Electronic Colloquium on Computational Complexity, Report No. 35 (2006)

The Descriptive Complexity of the Reachability Problem As a Function of Different Graph Parameters

Till Tantau

Universität zu Lübeck
Institut für Theoretische Informatik
23538 Lübeck, Germany
tantau@tcs.uni-luebeck.de

January 26, 2006

Abstract

The reachability problem for graphs cannot be described, in the sense of descriptive complexity theory, using a single first-order formula. This is true both for directed and undirected graphs, both in the finite and infinite. However, if we restrict ourselves to graphs in which a certain graph parameter is fixed to a certain value, first-order formulas often suffice. A trivial example are graphs whose number of vertices is fixed to n. In such graphs reachability can be described using a first-order formula with a quantifier nesting depth of $\log_2 n$, which is both a lower and an upper bound. In this paper we investigate how the descriptive complexity of the reachability problem varies as a function of graph parameters such as the size of the graph, the clique number, the matching number, the independence number or the domination number. The independence number turns out to be the by far most challenging graph parameter.

1 Introduction

The descriptive complexity of a problem quantifies the complexity of a minimal description of a problem. The idea is to fix a formalism, typically a logic like first- or second-order logic, and to then find a simple description of the problem, like a formula that "states" the problem. Which formulas are considered to be "simple" depends on the kind of descriptive complexity that we want to study: Important measures include the number of different variables in the formula, the number of quantifiers, the number of nested quantifiers, the number of quantifier alternations, and the length of the formula.

To take an example, consider the problem "Is there a path of length at most 2 from s to t is a directed graph G?" and fix the formalism to "first-order logic." In this formalism a graph $G = (V, E)$ is represented as a logical structure $\mathcal{G}$ over the signature containing a binary relational symbol E. The universe of the structure is the vertex set V, the interpretation $\mathrm{E}^{\mathcal{G}}$ of the relation symbol E is the edge relation E. In order to model the source and target vertices s and t, two additional constant symbols s and t can be used. A formula that describes reachability in two steps is the following: $\mathrm{s} = \mathrm{t} \lor \mathrm{E}(\mathrm{s}, \mathrm{t}) \lor \exists z \,.\, \mathrm{E}(\mathrm{s}, z) \land \mathrm{E}(z, \mathrm{t})$.

The descriptive complexity of a problem is not only of interest in the context of logic and model theory. It also tells us a lot about the computational complexity of the problem. For example, if we can describe a problem using q nested first-order quantifiers, we can decide the problem using $O(\log^q n)$ space and $O(q)$ parallel time (for appropriate models of parallel computation).

ISSN 1433-8092

It is well-known that the descriptive complexity of the reachability problem (without any bound on the distance) with respect to first-order formulas is "infinite": No first-order formula is true exactly for those graphs in which there is a path from s to t. The reachability problem is also known as the s-t-connectivity problem and also as the graph accessibility problem. However, if we have some additional information about the graphs for which we would like to describe the reachability problem, it is often possible to give an efficient description. For example, we might be promised that the graph has at most n vertices for some constant n. In this case, we can write down a "trivial" formula having n quantifiers that decides the reachability problem. It is even possible to design the formula in such a way that it uses only three different variables and has a nesting depth of $\log_2 n$ and it is known that both these measures (3 and $\log_2 n$) are optimal.

The present paper presents a systematic study of how the descriptive complexity of the reachability problem varies as a function of different graph parameters. However, only first-order formulas will be treated and the only formula complexity measure that will be studied is the quantifier complexity, which is also known as quantifier (nesting) depth or quantifier rank. The quantifier complexity, abbreviated $\mathrm{qc}(\phi)$, is zero for atomic formulas ϕ; negating a formula does not increase the quantifier complexity; $\mathrm{qc}(\psi \wedge \rho) = \max\{\mathrm{qc}(\psi), \mathrm{qc}(\rho)\}$; and the nesting depth of $\exists x \,.\, \psi$ is $1 + \mathrm{qc}(\psi)$. Note that the quantifier complexity does *not* measure the number of quantifier *alternations*, but the number of quantifier *nestings*. For example, $\mathrm{qc}\big((\exists x \,.\, x = x) \wedge (\forall \,.\, y = y)\big) = 1$ and $\mathrm{qc}(\exists x \exists y \,.\, x = y) = 2$.

Formally, let $p(G)$ be some graph parameter (like the size of the graph G or the independence number or the chromatic number). For each number k let $\text{REACH}_{p=k}$ denote the following promise problem: We are promised for a given directed graph G that $p(G) = k$ and we wish to find out whether there is a path from s to t in G. In this paper we are interested in the following values:

1. $\mathrm{qc}(\text{REACH}_{p=k})$ denotes the minimal quantifier complexity of first-order formulas ϕ for which for every (G, s, t), where G is a finite directed graph with $p(G) = k$, there is a path from s to t in G if and only if $(G, s, t) \models \phi$. If no such formula exists, let $\mathrm{qc}(\text{REACH}_{p=k}) = \infty$.

2. $\mathrm{qc}(\text{REACH}^{\infty}_{p=k})$ denotes the minimal quantifier complexity of first-order formulas ϕ for which for every (G, s, t), where G is a *finite or infinite* directed graph with $p(G) = k$, there is a path from s to t in G if and only if $(G, s, t) \models \phi$.

3. $\mathrm{qc}(\text{UREACH}_{p=k})$ denotes the restriction to finite undirected graphs (graphs with a symmetric edge relation).

4. $\mathrm{qc}(\text{UREACH}^{\infty}_{p=k})$ denotes the restriction to undirected graphs (finite or infinite).

The results presented in this paper uncover an simple pattern. A promise $p(G) = k$ typically influences the quantifier complexity of the reachability problem in one of the following ways:

1. The promise does not make describing the reachability problem any easier (has no effect).

2. The promise makes describing the reachability problem trivial.

3. The promise causes the quantifier complexity to become $\log_2 k$ (possibly plus or minus some small constant).

There is one graph parameter whose influence does not follow this pattern: the independence number $\alpha(G)$. It is the largest number of vertices that can be picked from a graph such that there is no edge between any two picked vertices. While for undirected graphs the promise

$\alpha(G) = k$ causes the quantifier complexity of the reachability problem to become $\log_2 k$ "as usual," for finite directed graphs the following two bounds holds:

$$c \log_2 k \leq \text{qc}(\text{REACH}_{\alpha=k}) \leq k + 3,$$

where $c = 2\log_3 2 > 1.26$. The upper bound was established in [NT05]. The lower bound is proved in the present paper and seems challenging to prove.

For most graph parameters, the upper bounds are established using simple observations on how these graph parameters relate to the maximum distance of vertices in the graph. For example, if we know that an undirected graph has a vertex cover of size k, then the distance of s and t cannot be more than $2k$, if they are connected at all.

All lower bounds are established using Ehrenfeucht–Fraïssé games. However, for most graph parameters it is not necessary to explicitly state how these games are played. Instead, we can use Hanf's theorem, which is based on Ehrenfeucht–Fraïssé games.

For the graph parameter "independence number" we cannot use Hanf's theorem. Hanf's theorem states, very roughly, that if the local neighborhoods of the vertices in two different graphs "look alike," then the two graphs cannot be distinguished using first-order formulas of a certain quantifier complexity. However, in graphs with bounded independence number the "local neighborhood" of any vertex is always the whole graph, rendering Hanf's theorem useless in this situation. Because of this it is necessary to devise appropriate Ehrenfeucht–Fraïssé games from scratch to prove the lower bound of $c \log_2 k$ for $\text{qc}(\text{REACH}_{\alpha=k})$.

This paper is organized as follows. In Section 2 basic definitions are given and the mathematical tools are presented that will be used in the main proofs. In Section 3 the "simple cases" are discussed, which refers to all graph parameters except for the independence number. This latter graph parameter is studied in Section 4, where the lower bound on the quantifier complexity of $\text{qc}(\text{REACH}_{\alpha=k})$ is proved.

2 Review of Model Theory Tools

In the present section the basic tools of model theory that will be used in the later sections are reviewed. We start with fixing some notations and terminology. Next, Ehrenfeucht–Fraïssé games are defined and Hanf's theorem is stated.

Throughout this paper we will restrict our attention to directed or undirected graphs with two distinguished vertices called s and t. Such graphs will just be called "graphs" in the following. We view graphs to be logical structures $\mathcal{G}$ over the relational signature $(\mathrm{E}^2, \mathrm{s}^0, \mathrm{t}^0)$, where the universe of $\mathcal{G}$ is the vertex set of the graph and the binary relation $E^{\mathcal{G}}$ is the edge relation. Given a set A of graphs, we say that a (first-order) formula ϕ *describes* A if a tuple (G, s, t) is in A if and only if $(G, s, t) \models \phi$. The *quantifier complexity* of a formula ϕ was defined in the introduction: It is maximal number of "nested" quantifiers in the formula.

A *promise problem* is a pair (P, Q) consisting of a *promise* P and a *question* Q. In the present paper, Q and P will always be sets of graphs. A *solution* to a promise problem (P, Q) is a set A such that $A \cap P = Q \cap P$. A formula *describes a promise problem* if it describes a solution of the promise problem.

The most important tool for proving lower bounds on the quantifier complexity of formulas that describe a promise problem are Ehrenfeucht–Fraïssé games. For the special case of graphs with two distinguished vertices these games are played as follows: There are two players, called player I and player II (there are numerous other names in the literature). The playing field consists of two graphs G and $\bar{G}$. The game is played in r rounds. In each round, player I starts by picking up a coloured pebble from a pile of pebbles and placing the pebble on some vertex of either graph. Player II responds by picking up a new pebble of the same color from the pile and placing this pebble on some vertex of the other graph. Next, it is player I's turn once more, who now picks up a new pebble of a new color from the pile and places it on some vertex of either graph. Player II once more responds by putting a pebble of the same color on the other structure. The game ends after r rounds. To determine the winner, we remove all vertices from the graphs that do not have a pebble on them and we also leave the source and the target (as if there were a pebble on them). If the resulting coloured graphs are isomorphic, player II wins; otherwise player I wins.

The connection between Ehrenfeucht–Fraïssé games and first-order formulas is given by the following theorem:

Theorem 2.1 ([Ehr61, Fra54]). *Suppose player II has a winning strategy for the r round game on two graphs G and $\bar{G}$. Then for every first-order formula ϕ with $\operatorname{qc}(\phi) \leq r$ we have $G \models \phi$ if and only if $\bar{G} \models \phi$.*

Corollary 2.2. *Fix a graph parameter $p(G)$ and a number k. Let G and $\bar{G}$ be graphs with $p(g) = p(\bar{G}) = k$ and let $G \in \text{REACH}$ and $\bar{G} \notin \text{REACH}$. Suppose player II has a winning strategy for the r round game on G and $\bar{G}$. Then $\operatorname{qc}(\text{REACH}_{p=k}) > r$.*

In the following, whenever we consider pairs of graphs where one graph is an element of REACH while the other is not, the instance that is not an element of REACH is indicated by a bar above it. This is the same notation as the one used in the corollary.

The above corollary is a powerful tool. In order to prove, say, $\operatorname{qc}(\text{REACH}_{\alpha=1}) > 3$, all we have to do is to find two finite graphs G and $\bar{G}$ with independence number 1 (such graphs are also known as tournaments or, to be precise, as semicomplete graphs) such that in the first graph there is a path from s to t and in the other is not and player II has a winning strategy for three rounds. Since the graphs are finite, such a winning strategy can be tested (at least if G and $\bar{G}$ are reasonably small) using exhaustive search. In other words, if we are able to find appropriate graphs G and $\bar{G}$ we can use a computer to prove lower bounds on the quantifier complexity.

Unfortunately, finding small graphs G and $\bar{G}$ with the desired properties turns out to be a tricky business. Fortunately, once we have found them, they can often be reused in new proofs. For example, if player II has a winning strategy on G and $\bar{G}$, player II also has a winning strategy on the disjoint unions $G \mathbin{\dot\cup} H$ and $\bar{G} \mathbin{\dot\cup} H$, where H is some other graph. Several other operations on graphs also do not destroy player II's winning strategy: It is often difficult to construct the winning strategy of player II from scratch. Hanf's theorem provides us with an easy-to-check condition on G and $\bar{G}$ that, when satisfied, ensures that player II has a winning strategy.

For a graph H (with two distinguished vertices) we define its *isomorphism type* as the equivalence class under graph isomorphism that honour and the source and the target vertices. For a graph $G = (V, E)$ let G_{Gaifman} denote the underlying undirected graph of G (in more general contexts this graph is known as the Gaifman graph of the logical structure G, hence the name). For a vertex v and a radius r let $S(v, r) := \{u \in V \mid \mathrm{d}_{G_{\text{Gaifman}}}(v, u) \leq r\}$ denote the "sphere" around v or radius r. Note that the distance is measured in the undirected version of G. We can view the sphere as a substructure of G. The *r-type* of a vertex v is the isomorphism type of the structure $S(v, r)$.

Theorem 2.3 ([Han65]). *Let G and $\bar{G}$ be graphs and let r be a number. For each possible 2^r-type, that is, for each possible isomorphism type of spheres of radius 2^r, let G and $\bar{G}$ have the same number of vertices of this type. Then player II wins an r-round Ehrenfeucht–Fraïssé games on G and $\bar{G}$.*

Parameter $p(G)$	$\mathrm{qc}(\text{REACH}_{p=k})$	$\mathrm{qc}(\text{REACH}^{\infty}_{p=k})$	$\mathrm{qc}(\text{UREACH}_{p=k})$	$\mathrm{qc}(\text{UREACH}^{\infty}_{p=k})$
$\kappa > 0$ (connectivity)	0	0	0	0
$d < \infty$ (diameter)	0	0	0	0
$\chi = 1$ (chromatic number)	0	0	0	0
$\omega = 1$ (clique number)	0	0	0	0
number of components is 1	0	0	0	0
$\Delta \leq 1$ (maximum degree)	0	0	0	0
n (number of vertices)	$\log_2 k \pm O(1)$	$\log_2 k \pm O(1)$	$\log_2 k \pm O(1)$	$\log_2 k \pm O(1)$
maximum finite distance	$\log_2 k \pm O(1)$	$\log_2 k \pm O(1)$	$\log_2 k \pm O(1)$	$\log_2 k \pm O(1)$
vertex covering number	$\log_2 k \pm O(1)$	$\log_2 k \pm O(1)$	$\log_2 k \pm O(1)$	$\log_2 k \pm O(1)$
matching number	$\log_2 k \pm O(1)$	$\log_2 k \pm O(1)$	$\log_2 k \pm O(1)$	$\log_2 k \pm O(1)$
domination number	∞	∞	$\log_2 k \pm O(1)$	$\log_2 k \pm O(1)$
α (independence number)	$> 1.26 \log_2 k$ $\leq k + 3$	∞	$\log_2 k \pm O(1)$	$\log_2 k \pm O(1)$
$\kappa = 0$ (connectivity)	∞	∞	∞	∞
$d = \infty$ (diameter)	∞	∞	∞	∞
$\chi > 1$ (chromatic number)	∞	∞	∞	∞
$\omega > 1$ (clique number)	∞	∞	∞	∞
number of components ≥ 2	∞	∞	∞	∞
δ (minimum degree)	∞	∞	∞	∞
$\Delta > 1$ (maximum degree)	∞	∞	∞	∞

Table 1: The quantifier complexity of the reachability problem as a function of the different graph parameters.

3 Graph Parameters: Simple Cases

In the present section the descriptive complexity of the promise problem $\text{REACH}_{p=k}$ is studied for different graph parameters p. The presentation of the results is sorted such that the proofs are as easy as possible; a table summing up the results and sorting them according to the problem complexity is given on page 5.

The definitions of the different graph parameters are taken from Diestel's book [Die97], though the definition may be adapted for directed graphs if the original definition made sense only for undirected graphs.

3.1 Graph Parameter: Number of Vertices

Perhaps the most basic graph parameter is the number $n(G)$ of vertices. If we are promised that a given graph has exactly $n = k$ vertices, the problem of finding a first-order description becomes easy in the following sense: *Any* graph property that is invariant under isomorphisms can be described using an appropriate first-order formula when we are promised that the graph has exactly k vertices. The reason is that using k quantifiers it is possible to fix all vertices of the graph and to use a big conjunction to list all k-vertex graphs having the given property.

Thus, we can express reachability in k-vertex graphs using k nested quantifiers and k variables. However, it is well-known that we can do much better: $\log_2 k$ nested quantifiers suffice and this is optimal. The proof of this result can be found in standard textbooks, but it is included below since we will use similar arguments in later proofs.

Theorem 3.1. *For all positive k we have*

$$\mathrm{qc}(\textsc{reach}_{n=k}) = \mathrm{qc}(\textsc{reach}^{\infty}_{n=k}) = \mathrm{qc}(\textsc{ureach}_{n=k}) = \mathrm{qc}(\textsc{ureach}^{\infty}_{n=k}) = \log_2 k \pm O(1).$$

Proof. For the upper bound, let $\phi_1(u,v) \equiv u = v \vee \mathrm{E}(u,v)$ describe the graph property "there is a path of length at most 1 from u to v," and let $\phi_{2n}(u,v) \equiv \phi_n(u,v) \vee \exists z \,.\, \big(\phi_n(u,z) \wedge \phi_n(z,v)\big)$ describe the property "there is a path of length at most $2n$ from u to v." Then reachability can be expressed using the formula $\phi_k(\mathrm{s},\mathrm{t})$. Clearly, the quantifier complexity of ϕ_k is $\log_2 k$.

For the lower bound, define two graphs G and $\bar{G}$ as follows: Let G consist of an undirected cycle of length $2 \cdot 2^k + 4$ with s and t at opposite "ends" of the cycle, that is, at maximal distance from each other. Let $\bar{G}$ consist of two undirected cycles, each of length $2^k + 2$, and let s be in one of the cycles and t in the other one. Clearly, G is an instance of the (undirected and also the directed) reachability problem, while $\bar{G}$ is not. Both graphs have the same number of vertices of each 2^k-type. By Hanf's theorem the graphs are indistinguishable. □

3.2 Graph Parameter: Number of Components

A *component* of a graph is a maximal (strongly) connected subgraph of the graph. Suppose we are promised that the number of components in a graph is k. How does that help us with the reachability problem? Certainly, if we are promised that k is 1, then the reachability problem becomes trivial since there is always a path from s to t in this situation. However, knowing that a graph has, say, 25 components does not really help us with deciding whether there is a path from s to t. The following theorem formalizes this observation:

Theorem 3.2. *Let $p(G)$ denote the number of components in G. For all $k > 1$ we have*

$$\mathrm{qc}(\textsc{reach}_{p=1}) = \mathrm{qc}(\textsc{reach}^{\infty}_{p=1}) = \mathrm{qc}(\textsc{ureach}_{p=1}) = \mathrm{qc}(\textsc{ureach}^{\infty}_{p=1}) = 0,$$
$$\mathrm{qc}(\textsc{reach}_{p=k}) = \mathrm{qc}(\textsc{reach}^{\infty}_{p=k}) = \mathrm{qc}(\textsc{ureach}_{p=k}) = \mathrm{qc}(\textsc{ureach}^{\infty}_{p=k}) = \infty.$$

Proof. We only need to prove the second claim since the first is trivial. To prove the infinite lower bound, suppose one of these problems could be described using a formula of quantifier complexity r. Consider the following graphs: let both G and $\bar{G}$ consist of k undirected cycles of length $2 \cdot 2^r + 4$. In G the vertices s and t are in the same cycle, but at maximal distance from each other. In $\bar{G}$ they are in different cycles. Then G and $\bar{G}$ satisfy the promise, but Hanf's theorem tells us that they cannot be distinguished using a formula of quantifier depth r. □

3.3 Graph Parameter: Connectivity

The *connectivity* κ of a graph is the minimal number of vertices that we have to remove to make the graph disconnected. Thus, a disconnected graph has connectivity 0 and a clique of size n has connectivity $n-1$.

It seems quite natural that the connectivity of a graph should have a strong influence on the complexity of the reachability problem. Indeed, if $\kappa(G) > 0$ then the graph is connected and there is always a path from s to t. Thus, if we are promised $\kappa(G) = k$ for some $k > 0$ then the quantifier complexity of $\textsc{reach}_{\kappa=k}$ is zero.

If we are promised that $\kappa(G) = 0$, we just know that the graph is disconnected. This does not really help us with the question whether there is a path from s to t since we do not know whether s and t are in the some component or not. Formally, the graphs from Theorem 3.2 have $\kappa(G) = 0$ and show that no first-order formula can describe reachability in all of them.

Theorem 3.3. *For all $k > 0$ we have*

$$\mathrm{qc}(\textsc{reach}_{\kappa=0}) = \mathrm{qc}(\textsc{reach}^{\infty}_{\kappa=0}) = \mathrm{qc}(\textsc{ureach}_{\kappa=0}) = \mathrm{qc}(\textsc{ureach}^{\infty}_{\kappa=0}) = \infty,$$
$$\mathrm{qc}(\textsc{reach}_{\kappa=k}) = \mathrm{qc}(\textsc{reach}^{\infty}_{\kappa=k}) = \mathrm{qc}(\textsc{ureach}_{\kappa=k}) = \mathrm{qc}(\textsc{ureach}^{\infty}_{\kappa=k}) = 0.$$

3.4 Graph Parameter: Diameter and Maximum Finite Distance

The diameter of a graph is the maximum distance between any two vertices. It is infinite if the graph is not connected. The maximum finite distance in a graph is maximum distance taken over all pairs of connected vertices.

If the diameter of a graph is finite then, clearly, there is a path from s to t and the reachability problem becomes trivial. On the other hand, if the distance is infinite, this does not tell us anything about the question of whether s and t are in the same component. Formally, the graphs of Theorem 3.2 have infinite distance. This proves the following theorem: observation:

Theorem 3.4. *Let $p(G)$ denote the diameter of G. For all $k < \infty$ we have*

$$\mathrm{qc}(\text{REACH}_{p=k}) = \mathrm{qc}(\text{REACH}^{\infty}_{p=k}) = \mathrm{qc}(\text{UREACH}_{p=k}) = \mathrm{qc}(\text{UREACH}^{\infty}_{p=k}) = 0,$$
$$\mathrm{qc}(\text{REACH}_{p=\infty}) = \mathrm{qc}(\text{REACH}^{\infty}_{p=\infty}) = \mathrm{qc}(\text{UREACH}_{p=\infty}) = \mathrm{qc}(\text{UREACH}^{\infty}_{p=\infty}) = \infty.$$

For the maximum finite distance the following holds:

Theorem 3.5. *Let $p(G)$ denote the maximum finite distance in G. Then for all k we have*

$$\mathrm{qc}(\text{REACH}_{p=k}) = \mathrm{qc}(\text{REACH}^{\infty}_{p=k}) = \mathrm{qc}(\text{UREACH}_{p=k}) = \mathrm{qc}(\text{UREACH}^{\infty}_{p=k}) = \log_2 k \pm O(1).$$

Proof. Both for the upper bound and lower bound, just observe that the argument from Theorem 3.1 holds also for the maximum finite distance. ☐

3.5 Graph Parameter: Chromatic Number and Clique Number

The chromatic number of a graph is the number of colors that are needed to color the graph such that no edge is between vertices of the same color. The clique number is the size of the largest complete subgraph in a graph. Both the numbers have no effect on the reachability problem if they are at least 2. To see this, consider the graphs from Theorem 3.2, where player II had a winning strategy for higher and higher r. If we add a large clique to these graphs, the chromatic and clique numbers can be raised to an arbitrary level, without changing the fact that player II has a winning strategy. For a chromatic or clique number of 1 the graph must be discrete and reachability becomes trivial.

Theorem 3.6. *Let $p(G)$ denote the clique number or the chromatic number of G. For all $k > 1$ we have*

$$\mathrm{qc}(\text{REACH}_{p=1}) = \mathrm{qc}(\text{REACH}^{\infty}_{p=1}) = \mathrm{qc}(\text{UREACH}_{p=1}) = \mathrm{qc}(\text{UREACH}^{\infty}_{p=1}) = 0,$$
$$\mathrm{qc}(\text{REACH}_{p=k}) = \mathrm{qc}(\text{REACH}^{\infty}_{p=k}) = \mathrm{qc}(\text{UREACH}_{p=k}) = \mathrm{qc}(\text{UREACH}^{\infty}_{p=k}) = \infty.$$

3.6 Graph Parameter: Minimum and Maximum Degree

If for some graph G the maximum degree of the underlying undirected graph is 0 or 1, reachability becomes trivial. However, the graphs from Theorem 3.2, possibly with cliques added to increase the maximum degree, show that for larger k a maximum degree of k has no influence on the reachability problem.

The minimum degree of vertices in the graph has no influence on the reachability problem at all. However, to prove this, we cannot simply use the graphs from Theorem 3.2 nor graphs in which we add cliques since this does not raise the minimum degree. Fortunately, this is easy to fix: We can raise the minimum degree in the graphs from Theorem 3.2 by replacing each vertex by a clique of appropriate size. This will not change the fact that player II has a winning strategy.

Theorem 3.7. *Let $p(G)$ denote the minimum degree of G. The for all k we have*

$$\mathrm{qc}(\textsc{reach}_{p=k}) = \mathrm{qc}(\textsc{reach}^{\infty}_{p=k}) = \mathrm{qc}(\textsc{ureach}_{p=k}) = \mathrm{qc}(\textsc{ureach}^{\infty}_{p=k}) = \infty.$$

Let $p(G)$ denote the maximum degree of the underlying undirected graph of G. The for all $k > 1$ we have

$$\begin{aligned}
\mathrm{qc}(\textsc{reach}_{p=0}) = \mathrm{qc}(\textsc{reach}^{\infty}_{p=0}) = \mathrm{qc}(\textsc{ureach}_{p=0}) = \mathrm{qc}(\textsc{ureach}^{\infty}_{p=0}) = 0 \\
\mathrm{qc}(\textsc{reach}_{p=1}) = \mathrm{qc}(\textsc{reach}^{\infty}_{p=1}) = \mathrm{qc}(\textsc{ureach}_{p=1}) = \mathrm{qc}(\textsc{ureach}^{\infty}_{p=1}) = 0 \\
\mathrm{qc}(\textsc{reach}_{p=k}) = \mathrm{qc}(\textsc{reach}^{\infty}_{p=k}) = \mathrm{qc}(\textsc{ureach}_{p=k}) = \mathrm{qc}(\textsc{ureach}^{\infty}_{p=k}) = \infty.
\end{aligned}$$

3.7 Graph Parameter: Vertex Covering Number

A vertex cover of a graph is a set of vertices such that for every edge (u, v) of the graph either u or v or both are in the vertex cover. The vertex covering number is the size of the smallest vertex cover of a graph.

Theorem 3.8. *Let $p(G)$ denote the vertex covering number of the graph G. Then*

$$\mathrm{qc}(\textsc{reach}_{p=k}) = \mathrm{qc}(\textsc{reach}^{\infty}_{p=k}) = \mathrm{qc}(\textsc{ureach}_{p=k}) = \mathrm{qc}(\textsc{ureach}^{\infty}_{p=k}) = \log_2 k \pm O(1).$$

Proof. For the lower bound note that the graphs constructed in the proof of Theorem 3.1 have a vertex covering number of $\log_2 k - O(1)$.

For the upper bound note that if a graph has a vertex covering number of k, then the shortest path from s to t can have length at most $2k$. □

3.8 Graph Parameter: Matching Number

The matching number of a graph is the size of a maximum matching in the graph. For directed graph, we define this number to be the maximum size of a matching in the underlying undirected graph.

Theorem 3.9. *Let $p(G)$ denote the matching number number of the graph G. Then*

$$\mathrm{qc}(\textsc{reach}_{p=k}) = \mathrm{qc}(\textsc{reach}^{\infty}_{p=k}) = \mathrm{qc}(\textsc{ureach}_{p=k}) = \mathrm{qc}(\textsc{ureach}^{\infty}_{p=k}) = \log_2 k \pm O(1).$$

Proof. For the lower bound note that the graphs constructed in the proof of Theorem 3.1 have a matching of $\log_2 k - O(1)$.

For the upper bound note that if a graph has a matching number of k, then the shortest path from s to t can have length at most $2k$. □

3.9 Graph Parameter: Domination Number

A dominating set of graph $G = (V, E)$ is a set D of vertices such that for every vertex $v \in V$ either $v \in D$ or there is a $u \in D$ such that $(u, v) \in E$. The domination number of a graph is the size of the smallest dominating set of the graph.

Theorem 3.10. *Let $p(G)$ denote the domination number of the graph G. Then*

$$\begin{aligned}
\mathrm{qc}(\textsc{reach}_{p=k}) = \mathrm{qc}(\textsc{reach}^{\infty}_{p=k}) = \infty, \\
\mathrm{qc}(\textsc{ureach}_{p=k}) = \mathrm{qc}(\textsc{ureach}^{\infty}_{p=k}) = \log_2 k \pm O(1).
\end{aligned}$$

Proof. For the directed case, to prove the infinite lower bound consider the graphs from Theorem 3.2. If we add a single vertex to the each of these graphs and add edges from this vertex to every other vertex, the graphs have domination number 1. However, adding the vertices has no effect on the question of whether there is a path from s to t and it does not change the fact the player II has a winning strategy. By adding not one, but several new vertices and connecting them to all original vertices we can also increase the domination number to any given k.

For the undirected case, for the lower bound note that the graphs constructed in the proof of Theorem 3.1 have a dominating set of size $\log_2 k - O(1)$ (just take every second vertex).

For the upper bound in the undirected case observe that if an undirected graph has a dominating set of size k, then the shortest path from s to t can have length at most $2k$. To see this, note that a vertex in the dominating set can be adjacent to at most two vertices on the path; for if it were adjacent to three vertices on the path, then path could be shortened by going through the vertex in the dominating set. ☐

4 Graph Parameter: The Independence Number

For the graph parameters studied in the previous section we saw that the promise that a graph parameter p has a certain value k has one of the following three different kinds of effect:

1. The promise does not help us at all. For example, knowing that the chromatic number of a graph is exactly 1000 does not reduce the quantifier complexity of the reachability problem—neither for directed nor for undirected graphs.
2. The promise makes the reachability problem trivial. For example, if a graph has a chromatic number of 1, then the graph is discrete and there is a path between any two vertices if and only if they are identical.
3. The promise changes the quantifier complexity to $\log_2 k$ plus or minus some constant. For example, knowing the an undirected graph has a dominating set of size k allows us to describe reachability using $\log_2 k \pm O(1)$ quantifiers and this bound is tight.

In the present section we study a graph parameter that does not follow this pattern: the independence number $\alpha(G)$. An *independent set* in a graph is a set of vertices such that there is no edge between them (note that this definition also makes sense for directed graphs). The independence number of a graph is the size of its smallest independent set.

For undirected graphs, the influence of the independence number follows the familiar pattern:

Theorem 4.1.

$$\mathrm{qc}(\textsc{ureach}_{\alpha=k}) = \mathrm{qc}(\textsc{ureach}^{\infty}_{\alpha=k}) = \log_2 k \pm O(1).$$

Proof. For the lower bound, note that the graphs from Theorem 3.1 have independence number $\log_2 k - O(1)$ (just take every second vertex). For the upper bound, note that in a graph having independence number α the distance from s to t can be at most 2α since taking every second vertex on a shortest path yields an independent set. ☐

However, for directed graphs the situation is different as the next theorem shows. The upper bound in the theorem is proved in [NT05]. The new part is the lower bound.

Theorem 4.2.

$$\underbrace{2\log_3 2}_{>1.26} \log_2 k \leq \mathrm{qc}(\textsc{reach}_{\alpha=k}) \leq k + 3.$$

The theorem shows that $\text{qc}(\textsc{reach}_{\alpha=k})$ is neither infinity nor is it $\log_2 k + O(1)$. In the rest of the present section the lower bound from this theorem is proved.

4.1 Outline of the Proof of the Lower Bound

Our aim for the rest of this section is to show

$$\text{qc}(\textsc{reach}_{\alpha=k}) \geq 4 + 2\log_3 2\log_2 k = 4 + 2\log_3 k$$

for $k = 3^x$. The proof is based on an induction on k (more precisely on x).

1. We first prove the claim for $k = 1$. For this we construct two directed graphs G_1 and $\bar{G}_1$ such that
   (a) There is a path from s to t in G_1.
   (b) There is no path from s to t in $\bar{G}_1$.
   (c) Both graphs have independence number 1.
   (d) Player II has a 3-round winning strategy for these graphs.

   Once we have constructed graphs with these properties, we know $\text{qc}(\textsc{reach}_{\alpha=1}) > 3$. For $k = 1$ this is exactly the claim.

2. In the inductive step from $k = 3^x$ to $k' = 3k = 3^{x+1}$ we construct two new graphs G_{3k} and $\bar{G}_{3k}$ based the graphs G_k and $\bar{G}_k$. The new graphs will have the following properties:
   (a) There is a path from s to t in G_{3k}.
   (b) There is no path from s to t in $\bar{G}_{3k}$.
   (c) Both graphs have independence number $3k$.
   (d) Player II has a winning strategy for these graphs that lasts two rounds longer than the winning strategy for the graphs G_k and $\bar{G}_k$.

4.2 Start of the Induction: The Special Case of Tournaments

Our first task is to prove the claim for $\alpha = 1$. As stated before, we must construct two graphs G_1 and $\bar{G}_1$ such that player II has a 3-round winning strategy. Graphs with independence number 1 are also known as *tournaments*. (To be precise, directed graphs with $\alpha(G) = 1$ are called *semicomplete*. A tournament is a semicomplete graph whose edge relation is antisymmetric and antireflexive.)

The main problem with the construction of the two tournaments is that we cannot use Hanf's theorem to show that player II has a winning strategy: For a 3-round game, Hanf's theorem asks us to count the number of isomorphism types of spheres of radius 8 in the Gaifman graph of these tournaments. However, the Gaifman graph of a tournament is a clique and, hence, spheres even of radius 1 always encompass the whole tournament. Thus, Hanf's theorem essentially asks us to compare the isomorphism types of the whole tournaments—and these isomorphism types must be different since in one tournament there is a path from s to t and in the other one there is none. Thus, *we cannot apply Hanf's theorem, even if we try to make the tournaments arbitrarily large.*

This leaves us with the challenge of constructing two (preferably small) tournaments G and $\bar{G}$ and directly proving that player II has a 3-round winning strategy. The tournaments cannot be arbitrarily small since we can express reachability in 8 steps using 3 nested quantifiers, which shows that the minimum size of the tournaments must be 10.

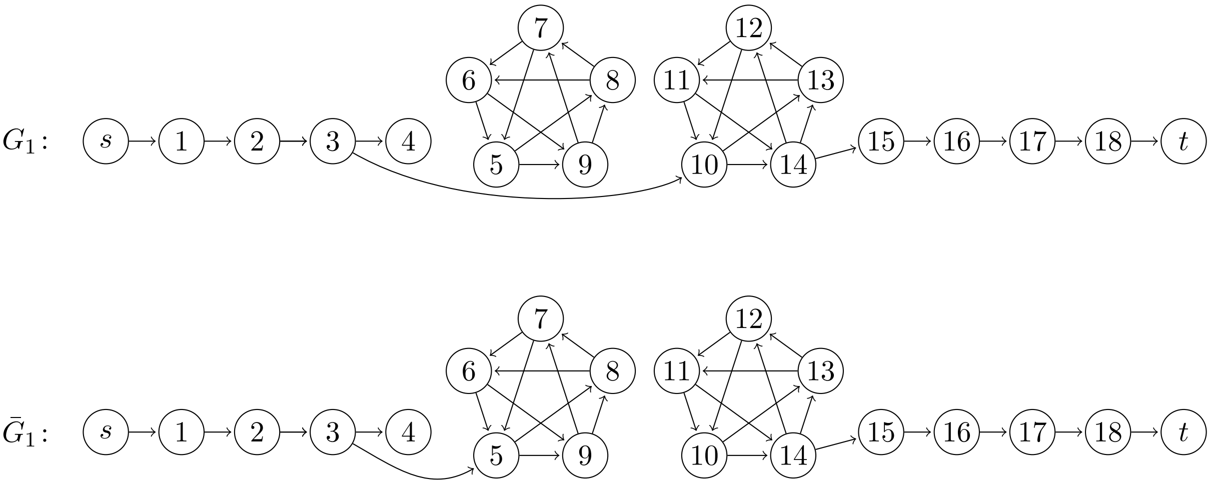

Figure 1: The tournaments G_1 and $\bar{G}_1$. For clarity, most edges are not shown. For any vertices that are not connected by an edge in the figure, the missing edge "points left." For example, there is an edge from vertex 8 to vertex 4 in both graphs since vertex 4 is more to the left than vertex 8. In the upper tournament there is a path from s to t, but in the lower there is not. Player II wins a depth 3 Ehrenfeucht–Fraïssé game on these tournaments.

Theorem 4.3. *Player II has a depth 3 winning strategy for the tournaments G_1 and $\bar{G}_1$ from Figure 1.*

Proof. The strategy of player II for the tournaments G_1 and $\bar{G}_1$ is hard to describe verbally. Most of the time, player II either duplicates the moves of player I identically or player II "mirrors" the two pentagons. However, there are many (hundreds) of cases when player II has to do an unintuitive special move at the end to ensure that the induced graphs are isomorphic.

Because of this, a computer program was used to verify that player II does, indeed, have a winning strategy for the first three rounds. The program did an exhaustive search over all possibly strategies of player I and verified that player II always wins. □

Corollary 4.4. $\mathrm{qc}(\text{REACH}_{\alpha=1}) > 3$.

Corollary 4.5. $\mathrm{qc}(\text{REACH}_{\alpha=1}) = 4 = \alpha + 3$.

4.3 The Inductive Step

Our next task is to prove the inductive step. Recall from the proof outline that our aim is to prove the following theorem:

Theorem 4.6. *Let G_k and $\bar{G}_k$ be graphs with independence number k. Let $G_k \in$ REACH and $\bar{G}_k \notin$ REACH and let player II have a winning strategy for an r-round Ehrenfeucht–Fraïssé game on these graphs. Then there exist graphs G_{3k} and $\bar{G}_{3k}$ such that*

1. *There is a path from s to t in G_{3k}.*
2. *There is no path from s to t in $\bar{G}_{3k}$.*
3. *Both graphs have independence number $3k$.*
4. *Player II has a winning strategy for an $(r+2)$-round Ehrenfeucht–Fraïssé game on these graphs.*

$$R\colon \qquad G_k \longrightarrow G_k \longrightarrow G_k \longrightarrow \cdots \longrightarrow G_k \longrightarrow G_k \longrightarrow G_k \longrightarrow \cdots \longrightarrow G_k \longrightarrow G_k \longrightarrow G_k$$

$$\bar{R}\colon \qquad G_k \longrightarrow G_k \longrightarrow G_k \longrightarrow \cdots \longrightarrow G_k \longrightarrow \bar{G}_k \longrightarrow G_k \longrightarrow \cdots \longrightarrow G_k \longrightarrow G_k \longrightarrow G_k$$

Figure 2: The row gadgets R and $\bar{R}$. There are l copies of G_k in each of the graphs, referred to as G_k^1 to G_k^l, except that the middle copy $G_k^{l/2}$ in $\bar{R}$ is replaced by $\bar{G}_k$. The edges from one copy to the next represent a single edge from the target of one copy to the source of the next copy. Between any two different copies all vertices are connected by edges pointing left.

Once we have proved this theorem, we can conclude that for $k = 3^x$ we have

$$\mathrm{qc}(\textsc{reach}_{\alpha=k}) \geq 4 + 2x = 4 + 2\log_3 k.$$

Proof. The graphs G_{3k} and $\bar{G}_{3k}$ are made up from a number of gadgets, whose construction is described in the following. For each gadget we first describe its construction and then prove some basic properties about it.

The Basic Row Gadget. The first gadget, called the *row gadget* R, is a long row of copies of G_k. In detail, the row gadget R is a graph with a source vertex s and a target vertex t that contains a large number G_k^1, G_k^2, ..., G_k^l of copies of G_k. The exact number l of copies is not important as long as there are "enough" copies, but at the end of the proof we will see that $l = 2^{r+4}$ suffices. In particular, this choice ensures that the condition $r < \log_2 l - 3$ holds.

The copies are connected as follows: We add a directed edge from the target of G_k^1 to the source of G_k^2. Next, there is an edge from the target of G_k^2 to the source of G_k^3; and so on. The source s of R is the source of G_k^1, the target t of R is the target of G_k^l. In addition to all these edges we add "backward edges" from all copies to all previous copies. Thus, for any two vertices u in G_k^i and v in G_k^j, where $i < j$, we add an edge from v to u.

We make the following claims about R:

1. There is a path from s to t in R.

2. The independence number of R is at most k.

The first claim is easy to prove: The path from s to t simply passes through all G_k^i, each of which is an element of $\textsc{reach}$ be assumption and each of which is connected to the next copy. For the second claim, just note that an independent set in R cannot contain vertices from two different copies since any two such vertices are connected by an edge. Thus, the independence number of R is the same as the independence number of G_k.

The Distance Hiding Strategy. For our next claim, consider an Ehrenfeucht–Fraïssé games played on two copies of R. Player II trivially has a winning strategy for this game, no matter how long we play. Now consider the same situation, but assume that one pebble pair has already been placed: In the first R there is a pebble on some vertex v of G_k in copy number $l/2$ of G_k. In the second R there is a pebble of the same color on the same vertex v of G_k but in copy number $l/2 + 1$. In this situation player I must have a winning strategy for a sufficiently large number of rounds since the two pebbles are at different distances from both s and t. However, we claim that player II can cheat about the distance for a certain number of rounds:

3. Suppose there is a pebble in the first R on some vertex v of G_k in copy number $l/2$ of G_k and in the second R there is a pebble of the same color on the same vertex v of G_k but in

copy number $l/2+1$. Then player II has a winning strategy for a game played for $\log_2 l-2$ rounds on these prepebbled graphs.

The strategy of player II might be called the "distance hiding strategy." It is a slight generalization of a well-known strategy that is used to show that reachability in 2^r steps really needs r quantifiers: The basic observation is that when we have r pebbles available, distances greater than 2^r are indistinguishable from infinity.

In detail, the distance hiding strategy works as follows: In round s, let p_{-1}, p_0, p_1, …, p_s be the vertices of the pebbles placed on the first R and let q_{-1}, q_0, q_1, …, q_s denote the corresponding vertices of the pebbles placed in the other R. The pebbles p_{-1} and q_{-1} are special; they are fixed to lie on the sources of the two R. Similarly, the pebbles p_0 and q_0 are fixed to lie on the targets. The pebbles p_1 and q_1 are the two pebbles for which we know that one lies in the "middle" copy of G_k and the copy next to this copy.

Each pebble lies in some copy of G_k. Let p_i lie in copy $\hat{p}_i$ and, likewise, let q_i lie in $\hat{q}_i$. The objective of player II is to maintain the following invariants: First, a pebble p_i and the corresponding pebble q_i may lie in different copies of G_k, but *inside* the two copies they lie on the same vertex of G_k. Second, for all pebble pairs (p_i, p_j) in the first R and the corresponding pair (q_i, q_j) in the second R, the differences $\hat{p}_i - \hat{p}_j$ and $\hat{q}_i - \hat{q}_j$ are

- the same or
- their absolute values are both larger than $2^{(\log_2 l-2)-s}$.

In order to maintain the invariant, player II plays as follows: Whenever player I places a pebble in the first R (the other case is symmetric) on some vertex p_{s+1} in copy $\hat{p}_{s+1}$ of G_k, player II places q_{s+1} on the same vertex inside copy number $\hat{q}_{s+1}$ of G_k. It remains to explain how $\hat{q}_{s+1}$ is chosen. To determine this number, player II finds the number $\hat{p}_i$ that is closest to $\hat{p}_{s+1}$ and chooses $\hat{q}_{s+1}$ such that it has the same distance to $\hat{q}_i$. Formally, $\hat{q}_{s+1} = \hat{q}_i + \hat{p}_{s+1} - \hat{p}_i$.

It is now easy to see, see for example [Imm98, page 97] or [EF91, page 23] for a detailed proofs, that player II can maintain the invariants for $\log_2 l - 2$ rounds.

The Row Gadget Pair. Based on the construction of the gadget R, we can define a new gadget $\bar{R}$ as follows: the gadget $\bar{R}$ is identical to the gadget R, except that the middle copy $G_k^{l/2}$ of G_k in R is replaced by a copy of $\bar{G}_k$ in $\bar{R}$. The following claims concerning $\bar{R}$ follow easily from the definitions:

4. There is no path from s to t in $\bar{R}$.
5. The independence number of $\bar{R}$ is at most k.

We can now ask how many rounds player II will win an Ehrenfeucht–Fraïssé games played on R and $\bar{R}$. A simple strategy for player II is to play as follows: Whenever player I places a pebble p_i on any of the copies of G_k *except for the middle one*, place the pebble q_i at the exact same position in $\bar{R}$; and vice versa. However, when player I places a pebble on the middle copy of R or $\bar{R}$, answer with a pebble in the middle copy of G_k in $\bar{R}$ or R *chosen according to player II's winning strategy for the graphs G_k and $\bar{G}_k$.* Clearly, this ensures that player II has an r-round winning strategy for R and $\bar{R}$:

6. Player II has an r-round winning strategy for R and $\bar{R}$.

A Strategy for the Row Gadget Pair. For our next claim, we once more consider the "prepebbled game" played on R and $\bar{R}$.

7. Suppose there is a pebble in the middle copy $G_k^{l/2}$ of R and a pebble of the same color in $\bar{R}$ at the same position inside G_k, but not in the middle copy of G_k (which has been replaced by $\bar{G}_k$), but rather in the copy $G_k^{l/2+1}$. Then player II has a winning strategy for a game of r rounds played on these prepebbled graphs, provided $r < \log_2 l - 2$.

The claim is shown by combining the distance hiding strategy and the winning strategy for G_k and $\bar{G}_k$: Whenever player I has placed a new pebble, player II uses the distance hiding strategy to determine the copy G_k into which she should place her pebble in answer. This results in the numbers $\hat{p}_{s+1}$ and $\hat{q}_{s+1}$ of copies of G_k or, possibly, $\bar{G}_k$. One is the number of the copy where player I has placed his pebble, the other is the number where player II will place her pebble in answer. Player II must now decide where to place her pebble inside the copy of G_k or $\bar{G}_k$. If the copies number $\hat{p}_{s+1}$ and $\hat{q}_{s+1}$ are both "normal" copies, then player II just places her pebble in the same way as player I did. However, if $\hat{q}_{s+1}$ happens to be the number of the copy that was replaced by $\bar{G}_k$, then player II places her pebble according to the winning strategy for G_k and $\bar{G}_k$.

The Fork Gadget Pair. We are now ready to introduce a new pair F and $\bar{F}$ of gadgets. These *fork gadgets* are defined as follows: F consists of two disjoint copies of $\bar{R}$ and a copy of R. Let us call the first two copies the *first two rows*, the third, different, copy is the *third row*. The sources of all three copies are merged and the targets are also merged. No other edges are added. The gadget $\bar{F}$ is constructed in a similar way, only this time three copies of $\bar{R}$ are used instead of two copies of $\bar{R}$ and a single copy of R. We also call these copies the three rows of $\bar{R}$. We make the following claims about the gadget pair:

8. There is a path from s to t in F, but no path in $\bar{F}$.

9. The independence numbers of F and of $\bar{F}$ are $3k$.

To see that these claims hold, first note that there is, indeed, a path from s to t in F: We just go through the copy of R inside F. On the other hand, there is not path in $\bar{F}$ since there is no path from s to t in any of the $\bar{R}$. The independence number of both F and $\bar{F}$ are $3k$ since picking any $3k + 1$ vertices from F or $\bar{F}$ will cause $k + 1$ vertices to be picked from the same copy of R or $\bar{R}$. Since the independence numbers of these graphs are k, neither they nor the original $3k + 1$ vertices can be independent.

Strategies for the Fork Gadget Pair. As before, we show that player II has a winning strategy in a certain prepebbled game:

10. Suppose there is a pebble in F's first row and a pebble of the same color at the same position in $\bar{F}$. Then player II has a winning strategy for a game of $r + 1$ rounds played on these prepebbled graphs, provided $r < \log_2 l - 2$.

To prove this claim, we consider the different places where player I can place his pebble in the first of the $r + 1$ rounds. We go over them one-by-one:

- Player I places the first pebble (of the $r + 1$ pebbles) on a vertex anywhere in the first two rows of F or $\bar{F}$ or in the third row of F or $\bar{F}$, but not in the critical middle copy $G_k^{l/2}$ of G_k. Player II responds by placing her on the same vertex in the other graph.

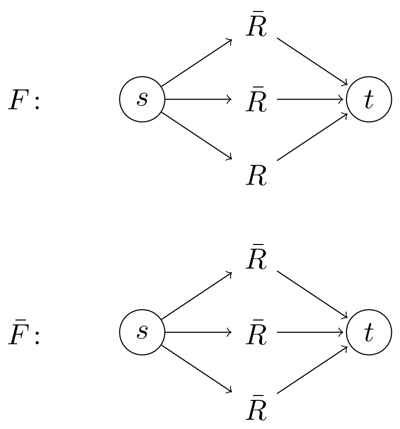

Figure 3: The fork gadgets F and $\bar{F}$. The sources of the row gadgets have been merged into the single vertex s and the targets are also merged into the single vertex t.

The situation we now face is that the graphs F and $\bar{F}$ are pebbled identically and there are no pebbles in the critical copy of G_k where F and $\bar{F}$ differ.

The winning strategy of player II for the remaining r rounds can now be assembled from individual the winning strategies for each row. For the first two rows player II's winning strategy is the trivial duplication strategy since these rows are identical in both graphs. For the third rows of F and $\bar{F}$, player II also plays a trivial duplication strategy, except when some (or all) of the r pebbles are placed in the middle copy $G_k^{l/2}$ of G_k. Here we do the same as we did in the strategy from claim 6: We play according to player II's winning strategy on G_k and $\bar{G}_k$.

- Player I places his first pebble (of the $r + 1$ pebbles) on a vertex in critical copy of $\bar{G}_k$ in the third row of $\bar{F}$. In this case, player II mentally changes what she considers to be the "second" row and what is the "third" row of $\bar{F}$. This change is, indeed, just a "mental" change since all rows of $\bar{F}$ are identical and we can label them in whatever way we like. After the change, player II faces the same situation as above, namely where player I has just placed a pebble in the second row, and player II can answer as above.

- Player I places his first pebble on a vertex of the critical copy $G_k^{l/2}$ of G_k in the third row of F. Player II responds by placing a pebble on the following vertex of $\bar{F}$: This vertex is located in the third row in copy $G_k^{l/2+1}$ of G_k (and not in copy $G_k^{l/2}$ since this has been replaced by $\bar{G}_k$). The position inside $G_k^{l/2+1}$ chosen by player II is the same as the position of player I's pebble inside $G_k^{l/2}$.

  Player II's winning strategy can now be assembled as follows: For the first two rows player II can uses the trivial duplication strategy. For the third row player II uses the strategy from claim 7.

A useful corollary of claim 11 is the following observation:

11. Player II has a winning strategy for a game of $r + 1$ rounds played on F and $\bar{F}$, provided $r < \log_2 l - 2$.

The Final Graphs. We now have all the basic building blocks to assemble the final graphs G_{3k} and $\bar{G}_{3k}$. They are constructed similarly to the row gadgets, only we now use F and $\bar{F}$

$$G_{3k}: \qquad F \longrightarrow F \longrightarrow F \longrightarrow \cdots \longrightarrow F \longrightarrow F \longrightarrow F \longrightarrow \cdots \longrightarrow F \longrightarrow F \longrightarrow F$$

$$\bar{G}_{3k}: \qquad F \longrightarrow F \longrightarrow F \longrightarrow \cdots \longrightarrow F \longrightarrow \bar{F} \longrightarrow F \longrightarrow \cdots \longrightarrow F \longrightarrow F \longrightarrow F$$

Figure 4: The final graphs G_{3k} and $\bar{G}_{3k}$. The situation is the same as in Figure 2, only G_k is replaced by F.

as building blocks instead of G_k and $\bar{G}_k$. In detail, G_{3k} consists of l copies F^1, F^2, ..., F^l of F. The target of each F^i is connected to the next F^{i+1} and we add "backward edges," which leading from every vertex in every F^i to every vertex in every F^j with $j < i$. The graph $\bar{G}_{3k}$ is constructed in the same way, only for the middle copy $F^{l/2}$ of F we use $\bar{F}$ instead of F.

It remains to prove three claims to complete the proof of the theorem.

12. There is path from s to t in G_{3k}, but there is no such path in $\bar{G}_{3k}$.

Inside G_{3k} we can get from s to t by following the paths through each F. Such a path exists in each F by claim 8. However, there is no way "to get through" the middle copy of $\bar{F}$ inside $\bar{G}_{3k}$, see claim 8 once more.

13. The graphs G_{3k} and $\bar{G}_{3k}$ have independence number $3k$.

Consider an independent set. All vertices of this set must have been chosen from the same copy of F or $\bar{F}$ since vertices from different copies are always connected by an edge. However, since both F and $\bar{F}$ have independence number $3k$ by claim 9, the independent set can have size at most $3k$.

14. Player II has an $(r+2)$-round winning strategy for G_{3k} and $\bar{G}_{3k}$, provided $r < \log_2 l - 3$.

To prove this final claim, we consider what player I might do with his first pebble:

- Player I places the first pebble anywhere inside G_{3k} or $\bar{G}_{3k}$, but not in the critical middle copies. Player II answers by placing a pebble on the same vertex in the other graph.

  The winning strategy for player II for the remaining $r+1$ rounds works as follows: Pebbles placed anywhere but on the critical copies of F and $\bar{F}$ are answered identically in the other graph. For pebbles placed on the two critical copies of F and $\bar{F}$, we use the strategy of claim 11.

- Player I places the first pebble inside G_{3k} and, there, inside the critical copy $F^{l/2}$ of F. Player II responds by placing her pebble on the vertex of $\bar{G}_{3k}$ that is at the same position inside F, but in copy $F^{l/2+1}$, not in the (replaced) copy $F^{l/2}$.

  For the remaining $r+1$ rounds player II plays according to a combination of the distance hiding strategy and the strategy from claim 11. More precisely, the argument is now exactly the same as the one we used in the proof of claim 7, only G_k is replaced by F and $\bar{G}_k$ is replaced by $\bar{F}$ and player II plays the distance hiding strategy for $r+1$ rounds instead of r rounds. Since $r+1 < \log_2 k - 2$ by assumption, player II will succeed in hiding the distance to the source and target for $r+1$ rounds.

- Player I places the first pebble inside the critical middle copy of $\bar{F}$ inside $\bar{G}_{3k}$. If the pebble has not been placed in the first row of $\bar{F}$, as in the proof of claim 10 player II exchanges the roles of the row the pebble is in and the first row. Thus, we may assume that the pebble has been placed in the first row of $\bar{F}$. Player II then places her pebble on the same vertex in G_{3k}, that is, in the first row of the middle copy of F.

The winning strategy for player II for the remaining $r+1$ rounds works as follows: Whenever player I places pebbles outside of the critical copies, player II uses the trivial duplication strategy. When player I places a pebble inside the critical copies, player II uses the strategy from claim 10.

To get the claim of the theorem, we invoke claim 14 for $l = 2^{r+4}$. $\square$

5 Conclusion

The first-order quantifier complexity of the reachability problems for graphs is related to many natural graph parameters. Typically, the first-order quantifier complexity of the reachability problem restricted to the class of graphs for which some graph parameter has a certain value k is either 0 or $\log_2 k \pm O(1)$ or ∞. This is true both in the finite and in the infinite, both in the directed and in the undirected case.

In the present paper one exception to this rule was presented: the graph parameter α. For finite directed graphs we showed that the first-order quantifier complexity of the reachability problem is somewhere between $1.26 \log_2 k$ and $k+3$ and, thus, neither of the form $\log_2 k + O(1)$ nor ∞.

An obvious remaining open problem is finding a way to narrow this gap. We can both try to raise the lower bound and also to lower the upper bound. For $\alpha = 1, 2, 3$ the lower and upper bounds do already match, which suggests that the following conjecture is true:

Conjecture 5.1. $\mathrm{qc}(\textsc{reach}_{\alpha=k}) = k + 3$.

Another open problem is finding another "interesting" graph parameter. A candidate for such a parameter is the spectral gap of a graph. This parameter is related to the distance in a non-trivial way and it is possible to prove non-trivial upper bounds on the first-order quantifier complexity of the reachability problem as a function of the spectral gap. Finding a matching lower bound appears to be a challenging problem.

References

[Die97] Reinhard Diestel. *Graph Theory.* Springer-Verlag, 1997.

[EF91] Heinz-Dieter Ebbeinhaus and Jörg Flum. *Finite Model Theory.* Springer-Verlag, 1991.

[Ehr61] Andrzej Ehrenfeucht. An application of games to the completeness problem for formalized theories. *Fundamenta Mathematicae*, 49:129–141, 1961.

[Fra54] R. Fraïssé. Sur quelques classifications des systèmes de relations. *Publ. Sci. Univ. Alger. Sér. A*, 1:35–182, 1954.

[Han65] W. Hanf. Model-theoretic methods in the study of elementary logic. In J. Addison, L. Henkin, and A. Tarski, editors, *The Theory of Models*, pages 132–145. North Holland, 1965.

[Imm98] Neil Immerman. *Descriptive Complexity.* Springer-Verlag, 1998.

[NT05] Arfst Nickelsen and Till Tantau. The complexity of finding paths in graphs with bounded independence number. *SIAM Journal on Computing*, 34(5):1176–1195, 2005.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de/