose maximum we take to define $\tau_i(\alpha) = \tau(\alpha, \tau_{i-1}(\alpha))$, which contradicts to the fact that the maximum was $\frac{r}{s} < \frac{r^+}{s^+}$. The second statement is obvious from the first statement of this lemma and the first statement of Lemma 3.1.10. □

Using the above two lemmas we will prove a crucial connection between the one down sequence and the strong approximation sequence:

Lemma 3.1.13. *For any* $0 \leq \frac{p}{q} < 1$ *and any* $i \geq 0$ *the number* $\tau_i(\frac{p}{q})$ *is an element of the one down sequence of* $\frac{p}{q}$.

Proof. For $i = 0$ the statement is obvious. For $i > 0$ if $\frac{p}{q} = \tau_i(\frac{p}{q})$ then we are done. Otherwise, let us denote $\frac{r}{s} = \tau_i(\frac{p}{q})$. From the previous lemma $\frac{r^+}{s^+} = OU(\frac{r}{s}) > \frac{p}{q}$. If $\frac{r}{s}$ is not in the one down sequence then there is a last element $\frac{p'}{q'}$ of the sequence such that $\frac{p'}{q'} > \frac{r}{s}$. Let the next element in the sequence be $\frac{p''}{q''} < \frac{r}{s}$. But then the four rationals $\frac{p''}{q''} < \frac{r}{s} < \frac{p'}{q'} < \frac{r^+}{s^+}$ contradict Lemma 3.1.11. □

Reversing the one down sequence allows us to extend the definition to irrational numbers:

Definition 3.1.14. For a rational α we get the *reversed one down sequence* of α by reversing its one down sequence, that is it is the sequence $\beta_1 = 0$, β_2, ..., $\beta_k = \alpha$ where $\beta_k = \alpha$, β_{k-1}, ..., $\beta_1 = 0$ is the one down sequence of α. For an irrational $\alpha \in [0, 1)$ the reverse one down sequence is the infinite sequence $\beta_1 = 0$, β_2, ... where we get β_j in the following way. Find an n for which the reversed one down sequence $\beta'_1 = 0$, β'_2, ..., $\beta'_{k'} = \tau_n(\alpha)$ of $\tau_n(\alpha)$ is long enough, that is $k' > j$ and set $\beta_j = \beta'_j$.

The definition is good, that is it does not depend on the choice of n. Indeed as established in Lemma 3.1.13 the reversed one down sequence of $\tau_i(\alpha)$ contains $\tau_j(\alpha)$ if $i \geq j$ as $\tau_j(\alpha) = \tau_j(\tau_i(\alpha))$. Thus the reversed one down sequence of $\tau_i(\alpha)$ starts with the reversed one down sequence of $\tau_j(\alpha)$, so they have the same elements on the common positions. Also notice that the length of the one down sequence of $\tau_n(\alpha)$ is at least $n+1$, so there exists an n large enough. With this definition the generalization of Lemma 3.1.13 to irrationals is obvious:

Lemma 3.1.15. *For any* $\alpha \in [0, 1)$ *the strong approximation sequence of* α *is a subsequence of the reversed one down sequence of* α.

Remark 3.1.16. We remark here that the one down sequence is closely related to the Stern-Brocot tree as defined independently by Moritz Stern ([8]) and Achille Brocot ([2]). It is a nice arrangement of all possible positive rationals in an infinite binary search tree. One can find a good discussion about these trees at [1]. Here we only want to point out the relation of the one down and the analogously definable one up sequences to the Stern-Brocot tree. For any (rational or irrational) $0 < \alpha < 1$ let us consider the sequence of rationals that we get by starting from the root of the tree and searching for α as we do search in a binary search tree and writing down the rationals we see at each node. It is a finite sequence for a rational α as we find α sooner or later and infinite otherwise. The fact is that this sequence is a merge of the elements of the reversed one down sequence without the 0 and the elements of the reversed one up sequence. That is all its elements are from one of the two sequences, and all elements of the reversed one down and one up sequences are present in their original order. Obviously the elements larger then α belongs to the one up sequence, the other smaller elements belong to the one down sequence.

Next we prove that the $q\alpha - p$ quantity strictly monotonically decreases on the $\frac{p}{q}$ elements of the reversed one down sequence of α:

Lemma 3.1.17. *If $\frac{p_0}{q_0} = 0$, $\frac{p_1}{q_1}$, ... is the (finite or infinite) reversed one down sequence of the (rational or irrational) $\alpha \in [0, 1)$ then for any two indices i, j of this sequence if $i > j$ then $\alpha q_i - p_i < \alpha q_j - p_j$.*

Proof. For any $i > 0$ index of the reversed one down sequence by Lemmas 3.1.12 and 3.1.9 we know that $\frac{p_i - p_{i-1}}{q_i - q_{i-1}} = OU(\frac{p_i}{q_i}) > \alpha$. As $q_i > q_{i-1}$ multiplying by $q_i - q_{i-1}$ implies $\alpha q_{i-1} - p_{i-1} > q_i\alpha - p_i$ which proves the lemma. □

Finally, we have all the tools needed to prove the main theorem.

Proof of the main theorem. We will prove by induction on i the stronger statement that every element of the one down sequence of $\tau_i(\alpha)$ is contained in $H_i(\alpha)$. Observe that by Lemma 3.1.15 the reversed one down sequence of $\tau_i(\alpha)$ is always the beginning of the reversed one down sequence of α. For $i = 0$ the statement is trivial. For $i = 1$, let q be the smallest positive integer for which $\frac{1}{q} \leq \alpha$. Then we have $\tau_1(\alpha) = \frac{1}{q}$, $H_1(\alpha) = \{0, \frac{1}{q}\}$ and $OD(\frac{1}{q}) = 0$ so the statement is true. Now assume that for $i \geq 1$ the one down sequence of $\tau_i(\alpha)$ is contained in $H_i(\alpha)$. If $\tau_i(\alpha) = \tau_{i+1}(\alpha)$, meaning that we have already reached α, then we are done.

Otherwise, let $\frac{p}{q} > \tau_i(\alpha)$ be an element of the one down sequence of $\tau_{i+1}(\alpha)$. (The smaller elements of the one down sequence of $\tau_{i+1}(\alpha)$ are already in $H_i(\alpha)$.) First observe that $*\frac{p-1}{q}* \leq \tau_i(\alpha)$. Indeed, for $\frac{p'}{q'} = \tau_{i+1}(\alpha)$ we have $*\frac{p'-1}{q'}* \leq \tau_i(\alpha)$ by definition and it is very easy to see that if $\frac{r}{s} = OD(\frac{u}{v})$ then $*\frac{r-1}{s}* < *\frac{u-1}{v}*$. So there are two consecutive elements $\frac{u_1}{v_1}, \frac{u_2}{v_2}$ of the one down sequence of $\tau_i(\alpha)$ for which $\frac{u_1}{v_1} \leq *\frac{p-1}{q}* \leq \frac{u_2}{v_2}$. But these two rationals and $\frac{p}{q}$ are all in the reversed one down sequence of α, so we know by Lemma 3.1.17 that $\alpha q - p < \alpha v_j - u_j$ for $j = 1, 2$. By Lemma 3.1.8 we have that $\frac{p}{q} \in H(\{\frac{u_1}{v_1}, \frac{u_2}{v_2}\}, \alpha) \subseteq H(H_i, \alpha) \subseteq H_{i+1}(\alpha)$ which we wanted to prove. □

Observe that we proved a somewhat stronger statement in that we do not really need all the elements of H_{i-1} to get τ_i and the new elements of its one down sequence, we need only those that are the elements of the one down sequence of τ_{i-1}. It will be convenient to use this property, so we state it more precisely:

Lemma 3.1.18. *For any $i \geq 1$ if the set A contains all the elements of the one down sequence of $\tau_{i+1}(\alpha)$ which are at least $\tau_{i-1}(\alpha)$ and at most $\tau_i(\alpha)$ then $H(\alpha, A)$ contains all the elements $\frac{p}{q}$ of the one down sequence of $\tau_{i+1}(\alpha)$ for which $\tau_i(\alpha) < \frac{p}{q}$.*

Proof. At the end of the proof of the theorem above we found two elements $\frac{u_1}{v_1}, \frac{u_2}{v_2}$ of the one down sequence of $\tau_i(\alpha)$ such that $\frac{u_1}{v_1} \leq *\frac{p-1}{q}* \leq \frac{u_2}{v_2}$ and we used these to α-reach $\frac{p}{q}$. But we had $\frac{p}{q} > \tau_i(\alpha)$ so by the definition of τ we have $*\frac{p-1}{q}* > \tau_{i-1}(\alpha)$. Thus $\frac{u_1}{v_1}$ must be at least $\tau_{i-1}(\alpha)$ and trivially $\frac{u_2}{v_2}$ is at most $\tau_i(\alpha)$, so if we have all the elements of the one down sequence between $\tau_{i-1}(\alpha)$ and $\tau_i(\alpha)$ we can α-reach all other elements up to $\tau_{i+1}(\alpha)$. □

3.2 Speed of the approximation

Although not necessary for the main goal of this thesis, here we will show some results about the speed of our approximation sequences. First we observe that we can give an upper bound to the distance of α and $\tau_i(\alpha)$ using the denominator of $\tau_i(\alpha)$. This motivates us to concentrate on how fast the denominators grow in the rest of the section.

Lemma 3.2.1. *Let* $\alpha < 1$, $\frac{p}{q} = \tau_i(\alpha)$, *then* $*\frac{p+1}{q}* > \alpha$, *thus* $\alpha - \tau_i(\alpha) < \frac{1}{q}$.

Proof. Let $\frac{p'}{q'} = OU(\frac{p}{q})$. Then $\frac{p'}{q'} \leq *\frac{p+1}{q}*$ as:

$$(p+1)q' - p'q = pp' - p'q + q' = q' - 1 \geq 0$$

But according to Lemma 3.1.12 we have $\frac{p'}{q'} > \alpha$. □

Lemma 3.1.15 showed that the strong approximation sequence is a subsequence of the reversed one down sequence. Now we characterize this connection more precisely:

Lemma 3.2.2. *Let* $\frac{p_0}{q_0} = \frac{0}{1}$, $\frac{p_1}{q_1}$, ... *be the reversed one down sequence of* α. *Assume for some* $i \geq 1$ *that* $\tau_{i-1}(\alpha) < \alpha$ *and let* $\tau_{i-1}(\alpha) = \frac{p_j}{q_j}$ *for some* $j \geq 0$. *Then* $\tau_i(\alpha) = \frac{p_k}{q_k}$ *where* $k = max\{t \mid l(\frac{p_j}{q_j}, \frac{p_t}{q_t}) \leq q_j\}$.

Proof. Observe that $*\frac{u-1}{v}* \leq \frac{r}{s}$ if and only if $l(\frac{r}{s}, \frac{u}{v}) \leq s$. As we already know that $\tau_i(\frac{p}{q})$ is one of the elements of the one down sequence, then it must be the largest of those satisfying $l(\frac{p_j}{q_j}, \frac{p_t}{q_t}) \leq q_j$. □

A statement strongly related to Lemma 3.1.17 is the following:

Lemma 3.2.3. *Let* $\frac{p_0}{q_0} < \frac{p_1}{q_1} < ...$ *be a (finite or infinite) reversed one down sequence. Then for any* i, j, k *indices of the this reversed one down sequence if* $i > j$ *then* $l(\frac{p_k}{q_k}, \frac{p_i}{q_i}) > l(\frac{p_k}{q_k}, \frac{p_j}{q_j})$.

Proof. Lemma 3.1.17 implies the current lemma if both i and j are less then k with the substitution $\alpha = \frac{p_k}{q_k}$ and with the trivial observation that $l(\frac{p_k}{q_k}, \frac{p_i}{q_i}) > l(\frac{p_k}{q_k}, \frac{p_j}{q_j})$ if and only if $\frac{p_k}{q_k}q_i - p_i < \frac{p_k}{q_k}q_j - p_j$. If $j \leq k \leq i$ then the statement is trivial by the relation of the sign of the value of l to the ordering of the parameters. If both i and j are at least k, then $\frac{p_j}{q_j} < \frac{p_i}{q_i}$ and also $q_j < q_i$, so:

$$l(\frac{p_k}{q_k}, \frac{p_j}{q_j}) = q_k q_j (\frac{p_j}{q_j} - \frac{p_k}{q_k}) < q_k q_i (\frac{p_i}{q_i} - \frac{p_k}{q_k}) = l(\frac{p_k}{q_k}, \frac{p_i}{q_i})$$

which proves the remaining case of our lemma. □

Next we give an upper bound to the denominator of the kth element of the strong approximation sequence.

Lemma 3.2.4. *Let* $\frac{p}{q} = \tau_k(\alpha)$, $\frac{p'}{q'} = OU(\frac{p}{q})$. *Then* $q \leq (q'+1)^k$.

Proof. Not surprisingly, we prove by induction on k. Let $\frac{u}{v} = \tau_{k-1}(\alpha)$, $\frac{u'}{v'} = OU(\frac{u}{v})$. We know that walking down on the one down sequence of $\frac{p}{q}$, we reach $\frac{u}{v}$ sooner or later. By the strict monotonicity of l established in Lemma 3.2.3 and by the characterization given in Lemma 3.2.2, we also know that we can have at most v steps. In one step in a one down sequence from element a the denominator decreases by the denominator of $OU(a)$ (Lemma 3.1.9). We also know from lemma 3.1.10 that the denominator of the corresponding OU's monotonically decreases along a one down sequence. Putting all together, we have that we decreased the denominator at most v times with at most q', thus $q \leq v + q'v = (q'+1)v \leq (q'+1)(v'+1)^{k-1} \leq (q'+1)^k$. □

Next we characterize what are the possible next elements from a given rational in a reversed one down sequence.

Lemma 3.2.5. *Let $\frac{p}{q}$ be a rational and let $\frac{r}{s} = OU(\frac{p}{q})$. Then the set of all rationals x having $l(\frac{p}{q}, x) = 1$ is $\{\frac{r+tp}{s+tq} \mid t \in \mathbb{N}\}$. Specially the set of rationals y for which $OD(y) = \frac{p}{q}$ is $\{\frac{r+tp}{s+tq} \mid t \in \mathbb{Z}^+\}$. Finally, $OU(\frac{r+tp}{s+tq}) = \frac{r+(t-1)p}{s+(t-1)q}$ for all $t \in \mathbb{Z}^+$.*

Proof. If a rational $\frac{u}{v}$ has $l(\frac{p}{q}, \frac{u}{v}) = 1$ then v is a solution to the linear congruence:

$$px \equiv -1 \ (q)$$

As established in the proof of 3.1.9 the integer s is the unique solution of this congruence between 1 and $q-1$. Thus the set of all positive solutions is $tq + s$. If $v = tq + s$ then u has to be $r + tp$ and $\frac{r+tp}{s+tq}$ is indeed a good rational, which proves the first statement. Notice that in $\frac{r+tp}{s+tq}$ * was not forgot accidently: $gcd(r + tp, s + tq) = 1$ indeed for any t. The second statement is a trivial consequence of the first as $\frac{r}{s}$ is the only element of the set $\{\frac{r+tp}{s+tq} \mid t \in \mathbb{N}\}$ with denominator not larger then q. By

$$(r + (t-1)p)(s + tq) - (r + tp)(s + (t-1)q) = rq - ps = 1$$

$l(\frac{r+tp}{s+tq}, \frac{r+(t-1)p}{s+(t-1)q}) = 1$ and $s + tq > s + (t-1)q$ trivially, so the the last statement is also true. □

To establish a lower bound on the growing speed of the denominators of the strict approximation sequence, we will do one last induction:

Lemma 3.2.6. *Let $\frac{p_0}{q_0}$, $\frac{p_1}{q_1}$, ..., $\frac{p_n}{q_n}$ be a segment of a reversed one down sequence, that is a sequence of rationals such that $q_0 < q_1 < ... < q_n$ and $l(\frac{p_i}{q_i}, \frac{p_{i+1}}{q_{i+1}}) = 1$ for all $i < n$. Then we have:*

a) $q_i \geq q_0 + l(\frac{p_0}{q_0}, \frac{p_i}{q_i})q'_0$

b) $q'_i \geq l(\frac{p_0}{q_0}, \frac{p'_i}{q'_i})q'_0$

Proof. For $i = 0$ both inequalities hold trivially, with equality. Let us assume both equalities for $i = k$. Using the the previous lemma there is a positive integer t for which $\frac{p_{k+1}}{q_{k+1}} = \frac{p'_k+tp_k}{q'_k+tq_k}$ and $\frac{p'_{k+1}}{q'_{k+1}} = \frac{p'_k+(t-1)p_k}{q'_k+(t-1)q_k}$. Then

$$q_0 + l\left(\frac{p_0}{q_0}, \frac{p_{k+1}}{q_{k+1}}\right) q_0' = q_0 + \left(l\left(\frac{p_0}{q_0}, \frac{p_k'}{q_k'}\right) + tl\left(\frac{p_0}{q_0}, \frac{p_k}{q_k}\right)\right) q_0' \leq$$
$$\leq q_0 + (q_k'/q_0' + t(q_k - q_0)/q_0')\, q_0' = q_{k+1} + (1-t)q_0 \leq q_{k+1}$$

For the first equation we used Lemma 3.1.7, for the first inequality we used the inductive hypothesis. The same way:

$$l\left(\frac{p_0}{q_0}, \frac{p_{k+1}'}{q_{k+1}'}\right) q_0' = \left(l\left(\frac{p_0}{q_0}, \frac{p_k'}{q_k'}\right) + (t-1)l\left(\frac{p_0}{q_0}, \frac{p_k}{q_k}\right)\right) q_0' \leq$$
$$\leq (q_k'/q_0' + (t-1)(q_k - q_0)/q_0')\, q_0' = q_{k+1}' + (1-t)q_0 \leq q_{k+1}'$$

□

Corollary 3.2.7. *Let $\frac{p}{q} = \tau_i(\alpha)$, $\frac{u}{v} = \tau_{i+2}(\alpha)$, $\frac{p'}{q'} = OU(\frac{p}{q})$. Then $v \geq q(q'+1)$.*

Proof. By Lemma 3.2.2, $l(\frac{p}{q}, \frac{u}{v}) \geq q$ (it is already true if we put the element right after $\tau_{i+1}(\alpha)$ in the reversed one down sequence instead of $\frac{u}{v}$). By the first statement of the previous lemma, $v \geq q + qq' = q(q'+1)$. □

Putting the two bounds together, we get our main result about the convergence speed of the strong approximation sequence:

Theorem 3.2.8. *If $k \geq 0$, $\alpha < 1$, $\frac{p}{q} = \tau_k(\alpha)$, $\frac{u}{v} = \tau_{3k}(\alpha)$ then $v \geq q^2$.*

Proof. Let $\frac{p'}{q'} = OU(\frac{p}{q})$. By Lemma 3.2.4 we know that $q \leq (q'+1)^k$. By corollary 3.2.7 and by the fact that the denominators of the corresponding one ups grow monotonically along a strong approximation sequence (as they do so along a reversed one down sequence), we have that the denominator of $\tau_{k+2i}(\alpha))$ is at least $q(q'+1)^i$, thus $v \geq q(q'+1)^k \geq q^2$ indeed. □

3.3 Accuracy of the approximation relative to the denominator

While we will not use the results in the previous section, we will use those in this section. It will turn out in Chapter 5 that using a well chosen graph extension of size (v, e) where v and e are coprimes we will be able to characterize arbitrary relations on sets of size polynomial in $\left\lfloor \frac{1}{\alpha e - v} \right\rfloor$. On the other hand we will need to use relations on sets of size $v + h$ for some fixed constant h. As for the interesting extensions we will have $\frac{v}{e} = \tau_k(\alpha)$ it will be enough to have that $v + h$ is polynomial in $\left\lfloor \frac{1}{\alpha e - v} \right\rfloor$ for these numbers. This is the goal of this section.

First we state a trivial consequence of Lemma 3.1.17:

Lemma 3.3.1. *Let $0 = \frac{p_0}{q_0}, \frac{p_1}{p_1}, \frac{p_2}{q_2}, \ldots$ be the reversed one down sequence of $\alpha \in [0,1)$. Then for any $i < j$ indices of the above sequence we have $\left\lfloor \frac{1}{\alpha q_i - p_i} \right\rfloor \leq \left\lfloor \frac{1}{\alpha q_j - p_j} \right\rfloor$.*

Let us next investigate what coefficients are possible when α-reaching a number.

Lemma 3.3.2. *Let $\frac{p_1}{q_1}, \frac{p_2}{q_2}, \frac{p}{q} \le \alpha$. Assume $\frac{p}{q}$ is α-reached from $\frac{p_1}{q_1}$, $\frac{p_2}{q_2}$ with coefficients k_1 and k_2. Then $k_i < \frac{1}{q_i\alpha - p_i} + 1$*

Proof. By definition $\frac{p_1(k_1-1)+p_2k_2+1}{q_1(k_1-1)+q_2k_2} > \alpha$. As either $k_2 = 0$ or $\frac{p_2k_2}{q_2k_2} = \frac{p_2}{q_2} < \alpha$, the above inequality also implies:

$$\frac{p_1(k_1-1)+1}{q_1(k_1-1)} > \alpha$$

$$k_1 - 1 < \frac{1}{q_1\alpha - p_1}$$

$$k_1 < \frac{1}{q_1\alpha - p_1} + 1$$

which yields the statement for k_1. The same argument works for k_2. □

Lemma 3.3.3. *For every n and h non-negative constants there is a $c = c(n,h)$ such that for any $0 < \alpha < \frac{1}{2}$ if $\frac{p}{q} = \tau_n(\alpha)$ then $q + h \le \left\lfloor \frac{1}{q\alpha - p} \right\rfloor^c$.*

Proof. We first prove by induction on n that for $h = 0$ the choice $c(n,0) = 3n$ is good. As $\alpha < \frac{1}{2}$ we have $\left\lfloor \frac{1}{1\alpha - 0} \right\rfloor > 2$, therefore $\left\lfloor \frac{1}{p\alpha - q} \right\rfloor > 2$ by Lemma 3.3.1, which yields $\left\lfloor \frac{1}{p\alpha - q} \right\rfloor^3 > 2(\frac{1}{p\alpha - q} + 1)$. As we saw in the proof of the equivalence of the two approximation sequences, there are two elements $\frac{p_1}{q_1}$ and $\frac{p_2}{q_2}$ of the one down sequence of $\frac{r}{s} = \tau_{n-1}(\alpha)$ such that $p = k_1p_1 + k_2p_2 + 1$ and $q = k_1q_1 + k_2q_2$. By Lemma 3.3.2 and by the fact that $q > s$, $s \ge q_i$ and $q\alpha - p < s\alpha - r \le q_i\alpha - p_i$:

$$q \le \left(\frac{1}{q_1\alpha - p_1} + 1\right) q_1 + \left(\frac{1}{q_2\alpha - p_2} + 1\right) q_2 \le 2\left(\frac{1}{q\alpha - p} + 1\right) s <$$

$$< \left\lfloor \frac{1}{p\alpha - q} \right\rfloor^3 \left\lfloor \frac{1}{s\alpha - r} \right\rfloor^{c(n-1)} \le \left\lfloor \frac{1}{p\alpha - q} \right\rfloor^{(3n-3)+3} = \left\lfloor \frac{1}{p\alpha - q} \right\rfloor^{3n}$$

By $q \ge 1$ and $\left\lfloor \frac{1}{p\alpha - q} \right\rfloor > 2$ it is obvious that $c(n,h) = c(n,0) + \lceil \log_2(h+1) \rceil$ is good. □

Chapter 4

Capturing the density of a random graph with first order formulae

In this chapter, our aim is to create a first order "handle" on the value of α in an α-graph. That is, we are going to be able to create first order formulae which characterize extensions in an α-graph whose size is (v, e) where $gcd(v, e) = 1$ and $\tau_i(\alpha) = \frac{v}{e}$. Of course these formulae will not depend on the exact value of α, this is the very purpose of this construction. But they do depend on which of the intervals $[\frac{1}{k}, \frac{1}{k-1}]$ for $k \geq 3$ contains α. This will not cause any problem as by the Very Dense Condition we will only need to deal with finitely many such intervals, so we can combine the formulae for the individual intervals into one big formula which works in the whole $[0, \frac{1}{2}]$ interval.

So we fix for this whole chapter (and as a matter of fact for the next, too) an integer $k \geq 3$. We also fix an irrational $\frac{1}{k} < \alpha < \frac{1}{k-1}$ and finally an α-graph G.

4.1 Rooted graphs and their hybrids

We call a graph H together with $k+1$ distinct designated vertices x_1, x_2, ..., x_{k-2}, y, z, and w a *k-rooted graph* or *k-rgraph* if there is no edge connecting two of the first k designated vertices. We shell often omit the k as it will be fixed except for the last chapter. We will use the notation $\underline{x} = (x_1, ..., x_{k-2})$ and denote the above rooted graph by $\mathcal{H} = (H, \underline{x}, y, z, w)$. We will also use $X = \{x_1, ..., x_{k-2}\}$. We call the x_1, x_2, ..., x_{k-2}, y, z designated vertices the *base vertices* and often regard a rooted graph $(H, \underline{x}, y, z, w)$ as a graph extension $(X \cup \{y, z\}, H)$. We call the last designated vertex of an rgraph the *counting vertex*. Figure 4.1 shows an example rooted graph.

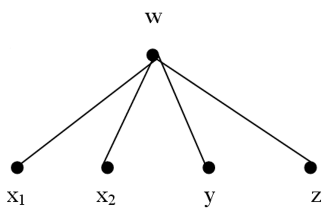

Figure 4.1: A very simple rooted graph for $k = 4$.

We call two rgraphs $(H, \underline{x}, y, z, w)$ and $(H', \underline{x}', y', z', w')$ *isomorphic* if an isomorphism from H to H' maps x_i to x'_i, y to y', z to z' and w to w'. A *subrgraph* of an rgraph $(H, \underline{x}, y, z, w)$ is another rgraph $(H', \underline{x}, y, z, w)$ where H' is a subgraph of H containing x_1,..., x_{k-2}, y, z and w. We call $\mathcal{H}'$ a proper subrgraph of $\mathcal{H}$ if $\mathcal{H}'$ is a subrgraph of $\mathcal{H}$ but $\mathcal{H}' \neq \mathcal{H}$. We define the size of a rooted graph to be the size of the corresponding extension, i.e., it is (v, e) where v is the number of vertices excluding the base vertices (but not excluding w) and e is the number of edges.

As we have fixed an $\frac{1}{k} < \alpha < \frac{1}{k-1}$ irrational the notion of sparse, dense, safe, and rigid extensions are defined (see Section 2.1). When using the words sparse, dense, safe, and rigid for rooted graphs we mean that the corresponding extensions are such.

Definition 4.1.1. We call the k-rooted graph $\mathcal{H}$ *valid* if

1. $\mathcal{H}$ is rigid and
2. each proper subrgraph of $\mathcal{H}$ is safe and
3. no base vertex of $\mathcal{H}$ is isolated and
4. all automorphisms of the underlying graph of $\mathcal{H}$ fixing each of the base vertices also fixes the counting vertex.

Notice that in this definition rigid and safe can be equivalently replaced by dense and sparse. Also notice that item 3 is equivalent to saying that for $\mathcal{H} = (H, \underline{x}, y, z, w)$ all the extensions (A, H) are safe for all proper subset A of $X \cup \{y, z\}$.

Also observe that if we remove the base vertices from the underlying graph of a valid rooted graph then it remains connected. Indeed, otherwise there would be two disjoint unconnected non-empty vertex sets L_1, L_2 of $H \setminus (X \cup \{y, z\})$ such that $L_1 \cup L_2$ contains all non-base vertices. Then let H_1 be the induced subgraph of H spanned by $L_1 \cup X \cup \{y, z\}$ and H_2 be the induced subgraph of H spanned by $L_2 \cup X \cup \{y, z\}$. By item 2 of validity both $(X \cup \{y, z\}, H_1)$ (of size (v_1, e_1)) and $(X \cup \{y, z\}, H_2)$ (of size (v_2, e_2)) are sparse, thus $(X \cup \{y, z\}, H)$ of size $(v_1 + v_2, e_1 + e_2)$ is also sparse, contradicting item 1.

We will build larger rooted graphs from two smaller ones via the following construction.

Definition 4.1.2. $\mathcal{H}' = (H', \underline{x}', y', z', w')$ and $\mathcal{H}'' = (H'', \underline{x}'', y'', z'', w'')$ be two rooted graphs. For $0 \leq l < k$ the *+l-hybrid* of these rooted graphs with non-negative integer multiplicities m', m'' is the following rooted graph $(H, \underline{x}, y, z, w)$. First we take the union of m' isomorphic copies of H' and m'' isomorphic copies of H'' such that these copies are pairwise disjoint except for following cases. For any $0 \leq j \leq k-2$ the image of x'_j and x''_j is the same vertex x_j for each copy, the image of y' is the same vertex w for each copy of H', the image of z'' is also the above vertex w for each copy of H'', the image of y'' is the vertex y in each copy of H'', finally the image of z' is the same vertex z in each copy of H'. If (because m' and/or m'' is 0) any of x_i, y, z or w was not created above, we just add those as isolated vertices. Finally we add l extra edges to this rooted graph: we connect w with x_1, ..., x_{l-1}, and we connect w with x_l if $1 \leq l \leq k-2$ or with y if $l = k-1$. We will denote the above hybrid as $Hib(l, \mathcal{H}', \mathcal{H}'', m', m'')$.

We study first when the hybrid construction on graphs preserves validity.

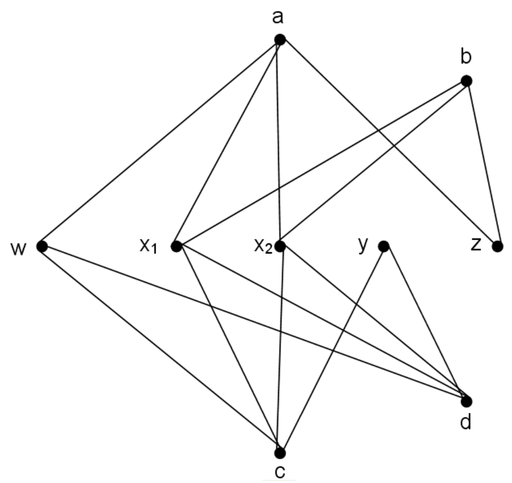

Figure 4.2: The hybrid $Hyb(0, \mathcal{H}, \mathcal{H}, 2, 2)$ where $\mathcal{H}$ is the rooted graph on Figure 4.1.

Lemma 4.1.3. *Let $\mathcal{H}'$ and $\mathcal{H}''$ be two valid graphs of size (v', e') and (v'', e'') respectively. The $+l$-hybrid $\mathcal{H}$ of these graphs with positive multiplicities m', m'' is of size $(v, e) = (1 + m'v' + m''v'', m'e' + m''e'' + l)$. It is valid if and only if $v/e < \alpha$, $(v - v')/(e - e') > \alpha$ and $(v - v'')/(e - e'') > \alpha$. If $v/e > \alpha$ then $\mathcal{H}$ is safe. If $(v - v')/(e - e') < \alpha$ (resp. $(v - v'')/(e - e'') < \alpha$) then the hybrid of $\mathcal{H}'$, $\mathcal{H}''$ with multiplicities $m' - 1$, m'' (resp. m', $m'' - 1$) is dense.*

Proof. The size of the hybrid (even with non-negative multiplicities) is clearly as stated since it consists of m' copies of $\mathcal{H}'$, m'' copies of $\mathcal{H}''$ and these copies are disjoint except for the uncounted base vertices. The plus one in the formula for v comes from the vertex w which is counted in the hybrid but was not counted before. The plus l in the formula for e comes from the l extra edges added at the end of the construction.

Note that a rooted graph $\mathcal{G}$ is dense if and only if for its size $(v^{\mathcal{G}}, e^{\mathcal{G}})$ it holds that $v^{\mathcal{G}}/e^{\mathcal{G}} < \alpha$. This immediately implies the last statement and (since the hybrids with multiplicities $m' - 1$, m'' and m', $m'' - 1$ are proper subrgraphs of $\mathcal{H}$) also the only if part of the statement on the validity of $\mathcal{H}$.

We claim that if both $(v - v')/(e - e') > \alpha$ and $(v - v'')/(e - e'') > \alpha$ then each proper subrgraph $\mathcal{H}^*$ of $\mathcal{H}$ is sparse. For the size (v^*, e^*) of $\mathcal{H}^*$ we have to prove $v^* - \alpha e^* \geq 0$. By the validity of $\mathcal{H}'$ and $\mathcal{H}''$ we decrease $v^* - \alpha e^*$ by removing the non-base vertices from $\mathcal{H}^*$ of any copy of $\mathcal{H}'$ or H'' not completely contained in $\mathcal{H}^*$. Notice that if w is not in $\mathcal{H}^*$ then this includes the removal of all non-base vertices. Also, if $w \in V(H^*)$, we only increase this value if we remove any of the edges connecting w to the base vertices, so we can assume that all l edges are preserved. Thus the minimum will be obtained by $(v', e') = (0, 0)$ (if w and thus everything else was removed), or by a $+l$-hybrid of the graphs $\mathcal{H}'$, $\mathcal{H}''$ with multiplicities $n' \leq m'$ and $n'' \leq m''$. In the former case the inequality trivially holds. Among the graphs in the latter case the maximum is realized by one of the two hybrids where $n' = m'$ and $n'' = m'' - 1$ or $n' = m' - 1$ and $n'' = m''$ — where the inequality was assumed.

If $v/e > \alpha$ then $\mathcal{H}$ is sparse and so are its subrgraphs as proved above since in this case $(v - v')/(e - e') > \alpha$ and $(v - v'')/(e - e'') > \alpha$. Thus $\mathcal{H}$ is safe as claimed.

The if part of the statement on the validity of $\mathcal{H}$ also follows from the above claim. Notice that for $\mathcal{H} = (H, \underline{x}, y, z, w)$ any automorphism of H fixing x_1, ..., x_{k-2}, y and z has to fix w since w is a cutpoint in $H \backslash X$ separating y from z. □

4.2 Relations characterizing rgraphs

We use $(k+2)$-ary relations on the vertices of G to distinguish subgraphs isomorphic to a rooted graph $\mathcal{H}$. We start with some technical definitions. For a $(k+2)$-ary relation R we write $R(\underline{x}, y, z, w)$ as a shorthand for $R(\underline{x}, y, z, w, w)$. We define $R'(\underline{x}, y, z) = \{w \mid R(\underline{x}, y, z, w)\}$ and $R'(\underline{x}, y, z, w) = \{t \mid R(\underline{x}, y, z, w, t)\}$. We further define $R^{\underline{x}}(y, z)$ to be the relation $\exists w R(\underline{x}, y, z, w)$.

Definition 4.2.1. We call a k-tuple $(\underline{x}, y, z)$ of distinct vertices *R-separated* if the sets $R'(\underline{x}, y, z, w)$ are pairwise disjoint for all $w \in R'(\underline{x}, y, z)$.

Definition 4.2.2. We say that an rgraph $\mathcal{H}$ *present* in an rgraph $\mathcal{H}'$ if $\mathcal{H}'$ has a subrgraph isomorphic to $\mathcal{H}$.

Definition 4.2.3. We say that an rgraph $\mathcal{H}$ of size (v, e) is *isolated* in $(G, \underline{x}, y, z, w)$ if $(G, \underline{x}, y, z, w)$ has a subrgraph $\mathcal{H}' = (H', \underline{x}, y, z, w)$ such that $\mathcal{H}'$ is isomorphic to $\mathcal{H}$, and for any rigid extension H'' of H' in G with at most v vertices there is no edge in $E(H'') - E(H')$ having an endpoint in $V(H') \backslash (X \cup \{y, z, w\})$.

Notice that being isolated implies being present. Also observe that being isolated means being present in such a way that the corresponding extension in G is v-generic (see Definition 2.2.1).

Definition 4.2.4. We say that a $(k+2)$-ary relation $R \subset V(G)^{k+2}$ *characterizes* the finite rooted graph $\mathcal{H}$ if for any vertices $\underline{x}$, y, z and w of G both assertions below are satisfied:

a) If $R(\underline{x}, y, z, w)$ holds then $(G, \underline{x}, y, z, w)$ has a subrgraph $\mathcal{H}' = (H', \underline{x}, y, z, w)$ isomorphic to $\mathcal{H}$ and $R'(\underline{x}, y, z, w) = V(H') \setminus \{x_1, ..., x_{k-2}, y, z\}$.

b) If $\mathcal{H}$ is isolated in $(G, \underline{x}, y, z, w)$ then $R(\underline{x}, y, z, w)$ holds.

The two criteria clearly implies that $R(\underline{x}, y, z, w)$ has to be in between $\mathcal{H}$ being present and $\mathcal{H}$ being isolated in $(G, \underline{x}, y, z, w)$.

4.3 Counting and comparing using relations characterizing valid rgraphs

In the hybrid construction we need numerical parameters: the two multiplicities and the number of edges to inject. The last one causes no problems, as we know apriori that there are only a fixed, finite number (k, which is fixed) possibilities, which we will be able to encode to our formulae easily. But to deal with the first kind of parameters we will need to somehow represent numbers in graphs. If given a relation R characterizing a valid rooted graph, we would like to represent the number i by choosing a k-tuple $\underline{x}$, y, z such that $|\{w \mid R(\underline{x}, y, z, w)\}| = i$. First we study the limits of this approach.

Lemma 4.3.1. *Let R be a relation characterizing the valid rgraph $\mathcal{H} = (H, \underline{x}, y, z)$ of size (v, e). For any R-separated triple $(\underline{x}, y, z)$ one has $|R'(\underline{x}, y, z)| < k/(\alpha e - v)$. For any nonnegative integer $i < k/(\alpha e - v)$ there exist an R-separated k-tuple $(\underline{x}, y, z)$ with $|R'(\underline{x}, y, z)| = i$. The above R-separated k-tuple $(\underline{x}, y, z)$ can be chosen such a way that there are no edges among the base vertices and there are no extra edges in any copy of H, that is the graph spanned by the set $X \cup \{y, z\} \cup R'(\underline{x}, y, z, w)$ in G is isomorphic to H for any $w \in R'(\underline{x}, y, z)$.*

Proof. For a positive integer i let us call $\mathcal{H}^i$ the (not rooted) graph consisting of i isomorphic copies of H that are disjoint except for identifying the corresponding base vertices in each of the copies. $\mathcal{H}^i$ has $iv + k$ vertices and ie edges.

Let $(\underline{x}, y, z)$ be an R-separated k-tuple. Notice that with $i = |R'(\underline{x}, y, z)|$ the graph $\mathcal{H}^i$ appears as a subgraph of G: its set of vertices is the union of the sets $R'(\underline{x}, y, z, w)$ for $w \in R'(\underline{x}, y, z)$ and the set $X \cup \{y, z\}$. By the sparsity axiom we must have $iv + k > \alpha ie$ proving the first claim.

For the second claim let $i < k/(\alpha e - v)$ and consider the extension $\mathcal{H}^i$ over the empty graph. We claim that this extension is safe. To prove it we have to prove that for any subgraph of $\mathcal{H}^i$ with v' vertices and e' edges we have $v' - \alpha e' \geq 0$. Assume the contrary and fix a subgraph H^* violating this inequality. By the validity of $\mathcal{H}$ we decrease this formula by removing all non-base vertices of any copy of $\mathcal{H}$ not entirely contained in H^*. (Notice that here we not only use the second, but also the third condition of validity, which means as remarked there that the extension also becomes safe if we remove one or more base vertices.) Thus the minimum is either realized by the empty graph or the graph $\mathcal{H}^{i'}$ for some $1 \leq i' \leq i$. We get no negative values in any of these cases. Now the safe extension axiom $B^v_{\emptyset, \mathcal{H}^i}$ gives an isolated embedding of $\mathcal{H}^i$ in G. The images of the base points form an R-separated k-tuple $(\underline{x}, y, z)$ with $|R'(\underline{x}, y, z)| = i$. As any extra edge would form a small rigid extension, the last statement of the lemma is also true.

□

To use this kind of representation of numbers, we will at least need to compare cardinalities of finite sets. First we capture general binary relations.

Lemma 4.3.2. *Let R be a relation characterizing a valid rooted graph of size (v, e). Let x_1, x_2, ..., x_r be disjoint vertices of G for $0 \leq r \leq k - 3$. Let A and B be finite sets of vertices of G disjoint from each other and the x_i's, and let $T \subseteq A \times B$ be a binary relation. If $|T| < \frac{k-2-r}{\alpha e - v}$ then there exists $k - 2 - r$ distinct vertices x^T_{r+1}, ..., x^T_{k-2} in G such that for $\underline{x}^T = (x_1, x_2, ..., x_r, x^T_{r+1}, ..., x^T_{k-2})$ we have that $R^{\underline{x}^T}$ restricted to $A \times B$ is T.*

Proof. We use the safe extension axiom $B^v_{H_1, H_2}$ for the following graphs: Let H_1 be the induced subgraph spanned by $A \cup B \cup \{x_1, ..., x_r\}$ in G. To get H_2 add $k - 2 - r$ new vertices x_{r+1}, ..., x_{k-2} to H_1 and for all pairs $(y, z) \in T$ add a disjoint copy of $\mathcal{H}$ around the base vertices $\underline{x} = (x_1, x_2, ..., x_r, x_{r+1}, ..., x_{k-2})$, y and z. (The vertices in $H_2 \setminus H_1$ are abstract – not from G.) The size of this extension is $(k - 2 - r + v|T|, e|T|)$. The bound on $|T|$ ensures that the extension is sparse. Using that $\mathcal{H}$ is valid one can see that the extension is safe by the very same argument as in the above lemma.

Axiom $B^v_{H_1, H_2}$ gives a map $f : H_2 \to G$ that is identity on H_1. We claim that $x^T_i = f(x_i)$ for $r < i \leq k - 2$ is a good choice for the statement of the lemma. The construction of H_2 gives that $\mathcal{H}$ is present in $(G, \underline{x}^T, y, z, f(w_{y,z}))$ if $(y, z) \in T$ and $w_{y,z}$ is the counting vertex of the copy of $\mathcal{H}$ added around the base vertices $\underline{x}$, y and z. From the assertions of the axiom about the small rigid extensions of $f(H_2)$ in G one can see that $\mathcal{H}$ is isolated in the above rgraph and $\mathcal{H}$ is not even present in $(G, \underline{x}^T, y, z, w)$ for any w if $(y, z) \in A \times B \setminus T$. □

Observe that when $k = 3$, we have no choice, r has to be 0. But for larger k, r represents a trade-off. If we set r to small, then we are able to characterize larger relations, but the price is that we have to allow for the selection of larges tuples.

The above lemma gives us a very important tool that we next use for checking set size equalities. For vertex sets A and B we define $|A| \leq_R |B|$ if there are $(k-2)$-tuples $\underline{x}^{(i)}$ for $i = 1, \ldots, 5$ such that the union of the relations $R^{\underline{x}^{(i)}}$ restricted to $(A \setminus B) \times (B \setminus A)$ is an injection. We write $|A| =_R |B|$ as a shorthand for $|A| \leq_R |B|$ and $|B| \leq_R |A|$. Clearly $|A| \leq_R |B|$ implies $|A| \leq |B|$, but the converse is not true in general. The following observation claims the converse is also true if one of the sets is small enough.

Lemma 4.3.3. *Let R be a relation characterizing the valid rooted graph $\mathcal{H}$ of size (v, e), and let $b = \left\lfloor \frac{k-2}{\alpha e - v} \right\rfloor$. For every two finite sets A and B of vertices of G such that one of them has size at most $5b$ the relations $|A| \leq |B|$ and $|A| \leq_R |B|$ are equivalent.*

Proof. As mentioned before the lemma $\leq_R$ always implies $\leq$. To prove the converse in this special case we fix an injection from $A \setminus B$ to $B \setminus A$ regarded as a relation, partition the pairs in this relation into five classes of size at most b and apply the result of Lemma 4.3.2 with $r = 0$ to get $\underline{x}^{(i)}$ for $i = 1 \ldots 5$ so that $R^{\underline{x}^{(i)}}$ restricted to $(A \setminus B) \times (B \setminus A)$ give the five parts of the injection. □

The above lemma is enough to compare any set to another set characterized as in Lemma 4.3.1:

Lemma 4.3.4. *Let R be a relation characterizing the valid rgraph $\mathcal{H}$. For any R-separated k-tuple $(\underline{x}, y, z)$ and any set of vertices S the statements $|S| = |R'(\underline{x}, y, z)|$ and $|S| =_R |R'(\underline{x}, y, z)|$ are equivalent.*

Proof. Let (v, e) be the size of $\mathcal{H}$, let $\epsilon = \alpha e - v$ and let $b = \left\lfloor \frac{k-2}{\epsilon} \right\rfloor$. Notice that the validity of $\mathcal{H}$ implies $\epsilon < 1$ and remember $k \geq 3$ so $b = \frac{k-2}{\epsilon} > 1$ and $\frac{k}{\epsilon} \leq 3\frac{k-2}{\epsilon}$. Using that for any real $x > 1$ we have $\lfloor 3x \rfloor \leq 5\lfloor x \rfloor$ we have $\lfloor k/\epsilon \rfloor \leq 5b$. Now apply lemmas 4.3.1 and 4.3.3 to prove the nontrivial direction of the claim. □

4.4 Characterizing hybrids

Given relations characterizing two (not necessarily different) rgraphs, we want to give a first order relation characterizing the possible hybrids built from these graphs. We will use the definition below:

Definition 4.4.1. Let P, Q be $(k+2)$-ary relations and let $\underline{x}^P$, y^P, z^P and $\underline{x}^Q$, y^Q, z^Q be two k-tuples of vertices of G, and finally let $0 \leq l < k$. We define the $(k+2)$-ary *hybrid relation* $R = HybR(l, P, Q, \underline{x}^P, y^P, z^P, \underline{x}^Q, y^Q, z^Q)$ as follows. We set $R(\underline{x}, y, z, w, t)$ if all the following conditions are met:

i) x_1, ..., x_{k-2}, y, z, w and t are distinct, except possibly $w = t$,

ii) $|P'(\underline{x}, w, z)| =_P |P'(\underline{x}^P, y^P, z^P)|$,

iii) $|Q'(\underline{x}, y, w)| =_Q |Q'(\underline{x}^Q, y^Q, z^Q)|$,

iv) w is connected to x_1, ..., x_{l-1}; w is also connected to x_l if $1 \leq l \leq k-2$ and finally to y if $l = k-1$,

v) the sets $P'(\underline{x}, w, z, u)$ for $u \in P'(\underline{x}, w, z)$ and the sets $Q'(\underline{x}, y, w, u)$ for $u \in Q'(\underline{x}, y, w)$ and $X \cup \{y, z, w\}$ are pairwise disjoint,

vi) t is included in $P'(\underline{x}, w, z, u)$ for some $u \in P'(\underline{x}, w, z)$ or in $Q'(\underline{x}, y, w, u)$ for some $u \in Q'(\underline{x}, y, w)$ or $t = w$.

The definition of the hybrid relation is clearly first-order, if all the P and Q are first order defined with the help of parameters, then so is R. As promised, we can use the above definition to characterize hybrids:

Lemma 4.4.2. *Suppose the relations P and Q characterizes the valid rooted graphs $\mathcal{H}^P$ and $\mathcal{H}^Q$ respectively. For any two k-tuples of vertices $(\underline{x}^P, y^P, z^P)$ and $(\underline{x}^Q, y^Q, z^Q)$ and for any $0 \leq l < k$ the hybrid relation $R = HybR(l, P, Q, \underline{x}^P, y^P, z^P, \underline{x}^Q, y^Q, z^Q)$ characterizes the hybrid rgraph $\mathcal{H} = Hyb(l, H^P, H^Q, |P'(\underline{x}^P, y^P, z^P)|, |Q'(\underline{x}^Q, y^Q, z^Q)|)$.*

Proof. For part a) of the definition of R characterizing $\mathcal{H}$ suppose $R(\underline{x}, y, z, w)$ for some vertices $\underline{x}$, y, z and w of G. By the definition of R the set $R'(\underline{x}, y, z, w)$ consists of w and the disjoint subsets $P'(\underline{x}, w, z, u)$ and $Q'(\underline{x}, y, w, u)$ for the appropriate vertices u. Since P characterizes $\mathcal{H}^P$ and Q characterizes $\mathcal{H}^Q$ each of these sets give rise to an almost disjoint copy of $\mathcal{H}^P$ or $\mathcal{H}^Q$. We also guarantee the existence of the additional edges in point iv) of the definition of R. Together these copies and edges give an isomorphic copy of $\mathcal{H}$ as a subgraph of $(G, \underline{x}, y, z, w)$. By points ii) and iii) of the definition of R and by the fact that $=_S$ implies $=$ for any $(k+2)$-ary relation S we have $|P'(\underline{x}, w, z)| = |P'(\underline{x}^P, y^P, z^P)|$ and $|Q'(\underline{x}, y, w)| = |Q'(\underline{x}^Q, y^Q, z^Q)|$. Thus the multiplicities are as claimed. The set of vertices of the copy of $\mathcal{H}$ is exactly the disjoint union of the set $X \cup \{y, z\}$, the set $\{w\}$, the sets $P'(\underline{x}, w, z, u)$ and the sets $Q'(\underline{x}, y, w, u)$. The union of the set $\{w\}$, the sets $P'(\underline{x}, w, z, u)$ and the sets $Q'(\underline{x}, y, w, u)$ is $R'(\underline{x}, y, z, w)$ thus part a) holds.

For part b) suppose $\mathcal{H}$ is isolated in $(G, \underline{x}, y, z, w)$, and let $(H', \underline{x}, y, z, w)$ be the subrgraph of $(G, \underline{x}, y, z, w)$ isomorphic to $\mathcal{H}$. Let the size of $\mathcal{H}$ be (v, e). Let $m^P = |P'(\underline{x}^P, y^P, z^P)|$ and $m^Q = |Q'(\underline{x}^Q, y^Q, z^Q)|$. $\mathcal{H}$ consists of m^P copies of $\mathcal{H}^P$ and m^Q copies of $\mathcal{H}^Q$ plus l extra edges connecting w with some base points. The latter implies that point vi) holds, the previous implies that $\mathcal{H}^P$ must be present in $(G, \underline{x}, w, z, u)$ for m^P vertices u and $\mathcal{H}^Q$ must be present in $(G, \underline{x}, y, w, u)$ for m^Q vertices u.

We claim that $\mathcal{H}^P$ and $\mathcal{H}^Q$ must be isolated in each of these cases, to make $\mathcal{H}$ isolated in $(G, \underline{x}, y, z, w)$. Indeed, let $(J, \underline{x}, w, z, u)$ be one of the above copies of $\mathcal{H}^P$ in $(G, \underline{x}, w, z, u)$, and let (v^P, e^P) be the size of $\mathcal{H}^P$. Suppose there is a rigid extension (J, J') with at most v^P non-base vertices such that there is at least one edge $e \in E(J') \setminus E(J)$ adjacent to a vertex in $V(J) \setminus (X \cup \{w, z\})$. By the hybrid construction any edge in H' adjacent to $V(J) \setminus (X \cup \{w, z\})$ is also in J, so $e \notin E(H')$. But by the definition of rigidity $(H', H' \cup J')$ is also rigid, it has at most v^P thus at most v non-base vertices, and e is in this extension adjacent to a non-base vertex, so this contradicts with $\mathcal{H}$ being isolated. The same argument works for the copies of $\mathcal{H}^Q$.

This implies $|P'(\underline{x}, w, z)| \geq m^P$ and $|Q'(\underline{x}, y, w)| \geq m^Q$. We claim that equality holds in the above formulae, i.e., no unintended copies of $\mathcal{H}^P$ or $\mathcal{H}^Q$ appear. For this we use that $\mathcal{H}^P$ and $\mathcal{H}^Q$ are valid. Since they are rigid and $\mathcal{H}$ is isolated all copies of $\mathcal{H}^P$ must appear inside H'. Since valid graphs remain connected after removing the base vertices all copies of $\mathcal{H}^P$ must be contained in a single copy of $\mathcal{H}^P$ or $\mathcal{H}^Q$. But as the last base vertex is z in case of the copies of $\mathcal{H}^P$ and w in case of the copies of $\mathcal{H}^Q$, all copies of $\mathcal{H}^P$ must appear inside some already counted copy of $\mathcal{H}^P$, and by being equal size it must be the entire copy. Finally, since each automorphism of the underlying graph of H^P fixing the base vertices also fixes the counting vertex the contribution of a single copy of H^P in $P'(\underline{x}, w, z)$ is only one vertex. By symmetry, this also holds for the copies of $\mathcal{H}^Q$.

We now have $|P'(\underline{x}, w, z)| = m^P = |P'(\underline{x}^P, y^P, z^P)|$. By Lemma 4.3.3 this implies $|P'(\underline{x}, w, z)| =_P |P'(\underline{x}^P, y^P, z^P)|$ since $(\underline{x}, w, z)$ is P-separated. Notice that for this latter statement one also needs the argument in the previous paragraph. The same way $|Q'(\underline{x}, y, w)| =_Q |Q'(\underline{x}^Q, y^Q, z^Q)|$. □

We also need to show that for all ,,interesting" hybrids we can represent the multiplicities we need to get them:

Lemma 4.4.3. *Let P and Q be two relations characterizing the valid rgraphs $\mathcal{H}^P$ and $\mathcal{H}^Q$, $0 \leq l < k$, $0 \leq m^P$,m^Q be integers. Suppose that the hybrid $\mathcal{H} = Hyb(l, \mathcal{H}^P, \mathcal{H}^Q, m^P, m^Q)$ is valid or safe. Then there exists a P-separated k-tuple $(\underline{x}^{P,m^P}, y^{P,m^P}, z^{P,m^P})$ such that $|P'(\underline{x}^{P,m^P}, y^{P,m^P}, z^{P,m^P})| = m^P$ and also there exists a Q-separated k-tuple $(\underline{x}^{Q,m^Q}, y^{Q,m^Q}, z^{Q,m^Q})$ such that $|Q'(\underline{x}^{Q,m^Q}, y^{Q,m^Q}, z^{Q,m^Q})| = m^Q$, thus the hybrid relation $HybR(l, P, Q, \underline{x}^{P,m^P}, y^{P,m^P}, z^{P,m^P}, \underline{x}^{Q,m^Q}, y^{Q,m^Q}, z^{Q,m^Q})$ characterizes the hybrid $\mathcal{H}$.*

Proof. Consider the subrgraph $\mathcal{H}^*$ of $\mathcal{H}$ consisting of the m^P copies of $\mathcal{H}^P$. This is not a valid rgraph as the base point z is isolated thus it is either a proper subrgraph or $\mathcal{H}$ must be safe. $\mathcal{H}^*$ is safe in either case. Its size is $(m^P v^P + 1, m^P e^P)$ if the size of $\mathcal{H}^P$ is (v^P, e^P). Thus $m^P v^P + 1 - \alpha m^P e^P > 0$ so $m^P < 1/(\alpha e^P - v^P)$. Thus Lemma 4.3.1 gives the existence of $(\underline{x}^{P,m^P}, y^{P,m^P}, z^{P,m^P})$. The existence of $(\underline{x}^{Q,m^Q}, y^{Q,m^Q}, z^{Q,m^Q})$ can be proved the same way. □

4.5 First order defining validity

In the previous section we were able to characterize hybrids of already characterized valid graphs. But there was no requirement or guarantee about the validness of the resulting graph. To detect whether the result of a hybrid construction is valid we will use the criteria given in Lemma 4.1.3. But first we need the following observation:

Lemma 4.5.1. *Suppose the relation R characterizes an rgraph $\mathcal{H} = (H, \underline{x}, y, z, w)$. Then R satisfies the first order statement "for all k-tuple of distinct vertices $\underline{x}'$, y', z' there exists w' such that $R(\underline{x}', y', z', w')$" if and only if $\mathcal{H}$ is safe.*

Proof. The if part follows from the safe extension axiom $B^m_{X \cup \{y,z\},H}$ (see Theorem 2.2.2) where m is the number of vertices in H. It says in effect that for every k distinct vertices $\underline{x}'$, y' and z' one can find w such that $\mathcal{H}$ is isolated in $(G, \underline{x}', y', z', w')$.

The only if part follows from the safe extension axiom $B^m_{H_1,H_2}$, where H_2 consists of three isolated vertices, H_1 is the empty graph, and m is the number of vertices in $\mathcal{H}$. The axiom then claims the existence of distinct vertices $\underline{x}'$, y' and z' having no rigid extension on at most m vertices. If $\mathcal{H}$ is not safe, then it has a nontrivial rigid sub extension thus $\mathcal{H}$ cannot be present in $(G, \underline{x}', y', z', w')$ and so $R(\underline{x}', y', z', w')$ cannot hold for any w'. □

Now we are ready to give a first order characterization of validness:

Definition 4.5.2. Let P and Q be two $(k+2)$-ary relations characterizing two valid rooted graphs and $0 \le l < k$ be an integer. We say the $2k$-tuple of vertices $\underline{x}^P$, y^P, z^P, $\underline{x}^Q$, y^Q, z^Q is *good* for P, Q and l if the following conditions are met:

1. Let $R = HybR(l, P, Q, \underline{x}^P, y^P, z^P, \underline{x}^Q, y^Q, z^Q)$. Then there exists k different vertices $\underline{x}$, y, z such that $R(\underline{x}, y, z, w)$ does not hold for any vertex w.
2. There exists a k-tuple $(\underline{x}', y', z')$ of vertices such that they satisfy both statements below:
   a) $|P(\underline{x}', y', z')| =_P |P(\underline{x}^P, y^P, z^P)| - 1$
   b) Let $S = HybR(l, P, Q, \underline{x}', y', z', \underline{x}^Q, y^Q, z^Q)$. Then for any k-tuple of distinct vertices $\underline{x}$, y, z there exists a vertex w such that $S(\underline{x}, y, z, w)$ holds
3. There exists a k-tuple $(\underline{x}', y', z')$ of vertices such that they satisfy both statements below:
   a) $|Q(\underline{x}', y', z')| =_Q |Q(\underline{x}^Q, y^Q, z^Q)| - 1$
   b) Let $S = HybR(l, P, Q, \underline{x}^P, y^P, z^P, \underline{x}', y', z')$. For any k-tuple of distinct vertices $\underline{x}$, y, z there exists a vertex w such that $S(\underline{x}, y, z, w)$ holds

By Lemma 4.5.1, by the fact that $=_P$ is equivalent to $=$ in this case and because being safe and being sparse is the same thing for hybrids of valid graphs, the above conditions are indeed equivalent to those given in Lemma 4.1.3, so we have the following:

Lemma 4.5.3. *Let P and Q be two $k+2$-ary relations characterizing two valid rooted graphs and $0 \le l < k$ be an integer. $\underline{x}^P$, y^P, z^P, $\underline{x}^Q$, y^Q, z^Q are good for P, Q and l if and only if $HybR(l, P, Q, \underline{x}^P, y^P, z^P, \underline{x}^Q, y^Q, z^Q)$ is characterizing a valid rooted graph.*

4.6 Simulating the weak approximation

It is time to link the constructions of this chapter with the results of Chapter 3. The main link is as promised that the sizes of graphs resulting from the hybrid construction and α-reachability are strongly related.

Lemma 4.6.1. *Let $\frac{p_1}{q_1}$, $\frac{p_2}{q_2}$, ..., $\frac{p_n}{q_n}$ and $\frac{p}{q}$ be rationals less then α, and assume as always in Chapter 3 that $gcd(p_i, q_i) = 1$ and $gcd(p, q) = 1$. We also assume $q > q_i$. Let $\mathcal{H}_1$, $\mathcal{H}_2$, ..., $\mathcal{H}_n$ be valid rooted graphs of sizes (p_1, q_1), (p_2, q_2), ..., (p_n, q_n) respectively. If $\frac{p}{q} \in H(\alpha, \{\frac{p_1}{q_1}, ..., \frac{p_n}{q_n}\})$ then there is a valid +0-hybrid of size (p, q) obtained from at most two $\mathcal{H}_i$'s.*

Proof. By definition, if $\frac{p}{q} \in H(\alpha, \{\frac{p_1}{q_1}, ..., \frac{p_n}{q_n}\})$ then either there are indexes i_1, i_2 and positive coefficients k_1 and k_2 such that $p = k_1 p_{i_1} + k_2 p_{i_2} + 1$, $q = k_1 q_{i_1} + k_2 q_{i_2}$ and we also have $(k_1 - 1)p_{i_1} + k_2 p_{i_2} + 1 > \alpha((k_1 - 1)q_{i_1} + k_2 q_{i_2})$ and $k_1 p_{i_1} + (k_2 - 1)p_{i_2} + 1 > \alpha(k_1 q_{i_1} + (k_2 - 1)q_{i_2})$ or there is and index i and a positive coefficient k such that $p = kp_i + 1$, $q = kq_i$ and $(k-1)p + 1 > \alpha(k-1)q$.

In the first case by Lemma 4.1.3 the +0-hybrid of $\mathcal{H}_{i_1}$ and $\mathcal{H}_{i_2}$ with multiplicities k_1 and k_2 is valid and has size (p, q). In the latter case by $q > q_i$ we have $k \geq 2$. Again by Lemma 4.1.3 the hybrid $Hyb(0, \mathcal{H}_i, \mathcal{H}_i, 1, k-1)$ is valid and has size (p, q).

□

Using this we can prove the following theorem which is the main results of this chapter.

Theorem 4.6.2. *For any $k \geq 3$ and d positive integers we have non-negative integers a, n, a-ary first order formulae φ_1, ..., φ_n and $(a + k + 2)$-ary first order formulae ψ_1, ..., ψ_n such that for any $\frac{1}{k} < \alpha < \frac{1}{k-1}$ and any α-graph G we have the following three properties:*

(1) For any a-tuple v_1, ..., $v_a \in V(G)$ and any $1 \leq i \leq n$ if $\varphi_i(v_1, ..., v_a)$ holds in G then the $(k+2)$-ary relation $\psi_i(v_1, ..., v_a, \_, ..., \_)$ characterizes a valid k-rooted graph in G.

(2) There is an a-tuple v_1, ..., $v_a \in V(G)$ and an integer $1 \leq i \leq n$ such that $\varphi_i(v_1, ..., v_a)$ holds and the k-rooted graph characterized by $\psi_i(v_1, ..., v_a, \_, ..., \_)$ is of size (v, e) where $gcd(v, e) = 1$ and $\frac{v}{e} = \tau_d(\alpha)$.

(3) If for some vertices v_1, ..., v_a and integer $1 \leq i \leq n$ we have $\varphi_i(v_1, ..., v_a)$ and the size of the rooted graph characterized by $\psi_i(v_1, ..., v_a, \_, ..., \_)$ is (v, e) then $\frac{v}{e} \leq \tau_d(\alpha)$.

Proof. We are going to prove by induction on d. Instead of (2) we are going to show the stronger:

(2') For every $\frac{p}{q}$ ($gcd(p, q) = 1$) non-zero element of the one down sequence of $\tau_d(\alpha)$ there is an a-tuple v_1, ..., $v_a \in V(G)$ and an $1 \leq i \leq n$ such that $\varphi_i(v_1, ..., v_a)$ holds and the rooted graph characterized by $\psi_i(v_1, ..., v_a, \_, ..., \_)$ is of size (p, q).

Let us assume first that we already know the theorem for some $d \geq 2$ and let n and a be the constants and φ_1,...,φ_n,ψ_1,...,ψ_n be the formulae it guarantees for that d. To prove the statement for $d' = d + 1$ let us set $a' = 2a + 2k$ and $n' = \binom{n}{2} + 2n$. We are going to have two kind of formulae:

a) For any $1 \leq i \leq n$ we have the following pair of formulae:

$$\varphi'(v_1, ..., v_{2a+2k}) = \varphi_i(v_1, ..., v_a)$$

$$\psi'(v_1, ..., v_{2a+2k}, \underline{x}, y, z, w, t) = \psi_i(v_1, ..., v_a, \underline{x}, y, z, w, t)$$

b) For any $1 \leq i \leq j \leq n$ we have the following pair of formulae:

$$\begin{aligned}\varphi'(v_1, ..., v_{2a+2k}) =& \varphi_i(v_1, ..., v_a) \wedge \varphi_j(v_{a+1}, ..., v_{2a}) \wedge \\ & \wedge \text{“the } v_{a+1}, ..., v_{a+2k} \text{ 2k-tuple is good for} \\ & \psi_i(v_1, ..., v_a, \_, ..., \_), \psi_j(v_{a+1}, ..., v_{2a}, \_, ..., \_) \text{ and l=0”}\end{aligned}$$

$$\begin{aligned}\psi'(v_1, ..., v_{2a+2k}, \underline{x}, y, z, w, t) =& \text{“}(\underline{x}, y, z, w, t) \in HybR(0, \psi_i(v_1, ..., v_a, \_, ..., \_), \\ & \psi_j(v_{a+1}, ..., v_{2a}, \_, ..., \_), v_{2a+1}, ..., v_{2a+2k})\text{”}\end{aligned}$$

As the $HybR$ relation and also being good are first order definable the parts between quotation marks can indeed be formulated as first order formulae. There are n formula pairs of type a), which are basically the copies of the formulae on the previous level, they simply ignore the extra parameters. There are $\binom{n}{2}+n$ formula pairs of type b) which are responsible for characterizing new hybrids. By induction point (1) holds for the type a) pairs, and also by induction and by lemmas 4.4.2 and 4.5.3 it also holds for the set b) of formula pairs. Let φ'_1, ..., $\varphi'_{n'}$ and ψ'_1, ..., $\psi'_{n'}$ be some ordering of the above formulae (of course the ordering of the φ's and ψ's must be the same).

By lemmas 4.4.3 and 4.5.3 the b) type formulae guarantee that we have the characterization of all possible valid hybrids of all the rgraphs that we could characterize in the previous step. Let A be the set of rationals $\frac{p}{q}$, $gcd(p,q)=1$ for which there exists $1 \leq i \leq n$ and an a-tuple of vertices v_1, ..., v_a such that $\varphi_i(v_1,...,v_a)$ holds and $\psi_i(v_1,...,v_a,\_,...,\_)$ characterizes a valid rooted graph of size (p,q). The same way let B be the set of rationals $\frac{p}{q}$, $gcd(p,q)=1$ for which there exists $1 \leq i \leq n'$ and an a'-tuple of vertices v_1, ..., $v_{a'}$ such that $\varphi'_i(v_1,...,v_{a'})$ holds and $\psi'_i(v_1,...,v_{a'},\_,...,\_)$ characterizes a valid rooted graph of size (p,q). By Lemma 4.6.1 we have $A \cup H(\alpha, A) \subset B$. By induction all the non-zero elements of the one down sequence of $\tau_d(\alpha)$ are present in A. As $d \geq 2$ this means that we have all the elements of the one down sequence between $\tau_{d-1}(\alpha)$ and $\tau_d(\alpha)$. (Notice that for $d=1$ it would not be true: we would miss $\tau_0(\alpha)=0$.) By Lemma 3.1.18 this means that $H(\alpha,A)$ has all the elements of the one down sequence of $\tau_{d+1}(\alpha)$ which are larger then $\tau_d(\alpha)$, thus B does contain all non-zero elements of the one down sequence of $\tau_{d'}(\alpha)$. This proves (2') for d'.

By induction (3) trivially holds for the formulae of type a). For type b) formulae consider any two valid rooted graphs $\mathcal{H}_1$ and $\mathcal{H}_2$ of sizes (p_1,q_1), (p_2,q_2) respectively which can be characterized using the formulae from the previous step. If $\mathcal{H} = Hyb(0,\mathcal{H}_1,\mathcal{H}_2,m_1,m_2)$ then the size of $\mathcal{H}$ is $(p,q)=(m_1p_1+m_2q_2+1, m_1q_1+m_2q_2)$. So $\frac{p-1}{q} = \frac{m_1p_1+m_2p_2}{m_1q_1+m_2q_2}$. This is at most $\frac{p_i}{q_i}$ for $i=1$ or $i=2$. But by induction this implies $\frac{p-1}{q} \leq \tau_d(\alpha)$ and if $\mathcal{H}$ is valid we also have $\frac{p}{q} < \alpha$. Thus by the definition of τ we have $\frac{p}{q} \leq \tau_{d+1}(\alpha)$. As all of the rgraphs characterized by type b) formulae are valid hybrids as considered above this proves (3) and completes the induction step.

To start up our induction we need some tricks to get around the disability to create a rooted graph of size $(0,1)$. First for $d=1$ let us have $n^{(1)}=1$ and $a^{(1)}=0$. φ_1^1 is constant true, and $\psi_1^1(\underline{x},y,z,w,t)$ holds if:

1. x_1, ..., x_{k-2}, y, z and w are distinct vertices,

2. $t=w$ and

3. w is connected to x_1, ..., x_{k-2}, y and z.

This is clearly first order and characterizes the valid rooted graph which has only one non-base vertex which is connected to all the base vertices. Thus (1) holds. We will call this rgraph $\mathcal{B}$. The size of this rgraph is $(1,k)$, and indeed for $\frac{1}{k} < \alpha < \frac{1}{k-1}$ we have $\tau_1(\alpha)=\frac{1}{k}$. Thus (3) holds and as $OD(\frac{1}{k})=0$ claim (2') also holds for this pair of formulae.

By Lemma 3.1.18, to fulfill (2') for $d=2$ it is enough to characterize all rationals $\frac{p}{q} \in H(\alpha, \{\frac{0}{1}, \frac{1}{k}\})$ for which $\frac{p}{q} > \frac{1}{k}$ as these are the elements that can potentially be elements of the one down sequence of $\tau_2(\alpha)$. The rationals in $H(\alpha, \{\frac{0}{1}, \frac{1}{k}\})$ are in the form $\frac{m_2+1}{m_1+km_2}$. First observe that if $m_1 \geq k$ then $\frac{m_2+1}{m_1+km_2} < \frac{1}{k}$, so we do not need to care about this case. Also observe that if $m_2 = 1$ and $m_1 \leq k-2$ then $\frac{m_2+1}{m_1+km_2} \geq \frac{2}{k-2+k} = \frac{1}{k-1}$ so it cannot be an element of $H(\alpha, \{\frac{0}{1}, \frac{1}{k}\})$. Finally if $m_2 = 0$ then we do not get any new rational (we get only $\frac{1}{k}$ that we already had in the first round). So we are going to have two distinct cases. One is when $m_1 = k-1$ and $m_2 = 1$. This case we will handle by characterizing a concrete rgraph as we did for $d = 1$. The second case is $0 \leq m_1 < k$ and $m_2 \geq 2$. This we will handle using the hybrid construction.

Now we give the formula pairs that prove the theorem for $d = 2$. We set $n^{(2)} = k+2$ and $a^{(2)} = 2k$ We will have the following three kind of formulae:

a) $\varphi_1^2(v_1, ..., v_{a^{(2)}})$ is constant true and $\psi_1^2(v_1, ..., v_{a^{(2)}}, \underline{x}, y, z, w, t) = \psi_1^1(\underline{x}, y, z, w, t)$

b) $\psi_2^2(v_1, ..., v_{a^{(2)}}, \underline{x}, y, z, w, t)$ holds if:

1. x_1, x_2, ..., x_{k-2}, y, z, w are distinct vertices,
2. w is connected to x_1, ..., x_{k-2} and to y,
3. there exists a unique vertex v distinct from x_1, x_2, ..., x_{k-2}, y, z, w which is connected to x_1, ..., x_{k-2}, z and w and
4. t is either w or the above mentioned v.

Let R be the $(k+2)$-ary relation $\psi_2^2(v_1, ..., v_{a^{(2)}}, \_, ..., \_)$. Define $\varphi_2^2(v_1, ..., v_{a^{(2)}})$ to hold if there are distinct vertices x_1, ..., x_{k-2}, y, z for which there is no w such that $R(\underline{x}, y, z, w)$

c) For any $0 \leq l \leq k-1$ we have the following pair of formulae:

$$\varphi_{l+3}^2(v_1, ..., v_{2k}) = \text{“the } v_1, ..., v_{2k} \text{ 2k-tuple is good for } \psi_1^1(\_, ..., \_), \psi_1^1(\_, ..., \_) \text{ and } l\text{”}$$

$$\psi_{l+3}^2(v_1, ..., v_{2k}, \underline{x}, y, z, w, t) = \text{“}(\underline{x}, y, z, w, t) \in HybR(l, \psi_1^1, \psi_1^1, v_1, ..., v_{2k})\text{”}$$

Point a) just copies φ_1^1 and ψ_1^1 as we did it in the proof of the induction step above. (1) and (3) trivially holds.

Point b) takes care of the case where $m_1 = k-1$ and $m_2 = 1$. As $\frac{m_2+1}{m_1+km_2} = \frac{2}{2k-1} \in H(\alpha, \{\frac{0}{1}, \frac{1}{k}\})$ if and only if $\frac{2}{2k-1} < \alpha$, we need to characterize this rational only in this case. ψ_2^2 characterizes the rooted graph with two non-base vertices v and w where v and w are connected, v is connected to all base vertices but y and w is connected to all base vertices but z. The size of this graph is $(2, 2k-1)$ so it indeed corresponds to the rational $\frac{2}{2k-1}$. It is easy to see that for $\frac{1}{k} < \alpha < \frac{1}{k-1}$ this graph is valid if and only if it is dense, otherwise it is safe. It is dense if and only if $\frac{2}{2k-1} < \alpha$, so this graph is valid exactly when we need the rational $\frac{2}{2k-1}$. The choice of φ_2^2 takes care of this using Lemma 4.5.1: it is true for any set of parameters if the above explained rgraph is valid and it is false for any set of parameters otherwise. By the above argument, (1) holds for this pair. As $\frac{2-1}{2k-1} < \frac{1}{k} = \tau_1(\alpha)$ and if φ_2^2 is ever true then $\frac{2}{2k-1} < \alpha$ so in this case we have $\frac{2}{2k-1} \leq \tau_2(\alpha)$, so (3) holds.

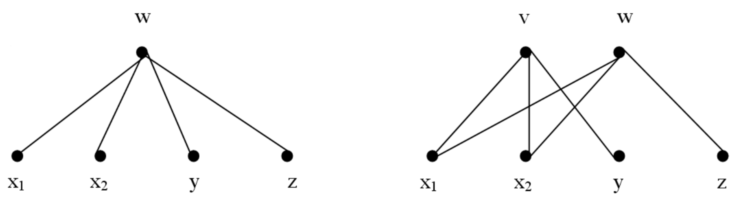

Figure 4.3: The rooted graph characterized by $\psi_1^1(\_, ..., \_)$ (on the left) and the rooted graph characterized by $\psi_2^2(v_1, ..., v_{a^{(2)}}, \_, ..., \_)$ for any vertices v_1, ..., $v_{a^{(2)}}$ (on the right) for $k = 4$.

Point c) handles the case where $0 \leq m_1 \leq k-1$ and $m_2 > 2$. These pairs satisfy (1). the same way point b) did in the induction step. With $l = m_1$ the hybrid rgraph $Hyb(l, \mathcal{B}, \mathcal{B}, 1, m_2 - 1)$ has size $(m_2 + 1, km_2 + m_1)$. Observe that $\frac{m_2+1}{km_2+m_1} \in H(\alpha, \{\frac{0}{1}, \frac{1}{k}\})$ if and only if $\frac{m_2+1}{km_2+m_1} < \alpha$ and $\frac{m_2-1+1}{k(m_2-1)+m_1} > \alpha$ as the latter inequality also implies $\frac{m_2+1}{km_2+m_1-1} > \alpha$ if $m_1 \geq 1$. This is exactly the condition when the above explained rgraph is valid. In this case by Lemma 4.4.3 and 4.5.3 there will be a $2k$ tuple v_1, ..., v_{2k} which is good for ψ_1^1, ψ_1^1 and l and for which $HybR(l, \psi_1^1, \psi_1^1, v_1, ..., v_{2k})$ characterizes $Hyb(l, \mathcal{B}, \mathcal{B}, 1, m_2 - 1)$. As $\frac{m_2}{km_2+m_1} \leq \frac{1}{k} = \tau_1(\alpha)$ claim (3) also holds.

Altogether these formulae satisfy (2') as all elements of $H(\alpha, \{\frac{0}{1}, \frac{1}{k}\})$ that are above $\frac{1}{k}$ can be captured by point b) or point c) and $\frac{1}{k}$ itself is captured by point a). This completes the proof of the theorem. □

We will need a trivial modification to the above theorem to make the application easier:

Theorem 4.6.3. *For any $k \geq 3$ and d positive integers we have an integer a', an a'-ary first order formula φ and an $(a' + k + 2)$-ary first order formulae ψ such that for any $\frac{1}{k} < \alpha < \frac{1}{k-1}$ and any α-graph G we have the following three properties:*

(1) For any a'-tuple v_1, ..., $v_{a'} \in V(G)$ if $\varphi(v_1, ..., v_{a'})$ holds then the $(k+2)$-ary are relation $\psi(v_1, ..., v_{a'}, \_, ..., \_)$ characterizes a valid k-rooted graph.

(2) There is an a'-tuple v_1, ..., $v_{a'} \in V(G)$ such that $\varphi_i(v_1, ..., v_{a'})$ holds and the k-rooted graph characterized by $\psi(v_1, ..., v_{a'}, \_, ..., \_)$ is of size (v, e) where $gcd(v, e) = 1$ and $\frac{v}{e} = \tau_d(\alpha)$.

(3) If for some vertices v_1, ..., $v_{a'}$ such that $\varphi(v_1, ..., v_{a'})$ holds the size of the rooted graph characterized by $\psi(v_1, ..., v_{a'}, \_, ..., \_)$ is (v, e) then $\frac{v}{e} \leq \tau_d(\alpha)$.

Proof. Let a, n be the constants and φ_1, ..., φ_n,ψ_1, ...,ψ_n be the formulae guaranteed by the previous theorem for d and k. Let $t = \lceil log_2(n) \rceil$ and $a' = a + t + 1$. Let us define the formulae $\beta_0^j = (y_j = y_b)$ and $\beta_1^j = (y_j \neq y_b)$. Finally let $g_i(j)$ for $1 \leq i \leq n$ and $1 \leq j \leq t$ be the jth digit of the number $i - 1$ written as a t long binary number. Then the following formulae will be good:

$$
\begin{aligned}
\varphi(y_b, y_1, ..., y_t, x_1, ..., x_a) = &(\beta^1_{g_1(1)} \wedge \beta^2_{g_1(2)} \wedge ... \wedge \beta^t_{g_1(t)} \wedge \varphi_1(x_1, ..., x_a)) \vee \\
&(\beta^1_{g_2(1)} \wedge \beta^2_{g_2(2)} \wedge ... \wedge \beta^t_{g_2(t)} \wedge \varphi_2(x_1, ..., x_a)) \vee \\
&... \\
&(\beta^1_{g_n(1)} \wedge \beta^2_{g_n(2)} \wedge ... \wedge \beta^t_{g_n(t)} \wedge \varphi_n(x_1, ..., x_a)) \\
\psi(y_b, y_1, ..., y_t, x_1, ..., x_{a+k+2}) = &(\beta^1_{g_1(1)} \wedge \beta^2_{g_1(2)} \wedge ... \wedge \beta^t_{g_1(t)} \wedge \psi_1(x_1, ..., x_{a+k+2})) \vee \\
&(\beta^1_{g_2(1)} \wedge \beta^2_{g_2(2)} \wedge ... \wedge \beta^t_{g_2(t)} \wedge \psi_2(x_1, ..., x_{a+k+2})) \vee \\
&... \\
&(\beta^1_{g_n(1)} \wedge \beta^2_{g_n(2)} \wedge ... \wedge \beta^t_{g_n(t)} \wedge \psi_n(x_1, ..., x_{a+k+2}))
\end{aligned}
$$

The first $t+1$ parameters are only used to select which original formula to use for the last a (or $a+k+2$) parameters. It is obvious that all formulae can be addressed by the right choice of the first $t+1$ parameters. By the properties of the original formulae it is easy to see that the new formulae indeed satisfies all the requirements. □

Chapter 5

Second order logic on small vertex sets

As in the previous chapter we fix α, $k \geq 3$ such that $\frac{1}{k} < \alpha < \frac{1}{k-1}$. We also fix an α-graph G. Additionally here we also fix $k-1$ distinct vertices of G: x_1, ..., x_{k-3}, $vtrue$, $vfalse$. To be able to refer to the set of all fixed vertices, let us introduce $F = \{x_1, ..., x_{k-3}, vtrue, vfalse\}$.

5.1 Representing multivariate functions

Suppose we are given disjoint finite vertex sets D, C_1, ..., C_n of G. We would like to represent all possible functions $f : C_1 \times ... \times C_n \to D$ with some kind of representative points using a relation characterizing a rooted graph. Of course we will not be able to do that for any set sizes. It will turn out that we have to choose the sizes of the C's very accurately. We will have $|C_i| \geq \left\lfloor \frac{1}{\alpha e - v} \right\rfloor$, where (v, e) is the size of the used rooted graph, which will be enough for our purposes.

For a one variable function, we already know the solution: Lemma 4.3.2 allows us to represent any binary relation of size at most $\left\lfloor \frac{1}{\alpha e - v} \right\rfloor$, so we can represent functions where the domain size is at most $\left\lfloor \frac{1}{\alpha e - v} \right\rfloor$. Observe that the size of the range does not matter at all, we only need it to be finite. The idea for representing multivariate functions is to think of an i-variate function $f : C_1 \times ... \times C_i \to D$ as a one variable function whose domain is C_i and whose range is the set of all possible $(i-1)$-variate functions $g : C_1 \times ... \times C_{i-1} \to D$. If we can represent all $(i-1)$-variate functions with vertices, then we just need to represent another function that maps C_i to the set of all possible representing points. Unfortunately we have a serious problem. The set of all possible representing points are not finite: if a function can be represented at all, then it can be represented with infinitely many points. So we cannot prove that the above idea works just by repeatedly applying Lemma 4.3.2. Nevertheless it does work, but we will need to work much more to prove it.

First we give a more precise formalization of the above notions.

Definition 5.1.1. Let R be a $(k+2)$-ary relation on the vertices of G. Let D, C_1, ..., C_n be finite vertex sets, disjoint from each other and from F. For $j \geq 0$ we will define a unary relation $R_D[C_1, ..., C_j]$. The elements of $R_D[C_1, ..., C_j]$ will be the vertices that *represent* some j-ary function $C_1 \times ... \times C_j \to D$. We also define a function $\hat{R}_D[C_1, ..., C_j]$ mapping elements of $R_D[C_1, ..., C_j]$ to sets of vertices. For $y \in R_D[C_1, ..., C_j]$ we will call the elements of $\hat{R}_D[C_1, ..., C_j](y)$ the vertices of the *defining structure* of y.

The definition is recursive in j. We start with $R_D[](y)$ if and only if $y \in D$ while $\hat{R}_D[](y) = \emptyset$. For $j > 0$ we define $R_D[C_1, ..., C_j](y)$ to hold if and only if both following conditions are met:

1. $\forall x \in C_j \exists !(z, w)(R_D[C_1, ..., C_{j-1}](z) \wedge R(x_1, ..., x_{k-3}, x, y, z, w) \wedge (\hat{R}_D[C_1, ..., C_{j-1}](z) \cap R'(x_1, ..., x_{k-3}, x, y, z, w) = \emptyset))$.

2. The sets $\{x_1, ..., x_{k-3}\}$, $\{y\}$, the sets C_i for $1 \leq i \leq j$, the set D and for the triplets (x, z, w) with $x \in C_j$, $R_D[C_1, ..., C_{j-1}](z)$ and $R(x_1, ..., x_{k-3}, x, y, z, w)$ the sets $\hat{R}_D[C_1, ..., C_{j-1}](z)$ and $R'(x_1, ..., x_{k-3}, x, y, z, w)$ are all pairwise disjoint.

If $R_D[C_1, ..., C_j](y)$ holds we set $\hat{R}_D[C_1, ..., C_j](y)$ to be the union of all the disjoint sets in item 2 above except for the sets C_i for $1 \leq i \leq j$, D and $\{x_1, ..., x_{k-3}\}$.

Finally in case $R_D[C_1, ..., C_j](y)$ holds we define the map $R^y_D[C_1, ..., C_j] : C_1 \times \cdots \times C_j \to D$ the following way. Let $R^y_D[]() = y$ and for $j > 0$ and $a_i \in C_i$ for $1 \leq i \leq j$ let $R^y_D[C_1, ..., C_j](a_1, \ldots, a_j) = R^z_D[C_1, ..., C_{j-1}](a_1, \ldots, a_{j-1})$ where (z, w) is the unique pair whose existence for $x = a_j$ is stated in item 1 of the definition above. We call this mapping the mapping *represented* by y, or we say y represents $R^y_D[C_1, ..., C_j]$.

Notice that all the above defined functions and relations do depend on the choice of the x_1, ..., x_{k-3} vertices. But we exclude them from the notations as we will keep them fixed all the time.

We will prove below that if the sizes of the C_i's are properly chosen, then for any $f : C_1 \times ... \times C_n \to D$ we can found a vertex y representing it. Also we will see that sets with these proper sizes can be first order defined in an α-graph.

5.2 Function representing extensions

In this section we are going to recursively define a sequence of graphs and study their properties. These graphs corresponds to representations of functions as defined in the previous section.

We fix a valid k-rooted graph $\mathcal{H} = (H, \underline{\tilde{x}}, \tilde{y}, \tilde{z}, \tilde{w})$ of size (v, e) and a positive integer n. If f is an n-variate function by $f^i_{a_{i+1},...,a_n}$ we will denote the i-variate function for which $f^i_{a_{i+1},...,a_n}(a_1, ..., a_i) = f(a_1, ..., a_n)$.

Before turning to the actual extensions interesting to us, we define a sequence of rooted graphs using the hybrid construction. Let $\mathcal{H}_0 = \mathcal{H}$. For $i > 0$ let $\mathcal{H}_i = Hyb(0, \mathcal{H}_{i-1}, \mathcal{H}, l_i, 1)$ where l_i is the smallest non-negative integer making this hybrid dense. Let (v_i, e_i) be the size of $\mathcal{H}_i$.

Lemma 5.2.1. *The $\mathcal{H}_i$ as defined above is valid and l_i is a non-decreasing sequence and $l_1 = \left\lfloor \frac{1}{\alpha e - v} \right\rfloor$.*

Proof. Let $d_i = \alpha e_i - v_i$. The hybrid $Hyb(0, \mathcal{H}_{i-1}, \mathcal{H}, l, 1)$ is dense if and only if $\alpha(le_{i-1} + e) - (lv_{i-1} + v + 1) = ld_{i-1} + d_0 - 1 > 0$, thus $l_i = \left\lceil \frac{1-d_0}{d_{i-1}} \right\rceil$. For $i = 1$ this means

$l_1 = \left\lceil \frac{1-d_0}{d_0} \right\rceil = \left\lfloor \frac{1}{d_0} \right\rfloor$ which proves the last statement. We will prove by induction that $\mathcal{H}_i$ is valid and that d_i is decreasing. The second claim proves that l_i is non-decreasing.

We know that $\mathcal{H}_0$ is valid, the other statement is empty for $i = 0$. Let's assume we know the statements for $j - 1$. We have $d_j = l_j d_{j-1} + d_0 - 1$ which is positive by $\mathcal{H}_j$ being dense. As l_j is the minimal l that makes the above hybrid dense, we have $(l_j - 1)d_{j-1} + d_0 - 1 = d_j - d_{j-1} < 0$, thus $d_j < d_{j-1}$ as claimed. To prove the validity of $\mathcal{H}_j$ by Lemma 4.1.3 we only have left to show that $\alpha(l_j e_{j-1} + e_0 - e_0) - (l_j v_{j-1} + v_0 + 1 - v_0) = d_j - d_0 < 0$. But we have just established the monotonicity of d_i up to j, so $d_j < d_0$ which completes the proof. □

We will use some standard graph constructions. If G_1 and G_2 are graphs then $G_1 \cup G_2$ is the graph whose vertex set is the union of the vertex sets of G_1 and G_2 and its edge set is the union of the edge sets of G_1 and G_2. For a graph M and two distinct vertices u and u' of M the graph that we get from M by *attaching* u to u' (denoted as $M(u \mapsto u')$) is the following graph:

$$V(M(u \mapsto u')) = V(M) \setminus \{u\}$$
$$E(M(u \mapsto u')) = E(M - \{u\}) \cup \{\{u', w\} \mid \{u, w\} \in E(M)\}$$

For any $T \subset V(M)$ and vertex $u \notin V(M)$ the graph that we get from M by *contracting* T as u (denoted as $M(T/u)$) is the following graph:

$$V(M(T/u)) = V(M) \cup \{u\} \setminus T$$
$$E(M(T/u)) = E(M - T) \cup \{\{u, w\} \mid \{u', w\} \in E(M) \text{ for some } u' \in T\}$$

We will potentially handle many isomorphic copies of the same graph, so we need a special notation. We will use $M[\mathcal{L}]$ where $\mathcal{L}$ is a finite list of some objects to refer to graphs. It will always be true that $M[\mathcal{L}] \cong M[\mathcal{L}']$ for any $\mathcal{L}$ and $\mathcal{L}'$ lists. M will be used as a shorthand to $M[]$. We will use the notation $a|\mathcal{L}$ to refer to the list of length $|\mathcal{L}| + 1$ whose first element is a and the rest are the elements of $\mathcal{L}$ in the original order.

Let us first define $H[\mathcal{L}]$ to be a copy of H for any $\mathcal{L}$. For any $\mathcal{L} \neq \mathcal{L}'$ we choose $H[\mathcal{L}]$ and $H[\mathcal{L}']$ to be disjoint. To comply with our convention we choose $H[]$ to be H itself. The copies of the base vertices of $\mathcal{H}$ in $H[\mathcal{L}]$ are denoted as $\tilde{x}_1[\mathcal{L}]$, ..., $\tilde{x}_{k-2}[\mathcal{L}]$, $\tilde{y}[\mathcal{L}]$, $\tilde{z}[\mathcal{L}]$, $\tilde{w}[\mathcal{L}]$.

Let us now fix finite disjoint sets B, A_1, ..., A_n, X' of sizes b, l_1, ..., l_n, $k-3$ respectively such that these sets are disjoint from all the above defined copies of H. The l_i's are as defined above, b is an arbitrary positive integer. We denote the elements of X' with x'_1, ..., x'_{k-3}. Let E be the empty graph on the set $B \cup A_1 \cup ... \cup A_n \cup X'$. Also for all $\mathcal{L}$ let us have a vertex $d[\mathcal{L}]$ disjoint from all the copies of H, all the sets defined above and all other $d[\mathcal{L}']$.

For any $0 \leq i \leq n$ and any i-variate function f from $A_1 \times ... \times A_i$ to B and any list $\mathcal{L}$ we will define the graph $FullExt^i_f[\mathcal{L}]$ and a designated vertex $s^i_f[\mathcal{L}] \in V(FullExt^i_f[\mathcal{L}])$. For any $1 \leq i \leq n$ and any i-variate function f from $A_1 \times ... \times A_i$ to B, an element $a \in A_i$ and a list $\mathcal{L}$ we will define the graph $OneExt^i_{f,a}[\mathcal{L}]$ and a designated vertex $t^i_{f,a}[\mathcal{L}] \in V(OneExt^i_{f,a})[\mathcal{L}]$.

For an $f \in B$ constant regarded as a nullary function let $FullExt^0_f[\mathcal{L}] = E$ and $s^0_f[\mathcal{L}] = f$ for any $\mathcal{L}$. Observe below that during the construction E will always be a subgraph of the defined graphs.

For $i > 0$, $a \in A_i$, the function $f : A_1 \times ... \times A_i \rightarrow B$ and the list $\mathcal{L}$ to define $OneExt^i_{f,a}[\mathcal{L}]$ take the union of $FullExt^{i-1}_{f^{i-1}_a}[\mathcal{L}]$ and $H[\mathcal{L}]$. Then attach the following vertices of $H[\mathcal{L}]$ to vertices of $FullExt^{i-1}_{f^{i-1}_a}[\mathcal{L}]$. Attach $\tilde{x}_i[\mathcal{L}]$ to x'_i for $1 \leq i \leq k-3$, $\tilde{x}_{k-2}[\mathcal{L}]$ to a and $\tilde{z}[\mathcal{L}]$ to $s^{i-1}_{f^{i-1}_a}[\mathcal{L}]$. We set $t^i_{f,a}[\mathcal{L}]$ to the vertex $\tilde{y}[\mathcal{L}]$ of $H[\mathcal{L}]$.

For $i > 0$ and $f : A_1 \times ... \times A_i \rightarrow B$ and a list $\mathcal{L}$ to define $FullExt^i_f[\mathcal{L}]$ take the union of the graphs $OneExt^i_{f,a}[a|\mathcal{L}]$ for each $a \in A_i$. Then contract the set $\{t^i_{f,a}[a|\mathcal{L}] \mid a \in A_i\}$ as $d^{\mathcal{L}}$, and set $s^i_f[\mathcal{L}]$ to $d^{\mathcal{L}}$.

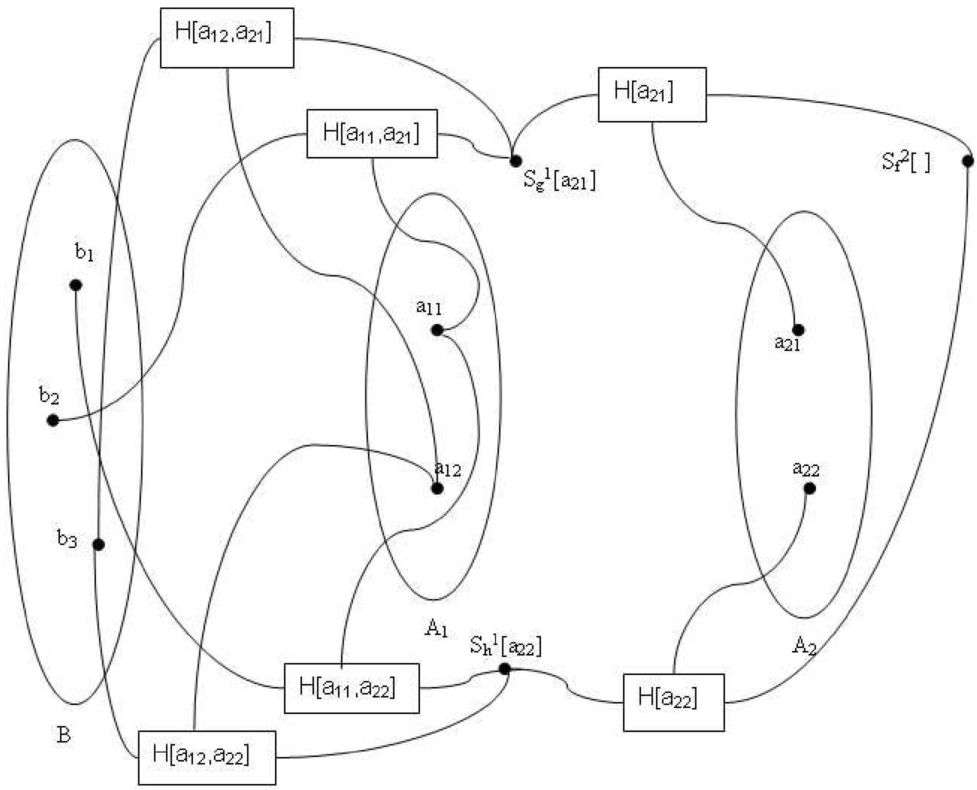

Figure 5.1: The graph $FullExt^2_f$. The function $g : A_1 \rightarrow B$ is defined as $g(a_{11}) = b_2$, $g(a_{12}) = b_3$. The functions $h : A_1 \rightarrow B$ is defined as $h(a_{11}) = b_1$, $h(a_{12}) = b_3$. The function $f : A_1 \times A_2 \rightarrow B$ is defined as $f(a_{21}, x) = g(x)$ and $f(a_{22}, x) = h(x)$. The points x_1, ..., x_{k-3} and their respective edges are omitted from this figure.

Notice the point of this whole construction: in the graph $FullExt^i_f[\mathcal{L}]$ the vertex $s^i_f[\mathcal{L}]$ represents the function f as defined in the previous section if R is a relation characterizing $\mathcal{H}$.

Using the above rgraph sequence we can now easily prove what we need to know about the extensions corresponding to function representations.

Lemma 5.2.2. *Let us fix a function $f : A_1 \times ... \times A_n \rightarrow B$ and elements $a_1 \in A_1$, ..., $a_n \in A_n$. Suppose the size of A_j is l_j for $1 \leq j \leq n$. Let us further denote $f_j = f^j_{a_{j+1},...,a_n}$ for $1 \leq j \leq n$. Then for $1 \leq i \leq n$ the extension $(E, FullExt^i_{f_i}[\mathcal{L}])$ is safe and the extension $(E \cup \{t^i_{f_i,a_i}\}, OneExt^i_{f_i,a_i}[\mathcal{L}])$ is rigid for any $\mathcal{L}$.*

Proof. Let $\underline{OneExt}^i_{f_i,a_i}[\mathcal{L}]$ be the graph that we get from $OneExt^i_{f_i,a_i}[\mathcal{L}]$ by contracting $A_1 \cup ... \cup A_n$ to a single new vertex a and contracting B to a single new vertex b. The crucial observation is that the rooted graph $(\underline{OneExt}^i_{f_i,a_i}[\mathcal{L}], x'_1, ..., x'_{k-3}, a, t^i_{f_i,a_i}[\mathcal{L}], b, s^{i-1}_{f_{i-1}}[\mathcal{L}])$ is isomorphic to $\mathcal{H}_{i-1}$. Indeed it is obvious for $i = 1$ and can be easily shown by induction for larger i using the construction of hybrids and the inductive construction of $OneExt$. Retracting unconnected base vertices does not change the size of an extension, thus it does not change the extension being dense/sparse/rigid/safe. Thus we have $(E \cup \{t^i_{f_i,a_i}[\mathcal{L}]\}, OneExt^i_{f_i,a_i}[\mathcal{L}])$ is rigid as we wanted. The same way let $\underline{FullExt}^i_{f_i}[\mathcal{L}]$ be the graph that we get from $FullExt^i_{f_i}[\mathcal{L}]$ by contracting $A_1 \cup ... \cup A_n$ to a single vertex a and contracting B to a single vertex b. From the above it is obvious that the rooted graph $(\underline{FullExt}^i_{f_i}[\mathcal{L}], x'_1, ..., x'_{k-3}, a, y', b, s^i_{f_i}[\mathcal{L}])$ where y' is a new isolated vertex is isomorphic to the hybrid $Hyb(0, \mathcal{H}_{i-1}, \mathcal{H}, l_i, 0)$, thus safe as a proper subgraph of the valid rgraph $\mathcal{H}_i$. This proves the statement about $FullExt^i_{f_i}[\mathcal{L}]$ that completes the proof. □

5.3 Existence of representations

Let R be a relation characterizing the valid rgraph $\mathcal{H}$ in the α-graph G. We fix n as above and we will use the integers b and l_i as defined in the above section. Notice that while b was chosen arbitrarily the value of l_i was determined by the choice of $\mathcal{H}$ and by α. We will also refer to all the sets and graphs defined in the previous section.

Lemma 5.3.1. *Let us fix finite sets of vertices of G: C_1, C_2, ..., C_n of sizes l_1, ..., l_n and D of size b such that these sets are pairwise disjoint from each other and from F. Let us fix bijections $\gamma_i : A_i \to C_i$ and $\delta : B \to D$. For a vertex y and for $0 \le j \le n$ assume $R_D[C_1, ..., C_j](y)$ holds. Let $f : A_1 \times ... \times A_j \to B$ be the function defined by:*

$$f(a_1, ..., a_j) = \delta^{-1}(R^y_D[C_1, ..., C_j](\gamma_1(a_1), ..., \gamma_j(a_j)))$$

Then G has a subgraph G' for which there is an isomorphism $\varphi : FullExt^j_f \to G'$ such that $\varphi(s^j_f) = y$, $\varphi(b) = \delta(b)$ for any $b \in B$, $\varphi(x_i) = x'_i$ for $1 \le i \le k-3$ and $\varphi(a) = \gamma_i(a)$ for any $1 \le i \le n$, and any $a \in A_i$. The vertex set of G' is:

$$D \cup C_1 \cup ... \cup C_n \cup \hat{R}_D[C_1, ..., C_j](y) \cup \{x_1, ..., x_{k-3}\}$$

Proof. We prove by induction on j. For $j = 0$ setting φ to be the union of the function δ and the functions γ_i will be good.

For $j > 0$ by definition of $R_D[C_1, ..., C_j]$ we know that there are vertices z_c, w_c for each vertex $c \in C_j$ such that $z_c \in R_D[C_1, ..., C_{j-1}]$ and there is a copy of $\mathcal{H}$ present in $(G, x_1, ..., x_{k-3}, c, y, z_c, w_c)$. We know by induction that there are isomorphisms:

$$\varphi_c : FullExt^{j-1}_{f^{j-1}_{\gamma_j^{-1}(c)}}[c] \to D \cup C_1 \cup ... \cup C_n \cup \hat{R}_D[C_1, ..., C_{j-1}](z_c) \cup \{x_1, ..., x_{k-3}\}$$

There are also isomorphisms φ'_c from $H[c]$ to the respective copy of $\mathcal{H}$ present in $(G, x_1, ..., x_{k-3}, c, y, z_c, w_c)$. By the extra conditions in the lemma on the isomorphism and by the 2. point of the definition of $R_D[C_1, ..., C_j]$ we know that all the functions φ_c

and φ'_c are compatible, that is if any two of them are defined on a common point then they give the same value. Thus the union of these isomorphisms give a homomorphism:

$$\varphi^* : \bigcup_{c\in C_j} FullExt^{j-1}_{f^{j-1}_{\gamma_j^{-1}(c)}}[c] \cup \bigcup_{c\in C_j} H[c] \to D \cup C_1 \cup ... \cup C_n \cup \hat{R}_D[C_1, ..., C_j](y) \cup \{x_1, ..., x_{k-3}\}$$

Furthermore we know that $\varphi'_c(\tilde{y}[c]) = y$ for all $c \in C_j$ and that $\varphi_c(x'_i) = \varphi'_c(\tilde{x}_i[c]) = x_i$ for $1 \leq i \leq k-3$, also $\varphi_c(\gamma_j^{-1}(c)) = \varphi'_c(\tilde{x}_{k-2}[c]) = c$ and $\varphi_c(s^{j-1}_{f^{j-1}_{\gamma_j^{-1}(c)}}) = \varphi'_c(\tilde{z}_i[c]) = z_c$. So φ^* induces a homomorphism:

$$\varphi : FullExt^j_f \to D \cup C_1 \cup ... \cup C_n \cup \hat{R}_D[C_1, ..., C_j](y) \cup \{x_1, ..., x_{k-3}\}$$

as $FullExt^j_f$ can be created from $\bigcup_{c\in C_j} FullExt^{j-1}_{f^{j-1}_c}[c] \cup \bigcup_{c\in C_j} H[c]$ by identifying vertices and any two vertices identified during the construction had the same image according to φ^*. But one can see that φ is a bijection, so it indeed is an isomorphism as wanted. □

Lemma 5.3.2. *Let G' be a subgraph of G such that there is an isomorphism φ from $FullExt^n_f$ to G' for some $f : A_1 \times ... \times A_n \to B$. Assume that for any rigid extension G'' of G' in G of at most $|V(FullExt^n_f)|$ extra vertices there is no edge $e \in E(G'') - E(G')$ which is adjacent to a vertex in $V(G') \setminus \varphi(B \cup A_1 \cup ... \cup A_n \cup X')$. Then for any $0 \leq j \leq n$ and $a_{j+1} \in A_{j+1}$, ..., $a_n \in A_n$ for $y = \varphi(s^j_{f^j_{a_{j+1},...,a_n}}[a_{j+1}, ..., a_n])$ we have $y \in R_{\varphi(B)}[\varphi(A_1), ..., \varphi(A_j)]$. Furthermore for any $a_1 \in A_1$, ..., $a_j \in A_j$ we have:*

$$\varphi^{-1}(R^y_{\varphi(B)}[\varphi(A_1), ..., \varphi(A_j)](\varphi(a_1), ..., \varphi(a_j)) = f^j_{a_{j+1},...,a_n}(a_1, ..., a_j)$$

Finally we have:

$$\hat{R}_{\varphi(B)}[\varphi(A_1),..., \varphi(A_j)](y) = \varphi(V(FullExt^j_{f^j_{a_{j+1},...,a_n}}[a_{j+1}, .., a_n])) \setminus \varphi(B \cup A_1 \cup ... \cup A_n \cup X')$$

Proof. We prove by induction on j. For $j = 0$ the lemma is trivial. Let $0 < j \leq n$. For any $a \in A_j$ let $z_a = \varphi(s^{j-1}_{f^{j-1}_{a,a_{j+1},...,a_n}}[a, a_{j+1}, ..., a_n])$ and $w_a = \varphi(\tilde{w}[a, a_{j+1}, ..., a_n])$. Then we have the following facts.

(1) By induction $z_a \in R_{\varphi(B)}[\varphi(A_1), ..., \varphi(A_{j-1})]$.

(2) By the facts that:

   a) the non-base vertices of $H[a, a_{j+1}, ..., a_n]$ are connected only to each other and to the respective base vertices in $FullExt^n_f$ and

   b) G', the image of $FullExt^n_f$, does not have small rigid extensions except for those of $\varphi(B \cup A_1 \cup ... \cup A_n \cup X')$

   we know that $\mathcal{H}$ is isolated in $G(x_1, ..., x_{k-3}, \varphi(a), y, z_a, w_a)$.

(3) We also know that the image of the set of the non-base vertices of $H[a, a_{j+1}, ..., a_n]$ are disjoint from the set:

$$\varphi(V(FullExt^{j}_{f^{j}_{a_{j+1},...,a_n}}[a_{j+1}, .., a_n])) \setminus \varphi(B \cup A_1 \cup ... \cup A_n \cup X')$$

simply by φ being an injection. The first set is exactly $R'(x_1, ..., x_{k-3}, \varphi(a), y, z_a, w_a)$ and by induction the second is $\hat{R}_{\varphi(B)}[\varphi(A_1), ..., \varphi(A_{j-1})](z_a)$.

Thus the first point of the definition of y being an element of $R_{\varphi(B)}[\varphi(A_1), ..., \varphi(A_j)]$ holds, except maybe for the uniqueness. If the uniqueness would also hold then by the inductive hypothesis on $\hat{R}$ and by the respective parts in $FullExt^{n}_{f}$ being disjoint we knew that the second point also holds. It is also easy to check using the respective inductive hypothesis on the functions represented by z_a, that y represents the function as claimed. So the only thing to check is that there are no other pair z'_a and w'_a for which:

$$R_{\varphi(B)}[\varphi(A_1), ..., \varphi(A_{j-1})](z'_a) \wedge R(x_1, ..., x_{k-3}, \varphi(a), y, z'_a, w'_a) \wedge$$
$$\wedge(\hat{R}_D[C_1, ..., C_{j-1}](z'_a) \cap R'(x_1, ..., x_{k-3}, \varphi(a), y, z'_a, w'_a) = \emptyset)$$

Assume the contrary and put together the previous lemma, the above condition on z'_a and w'_a and the fact that R characterizes $\mathcal{H}$ implies the existence of an isomorphism ψ from $OneExt^{j}_{g,a}$ for some $(j-1)$-ary function g to a subgraph G'' of G such that $\psi(E \cup \{t^{j}_{g,a}\}) \subseteq V(G')$. As by Lemma 5.2.2 $(\psi(E \cup \{t^{j}_{g,a}\}), G'')$ is a rigid extension we have that (G', G'') is also rigid. The number of non-base vertices of this extension is less then $|V(FullExt^{n}_{f})|$. By the construction of $FullExt^{n}_{f}$ one can also see that G'' must have an edge not present in G' with an endpoint in $V(G') - \varphi(B \cup A_1 \cup ... \cup A_n \cup X')$. But this contradict with the conditions of the lemma, which completes the proof. □

Theorem 5.3.3. *For any finite vertex sets C_1, ..., C_n, D of sizes l_1, ..., l_n, b respectively which are disjoint from each other and from F and for any $f : C_1 \times ... \times C_n \to D$ there exists a vertex $y \in R_D[C_1, ..., C_n]$ for which $R^{y}_{D}[C_1, ..., C_n] \equiv f$.*

Proof. The theorem follows from the above lemma and from the safe extension axiom $B^{|V(FullExt^{n}_{f})|}_{E,FullExt^{n}_{f}}$ applied to an isomorphism φ from $E \leq FullExt^{n}_{f}$ to the empty graph on $X \cup C_1 \cup ... \cup C_n \cup D$ for which $\varphi(x'_i) = x_i$, $\varphi(A_i) = C_i$ and $\varphi(B) = D$. Notice that by Lemma 5.2.2 the extension $(E, FullExt^{n}_{f})$ is indeed safe. □

5.4 Dressing up sets

Lemma 5.4.1. *If G is an α-graph then for any finite set T of its vertices $G - T$ is also an α-graph.*

Proof. We need to prove that both axiom schemes given in Theorem 2.2.2 holds in $G-T$. It is trivial that A_H holds for any dense H as there is no subgraph isomorphic to H in G, so obviously there is no such subgraph in $G - T$ either.

Let (H_0, H_1) be a finite safe extension and $k > 0$ an integer. We can assume H_1 (and thus H_0) being disjoint from G. For any graph M (not necessarily a subgraph of G) disjoint from T we denote with M^+ the graph that we get from M by adding the vertices in T as isolated vertices (that is $V(M^+) = V(M) \cup T$ and $E(M^+) = E(M)$). Observe that (H_0^+, H_1^+) is a safe extension. If there is an isomorphism φ from H_0 to a H_0' subgraph of $G - T$ then it can be extended to an isomorphism φ^+ from H_0^+ to $H_0'^+$ where φ^+ is the identity on T. Applying $B^k_{H_0^+, H_1^+}$ to this isomorphism we can extend it to an isomorphism ψ^+ from H_1^+ to a subgraph L of G. Observe that $\psi = \psi^+|_{V(H_1)}$ is an isomorphism from H_1 to $L - T$ which is a subgraph of $G - T$. By L being a k-generic extension of $H_0'^+$ in G it is obvious that $L - T$ is a k-generic extension of $H_0' = H_0'^+ - T$ in $G - T$, so $B^k_{H_0, H_1}$ holds in $G - T$ indeed. □

We will use the above lemma in the following way. When we know by one of our tools developed previously or directly by the safe extension axiom that some kind of structure exists in G (e.g. extension, subgraph) then we can always assume a copy disjoint to any given finite set.

In the previous section we showed how we can represent functions from Descartes products of small sets. In this section we want to use this tool to represent relations on larger sets. To capture at most d-ary relations we will first create a correspondence between the d-tuples of the big set and the elements of the Descartes product of cd small sets. Here c is used to compensate for the larger size of our big set. In a cd-tuple every c long block encodes one element in the d-tuple it corresponds to.

Definition 5.4.2. Let R be a relation characterizing a valid rooted graph $\mathcal{H}$. The $((k+2)(cd+1)+d)$-tuple $Dr = (\underline{x}_0, y_0, z_0, \underline{x}_1, y_1, z_1, ..., \underline{x}_{cd}, y_{cd}, z_{cd}, s_1, ..., s_d)$ is an (R, c, d)-*dress* of the vertex set S if the following properties hold. Let $R_0 = R$ and $R_i = HybR(0, R_{i-1}, R, \underline{x}_i, y_i, z_i, \underline{x}_0, y_0, z_0)$. Let $C_i = R_i'(\underline{x}_i, y_i, z_i)$ for $1 \leq i \leq cd$. We require that:

1. $|R'(\underline{x}_0, y_0, z_0)| = 1$.
2. The $2k$-tuple $\underline{x}_i$, y_i, z_i, $\underline{x}_0$, y_0, z_0 is good for R_{i-1}, R and $l = 0$, thus R_i characterizes a valid hybrid rooted graph.
3. The sets C_i are pairwise disjoint and also disjoint from F and S.
4. $s_i \in R_S[C_1, ..., C_{ic}]$
5. Let $f_i = R_S^{s_i}[C_1, ..., C_{ic}] : C_1 \times ... \times C_{ic} \to S$. There is an f_i' surjection from $C_{(i-1)c+1} \times ... \times C_{ic}$ to S such that for any $a_1 \in C_1$, ..., $a_{ic} \in C_{ic}$ we have $f_i(a_1, ..., a_{ic}) = f_i'(a_{(i-1)c+1}, a_{(i-1)c+2}, ..., a_{ic})$.

Observe that if R is first order defined then being a dress is also first order defined. This definition achieves the goals outlined before it by using vertices $\underline{x}_i$, y_i, z_i to define the small sets C_i and using s_i to define the correspondence between c-long blocks of tuples in $C_1 \times ... \times C_{cd}$ to elements of S.

Lemma 5.4.3. *Suppose R is a relation characterizing a valid rooted graph $\mathcal{H}$ of size (v, e). For any $c > 0$, $d > 0$ integers there exists an (R, c, d)-dress of the finite vertex set S if and only if there exists an $(R, c, 1)$-dress of S. If $|S| \leq \left\lfloor \frac{1}{\alpha e - v} \right\rfloor^c$ then for any $d > 0$ there exists an (R, c, d)-dress of S.*

Proof. Let us use the definitions of the previous sections for $n = cd$. Observe that the relation R_i in the definition of dress above captures the rooted graph $\mathcal{H}_i$ as defined in Section 5.2. This also implies $C_i = l_i$. By repeatedly applying Lemma 4.4.3 and Lemma 5.4.1 we can see that we always have $\underline{x}_0$, y_0, z_0, $\underline{x}_1$, y_1, z_1, ..., $\underline{x}_{cd}$, y_{cd}, z_{cd} which satisfies items 1., 2. and 3. As we know that $C_i = l_i$, by Theorem 5.3.3, we know that for any function $f : C_1 \times ... \times C_j \to S$ for any $1 \leq j \leq cd$ there is a vertex $z \in R_S[C_1, ..., C_j]$ such that $R^z_S[C_1, ..., C_j] \equiv f$. So s's satisfying 4. and 5. can be found if and only if we can found $f'_i : C_{(i-1)c+1} \times ... \times C_{ic} \to S$ surjections. This happens if and only if $|C_{(i-1)c+1}| \times ... \times |C_{ic}| = l_{(i-1)c+1} l_{(i-1)c+2} ... l_{ic} \geq |S|$. By the l_i being a non-decreasing sequence $l_{(i-1)c+1} l_{(i-1)c+2} ... l_{ic} \geq |S|$ for all $1 \leq i \leq d$ if and only if $l_1 l_2 ... l_c \geq |S|$. This proves the first statement. If $|S| \leq \left\lfloor \frac{1}{\alpha e - v} \right\rfloor^c$ then $l_1 l_2 ... l_c \geq l_1^c = \left\lfloor \frac{1}{\alpha e - v} \right\rfloor^c \geq |S|$ so the second statement is also true. □

Now we show how to use dresses to represent relations.

Definition 5.4.4. Let Dr be an (R, c, d)-dress of the finite set S. With the notations of Definition 5.4.2 we say the function $g_K : C_1 \times ... \times C_{rc} \to \{vtrue, vfalse\}$ is the *representing function* of the r-ary relation $K \subset S^r$ $(1 \leq r \leq d)$ if for any $a_1 \in C_1$, ..., $a_{rc} \in C_{rc}$ we have $g(a_1, ..., a_{rc}) = vtrue$ if and only if $(f_1(a_1, ..., a_c), f_2(a_{c+1}, ..., a_{2c}), ..., f_r(a_{(r-1)c+1}, ..., a_{rc})) \in K$.

It is obvious that for any relation uniquely exists a representing function.

Definition 5.4.5. Let Dr be an (R, c, d)-dress of the finite set S. With the notations of Definition 5.4.2 we say a vertex q *represents* $K \subset S^r$ if $q \in R_{\{vtrue,vfalse\}}[C_1, ..., C_{rc}]$ and $R^q_{\{vtrue,vfalse\}}[C_1, ..., C_{rc}] \equiv g_K$ where g_K is the representing function of K. We will denote the set of vertices representing any r-ary relation as defined above as $RelV^r_{S,Dr}$. The r-ary relation represented by a vertex $q \in RelV^r_{S,Dr}$ will be denoted as $Rel^r_{S,Dr}[q]$.

Notice that both $RelV^r_{S,Dr}$ and $Rel^r_{S,Dr}[q]$ are first order definable.

Lemma 5.4.6. *If Dr is an (R, c, d)-dress of the finite set $|S|$ then for any r-ary relation K for $1 \leq r \leq d$ there exists a vertex $q \in RelV^r_{S,Dr}$ such that $K = Rel^r_{S,Dr}[q]$.*

Proof. The lemma is obvious from the existence of the representing function and from the fact that for any $1 \leq j \leq cd$ and for any function $f : C_1 \times ... \times C_j \to \{vtrue, vfalse\}$ there is a vertex $z \in R_{\{vtrue,vfalse\}}[C_1, ..., C_j]$ such that $R^z_{\{vtrue,vfalse\}}[C_1, ..., C_j] \equiv f$. □

5.5 Converting second order formulae

Using the results above we can state the main result of this chapter which essentially says that we can first order simulate second order formulae on small vertex sets of G.

Theorem 5.5.1. *Let P be a fixed set of variables and suppose we are given first order P-formulae (of signature $(E/2)$) Δ, η, ι_1, ..., ι_m of arities 1, $k+2$, r_1, ..., r_m. We are also given a closed second order formula ψ of signature $(E/2, J_1/r_1, ..., J_m/r_m)$. Finally we are given a constant c. Then there exists a first order closed P-formula φ with the* *following properties. Assume $\sigma : P \to V(G)$ is any variable assignment. Let $S = \{v \in V(G) \mid G[\sigma] \models \Delta(v)\}$. Let us define the structure $\mathfrak{A} = (S, E^{\mathfrak{A}}, J_1^{\mathfrak{A}}, ..., J_m^{\mathfrak{A}})$ where:*

$$E^{\mathfrak{A}} = \{(v, w) \in S^2 \mid v \text{ and } w \text{ are connected in } G\}$$
$$J_i^{\mathfrak{A}} = \{(v_1, ..., v_{r_i}) \in S^{r_i} \mid G[\sigma] \models \iota_i(v_1, ..., v_{r_i})\}$$

First, if ψ is existential second order, then $\mathfrak{A} \not\models \psi$ implies $G[\sigma] \not\models \varphi$. Second, if σ is such that:

a) the relation $R = \{(v_1, ..., v_{k+2}) \in V(G)^{k+2} \mid G[\sigma] \models \eta(v_1, ..., v_{k+2})\}$ characterizes a valid rooted graph of size (v, e) (for some $v > 0$, $e > 0$ integers) and

b) the set S has an $(R, c, 1)$-dress

then $\mathfrak{A} \models \psi$ if and only if $G[\sigma] \models \varphi$. Finally, if a) holds but b) does not hold for σ, then $G[\sigma] \not\models \varphi$.

Proof. Let us first create a second order formula $Q_1R_1...Q_nR_n\psi'$ equivalent to ψ where Q_i is one of $\exists$ and $\forall$, R_i are relational variables and ψ' is closed first order (of signature $(E/2, J_1/r_1, ..., J_m/r_m, R_1/s_1, ..., R_n/s_n)$). We know that such rewrite is possible for every second order formula. If ψ is existential second order then $\psi' = \psi$ and $Q_i = \exists$ for all i. Set $d = max\{t_i\}$. The following formula will be good:

$$\begin{aligned}\varphi = \exists(Dr = (&\underline{x}_1, y_1, z_1, ..., \underline{x}_{cd}, y_{cd}, z_{cd}, s_1, ..., s_d))(\\ &\text{“}Dr \text{ is an } (R, c, d)\text{-dress for } S\text{”} \wedge \\ &Q_1q_1 \in RelV^{s_1}_{S,Dr}Q_2q_2 \in RelV^{s_2}_{S,Dr}...Q_nq_n \in RelV^{s_n}_{S,Dr}(\psi'') \\ &)\end{aligned}$$

where we get ψ'' from ψ' by substituting all occurrence of $R_i(t_1, ..., t_{s_i})$ for $1 \leq i \leq n$ with $Rel^{s_i}_{S,Dr}[q_i](t_1, ..., t_{s_i})$ and all occurrence of $J_i(t_1, ..., t_{s_i})$ for $1 \leq i \leq m$ with $\iota_i(t_1, ..., t_{s_i})$.

First, if ψ is existential second order and $\mathfrak{A} \not\models \psi$ then no choice of relations can satisfy ψ', thus ψ'' is also false for any possible values of $Rel^{s_i}_{S,Dr}[q_i](t_1, ..., t_{s_i})$, hence φ is false. Second, if a) holds but b) does not then by Lemma 5.4.3 no (R, c, d)-dress can be found, so the formula is going to be false. Finally using the fact that if both a) and b) holds then any s_i-ary relation on S can be represented as $Rel^{s_i}_{S,Dr}[q]$ for an appropriate q one can easily see that φ is indeed equivalent to ψ' and thus to ψ on S. □

We already have tools to first order define occurrences of various rooted graphs in our infinite random graph G. We also know that among these rooted graphs there is (at least) one whose size corresponds to the numerator and denominator of $\tau_l(\alpha)$ if we choose the characterizing formulae right for l. But we will need to actually find that specific occurrence among all the rooted graph occurrences that we can characterize. The distinctive feature of this specific rooted graph is that it has the largest $\frac{v}{e}$ ratio among all the characterizable rooted graphs. So all we need to do is to give a first order definition of one rooted graph occurrence being better then an other in the above sense.

Definition 5.5.2. Let $V_1 \subset W_1 \subset V(G)$ and $V_2 \subset W_2 \subset V(G)$ be vertex sets of G. Let $v_1 = |W_1 \setminus V_1|$, $v_2 = |W_2 \setminus V_2|$. Let e_1 be the number of edges in G between vertices of V_1, and the same way e_2 is the number of edges in G between vertices of V_2. We say that (V_1, W_1) is *better* then (V_2, W_2) if we have $\frac{v_1}{e_1} > \frac{v_2}{e_2}$.

When using this definition sets V_1 and V_2 will correspond to the base vertices of two rooted graph occurrence while W_1 and W_2 will correspond to all the vertices of the same occurrences.

Lemma 5.5.3. *Let P be a fixed variable set, and let ζ and ζ' be $(k+2)$-ary P-formulae and c be a positive integer constant. Then there exists a $(2k+2)$-ary P-formula μ with the following properties. Let $\sigma : P \to V(G)$ be a variable assignment and $a_1, ..., a_{k+1}, a'_1, ..., a'_{k+1}$ be vertices of G. Suppose that $R = \{(v_1, ..., v_{k+2}) \in V(G)^{k+2} \mid G[\sigma] \models \zeta(v_1, ..., v_{k+2}\}$ characterizes a valid rooted graph of size (v, e) and $R' = \{(v_1, ..., v_{k+2}) \in V(G)^{k+2} \mid G[\sigma] \models \zeta'(v_1, ..., v_{k+2}\}$ characterizes a valid rooted graph of size (v', e'). Let us denote $S = \{a \mid G[\sigma] \models \zeta(a_1, ..., a_{k+1}, a)\}$ and $S' = \{a \mid G[\sigma] \models \zeta'(a'_1, ..., a'_{k+1}, a)\}$. Also let $B = \{a_1, ..., a_k\}$ and $B' = \{a'_1, ..., a'_k\}$. Then we have:*

1. *If $(B', B' \cup S')$ is not better then $(B, B \cup S)$ then $G[\sigma] \not\models \mu(a_1, ..., a_{k+1}, a'_1, ..., a'_{k+1})$.*
2. *If $(B', B' \cup S')$ is better then $(B, B \cup S)$, $B \cup S$ has an $(R, c, 1)$-dress and $B' \cup S'$ has an $(R', c, 1)$-dress then $G[\sigma] \models \mu(a_1, ..., a_{k+1}, a'_1, ..., a'_{k+1})$.*

Proof. Apply Theorem 5.5.1 with variable set $P' = P \cup \{y_1, ..., y_{k+1}, y'_1, ..., y'_{k+1}\}$, the P'-formulae below:

$$\begin{aligned}
&\iota_1(z) = (z = y_1) \vee ... \vee (z = y_k)\\
&\iota_2(z) = (z = y'_1) \vee ... \vee (z = y'_k)\\
&\iota_3(z) = \zeta(y_1, ..., y_k, y_{k+1}, z) \vee \iota_1(z)\\
&\iota_4(z) = \zeta'(y'_1, ..., y'_k, y'_{k+1}, z) \vee \iota_2(z)\\
&\eta(z_1, ..., z_{k+2}) = \zeta(z_1, ..., z_{k+2})\\
&\Delta(z) = \iota_3(z) \vee \iota_4(z)
\end{aligned}$$

the constant $c' = c + 1$ and last but not least for the second order formula γ which has signature of $(E/2, J_1/1, J_2/1, J_3/1, J_4/1)$ and which essentially says that if V_i is the set for which the unary relation J_i holds then (V_2, V_4) is better then (V_1, V_3). Let the closed first order P'-formula given by the theorem be ν. Apply the theorem again with exactly the same parameters except for:

$$\eta(z_1, ..., z_{k+2}) = \zeta'(z_1, ..., z_{k+2})$$

Notice the only difference is that we have ζ' instead of ζ. Now let us call ν' the formula that the theorem gives. Finally set $\mu = \nu \vee \nu'$.

Observe that γ can be formulated as an existential second order formula. It essentially states the existence of an injection $f : (V_3 \setminus V_1) \times E' \to (V_4 \setminus V_2) \times E$ where E is the set of edges between vertices in V_3, E' is the set of edges between vertices in V_4 such that there is a pair $(v, e) \in (V_4 \setminus V_2) \times E$ which is not a value of f.

So, by Theorem 5.5.1, neither ν or ν' holds if $(B', B' \cup S')$ is not better then $(B, B \cup S)$, so item 1 holds. For item 2 one only have to notice that the size of the union of the vertex sets of the two occurrences is at most twice the size of the vertex set of the larger rooted graph. Thus by setting $c' = c + 1$ and by requiring that $B \cup S$ has an $(R, c, 1)$-dress and

$B' \cup S'$ has an $(R', c, 1)$-dress we ensured that $S \cup B \cup S' \cup B'$ either has an $(R, c', 1)$-dress or an $(R', c', 1)$-dress. Hence if γ holds on $S \cup B \cup S' \cup B'$ then at least one of ν and ν' holds, so μ holds as claimed. $\square$

Chapter 6

Putting all together

Finally we are ready to prove the main result of this thesis.

Theorem 6.0.4. *If a function $f : \mathbb{R}^+ \setminus \mathbb{Q} \to \{0,1\}$ satisfies the Very Dense Condition, the Locally Constant Condition and the Complexity condition then there is a formula ν such that $f|_{[0,1/2]} = f_\nu|_{[0,1/2]}$.*

Proof. As f satisfies the Very Dense Condition there is a positive integer $k_0 \geq 2$ such that f is constant on $[0, \frac{1}{k_0}]$. First we are going to separately construct formulae ν_k for $3 \leq k \leq k_0$ that has $f|_{[\frac{1}{k}, \frac{1}{k-1}]} = f_{\nu_k}|_{[\frac{1}{k}, \frac{1}{k-1}]}$.

By the Complexity Condition we know there is a PH algorithm $\mathcal{A}$ which calculates the value of f^- for any rational $\frac{p}{q}$ encoded as 0^p1^q. One can construct another PH algorithm $\mathcal{A}'$ which works the following way. It takes encoded structures of signature $(E/2, B/1)$. These structures corresponds to rooted graphs: E gives the edges and B marks the base vertices. The algorithm answers yes if and only if for the v number of non-base vertices and for the e number of edges $f^-(\frac{v}{e}) = 1$. Indeed, $\mathcal{A}'$ simply counts the edges and the non-base vertices, then invokes $\mathcal{A}$. As the preparation steps are clearly polynomial (actually linear) and as $\mathcal{A}$ is in PH and the input of the $\mathcal{A}$ invocation is smaller then the original input of $\mathcal{A}'$, the whole computation is clearly in PH. Also observe $\mathcal{A}'$ is obviously order independent as defined in Section 2.3.1.

According to Fagin's theorem (Theorem 2.3.2) there is a second order formula δ such that for a structure $\mathfrak{A}$ of signature $(E/2, B/1)$ we have $\mathfrak{A} \models \delta$ if and only if the above algorithm would accept it.

By the Locally Constant Condition, there is a constant l such that f is constant in $[\tau_l(\alpha), \alpha]$ for any α positive real number. Let us fix $3 \leq k \leq k_0$. Applying Theorem 4.6.3 with our k and $d = l + 1$ we will have formulae φ_k and ψ_k of arity n_k such that when for v_1, ..., v_{n_k} vertices of an α-graph G we have $G[\{x_1 \mapsto v_1, ..., x_{n_k} \mapsto v_{n_k}\}] \models \varphi_k(x_1, ..., x_{n_k})$ then the relation $\{(a_1, ..., a_{k+2}) \in V(G)^{k+2} \mid G[\{x_1 \mapsto v_1, ..., x_{n_k} \mapsto v_{n_k}; y_1 \mapsto a_1, ..., y_{k+2} \mapsto a_{k+2}\}] \models \psi_k(x_1, ..., x_{n_k}, y_1, ..., y_{k+2})\}$ characterizes a valid rooted graph and there are parameters v_1, ..., v_{n_k} with which the above relation characterizes a rooted graph of size (v, e) with v and e relatively prime and $\frac{v}{e} = \tau_{l+1}(\alpha)$.

Applying Lemma 3.3.3 to $n = l + 1$ and $h = k$ there is a constant c such that for any $0 < \alpha < \frac{1}{2}$ and $\frac{p}{q} = \tau_{l+1}(\alpha)$ with p and q relatively prime we have:

$$q + k \leq \left\lfloor \frac{1}{q\alpha - p} \right\rfloor^c \tag{6.1}$$

Apply Theorem 5.5.1 to the variable set $P = \{x_1, ..., x_{n_k}, y_1, ..., y_{k+1}\}$, the following P-formulae:

$$\iota_k(z) = (z = y_1) \vee ... \vee (z = y_k)$$
$$\Delta(z) = \psi_k(x_1, ..., x_{n_k}, y_1, ..., y_{k+1}, z) \vee \iota_k(z)$$
$$\eta(z_1, ..., z_{k+2}) = \psi_k(x_1, ..., x_{n_k}, z_1, ..., z_{k+2}),$$

the second order formula δ and the above defined c, and let δ' denote the first order closed P-formula given by the theorem.

Using Lemma 5.5.3 for the variable set $P' = \{x_1, ..., x_{n_k}, x'_1, ..., x'_{n_k}\}$ and the P'-formulae:

$$\zeta(z_1, ..., z_{k+2}) = \psi_k(x_1, ..., x_{n_k}, z_1, ..., z_{k+2})$$
$$\zeta'(z_1, ..., z_{k+2}) = \psi_k(x'_1, ..., x'_{n_k}, z_1, ..., z_{k+2})$$

we get the $(2k+2)$-ary P'-formula μ.

Let ν_k be the formula below:

$$\nu_k = \exists x_1, ..., x_{n_k}, y_1, ..., y_{k+1}(\tag{6.2}$$
$$\varphi_k(x_1, ..., x_{n_k}) \wedge \tag{6.3}$$
$$\psi_k(x_1, ..., x_{n_k}, y_1, ..., y_{k+1}, y_{k+1}) \wedge \tag{6.4}$$
$$\neg(\exists x'_1, ..., x'_{n_k}, y'_1, ..., y'_{n_k}(\varphi(x'_1, ..., x'_{n_k}) \wedge \mu(y_1, ..., y_{k+1}, y'_1, ..., y'_{k+1}))) \wedge \tag{6.5}$$
$$\delta') \tag{6.6}$$

To find out when ν_k is true let us investigate a specific assignment of the variables of the outermost quantification: let x_i be assigned to the vertex v_i for $1 \leq i \leq n_k$ and y_i be assigned to w_i. We assume $\frac{1}{k} < \alpha < \frac{1}{k-1}$. Let $R = \psi_k(v_1, .., v_{n_k}, \_, ..., \_)$ and if it characterizes a valid rooted graph then let us call that $\mathcal{H}$, and let (v, e) be the size of $\mathcal{H}$ and H be the underlying graph of $\mathcal{H}$. We also set $S = \{w_1, ..., w_k\} \cup \psi_k(v_1, ..., v_{n_k}, w_1, ..., w_{k+1}, \_)$. We call an assignment *good* if:

(i) $\varphi_k(v_1, .., v_{n_k})$ holds in G thus R characterizes a valid rooted graph and $\frac{v}{e} = \tau_{l+1}(\alpha)$

(ii) $\psi_k(v_1, .., v_{n_k}, w_1, ..., w_{k+1}, w_{k+1})$ holds and the induced subgraph of G spanned by the vertex set S is isomorphic to the H.

We remark that from (i) and from $\psi(v_1, .., v_{n_k}, w_1, ..., w_{k+1}, w_{k+1})$ we already know that the subgraph spanned by S contains H as a subgraph. (ii) also claims that there are no extra edges.

By Theorem 4.6.3 we know that there are vertices v_1, ..., v_n satisfying (i). By Lemma 4.3.1 for any v's satisfying (i) there are w's satisfying (ii) Thus there is a good assignment.

First let us investigate what happens if the assignment is good. Parts 6.3 and 6.4 of ν_k trivially hold. By the third point of Theorem 4.6.3 and by point 1. of Lemma 5.5.3 the formula $\mu(y_1, ..., y_{k+1}, y'_1, ..., y'_{k+1})$ can never hold if x'_1, ..., x'_{n_k} are assigned such way that $\varphi_k(x'_1, ..., x'_{n_k})$ holds. Thus part 6.5 also holds. Observe that by the choice of c and by Lemma 3.3.3 the set S has an $(R, c, 1)$-dress. Thus by Theorem 5.5.1 δ' holds if and only if $f^-(\tau_{l+1}(\alpha)) = 1$. Observe that as $\tau_{l+1} \in [\tau_l(\alpha), \alpha)]$ and f^- is continuous at the irrational point α using the Locally Constant Condition we have $f^-(\tau_{l+1}(\alpha)) = f(\alpha)$. Thus δ' holds if and only if $f(\alpha) = 1$.

We claim that if an assignment is not good then at least one of 6.3, 6.4, 6.5 or 6.6 will not hold. If R does not characterize a valid rooted graph then 6.3 fails. Otherwise if the rooted graph $\mathcal{H}$ characterized by R is not present in $(G, w_1, ..., w_k)$ with counting vertex w_{k+1} then 6.4 fail. If the set S does not have an $(R, c, 1)$-dress then 6.6 will fail by the last statement of Theorem 5.5.1. Otherwise if for the (v, e) size of $\mathcal{H}$ we have $\frac{v}{e} \neq \tau_{l+1}(\alpha)$ thus $\frac{v}{e} < \tau_{l+1}(\alpha)$ and/or there are extra edges in the induced subgraph spanned by S then 6.5 will fail by the existence of a good assignment which we can apply to the x''s and y''s.

Putting the above together we have that the ν_k holds if and only if $f(\alpha) = 1$ when $\frac{1}{k} < \alpha < \frac{1}{k-1}$ which implies $f|_{[\frac{1}{k}, \frac{1}{k-1}]} = f_{\nu_k}|_{[\frac{1}{k}, \frac{1}{k-1}]}$ as claimed.

For $k \geq 2$ let L_k be the graph on $k+1$ vertices $\{a, b_1, ..., b_k\}$ where a is connected to all others and no other edges are present. We can easily create a first order characterization of the rooted graph $\mathcal{L}_k = (L_k, b_1, ..., b_k, a)$. Thus by Lemma 4.5.1 we can create a closed first order formula β_k such that $G \models \beta_k$ if and only if $\mathcal{L}_k$ is safe. Observing that the $\mathcal{L}_k$ is safe if and only if $\alpha < \frac{1}{k}$ we get that $G \models \beta_k$ if and only if $\alpha < \frac{1}{k}$. Let us define the following formula:

$$\nu' = (\neg\beta_{k_0} \wedge \beta_{k_0-1} \wedge \nu_{k_0}) \vee (\neg\beta_{k_0-1} \wedge \beta_{k_0-2} \wedge \nu_{k_0-1}) \vee ... \vee (\neg\beta_3 \wedge \beta_2 \wedge \nu_3)$$

It is obvious from the properties of β_k and ν_k that $f|_{[\frac{1}{k_0}, \frac{1}{2}]} = f_{\nu'}|_{[\frac{1}{k_0}, \frac{1}{2}]}$ and that $f_{\nu'}$ is 0 outside $[\frac{1}{k_0}, \frac{1}{2}]$. Thus $\nu = \nu'$ is good if $f(x) = 0$ for $x \leq \frac{1}{k_0}$. Otherwise $\nu = \nu' \vee \beta_{k_0}$ is good. □

Bibliography

[1] Bogomolny, A. Stern-Brocot tree.
http://www.cut-the-knot.org/blue/Stern.shtml.

[2] Brocot, A. Calcul des rouages par approximation, nouvelle méthode. *Revue Chonométrique 3* (1861), 186–194.

[3] Fagin, R. Generalized first-order spectra and polynomial-time recognizable sets. In *Complexity of Computation, SIAM-AMS Proceedings 7* (1974), R. Karp, Ed., pp. 43–73.

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

[5] Spencer, J. Countable sparse random graphs. *Random Struct. Algorithms 1*, 2 (1990), 205–214.

[6] Spencer, J. Sparse random graphs: A continuum of complete theories. *Colloquia Math Soc. János Bolyai* (1991), 679–690.

[7] Spencer, J., and Tardos, G. Ups and downs of first order sentences on random graphs. *Combinatorica 20*, 2 (2000), 263–280.

[8] Stern, M. Über eine zahlentheoretische Funktion. *J. reine angew. Math. 55* (1858), 193–220.

[9] Wikipedia. Polynomial hierarchy.
http://en.wikipedia.org/wiki/polynomial_hierarchy.